JAWS2012

JAWS2012

Single Agent Optimal Pricing under Market Uncertainty by Using POMDP

Dong HAO — Graduate School of ISEE, Kyushu University
hao@agent.inf.kyushu-u.ac.jp

Atsushi IWASAKI — Graduate School of ISEE, Kyushu University
Iwasaki@inf.kyushu-u.ac.jp

Makoto YOKOO — Graduate School of ISEE, Kyushu University
yokoo@inf.kyushu-u.ac.jp

YongJoon JOE — Graduate School of ISEE, Kyushu University
yongjoon@agent.inf.kyushu-u.ac.jp

Michihiro KANDORI — Faculty of Economics, University of Tokyo
kandori@e.u-tokyo.ac.jp

Ichiro OBARA — Department of Economics, UCLA
iobara@econ.ucla.edu

keywords: monopoly market, dynamic pricing, demand uncertainty, informativeness, partially observed Markov decision processes (POMDP)

Summary

We develop a partial observed Markov decision process (POMDP) model, to investigate the optimal pricing policy for a monopoly agent, which faces uncertain market demand over an infinite time horizon. We analyze the informativeness of observation distributions and exam that, to get a more accurate knowledge about the market state, the monopoly agent should set its price relatively lower than the myopic monopoly price. On the basis of the POMDP model and the analysis of observation distributions, we conduct an experimental study by using POMDP solver, and find the optimal pricing policy for the monopoly agent. Our analysis results indicate that, when the monopoly agent is certain about the market state, choosing the monopoly price for the corresponding market state will maximize its profit. On the contrary, when the agent is uncertain about the market state, it should not stick to the myopic monopoly price. Instead, its optimal strategy is to choose a relatively lower price which can be used to exam the market and obtain a better knowledge of the market state.

1. Introduction

A monopoly agent is a single person or enterprise who is the only seller of a particular product. A monopoly market is where a monopoly agent exists. In the real world, there are several possible examples for monopoly market. Microsoft is a major monopoly in the software and operating system market. Another classical case is that the famous diamond company De Beers used to be pure monopoly for around one hundred years. Besides, AT&T used to be a pure monopoly in US telecommunication market. In the monopoly market, there is no other competitor to the monopoly agent and the monopoly agent's selling quantity of product is strongly depending on its pricing strategy. If the pricing repeats infinite times, the monopoly agent needs to dynamically adjust its price over time, in order to maximize its long-term expected profit. Such a problem falls into the category of dynamic pricing which has been attracting a lot of research effort [Aviv 05]. Many industries, such as airlines, hotels and various retailers, use dynamic pricing to match demand or maximize profit.

When the monopoly agent uses dynamic pricing, one key factor it needs to take into consideration is the market state. In the monopoly market, the market state (market condition) can be either good or bad. The demand curves (price versus selling quantity) for different market sates are not the same [Mirman 93]. The monopoly agent has corresponding optimal actions for each market state, which we called the "monopoly prices". The monopoly price for a good market is usually higher than the monopoly price for a bad market. For example, for the monopoly company MacDonald, its optimal action (monopoly price) when consumers have higher incomes is P_H; on the contrary, it will choose another best price P_M when consumers' incomes are low. Generally, in a good market state, the monopoly price is higher, which indicates that P_H is higher than P_M. Furthermore, because the demand curves for different market states are not the same, choosing a same price, the monopoly agent is more likely to realize a higher selling quantity in good market state than in bad market state. For example, choosing a same price, MacDonald will sell more BigMac hamburgers when it is in a good market.

It is clear that: if the monopoly agent is sure about which market state it is in, its price decision making is quit simple: it just need to choose the corresponding monopoly price P_H or P_M for a good market and a bad market, respectively. However, in the real world, unfortunately, these market states cannot be perfectly observed by the monopoly agent. In other words, after choosing a certain price, the observed selling quantity is stochastic and noisy. This makes the monopoly agent experience an ongoing uncertainty about which demand curve can best describes the demand at each stage of the infinite pricing process. The market state changes over time, and the transition between good state and bad state can be described by a stochastic matrix. For example, when MacDonald opens a new restaurant in one region, the market condition in this region is always changing due to the complicate economic and social factors. However MacDonald cannot know exactly whether the current market condition is good or bad. When it is very uncertain about the market state, it cannot simply set its price as monopoly prices P_H and P_M. Because a wrong price setting will sacrifice its profit even result in a financial loss. In this case, the monopoly agent cannot perfectly observe the market state. When choosing the best price, it also need to learn and infer whether the market state is good or bad.

Partially observed Markov decision process (POMDP) have been successfully studied in agent planning field that require balancing actions that increase an agent's knowledge and actions that optimize an agent's profit [Russell 09][Kandori 10]. The key advantages of the POMDP approach are that it is able to deal with uncertainty, and it is easy to specify. However, POMDP have not yet been widely introduced into the field of market design and profit management. In this paper, based on POMDP and the theory of statistic decision, we investigate the optimal pricing policy for the monopoly agent which faces uncertain market demand over infinite time horizon. We first propose a POMDP based model to catch the characteristic of the monopoly agent dynamic pricing problem. Then we analyze the informativeness of observation distributions under different price strategies. We exam that, to get a more accurate observation about the market state, the agent should set its price relatively lower than the myopic monopoly prices. On the basis of the POMDP model and the analysis of observation distributions, we finally conduct an experimental study by using POMDP solver, and find the optimal pricing policy for the monopoly agent. We also investigate the relationship between agent's belief and its optimal pricing strategies. Our analyze results indicate that, when the agent is certain about the market state, adopting the monopoly price for the corresponding market state will maximize its profit. On the contrary, when agent is unsure which state the market is in, it should not stick to the myopic monopoly prices. Instead, it's better to choose a relatively and reasonably lower price which can be used to obtain a better knowledge to infer the market state.

2. POMDP Model

Consider a monopoly agent facing an uncertain demand market pursues to maximizes its expected profit throughout the infinite horizon sales. The monopoly agent has incomplete knowledge about the market state, that is, the observation of the selling quantity is stochastic and noisy which makes it hard for the monopoly agent to infer which demand curve it should follow. By choosing a price, the agent not only receives a profit, but also obtain an observation of the selling quantities which can be used to update its belief and deal with the uncertainty of market state.

2·1 *Market demand uncertainty*

We assume there are two kinds of market states: $s = G$ is the good state and $s = B$ is the bad state. Let $\overline{x}_s(P)$ denote the average selling quantity when the monopoly agent set price P under market state s. Then there are two average demand curves for good and bad markets, which are $\overline{x}_G(P)$ and $\overline{x}_B(P)$, respectively. We illustrate the average demand curves under different market conditions (prices versus selling quantities) as in Figure 1. From the curves' slopes and their intercepts on y-axis, we can see that, in good market, to get a higher selling quantity, the firm does not need to decrease its price very much. However, in bad market, in order to sell more products, the firm has to decrease its price sharply.

In the real world, instead of obtaining a deterministic selling quantity, the agent can only observe a stochastic selling quantity which follows a certain probability distribution. This stochastic observation makes it difficult for the monopoly agent to know which state the market is in. We call this *imperfect observation* which results in an uncertainty of market state. To seize this point, we introduce

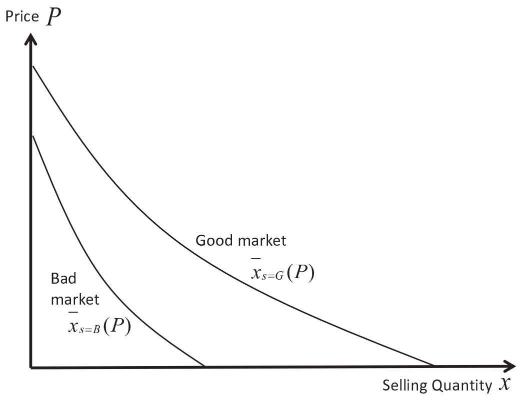

Fig. 1: Average demand.

such stochastic selling quantity in Figure 2. From this figure, we can see that, after choosing a price P, the realized selling quantity x follow probability distributions $R_s(x|P)$ which are shown as the two green curves. Note that one realized selling quantity that actually belongs to distribution $R_G(x|P)$, may also appears in the sample space of distribution $R_B(x|P)$. This is shown where the two distributions curves overlap each other. As a result, the monopoly agent can not perfectly distinguish which demand curve best describes the current market state.

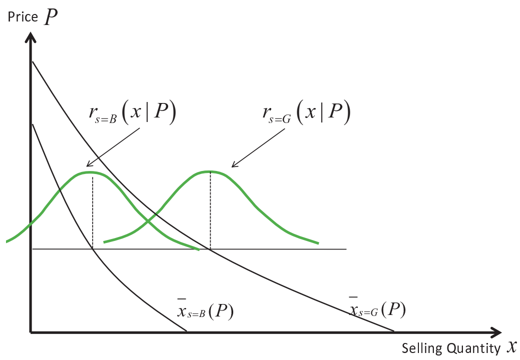

Fig. 2: Stochastic demands.

For numerical analysis, we consider the imperfect observation as in the following Figure 3. We concentrate on these two average demand curves, for good market state $s = G$ and bad market state $s = B$, respectively. We consider three candidate prices: low price P_L, middle price P_M and high price P_H. There are five selling quantity samples: low quantity $x_L = 0$, relatively low quantity $x_{LM} = 25$, middle quantity $x_M = 50$, relatively high quantity $x_{HM} = 75$, and high quantity $x_H = 100$. Then our study objects are the 15 black points in this figure.

In a word, for the monopoly agent optimal pricing problem, there are two physical market states. The agent has three actions, each of which induces a probability distri-

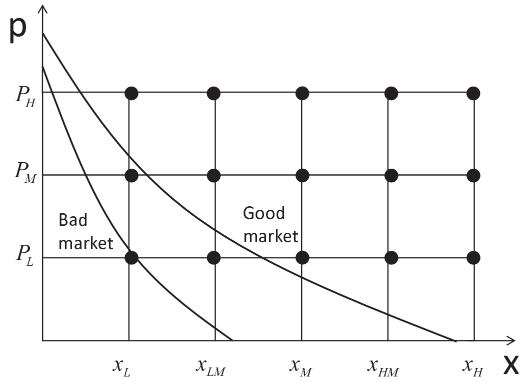

Fig. 3: 3 prices and 5 observations.

bution over five observation signals. The agent can learn about the market state by updating its belief, which is calculated from the observation distributions.

2·2 Belief of market state

At pricing period t, denote the monopoly agent's price strategy as P_t and its observed selling quantity as $x(P_t)$. This observed selling quantity $x(P_t)$ can be caused either by a good market market ($s = G$), or by a bad market ($s = B$). Set the probability that the market is in good and bad state as $\Pr(G)$ and $\Pr(B)$, respectively. Then, according to observation distribution $R_s(x|P_t)$ introduced in the last subsection, we have the conditional probabilities that: $\Pr(x(P_t)|G) = R_G(x|P_t)$ and $\Pr(x(P_t)|G) = R_B(x|P_t)$. Following Bayes rule, we derive the posterior probability as: $\Pr(G|x(P_t)) = \frac{\Pr(x(P_t)|G)\cdot\Pr(G)}{\Pr(x(P_t)|G)\cdot\Pr(G)+\Pr(x(P_t)|B)\cdot\Pr(B)}$. Let $\Pr(G) = b_t$ and $\Pr(G|x(P_t)) = b_{t+1}$, which are the agent's prior and posterior beliefs, respectively. Then we have

$$b_{t+1} = \frac{R_G(x|P_t)\cdot b_t}{R_G(x|P_t)\cdot b_t + R_B(x|P_t)\cdot(1-b_t)}. \qquad (1)$$

This posterior belief describes how much certainty the agent has that the current market is in good state ($s = G$). This belief is updated after every time the agent chooses its pricing strategy P_t and observes a selling quantity $x(P_t)$.

2·3 Profit function

Depending on the statement above, the monopoly agent's long-term expected profit can be calculated as:

$$\begin{aligned} & R_{ex}(P_t, b_t) \\ &= \sum_{t=1}^{\infty} \delta\cdot[b_t u_G(P_t) + (1-b_t)u_B(P_t)] \\ &= \sum_{t=1}^{\infty} \delta\cdot(P_t - c)[b_t \bar{x}_G(P_t) + (1-b_t)\bar{x}_B(P_t)], \end{aligned} \qquad (2)$$

where u_G and u_B are the deterministic profit when market state is good and bad, respectively. Factor c denotes the cost for producing one unit of product, and δ is the discount factor.

2·4 Formal POMDP model

The monopoly agent pricing problem is an single agent decision process in which it is assumed that the market dynamics are determined by an Markov decision process (MDP), but the monopoly agent can only indirectly learn about the market states. Thus, the monopoly agent pricing problem can be modeled as a partially observed Markov decision processes (POMDP).

Define the monopoly agent pricing problem as a six-tuple $\langle S, A, T, O, \Omega, B, R\rangle$, where S is the set of market states $s_t \in S = \{G, B\}$; A is the set of strategic prices in discrete horizon, and $P_t \in A$; T is the market state's transition matrix, that is $T = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix}$, where α and β both are decimals; O is the set of observed selling quantity, $x \in O$; Ω is the set of conditional observation probabilities. After choosing price P_t, the probability for the agent to observe a certain selling quantity $x(P_t)$ is denoted as $R_s(x|P_t)$; B is the set of beliefs about the current state, and $b_t(s_t|x(P_t)) \in B$; R is the set of expected profit, which is recorded as $R_{ex}(P_t, b_t) = \sum_{t=1}^{\infty} \delta \cdot (P_t - c)\left[b_t \overline{x}_G(P_t) + (1-b_t)\overline{x}_B(P_t)\right]$.

3. Observations and Uncertainty

In POMDP, the observation is a key point which has direct impact on the agent's profit. From this section, we begin to investigate the price strategies and observation for the POMDP pricing model. We will show that, observation under a higher price brings the agent with higher uncertainty. On the contrary decreasing the price will increase the agent's knowledge about market state.

3·1 Price strategy and observation accuracy

According to the demand curves proposed above, the monopoly agent can ensure that: By decreasing the price, the monopoly agent can push the mean demand curves further apart, which makes it easier for the monopoly agent to distinguish between the good and bad market states. The following Figure 4 shows this phenomenon. Recall that in Figure 2 we already discussed that the overlapping of the observation distributions makes it hard for the agent to infer which state the market is in. However, here we found that, by choosing a relatively lower price strategy, the agent can separate these two observation distributions (green lines) apart, which makes the overlapping area become smaller. The significance here is that, by changing the price strategy, the agent is possible to get a more accurate observed selling quantity, which can be a better indicator of the true market state.

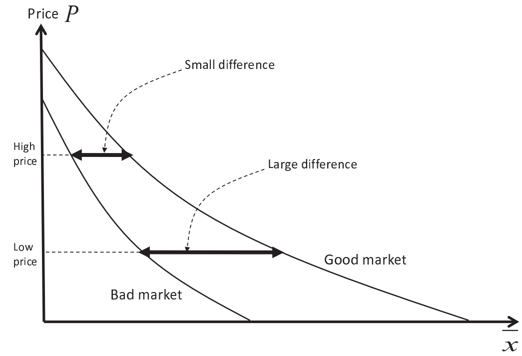

Fig. 4: Price VS. informativeness.

3·2 Relationship between observations

Assume that given a low price P_L, agent will observe a selling quantity $x'(P_L) \in (x_L, x_{LM}, x_M, x_{HM}, x_H)$. This observation quantity vector follows a probability distribution $R' = R'(x'|P_L)$ where $s \in \{G, B\}$. If the agent set another price $P_M > P_L$, there may be probability distribution $R = R(x|P_M)$. We say that the observation distribution $R'(x'|P_L)$ is more informative than $R(x|P_M)$, if and only if there exist a non-negative stochastic matrix $N = \left[N_{ij}\right]_{|R'|\times|R|}$, that is $R_j = \sum_i R'_i N_{ij}$. Note that $N = \left[m_{P_M,P_L}\left(x_j|x'_i\right)\right]_{i\leq 5, j\leq 5}$.

We call such a non-negative stochastic matrix the 'noise matrix', since there will be some noisy when we infer the unknown observation distribution under higher price, basing on the known observation distribution under a low price. We only know $R'_s(x'|\mathrm{P_L})$, what we need to do is to calculate $R_s(x|\mathrm{P_M})$ by using $R'_s(x'|\mathrm{P_L})$. The calculation utilized a noise matrix N, which makes $R'_s(x'|\mathrm{P_L})$ provide a better knowledge about market state than $R_s(x|\mathrm{P_M})$ (i.e., more informative [Agnieszka 03]). In our model, we define such a noise which ensures that

$$R_s(x|P_M) = R_s(x'|P_L)\left[m_{P_M,P_L}\left(x_j|x'_i\right)\right]_{i\leq 5, j\leq 5} \tag{3}$$

When we design the noise matrix N, we should consider about the following constrains: (1) Each element in the matrix $m_{P_M,P_L}\left(x_j|x'_i\right) > 0$. (2) Since the matrix N is stochastic, the sum of row elements should equal 1, that is $\sum_j m_{P_M,P_L}\left(x_j|x'_i\right) = 1$. (3) The sum of column elements is larger than 0, that is $0 < \sum_i m_{P_M,P_L}\left(x_j|x'_i\right) < \infty$. (4) The noise is independent of market states.

3·3 Noise matrix

Taking into consideration of the constrains summarized above, in this subsection, we put forward a noise matrix which will be used to describe the relationship of observations under different prices. If the agent observes a selling quantity x'_i under a low price P_L, it will have a conditional probability $m_{P_M,P_L}\left(x_j|x'_i\right)$ to observe another quantity x_j if it set a relatively higher price P_M. Given the two different

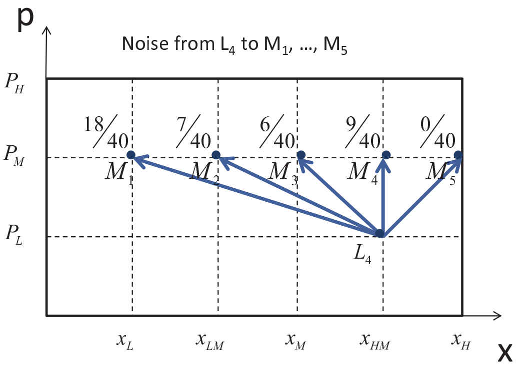

Fig. 5: Generate a sample noise matrix.

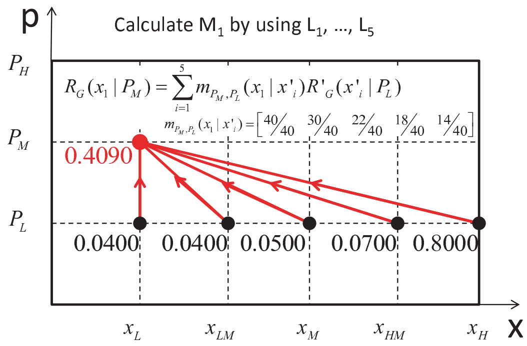

Fig. 6: Calculate distribution under middle price.

prices $P_M > P_L$, all different $m_{P_M,P_L}\left(x_j|x'_i\right)$ consist a noise matrix such that

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{30}{40} & \frac{10}{40} & 0 & 0 & 0 \\ \frac{22}{40} & \frac{6}{40} & \frac{5}{40} & \frac{7}{40} & 0 \\ \frac{18}{40} & \frac{7}{40} & \frac{6}{40} & \frac{9}{40} & 0 \\ \frac{14}{40} & \frac{8}{40} & \frac{7}{40} & \frac{6}{40} & \frac{5}{40} \end{bmatrix} \qquad (4)$$

In this matrix, rows are identified by index i while columns are identified by index j. Each number $m_{P_M,P_L}\left(x_j|x'_i\right)$ is a conditional probability that when quantity x'_i is observed under low price P_L, a corresponding quantity x_j will be observed under middle price P_M. Figure 5 illustrates how to assign the values in this noise matrix. Each blue arrow stands for a conditional probability that a selling quantity M_j will be observed under middle price P_M, given that selling quantity L_i has been observed under low price P_L. For example, the blue arrow from point L_4 to point M_1 denotes that: if a selling quantity x_{HM} is observed when the agent sets a low price P_L, when the agent sets a middle price $P_M > P_L$, it may observe selling quantity x_L with a conditional probability $m_{P_M,P_L}\left(x_L|x'_{HM}\right) = \frac{18}{40}$. For saving space, we only show the noises for the fourth observation signal: relatively high quantity x_{HM}.

4. Observation Distributions

In the last section, we proposed that, decreasing the price is a methodology that can reduce the agent's uncertainty about market state. To conduct a numerical study, in this section, we assign the values for the observation distributions accordingly.

4·1 Assign values for observation distributions

Based on the proposed noise matrix, in this section, we assume that, given that the agent chose a low price P_L, if the current market state is $s = G$, the probability distribution of the observed selling quantity follows

$$R'_G(x'|P_L) = [0.04, 0.04, 0.05, 0.07, 0.80]. \qquad (5)$$

On the contrary, if the current market state is $s = B$, we assume that the observation distribution follows

$$R'_B(x'|P_L) = [0.45, 0.15, 0.15, 0.15, 0.10]. \qquad (6)$$

By introducing the noise matrix we defined above, we can utilize the distributions under low price P_L to generate a new distributions under middle price P_M. Calculate the observation distribution under P_M in the following way:

$$\begin{aligned} R_G(x|P_M) &= R_G(x'|P_L) \times \left[m_{P_M,P_L}\left(x_j|x'_i\right)\right]_{5\times 5} \\ &= [0.4090, 0.1898, 0.1567, 0.1445, 0.1000] \end{aligned} . \qquad (7)$$

As well, using the same noise matrix, we can calculate the observation distribution under high price P_H as:

$$\begin{aligned} R_G(x|P_H) &= R_G(x'|P_M) \times \left[m_{P_H,P_M}\left(x_j|x'_i\right)\right]_{5\times 5} \\ &= [0.7376, 0.1162, 0.0588, 0.0749, 0.0125] \end{aligned} . \qquad (8)$$

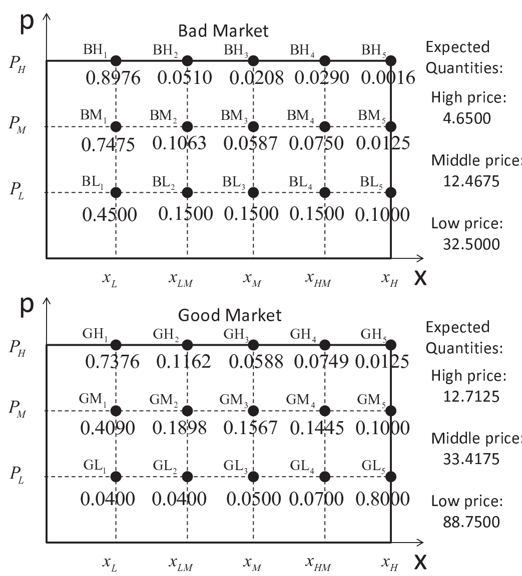

Fig. 7: Observation distributions.

4·2 Verification

By using the procedure for generating the observation distributions discussed above, we calculated the values for the probability distributions and expected selling quantities, which is summarized in Figure 7.

(1) The expected selling quantities under higher price is smaller (e.g., if the selling quantity samples are $x_L = 0$, $x_{LM} = 25$, $x_M = 50$, $x_{HM} = 75$ and $x_H = 100$, then in bad market, the expected quantities $32.5000 > 12.4675 > 4.6500$; in good market, $88.7500 > 33.4175 > 12.7125$). This reveals the truth that no matter which state the market is in, the agent can realize a higher selling quantity by decreasing its price.

(2) In general, the signal's probability is decreasing when the signal point is farther away from the average demand curve (e.g. $GM_1 > GM_2 > GM_3 > GM_4 > GM_5$ and $GH_1 > GH_2 > GH_3 > GH_4 > GH_5$). This is consistent with the reality that, the distribution density is higher where the observation signal is near the average demand curve, while it becomes smaller while the signal is leaving away from the average demand curve.

(3) The probability values on each same point under different market states have some similarities. And this similarities decrease when the price is lower. For example, $BL_1 = 0.4500$ is very different from $GL_1 = 0.0600$; but $BM_1 = 0.7475$ is relatively closer to $GM_1 = 0.4090$; furthermore, $BH_1 = 0.8976$ is much more closer to $GH_1 = 0.7376$. This makes the distributions correspond to the fact that, under higher price, it's more difficult for the agent to distinguish the current market state by using the observation distributions.

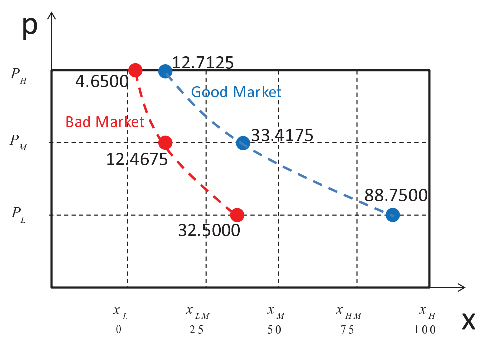

Fig. 8: Average demand and expected selling quantity.

In brief, choosing different price strategies brings the monopoly agent different observation distributions. And these distributions have underlying relationships which are characterized by a noise matrix. Choosing a low price makes the agent easier to infer the current market state.

5. Experiment and numerical results

In this section, we implement the above theoretical analysis result into a POMDP experiment and unitize the Cassandra POMDP solver to find the optimal pricing policy for the monopoly agent.

5·1 Experiment settings

In our POMDP file, there are three actions including high price P_H, middle price P_M and low price P_L. The observation distributions for each action in each market state is assigned according to the analysis in the previous section. The profits are assigned following the principle that: The monopoly (myopic) price for a good market is P_H while the monopoly (myopic) price for a bad market is P_M. For the market state transition matrix, we assume that it is independent from the action of the monopoly agent which means that the for each action, the state transition matrix is identical. Besides, we also consider that, if market is in bad state, it's more difficult for it to turn back to good state. Thus the state transitions are assigned as: $T_{good \to good} = 0.8$ and $T_{bad \to bad} = 0.9$.

5·2 Optimal pricing policy

Running the POMDP solver with the above parameter settings, we generate the optimal pricing policy for the monopoly agent as the following Table 1. In this table, numbers in the first column denote the sequence of the seven nodes which consist the optimal policy. The second column shows the optimal actions for each node. On each node, the observed selling quantity can be either X_L, X_{LM}, X_M, X_{HM} or X_H. Observing a selling quantity, the agent should transit (or stay) in order to maximize its profit. For instance, in node 1, the agent's best action is middle price P_M. Choosing P_M, if the agent observes a middle selling quantity X_M, it should transit to node 2, where its best action becomes P_L.

Table 1: Optimal pricing policy

Node	Action	X_L	X_{LM}	X_M	X_{HM}	X_H
0	1 (P_M)	1	1	2	1	4
1	1 (P_M)	1	2	2	2	4
2	2 (P_L)	1	1	1	1	5
3	2 (P_L)	1	1	1	2	6
4	2 (P_L)	1	1	2	2	6
5	2 (P_L)	1	2	2	2	6
6	0 (P_H)	5	6	6	6	6

From this optimal policy, it can be observed that: If large selling quantity is observed many times, it is more possible for the agent to transit to node 6, in which the best action is to choose high price P_H. While if small selling quantity is observed for many times, it is more possible for the agent to transit to nodes 1, in which the best action is to choose middle price P_M. This is consistent with the reality. Because when the agent is examing the market state by choosing a low price, the more times it observes a high selling quantity, the higher belief it will have that the market is in good state. Thus it is more likely to transit to node 6. On the contrary, if the agent observes small selling quantity for many times, it will convince itself that the market is in bad state, therefore it should choose the monopoly price for bad market which is P_M. For example, if the agent observes quantity X_L for many times, it will easily transit to node 1 and choose middle price P_M. However, if the agent observes middle selling quantity for many times, it could not distinctly see what is the current market state. In this case, the agent should not recklessly choose any monopoly prices, but cautiously exam and learn about the market state. To get more information about the market state, what the agent needs to do is decreasing the price to a reasonable lower value. The significance of such more informative observations under lower price is that they can make the agent have a better distinction between the good market and bad market. Once the agent has obtained enough information and its belief has been updated to a enough large (or small) value, it will be confident that the market is in good or bad state, and can switch its action to P_H (or P_M) accordingly.

Although there are seven nodes in the optimal policy, they can be divided into three groups: the nodes with action P_M, nodes with action P_L and nodes with action P_H. Furthermore, such policy is a finite state automaton. Therefore, we illustrate the refined policy as the following FSA.

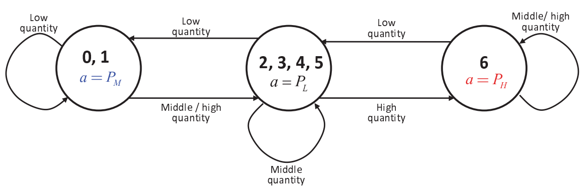

Fig. 9: Optimal pricing policy.

5·3 *Belief and profits*

According to the discussion in section 2, we know that the monopoly agent's belief about the market state will be updated while the pricing repeats over infinite horizon. Let b_t denote the monopoly agent's belief that the current market is in good state, and let u_G (u_B) denote the monopoly agent's profit if the market is good (bad) state. Then the agent's expected profit function in POMDP is $R_{ex}(P_t, b_t) = \sum_{t=1}^{\infty} \delta \cdot [b_t \overline{u}_G(P_t) + (1-b_t)\overline{u}_B(P_t)]$. Note that this value function is a linear function of the belief b_t.

After running the POMDP solver, we can get the optimal pricing policy with seven nodes, which we have already introduced in the last subsection. With accompany of these seven nodes in the policy, there are also seven linear value functions. And each node in the policy is characterized by one value function. In POMDP, the monopoly agent's optimal profit function is the upper envelop of those seven linear value functions, which reflects the maximized profit over the whole belief horizon $b \in [0, 1]$. The seven value functions are illustrated in the following table.

Table 2: Profits for the policy

Node	Action	profit ($s = G$)	profit ($s = B$)
0	P_M	613.1709	463.2531
1	P_M	614.4143	463.2474
2	P_L	614.4600	462.9942
3	P_L	614.5291	462.9595
4	P_L	614.5599	462.9334
5	P_L	614.5832	462.9064
6	P_H	614.6289	462.7351

This table shows the agent's profit values when the belief is at the endpoint ($b = 1$ and $b = 0$) where the market state is deterministic. The third and fourth column summarize the values of the seven nodes when market state is $s = G$ and $s = B$, respectively. The value function of each node is represented as a line linking two endpoints where $b = 0$ and $b = 1$. These functions are shown in Figure 10.

In each sub figure, the x-axis is the value of belief $b \in [0, 1]$. The y-axis is the profit values for each of the seven nodes. Each line in the figures illustrates a value function for one node. For example, in the first sub-figure Case-1, from top to bottom, the red line depict the value function for node 1 (where action is P_M); the four black lines denote the value functions for node 2, 3, 4, 5; and the blue line represents the value function for node 6. All the profit functions in the figure is piecewise-linear and convex (PWLC), and the monopoly agent's optimal profit is an upper envelop which consists of seven segments from the seven different profit functions. From the first sub-figure Case-1, we can observe that, when b is very close to 0, the red line (for node 1) is the value function which has the maximized value. This indicates that, when the belief is very small, the agent strongly believes that the market is in bad state, thus choosing middle price P_M is the best action. From the third sub-figure Case-3, where b is close

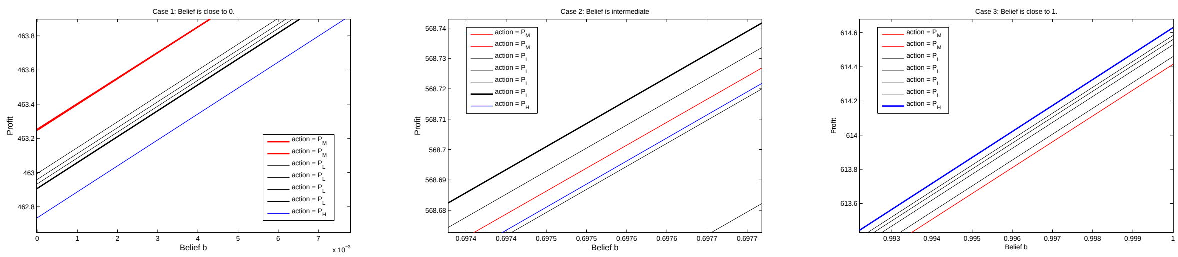

Fig. 10: Value functions versus belief.

to 1, the blue line (for node 6) has the maximum value, which means that, when belief is high, the agent is certain that the market is in good state. Therefore, choosing high price P_H is the best action. Besides, for the second sub-figure, where the belief is a intermediate value, price P_L (denoted by black lines) is the optimal action. Actually, from POMDP, we can know that, each node appearing in the optimal policy, will bring the agent with maximum value for a certain range of belief. This means that, the final optimal profit is upper envelop which consists of seven segments from the seven value functions. The optimal profit function R_{ex} embodies the following features: (1) It coincides with the value function of node 1 (P_M) for small beliefs. (2) It coincides with the value functions of node 2, 3, 4, 5 (P_L) for middle beliefs. (3) It coincides with the value functions of node 6 (P_H) for high beliefs.

These features are illustrated in the following abstracted Figure 11. Furthermore, based on the analytical and experimental result, we have the following conclusion:

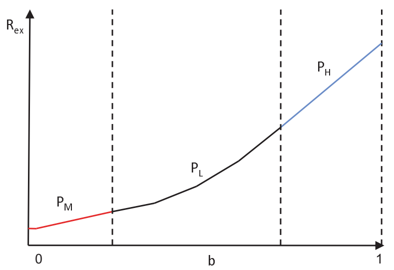

Fig. 11: Abstracted profit function, which is the upper envelop of the value functions from the seven nodes.

(1) When belief is relatively high $b \to 1$, it means that the monopoly agent has strong confidence that the market is in good state ($s = G$). In this case, to maximize its expected profit $R_{ex}(P_t, b_t)$, the agent is to choose the P_H (blue line), which is the good market monopoly price .

(2) When belief is relatively low $b \to 0$, the monopoly agent will be very confident that the market is in bad state ($s = B$). In this case the its best action is to choose P_M (red line), which the bad market monopoly price.

(3) When belief is intermediate, the monopoly agent will be very uncertain about the market state. In this case, recklessly choosing any monopoly price, the agent is possible to suffer some loss caused by an inaccurate price setting. Therefore, the agent should not myopically choose any monopoly price that may maximize the short-term immediate profit. Instead, it should exam and learn, and get more rich knowledge about the real market state. Regarding the exam and learn about the hidden market state, as the analysis in the previous section says: the agent should decrease its price to a relatively and reasonably lower level (P_L), which provides a more accurate observation of the market state. This corresponds to the result that, while belief is intermediate, the best action is P_L (black lines).

6. Conclusion

We propose a POMDP model, to investigate the optimal pricing policy for a monopoly agent, which faces uncertain market demand and maximizes its long-term overall profit. We exam that, by choosing a relatively lower price, the agent can obtain a more accurate observation of the market state. Consequently, such deviating from the myopic monopoly price can bring the agent with a higher profit. The POMDP experiment result shows that, when the agent has enough high (low) belief about the market state, choosing the myopic monopoly prices will be optimal. However, when it is very uncertain about the market state, its best choice is to set a price lower than the myopic monopoly prices.

◊ References ◊

[Agnieszka 03] Agnieszka, B. K.: Sufficiency in Blackwell's theorem., *Mathematical Social Sciences*, Vol. 46, pp. 21–25 (2003)

[Aviv 05] Aviv, Y. and Pazgal, A.: A Partially Observed Markov Decision Process for Dynamic Pricing, *Management Science*, Vol. 51, pp. 1400–1416 (2005)

[Kandori 10] Kandori, M. and Obara., I.: Towards a Belief-Based Theory of Repeated Games with Private Monitoring: An Application of POMDP., *mimeo* (2010)

[Mirman 93] Mirman, L. J., Samuelson, L., and Urbano, A.: Monopoly Experimentation, *International Economic Review*, Vol. 34, pp. 549–563 (1993)

[Russell 09] Russell, S. and Norvig, P.: *Artificial Intelligence: A Modern Approach (3rd Edition)*, Prentice Hall (2009)