



The University of Adelaide
School of Economics

Research Paper No. 2013-22
November 2013

Learning from Inferred Foregone Payoffs

Ralph-C Bayer and Hang Wu



Learning from inferred foregone payoffs

Hang Wu* and Ralph-C Bayer†

Abstract

A player's knowledge of her own actions and the corresponding own payoffs may enable her to infer or form belief about what the payoffs would have been if she had played differently. In studies of low-information game settings, however, players' ex-post inferences and beliefs have been largely ignored by quantitative learning models. For games with large strategy spaces, the omission may seriously weaken the predictive power of a learning model. We propose an extended payoff assessment learning model which explicitly incorporates players' ex-post inferences and beliefs about the foregone payoffs for unplayed strategies. We use the model to explain the pricing and learning behavior observed in a Bertrand market experiment. Maximum likelihood estimation shows that the extended model organizes the data remarkably well at both aggregate level and individual level.

JEL Numbers: C73, D43, D83, L13

Keywords: Learning, Ex-Post Inference, Partial Information, Bertrand Duopoly

*Corresponding author. School of Economics, The University of Adelaide, Adelaide, SA 5005, Australia. E-mail address: hang.wu@adelaide.edu.au. Phone: +61 4359 57226.

†School of Economics, The University of Adelaide, Adelaide, SA 5005, Australia. E-mail address: ralph.bayer@adelaide.edu.au.

1 Introduction

This paper studies learning in games where the players have large strategy spaces. We focus on a partial-information strategic environment: the underlying game structure and the corresponding payoff function are made explicit to all players, but after each round the players can observe only their own payoffs. To explain the learning behavior in such low-information settings we cannot use learning algorithms designed for high-information environments such as weighted fictitious play (Cheung & Friedman 1997) and experience weighted attraction learning (EWA, Camerer & Ho 1999). Moreover, since the players have large strategy spaces, the simple stimulus-response models such as choice reinforcement learning (CRL) of Erev & Roth (1998) and payoff assessment learning (PAL) of Sarin & Vahid (1999) are not directly applicable. This is because the simple stimulus-response models update only the propensities (CRL) or payoff assessments (PAL) of the chosen strategies and ignore an important factor that has potentially a large influence on learning: players' ex-post inferences or beliefs about foregone payoffs of unchosen strategies. Consequently, the aim of this paper is to extend the simple stimulus-response learning models such that learning is not only driven by realized payoffs but also by believed foregone profits of non-played strategies. Consider the following example to illustrate our point. Let us think about a seller who wants to sell one unit of a good on a posted offer market. After posting a price in the morning he did not sell the product. In traditional reinforcement learning models the seller would use this information by reasoning like something like: "don't choose this price again as you do not sell at this price." Clearly the seller can make a few additional inferences, like: "I would not have sold either if I had chosen an even higher price" and "with a lower price I might have sold." Our extended learning model takes these inferences into account. Note that belief-based learning models such as EWA and weighted fictitious play are not applicable here as they require information about the prices set by the competitors.

Our study is based on observations from a partial-information Bertrand duopoly experiment studied by Bayer & Ke (2011). In the experiments, subjects were assigned to fixed roles as sellers and buyers, and randomly re-matched in each period to form market groups. In total subjects played fifteen periods. Over-all 186 subjects participated and in each period 62 markets took place. A market consisted of a standard Bertrand price competition game. Two sellers simultaneously set integer prices between 30 (marginal cost) and 100 (reservation value for the buyers) to compete for a buyer who wishes to buy one unit of the good in question. Buyers initially saw one of the two prices and could without incurring any cost uncover the other price with a simple mouse-click. Then buyers could either choose the seller to buy from or exit the market without buying.¹ The payoff for successful sellers was the difference between price set and cost, while an unsuccessful seller earned zero. The buyers' payoffs were calculated as their reservation value minus the price they paid if they bought from a seller, and were equal to zero if they did not buy. Before the experiments, participants were provided with detailed instructions. After each repetition, the only feedback revealed to the participants were their own payoffs which meant that sellers could infer only if they sold or not but not the exact price charged by the competitor.

The Bertrand-Nash equilibrium for the stage game requires sellers to set prices equal to 30 (marginal cost) or just above. The experimental results, however, differed markedly from the Bertrand-Nash prediction. The main features of the experimental results can be briefly summarized as follows: (1) Prices were persistently dispersed above marginal cost and the Bertrand-Nash prices were rarely observed; (2) The mean and median prices as well as the corresponding price variance decreased with repetition; (3) The sellers tended to increase or keep their prices unchanged after a successful sale and to reduce prices after failing to sell.

¹ In more than 99% of the cases the buyers sampled both prices and bought from the seller offering the lower price.

Based on these observations, it is clear that the simple stimulus-response models (CRL or PAL) are not adequate. To illustrate, consider the case of a seller who has incurred a sales failure and obtained a zero profit. According to the assessment updating rules, for CRL the attractiveness for all alternatives are depreciated by the same proportion, and for PAL the only change is that the payoff assessment for the chosen price is adjusted down. As a result, there would be little change in the seller's choice probabilities. This is obviously inconsistent with the experimental evidence.

We address the problem by incorporating sellers' ex-post inferences and beliefs into the payoff assessment learning model . We assume that the players update the assessments not only for the chosen prices using the realized payoffs, but also for all the unchosen prices using the foregone payoffs. More specifically, knowing their own prices, an unsuccessful seller can easily infer that all higher prices would also have led to failing to sell with a profit of zero. Likewise, a successful seller can infer that all lower prices would also have led to a successful sale and hence the foregone payoffs for lower prices are simply the corresponding markups. Both successful and unsuccessful sellers can never know all foregone profits with certainty though. A successful seller who does not know the price of the competitor cannot be sure what the profit for prices higher than that posted would have been, as she does not know if a sale would have happened. In the same way an unsuccessful buyer does not know the foregone profits for prices lower than the one posted. In order to make an assessment a player requires to have beliefs about the price set by the opponent conditional on the sale result. We propose a parametric form for these beliefs and estimate the parameters straight from the data.

For comparison, we include in our analysis three alternative models of learning: first, the original payoff assessment learning (OPAL, Sarin & Vahid 1999) model which is appropriate for games of small strategy spaces; second, the payoff assessment learning with strategy similarity (LSS, Sarin & Vahid 2004) which is designed for extremely limited information settings

where subjects have no information about the underlying decision problem and can observe only their own payoffs; third, a new learning model with adaptive assessments (LAA), which extends the payoff assessment learning model for full-information settings where players know the underlying game, and are able to observe other players' past strategies. As we will show, the model incorporating players' ex-post inferences and beliefs, as expected, provides the best statistical fit to the behavior observed in the partial-information experiment. Moreover, it organizes the data remarkably well at both aggregate level and individual level.

Our study is closely related to the learning direction theory (cf. Selten & Stoecker 1986 and Selten & Buchta 1999), which is a qualitative approach postulating that subjects ex-post rationally assess in which direction foregone payoffs would have been higher than the payoff actually realized. The theory points out that the realized payoff per se might not be the deciding factor in shaping how people adjust their actions. Instead, subjects adjust their behavior according to their beliefs about how they can increase their payoffs. The theory therefore predicts that when the decision makers can infer from the feedback they received which actions could have led to higher payoffs, they are likely to adjust their behavior in the direction of better performance. Strategic adjustments following this ex-post rationality principle have been shown to exist and persist in many experimental studies (see, e.g., Cason & Friedman 1997; Selten et al. 2005; and Bruttel 2009).

Grosskopf (2003) stressed the importance of combining stimulus-response adjustment and directional learning when modeling learning behavior both quantitatively and qualitatively is the aim. The paper presents two sets of ultimatum game experiments, one in traditional form with one proposer and one responder, and the other with responder competition. The results show that the proposers adjusted their demands following an ex-post rational principle in both experiments. Moreover, such individual directional adjustments were persistent throughout the experiments. Using the results obtained from the experiments, Grosskopf (2003) demonstrated that using reinforcement learning alone and ignoring individual's round-

to-round directional adjustments is likely to be inadequate in explaining learning.

The current paper combines the qualitative directional learning with a quantitative stimulus-response learning model. As we will show, our model has some additional desirable features. It also has the features that actions which a subject believes would have been better in the past are more likely to be played at present, and more importantly, it allows us to estimate the degree of players' ex-post rationality based on the data.

The rest of the paper is structured as follows. Section 2 summarizes the Bertrand market experiments studied in Bayer & Ke (2011). Section 3 presents the general structure of the original payoff assessment learning model and the three extended models. Section 4 constructs a foregone payoff function for the partial-information Bertrand market game based on sellers' ex-post inferences and beliefs. Section 5 uses the experimental data to structurally estimate the parameters of the models. Furthermore, it contains a discussion of the estimation results and compares the performance of the different models. Section 6 concludes.

2 Experiment

The partial-information Bertrand market experiment of Bayer & Ke (2011) was conducted in the Adelaide Laboratory for Experimental Economics (AdLab) at the University of Adelaide. Participants were mainly university students studying for undergraduate or postgraduate degrees in a variety of disciplines. In total 186 participants were recruited. Participants were asked to play fixed roles as sellers and buyers in a Bertrand duopoly market game for thirty periods. The purpose of Bayer & Ke (2011) was to investigate the effects of exogenous cost shocks on market price. So they conducted a two-phase experiment, with 15 periods for each phase. After the first phase an unanticipated cost shock occurred. Since the cost shock effects are beyond the interest of this paper, we focus on the first phase (15 periods) of

the experiment where the cost shocks had not been imposed. Note that the cost shock was unanticipated such that it should not have any impact on play in the first periods. However, subjects knew that the experiment would run for 30 periods.

At the beginning of each period, sellers and buyers were randomly matched to form markets. There were 62 markets and each market consisted of two sellers and a buyer. In each period, sellers with unit cost $c = 30$ simultaneously and independently post a price from the price set $P \equiv \{30, 31, \dots, 100\}$ at which they offer to sell the product. Afterwards, the buyer with unit demand and reservation value $v = 100$ enters the market, learns one price and can click to see the other price without any cost. Then the buyer has to decide either to buy from one of the sellers or to leave without buying. After that a period ends. The payoffs for successful sellers are equal to their prices minus the cost, while the profits of unsuccessful sellers are zero. For a buyer, the payoff is v minus the price she paid if she bought and zero if she did not buy. All the participants were given detailed instructions on the structure of the game and the corresponding payoff functions. After each period, the only information revealed to participants were their own realized payoffs. At the end of the experiments, the payoffs were aggregated over all periods and exchanged for real money at a fixed exchange rate. On average, the participants earned about twenty Australian Dollars for about one hour of their time.

Figure 1 shows the frequency distributions of the posted prices for all periods. Each bar represents the relative frequency for the range of five consecutive prices to be charged as labeled at the ticks.² Prices are significantly dispersed above marginal cost for all periods and the Bertrand-Nash equilibrium prices are rarely observed throughout. Further, the price distributions exhibit remarkable temporal variation, as there is an obvious tendency of prices steering away from prices above 60 toward those below 60.

² The lowest bin contains six prices as the number of strategies was not divisible by five.

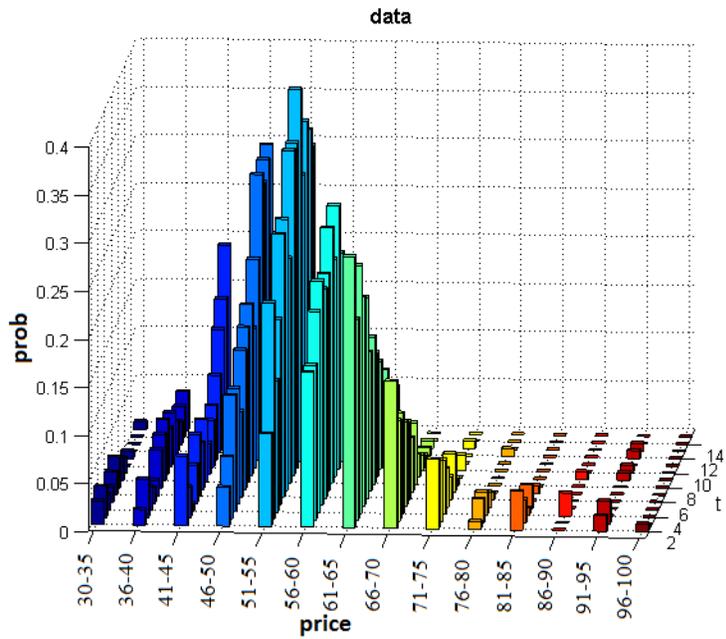


Figure 1: Price distributions by period

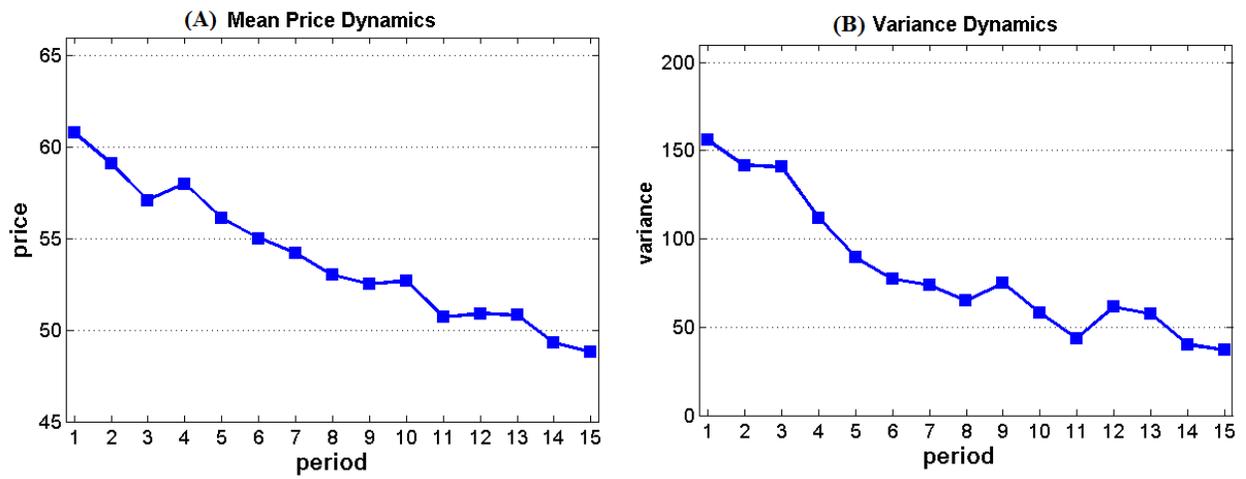


Figure 2: (A) Time series for average price; (B) Time series for price variance

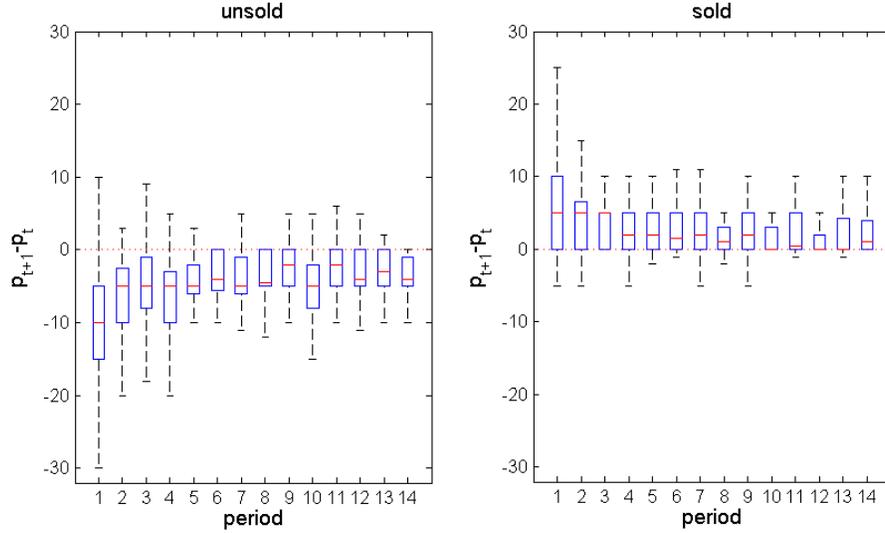


Figure 3: Box plots of price adjustments by sales outcomes

The tendency of price adjustment becomes clearer when plotting the dynamics of the mean and median market prices in panel (A) of Figure 2. Both mean and median prices start off at about 60 and decrease gradually with repetition. After nine periods, the speed of price reductions decreased markedly. Actually, the median price remains at 50 from period 11 to period 14. There is a similar trend for the variance of the prices (see panel B of Figure 2). The variance decreases quickly from 156 in period one to 77 in period six, and then slowly to 37 in the last period.

At the individual level, the most striking feature is that the sellers' price adjustments depend heavily on whether they sold their unit or not in the previous period. Figure 3 shows the interquartile box plots of the sellers' intertemporal price changes in all periods for unsuccessful (left panel) and successful (right panel) sellers, respectively. The red bars in the boxes depict the median price adjustments. As can be seen from the plots, the sellers, who were not able to sell in the period before, usually reduced their prices. Those sellers, who experienced a sale success typically adjust prices upwards or keep their prices unchanged. With other words, the intertemporal price adjustments observed present strong evidence for

the sellers being ex-post rational in the sense of direction learning theory.

3 Payoff Assessment Learning Models

This section presents the models. We will first briefly explain the well established payoff assessment learning (Sarin & Vahid 1999) and learning with strategy similarity (Sarin & Vahid 2004) models. Then we introduce two new extensions of the payoff assessment learning model: learning with adaptive assessments and learning with ex-post inference and beliefs. The three extended models are designed for three different information settings. First, the learning with strategy similarity (LSS) model is designed for extremely limited information settings where players do not know the game and the strategies of others. Second, learning with adaptive assessments (LAA) is developed for full-information settings where players know the game and the strategies played by others. We will use this model as some kind of a robustness test, as it should not fit at all well in our information environment. Finally, the purpose of learning with ex-post inference and beliefs (LEIB), which carries the main idea of this study, is appropriate for explaining learning under partial-information settings where the players know the game structure, but are uninformed about others' strategies.

Let us start with notation. A set $I \equiv \{1, 2, \dots, N\}$ of players engage repeatedly in game Γ . Each player $i \in I$ has a strategy set s^i of J actions: $s^i \equiv \{s^i(1), \dots, s^i(J)\}$. Let $A_t^i \equiv (A_t^i(1), \dots, A_t^i(J))$ denote the vector of player i 's subjective payoff assessments in period t , where the element $A_t^i(j) \in A_t^i$ represents the payoff that player i anticipates to receive by playing strategy $s^i(j) \in s^i$. At each period, a player chooses the strategy that she assesses to have the highest payoff.

3.1 Original Payoff Assessment Learning (OPAL)

The original payoff assessment learning (OPAL) model of Sarin & Vahid (1999) is suitable

for explaining learning behavior in games with small strategy spaces. According to OPAL, after period t player i learns her payoff $\pi_t^i(j)$ from the chosen strategy $s^i(j) \in s^i$, and updates her subjective payoff assessments using the following adaptive rule:

$$A_{t+1}^i(k) = \begin{cases} (1 - \phi)A_t^i(k) + \phi\pi_t^i(j), & \text{if } k = j \\ A_t^i(k), & \text{if } k \neq j \end{cases}. \quad (3.1)$$

where j is the action chosen by seller i in period t . For the chosen strategy j , the new payoff assessment is defined as the weighted sum of its previous payoff assessment and the realized payoff $\pi_t^i(j)$. The parameter ϕ ($0 < \phi < 1$) measures the weight the players assign to the realized payoffs of the new period, and $1 - \phi$ stands for the weight being assigned the payoff assessments of the previous period.³ Accordingly, the payoff assessment for the chosen strategy is adjusted towards the currently realized payoff. The payoff assessments for all unchosen strategies, in contrast, are not updated.

3.2 Learning with Strategy Similarity (LSS) under Limited Information

The OPAL model may lack explanatory power in games with large strategy spaces. To see this, consider an action which has led to a payoff that is much lower than a player's payoff assessment of that action. In this case payoff assessment reduces the likelihood of this strategy to be chosen again, without changing the relative assessments of other strategies. In a game with only a few strategies this is appropriate, as in such games strategies are often unrelated and unordered, which does not allow for any inference about what would have happened if other strategies were played.

³ For simplicity, ϕ will serve the same role in all of the following updating rules as in Equation (3.1).

In games with large strategy spaces just changing the assessment of the actually played strategy does not take into account that in these games some strategies are similar and inferences about likely payoffs from similar strategies are possible. For this reason, Sarin & Vahid (2004) incorporate the idea of strategy similarity into the standard payoff assessment learning model for games with large strategy spaces under limited information. The idea is that, although a player does neither know the game and the corresponding payoff structure, nor the actions taken by other players, she may expect similar actions to yield similar results as long as there is an order of strategies and she knows this. For such environments, Sarin & Vahid (2004) suggest that after each period, observing the realized payoff from the chosen strategy, a player updates her payoff assessments for both the chosen strategy and the strategies that are similar or close to the chosen one. Strategic similarity between two strategies, j and k , can be formalized using the following Bartlett similarity function:⁴

$$f(i, k; d) = \begin{cases} 1 - |j - k|/d & \text{if } |j - k| \leq d \\ 0 & \text{otherwise.} \end{cases} \quad (3.2)$$

The parameter d determines the $d - 1$ unchosen strategies on either side of the chosen strategies whose assessments are updated. We assume that the similarity function does not vary over time. Given the similarity function, for player i who chooses action j and receives payoff $\pi_t^i(j)$ in period t , the law of motion for the payoff assessments of strategy k is:

$$A_{t+1}^i(k) = [1 - \phi f(j, k; d)] A_t^i(k) + \phi f(j, k; d) \pi_t^i(j), \quad \forall k \in J. \quad (3.3)$$

Note that, like the OPAL model, the strategy similarity learning model is stimulus-based

⁴ see Chen & Khoroshilov (2003) and Bayer et al. (2013) for applications of learning with strategy similarity models using the Bartlett similarity function.

and only appropriate for games with limited information where the players do not know the structures and the underlying payoff functions of the games. Next we propose two new extensions of the payoff assessment learning model which are designed for the more realistic environments where the players at least know the games they are playing and the corresponding payoff-determining mechanisms.

3.3 Learning with Adaptive Assessments (LAA) under Full Information

Now consider a setting where players are getting full feedback about the actions taken by their opponents. In addition, the players clearly understand the game Γ they are playing. Knowing the underlying payoff function and the strategies taken by others, players can infer the foregone payoff for each strategy they did not play. Then, the players can not only update the payoff assessments for the chosen strategies, but also update the assessments for all unchosen strategies. Define the information that player i possesses after period t as $H_t^i \equiv (\Gamma, s_t)$, where Γ stands for the game being played and $s_t \equiv \{s_t^1, \dots, s_t^I\}$ represents the vector of all players' strategies in period t . Let $\tilde{\pi}_t^i(k|H_t^i)$ denote the player i 's ex-post assessment of the foregone (expected) payoff for strategy k in period t , conditional on H_t^i . Player i then updates her subjective assessments for all J strategies using the following adaptive updating rule:

$$A_{t+1}^i(k) = (1 - \phi) A_t^i(k) + \phi \tilde{\pi}_t^i(k|H_t^i), \quad \forall k \in J. \quad (3.4)$$

The learning with adaptive assessments (LAA) model of Equation (3.4) is closely related to the weighted fictitious learning models (cf. Cheung & Friedman 1997; Fudenberg & Levine 1995) in which the players play best response to their posterior beliefs formed by taking a weighted average of the empirical distribution of the opponents' strategies in all the past

periods. In contrast to the weighted fictitious play, Equation (3.4) assumes that a player's payoff assessments in period $t + 1$ is formed by taking a weighted average of the payoff assessments in period t and the foregone (expected) payoffs of period t . A disadvantage of the LAA model, like in the weighted fictitious learning models, is that the foregone payoffs are determined purely by the observed actions of others, while a player's own experience does not play a direct role in learning. Note that this kind of model requires more information than what our subjects has. Therefore we will use this model as a robustness test, as it should not fit well.

3.4 Learning with Ex-post Inference and Belief (LEIB) under Partial Information

So far we have extended the original payoff assessment learning model for games with large strategy sets and two extreme informational structures: learning with strategy similarity (LSS) for limited information games and learning with adaptive assessments (LAA) for full-information games. Many studies in the learning literature have investigated the above two information settings. Next we turn to the important but largely ignored intermediate case: games with partial information. The partial-information settings are characterized by players having complete prior information about the game environment, but are observing only their own payoffs after each round of play. In such settings, although no other players' chosen strategies are revealed, a player can infer something about the foregone payoff from other strategies from her own strategy and the corresponding payoff. For example, in our Bertrand duopoly experiment with partial information, if a seller failed to sell at a price, she can infer that her price was higher than that of the competitor, so a higher price would also have yielded a failure to sell. Likewise, in a first-price sealed auction, a bidder can infer that her bid was lower than others' bids if she did not win the auction. Being able

to draw such obvious inferences is likely to greatly influence future behavior and should therefore be included in learning models for environments such as ours. Despite of the fact that the importance of the ex-post inferences and beliefs on learning has been stressed by the works on learning direction theory (cf. Selten & Stoecker 1986 and Selten et al. 2005), it has largely been ignored by quantitative learning models. Thus, we propose a rule that explicitly incorporates players' ex-post inferences and beliefs into the payoff assessment learning model.

Let $\hat{H}_t^i \equiv (\Gamma, \pi_t^i(j))$ denote the information player i receives in period t , where Γ stands for the game being played and $\pi_t^i(j)$ represents player i 's realized payoff from choosing strategy j . Denote $\hat{\pi}_t^i(k|\hat{H}_t^i)$ as the inferred or believed foregone payoff unchosen strategy k would have yielded. The updating of players payoff assessments applies to both the assessments of the chosen strategies (using the realized payoffs) and the assessments of all unchosen strategies (using the foregone payoffs). Formally, the updating is governed by the following rule,

$$A_{t+1}^i(k) = \begin{cases} (1 - \phi)A_t^i(k) + \phi\pi_t^i(j), & \text{if } k = j \\ (1 - \phi)A_t^i(k) + \phi\hat{\pi}_t^i(k|\hat{H}_t^i), & \text{if } k \neq j \end{cases}, \forall k \in J. \quad (3.5)$$

Crucially, we need to find a satisfactory way of modeling how a player forms beliefs about the foregone payoff $\hat{\pi}_t^i$ from her information \hat{H}_t^i . We will discuss this in the next Section.

4 Ex-post inferences and beliefs in the Bertrand experiment

In this section we specify a functional form for the ex-post inferred or believed foregone payoffs of unchosen prices for the experimental Bertrand game. Denote the sellers' strategy set as $P \equiv \{c, c + 1, \dots, v\}$ where $c = 30$ and $v = 100$. Let $p_t^i \in P$ be the price seller i

chooses in period t . Denote $\hat{\pi}_t^i(p|\hat{H}_t^i)$ as the foregone payoff that seller i believes she would have obtained if she had posted price p instead of p_t^i . For the Bertrand game, the law of motion of the payoff assessments in (3.5) can be specified as

$$A_{t+1}^i(p) = \begin{cases} (1 - \phi)A_t^i(p) + \phi\pi_t^i(p), & \text{if } p = p_t^i \\ (1 - \phi)A_t^i(p) + \phi\hat{\pi}(p|\hat{H}_t^i), & \text{if } p \neq p_t^i \end{cases}. \quad (4.1)$$

Next we characterize the foregone payoff function $\hat{\pi}(p|\hat{H}_t^i)$. The first property we impose on $\hat{\pi}_t^i(p|\hat{H}_t^i)$ is that it captures the inference that the consumer buys from the cheaper seller. So if seller i has posted price p_t^i and observed that her profit is $\pi_t^i(p_t^i) = p_t^i - c$, she should infer that her competitor's price was higher than p_t^i . Therefore, all lower prices would also have led to a successful sale and would have yielded profits equal to the markup. The foregone payoff of a higher price, however, is not certain and depends on a seller's subjective assessment of the competing seller's price. Similarly, if seller i failed to sell at p_t^i and received zero profit, she can infer that all prices higher than the chosen one would have also led to zero profits. However, the probability of selling at lower prices is not pinned down.

The information contained in $\hat{H}_t^i \equiv (\Gamma, \pi_t^i(p_t^i))$ can be summarized by a simple indicator function $h(p_t^i)$, with $h(p_t^i) = 1$ indicating a sale at p_t^i with $\pi_t^i(p_t^i) = p_t^i - c$, while $h(p_t^i) = 0$ indicates a sale failure and $\pi_t^i(p_t^i) = 0$. Denote p_t^{-i} as the price posted by seller i 's actual competitor in period t . Then we can define the foregone payoff function for seller i as

$$\hat{\pi}_t^i(p|h(p_t^i)) = (p - c) \cdot \Pr(p < p_t^{-i}|h(p_t^i)). \quad (4.2)$$

Here $\Pr(p < p_t^{-i}|h(p_t^i))$ is seller i 's ex-post believed probability that she would have sold at price p conditioning on the observed sales result $h(p_t^i)$. A well defined form of $\Pr(p < p_t^{-i}|h(p_t^i))$ should also satisfy the following properties:

- P1.** $\Pr(p < p_t^{-i} | h(p_t^i) = 1) = 1$ for $p \leq p_t^i$: if seller i succeeded at p_t^i , then a lower price would also have yielded a sale.
- P2.** $\Pr(p < p_t^{-i} | h(p_t^i) = 0) = 0$ for $p \geq p_t^i$: if seller i failed to sell at p_t^i , then a higher price would also have failed.
- P3.** $\Pr(p < p_t^{-i} | h(p_t^i)) = 0$ for $p \geq v$: a seller would expect the probability of a sale to be zero if a price of v or above is charged.
- P4.** $\Pr(p < p_t^{-i} | h(p_t^i)) = 1$ for $p \leq c$: a seller would expect the probability of a sale to be one if a price of c or below is charged.
- P5.** For all $p \in P$, we have $\partial \Pr(p < p_t^{-i} | h(p_t^i)) / \partial p \leq 0$. That is, the sellers' believed sales probability satisfy the simple restriction that it decreases weakly with a higher price.

Now we need to specify the ex-post expected sales probability for the two cases and price regions where it is not pinned down. A natural candidate probability function which satisfies the above requirements for both circumstances (price above successful p_t and price below unsuccessful p_t), is the triangular distribution with the mode either at p_t after a sale or at c after failing to sell. In this case, a seller's believed probability is linearly decreasing in price p over the range where it is not pinned down.

We take a generalized version of the described probability function, which allows for one degree of freedom and is flexible with respect to the reactivity of the beliefs to price changes. We also allow for a different probability function after successful sales and after failures. We start with the beliefs after successful sales. Specifically, we assume $\Pr(p < p_t^{-i} | h(p_t^i) = 1) = [(v - p)/(v - p_t^i)]^\alpha$ for $p_t^i < p < v$ where $\alpha \in [0, \infty)$. The parameter α captures the degree of reactivity of the beliefs for prices above the chosen price where a sale occurred. Figure 4 provides a graphic example of $\Pr(p < p_t^{-i} | h(p_t^i) = 1) = [(v - p)/(v - p_t^i)]^\alpha$ with $v = 100$

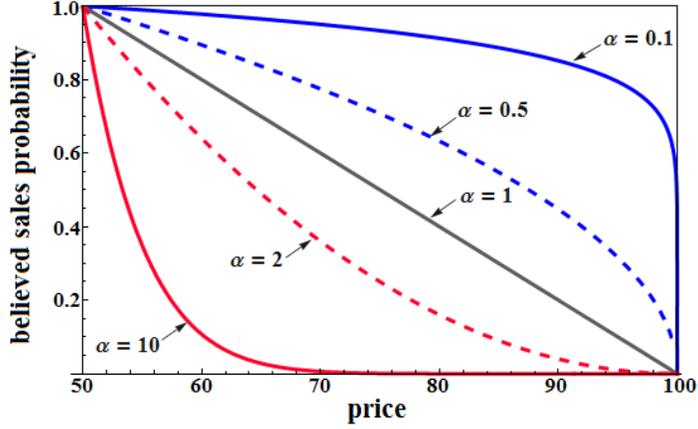


Figure 4: $\Pr(p < p_t^{-i} | s(p_t^i) = 1) = [(v - p)/(v - p_t^i)]^\alpha$ with $p_t^i = 50$ and different α values.

and $p_t^i = 50$ for different values of α . When $\alpha = 0$, the sellers are extremely insensitive and believe that they would have sold at any higher price. On the other hand, when $\alpha \rightarrow \infty$, they are extremely sensitive and believe that any price increase would have led to a failure to sell with probability one.

Likewise, for unsuccessful sellers, the sales probability function is defined as $\Pr(p < p_t^{-i} | h(p_t^i) = 0) = [(p_t^i - p)/(p_t^i - c)]^\beta$ for $c < p < p_t^i$. The parameter $\beta \in [0, \infty)$ serves to measure the sensitivity of an unsuccessful sellers' beliefs to a small price reduction. As β approaches zero, a seller tends to believe that a slight reduction in price would have yielded a successful sale with certainty. In contrast, for $\beta \rightarrow \infty$ a seller expects that any lower price would also have led to zero profit.

To summarize:

Definition. With $\alpha, \beta \in [0, \infty)$, $\forall i \in N$ and $\forall t \in T$, if the sellers' subjective foregone payoff function $\hat{\pi}_t^i(p | h(p_t^i))$ is given by :

$$\hat{\pi}_t^i(p | h(p_t^i) = 1) = \begin{cases} p - c & \text{if } p \leq p_t^i \\ (p - c) \left(\frac{v - p}{v - p_t^i} \right)^\alpha, & \text{if } p > p_t^i \end{cases} \quad (4.3)$$

and

$$\hat{\pi}_t^i(p|h(p_t^i) = 0) = \begin{cases} 0, & \text{if } p \geq p_t^i \\ (p - c) \left(\frac{p_t^i - p}{p_t^i - c}\right)^\beta, & \text{if } p < p_t^i \end{cases}, \quad (4.4)$$

then $\hat{\pi}_t^i(p|h(p_t^i))$ satisfies properties P1-P5.

It is obvious that with α and β being positive constants, the believed sales probabilities for both groups of sellers are weakly decreasing in p . Given the above specifications of $\hat{\pi}_t^i(p|h(p_t^i))$, a smaller α or β implies larger expected foregone profits of $\hat{\pi}_t^i(p|h(p_t^i))$ for prices in the “could-have-been-better” direction compared to the other prices. In other words, we are more likely to observe successful sellers increasing their prices and unsuccessful sellers reducing their prices if the values of α and β are small. To see this, consider the extreme condition with α and β equal to zero. For the successful sellers $\hat{\pi}_t^i(p|h(p_t^i) = 1)$ will be equal to $p - c$ for all prices and thus higher prices will have higher increases in the payoff assessments. This is because higher prices lead to higher margins. For unsuccessful sellers, for $p \geq p_t^i$ the attractions will not be changed, but for $p < p_t^i$, the change in its attraction will be $\hat{\pi}_t^i(p|h(p_t^i) = 0) = p - c$, i.e., the highest possible. Increasing values dampen the directional learning effect, since then the changes in the payoff assessments become smaller.

5 Parameter Estimation

In order to facilitate parameter estimation, we first transform the payoff assessments into choice probabilities using the standard logit probabilistic choice rule. Then we estimate the parameters using the maximum likelihood method. That is, we search for the model

parameter values that maximize the log-likelihood of observing the data our experiment produced with our learning model. Finally we conduct a horse race between our model and the conventional models. This will allow us to test if extending the learning models for partial information by adding inferences about foregone profits where possible, improves our ability to explain observed behavior.

5.1 Probabilistic Choice Rule

If a seller's price choices are not affected by random factors such as mood shocks or noise, then she would always choose the strategies that she assesses to have the highest payoffs. Now suppose that at period t , the seller experiences identically and independently distributed mood shocks $\varepsilon_t = (\varepsilon_t(c), \dots, \varepsilon_t(p))$. Denote the shock-distorted payoff assessments as $\tilde{A}_t^i(p) \equiv A_t^i(p) + \varepsilon_t^i(p)$, $\forall p$. At period t , instead of selecting the price that maximizes $A_t^i(p)$, we assume that the seller selects a price p if

$$\tilde{A}_t^i(p) > \tilde{A}_t^i(m), \forall m \neq p, \forall m \in P. \quad (5.1)$$

We use the logit probabilistic choice model to allow for this kind of noise. With this we implicitly assume that each $\varepsilon_t^i(p)$ follows a Gumbel or type I extreme value distribution.⁵ With this specification, at period t , the probability of seller i choosing price p is

$$f_t^i(p) = \frac{\exp[\mu_t \cdot A_{t-1}^i(p)]}{\sum_{k=c}^v \exp[\mu_t \cdot A_{t-1}^i(k)]} \quad (5.2)$$

⁵ With the Gumbel extreme value distribution (see Akiva & Lerman 1985), the density function for each $\varepsilon_t^i(p)$ is $f(\varepsilon_t^i(p)) = \exp[-\varepsilon_t^i(p) - \exp(-\varepsilon_t^i(p))]$ and the associated cumulative distribution is $F(\varepsilon_t^i(p)) = \exp[-\exp(-\varepsilon_t^i(p))]$.

where

$$\mu_t \equiv \lambda + (t - 1)\gamma. \tag{5.3}$$

Parameter μ_t is used to measure how sensitive the sellers are to the differences in payoff assessments of different prices, or in the mood shock formulations how strong the mood shocks are. When μ_t is positive, the logit model has an intuitively appealing property for modeling decisions: the strategies that being assessed to have higher payoffs are more likely to be selected. As μ_t approaches infinity, the firms will always choose the prices with the highest payoff assessment. On the other hand, when μ_t equals to zero, a seller becomes fully ignorant or confused about the game she is playing and makes assessment-unrelated, purely random choices. In contrast to most other applications of the logit choice rule we allow μ to have a time trend. We use μ_t as $\mu_t \equiv \lambda + (t - 1)\gamma$ for the following reasons.⁶ First, while playing the game repeatedly the players may accumulate experience so they are likely to make more precise choices over time. It is common in experimental applications that the period-specific estimations of μ_t often yield higher values for later periods than for earlier ones (see McKelvey & Palfrey 1995 and Dufwenberg et al. 2007). Second, allowing μ_t to vary over time may help us to disentangle two different learning effects: the learning effect of accumulating experience and making more precise decisions, and the effect of the payoff assessment learning following Equation (4.1).

5.2 Initial payoff assessments

The LAA model relies on a rather strong and unrealistic assumption that the players'

⁶ We also tried a nonlinear version using a harmonic function such as $\mu_t \equiv \lambda + \gamma \sum_1^t \frac{1}{t}$, which assumes that a player's sensitivity toward the payoff assessment differences is an increasing/decreasing function of t at a diminishing rate. It makes little difference to the estimates.

ex-post beliefs of the price distribution are consistent with the prices actually posted by the sellers. Therefore, for the LAA model we simply set the initial payoff assessment of a price to the expected foregone payoff had that price been played against the first period’s price distribution.

Finally we have to choose the initial payoff assessment vector $A_{t=0}^i$ for the individual players in the OPAL, LSS and LEIB models. For games with small strategy sets $A_{t=0}^i$ could be estimated directly through the model (e.g., Camerer & Ho (1999); Erev & Roth 1998). With large strategy spaces as ours, however, estimating the initial assessments introduces massive computational complexity as well as a large number of degrees of freedom, which will almost certainly lead to over-fitting. A simple, widely used alternative (e.g., Chen & Tang 1998; Chen & Khoroshilov 2003) is to assign a uniform initial assessment to all pure strategies that is equal to the average payoff earned by all players in period one. However, in our case, the null hypothesis of a uniform initial choice distribution being compatible with the first period behavior is rejected by a Kolmogorov–Smirnov(K-S) test. Ho et al. (2007) adopt the idea of cognitive hierarchy theory (Camerer et al. 2004) where players are categorized as step k thinkers: step 0 players randomize, step 1 players best respond to step 0, and step k players best responds to the believed (Poisson) distribution of step 0 to step $k - 1$ types. Since experimental research regularly finds that subjects can on average be classified as exhibiting between level one and two, Ho et al. (2007) set the initial payoff assessment of a strategy to its expected payoff from being played against level 1.5 subjects. In our context, however, this approach is also not ideal. Best responses played against a uniform distribution (step 1) lead to a payoff distribution that is nice and smooth, and that can be easily derived. For step 2 though, the price distribution becomes discrete and prices less than 65 (the price that step 1 players choose) yield expected payoffs equal to the markups and prices above 65 yield zero expected payoffs.

As the traditional approaches are not desirable in our case, we use a reverse engineering

approach in setting $A_{t=0}^i$. We are looking for an appropriate initial price distribution to which the best response is the observed modal price in period one. It turns out that a Poisson distribution serves our purpose quite well. The Poisson distribution has the appealing property of having only one parameter κ . Let the markup of a price be $m \equiv p - c \in \{0, 1, \dots, v - c\}$, then the probability function of the Poisson distribution with mean κ is given by

$$f(m) = \frac{\kappa^m e^{-\kappa}}{m!}, \forall m. \quad (5.4)$$

For the first period, the mode price of the data was 60 (so, $p - c = 30$) and it is a best response to a Poisson price distribution with $\kappa = 37$ (so the mean price of the distribution is $c + \kappa = 30 + 37 = 67$). Hence, the initial payoff assessment for price $p = m + c$ is defined as its expected payoff for being played against the Poisson distribution with $\kappa = 67$:

$$A_{t=0}^i(p) = (p - c) \left[1 - \sum_{q=0}^{p-c} \frac{\kappa^q e^{-\kappa}}{q!} \right], \forall p \in P. \quad (5.5)$$

This approach guarantees that the model and the data share the same modal price in the first period. Moreover, our estimation shows that the model using this approach fits the data way better than with all the above mentioned approaches. As we do not have any additional information on heterogeneity of individual subjects that could be usefully exploited, we assign the same described initial assessment vector to all subjects.

5.3 Maximum likelihood estimation

For our estimation we search for the values of the parameters that maximize the log-likelihood of observing the experimental data conditional on a model being correct.⁷ For-

⁷ Matlab codes are available from the corresponding author.

mally, we determine the parameter values for the different models that that maximize

$$\log(L) = \log \left[\prod_{t=2}^{15} \prod_{n=1}^N f_{i,t}(p) \right] = \sum_{t=2}^{15} \sum_{n=1}^N \log \left[\frac{\exp [\mu_t \cdot A_{t-1}^i(p)]}{\sum_{k=c}^v \exp [\mu_t \cdot A_{t-1}^i(k)]} \right] \quad (5.6)$$

where N is the total number of sellers in the experiments and A_t^i are the model-specific assessments.

Table 1: Maximum Likelihood Estimates

Model	parameter estimates					$\log(L)$	BIC
	ϕ	λ	γ	α	β		
LEIB	0.162 (0.02)	0.079 (0.02)	0.034 (0.01)	1.303 (0.07)	0.236 (0.03)	-5674	5693
OPAL	0.000 (0.71)	0.122 (0.94)	-0.002 (0.71)			-6549	6560
LSS ($d=12$)	0.097 (0.03)	0.085 (0.03)	0.023 (0.02)			-6534	6545
LAA	0.085 (0.03)	0.118 (0.03)	0.021 (0.03)			-6037	6048

Note: In parentheses are standard errors derived by numerical differentiation.

Table 1 reports the maximum likelihood estimates. The numbers in the parenthesis are the standard errors obtained by numerical differentiation. The Table also includes the log-likelihood $\log(L)$ and Bayesian Information Criterion (BIC) for the estimated models.⁸ As expected, learning with ex-post inference and beliefs (LEIB) fits the data best. The BIC value of LEIB is 5693, which is much smaller than the the BIC values of the LAA ($BIC = 6048$), LSS ($BIC = 6545$), and OPAL ($BIC = 6560$) models. The original payoff

⁸ BIC is used to compare the relative goodness of fit of different models. It penalizes models with additional parameters. A model with lower BIC is preferred. In our analysis, BIC is defined as $BIC = \frac{k}{2} \ln(N * T) - \log(L)$, where k is the number of parameters, N is the number of sellers, and T is the number of periods considered.

assessment learning model (OPAL) not only shows the worst fit in terms of BIC , it also produces rather unrealistic parameter values of $\hat{\phi} = 0$ and $\hat{\gamma} = -0.002$. We conclude that for the partial-information Bertrand experiments, incorporating the sellers' ex-post inferences and beliefs can significantly improve the explanatory power of the payoff assessment learning model.

Table 2: Maximum likelihood estimates of LEIB with truncated data sets

Data	parameter estimates					$\log(L)$
	ϕ	λ	γ	α	β	
Subset I (84 sellers)	0.153 (0.02)	0.063 (0.02)	0.041 (0.02)	1.254 (0.06)	0.332 (0.04)	-3784
Subset II (40 sellers)	0.187 (0.05)	0.093 (0.05)	0.026 (0.04)	1.392 (0.09)	0.129 (0.05)	-1873

Note: Standard errors in parentheses.

We now turn our attention to the estimates of the LEIB model. In order to examine the robustness of the estimates, we conduct estimations not only for the whole data set ($N = 124$, nine sessions⁹), but also for two subsets with 40 sellers (three sessions) and 84 sellers (six sessions), respectively. The results shows that the estimates are quite robust. As can be seen from Table 2, the estimates derived from the two truncated data sets are similar to these estimated from the whole data set. Actually, the estimates from the whole data set are roughly the average of those from the two truncated data sets. The weighting parameter $\hat{\phi}$ is about 0.16 which implies that the new experience attracts a significant weight and thus affect the price decision substantially. The initial bounded rationality parameter $\hat{\lambda} = 0.079$ is small and but significantly positive. As the game proceeds, the bounded rationality parameter μ_t increases ($\hat{\gamma} = 0.034$), indicating that the precision of player choice increased over time.

⁹ There were between 12 and 18 sellers in a session.

The sensitivity estimates of ex-post beliefs, $\hat{\alpha} = 1.303$ and $\hat{\beta} = 0.236$, turn out to be small enough to account for the directional learning price adjustments observed. For unsuccessful sellers, based on the model, downward adjustment of prices is obvious because only the lower prices may attract positive foregone payoffs. For successful sellers, the estimate $\hat{\alpha} = 1.303$ implies on average, a trend of upward adjustment of the prices, which is less strong than the downward adjustment after failing to sell. Recall that this is exactly what we observe (Figure 3). In what follows we are going to explain what drives the adjustment given our estimates.

Consider an example in which seller i has successfully sold at $p_t^i = 60$. In this case, with $\hat{\alpha} = 1.303$ the sum of $\hat{\pi}_t^i(p|s(60) = 1)$ for prices higher than 60 is $\sum_{p=61}^{100} \hat{\pi}_t^i(p|s(60) = 1) = 746.4$, which is much higher than the sum of $\hat{\pi}_t^i(p|s(60) = 1)$ for the lower prices ($\sum_{p=30}^{59} \hat{\pi}_t^i(p|s(60) = 1) = 435$). This indicates that for seller i the impulse to increase the price is higher than the impulse to reduce the price. This is the case for all successful prices that are below 70. Since we rarely observe sales at prices greater than 70 in the experiments, we can safely say that $\hat{\alpha} = 1.303$ conforms to the observed directional learning price adjustments by successful sellers.

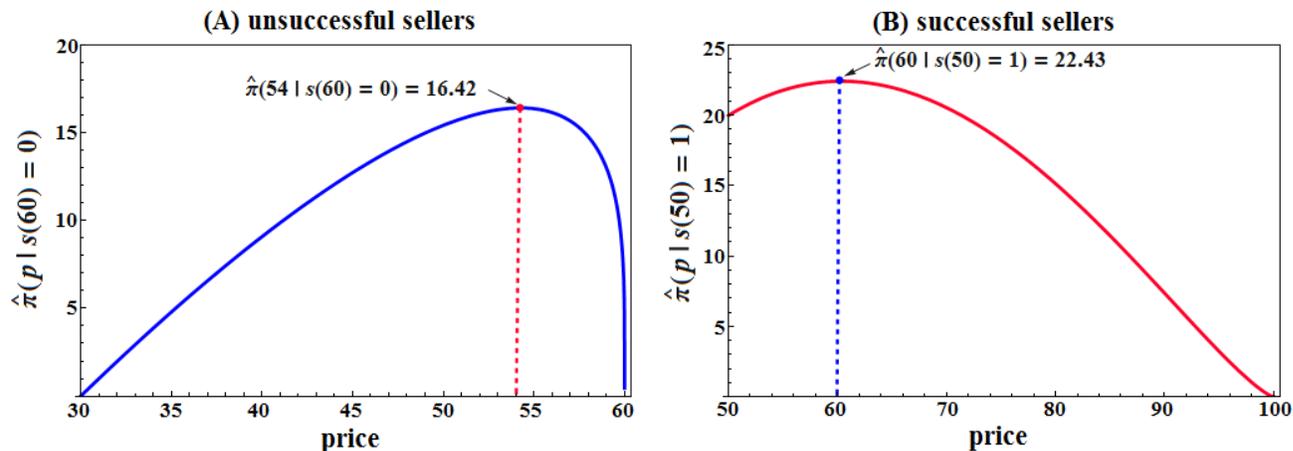


Figure 5: Examples of foregone payoffs at $\hat{\alpha} = 1.303$ and $\hat{\beta} = 0.236$

To illustrate how strong the sellers' price adjustments are for our estimates, we shown in Figure 5 two examples for the expected foregone profit implied by our model. Panel (A) shows the expected foregone payoffs for prices lower than 60 after incurring a sales failure by at $p = 60$ (60 was the mode of prices resulting in a failure to sell). Panel (B) plots the expected foregone payoffs for prices greater than 50 after a successful sale at $p = 50$ (50 was the mode of prices resulting in a sale). The plots show that, on average, both unsuccessful and successful sellers believed that the prices they chose were quite close to the optimal choices. For the unsuccessful sellers, reducing the price marginally to $p = 58$ would result in a foregone payoff of 14.78, which is within 2 units of the best expected foregone payoff 16.42 at $p = 54$. Similarly, on average a successful seller who has succeeded at $p = 50$ believes according to our estimates that the highest expected foregone payoff was 22.43 at $p = 60$, which is only slightly greater than the realized payoff.

5.4 How well does LEIB fit the data?

Next we evaluate the fit of the proposed LEIB model in greater detail. Figure 6 shows the price frequencies predicted by the model. By comparing Figure 6 to Figure 1, we can see that, overall, the model organizes the data remarkably well. The model precisely predicts the tendencies of relative frequency changes for most of the price clusters over time. The largest prediction error of the model is that it overpredicts the frequencies in the price range 46 to 50 and underpredicts those in the range 51 to 55. This discrepancy becomes more pronounced in later periods.

Panel (A) of Figure 7 shows the time series of the mean price estimated by the model, together with the real dynamics of average prices. Panel (B) shows the corresponding dynamics of price variances. We can see that the model does rather well in tracking both the means and variances of the prices.

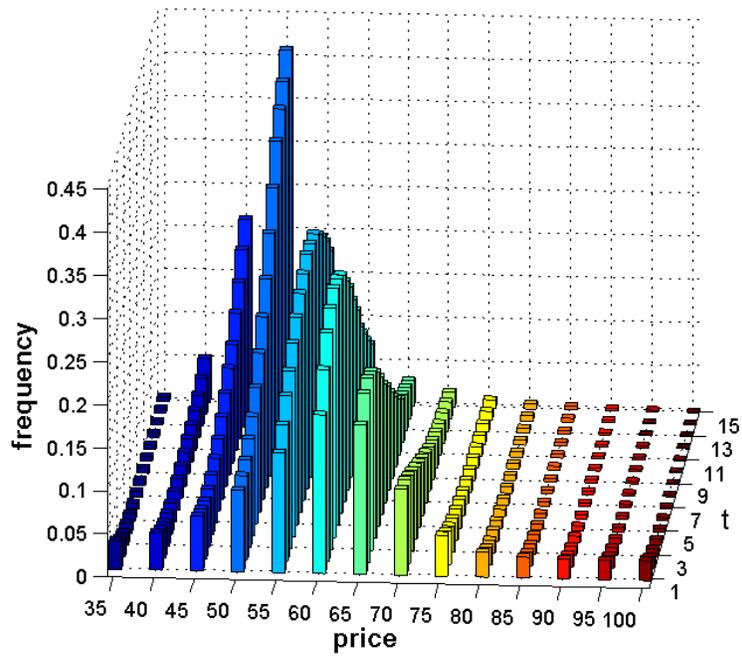


Figure 6: Price Dispersion: Data observation and Model Prediction

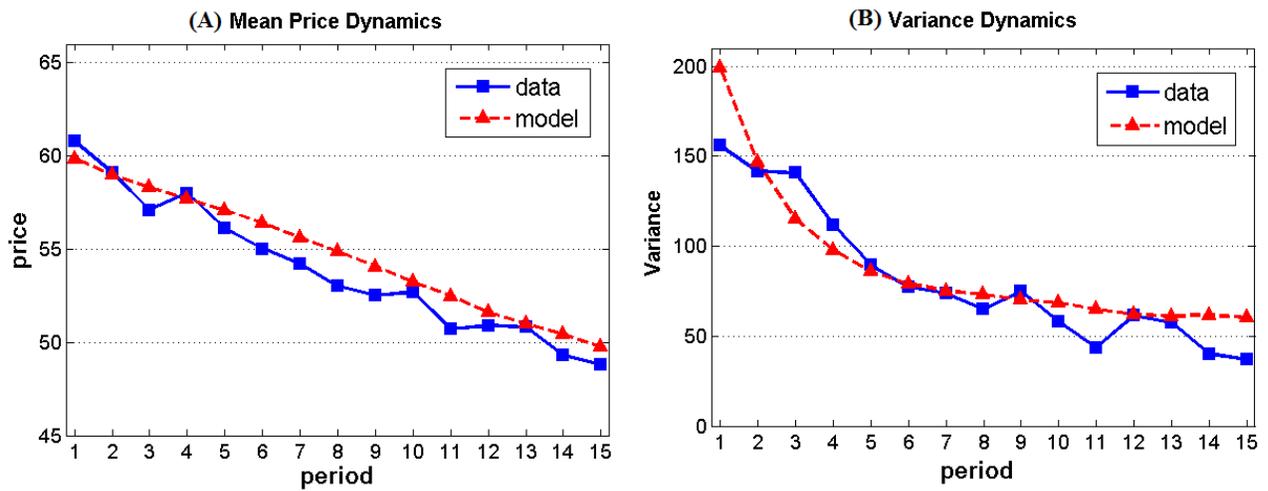


Figure 7: Mean (A) and Variance (B) dynamics: Data Observation and Model Prediction

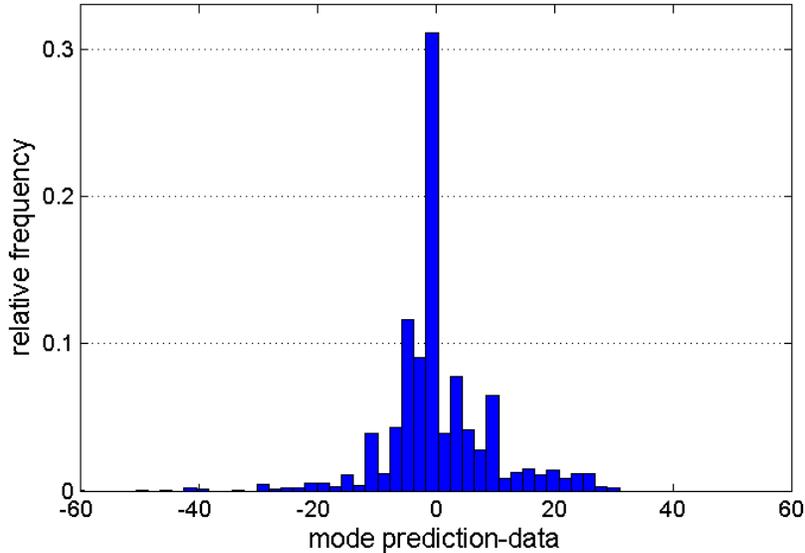


Figure 8: Histogram for differences between predicted mode prices and posted prices

The payoff assessment learning models (except the LAA model) have advantages over the belief-learning models for allowing players' decisions to depend directly on their own past choices and payoffs. The LEIB model enables us to predict players' individual strategies for all but the first periods and evaluate the model's goodness of fit at an individual level.¹⁰ To assess how well the model is tracking individual price choices, we compare for all sellers the prices they actually posted and the mode prices predicted by the model. There are 1860 ($N \times T$) pairs of posted prices and corresponding mode predictions. We take the differences of each pair (mode prediction minus observed price) and plot the histogram of the frequencies for the pooled differences. Figure 8 shows the results. Overall, the mode predictions of the model fit the posted prices excellently. The mean of the differences is 0.52 and the median

¹⁰ Wilcox (2006) shows that, if the comparison between reinforcement learning models and belief learning models are based only on the overall goodness of fit, then the results will be biased in favor of the reinforcement learning models. This is because the reinforcement learning models manage to carry idiosyncratic information of players into the estimation, while the belief learning models cannot.

is zero. For about 41% of the cases, the predicted modes are within 2 units of the actually posted prices. Therefore, we feel confident to conclude that our model also organizes the experimental data well at an individual level.

6 Conclusion

This paper contributes to the learning literature in two important ways. Firstly, it shows that the simple stimulus-response learning models may not be adequate in predicting the behavior in environments where the players' strategy sets are large. Like many other studies on learning (e.g., Camerer & Ho 1999; Sarin & Vahid 2004; Selten et al. 2005), the paper stresses the importance of incorporating foregone payoffs of unplayed strategies in modeling learning in large-strategy-space games. Secondly, the paper illustrates using the payoff assessment learning model as the point of departure (Sarin & Vahid 1999) how models can be extended such that foregone payoffs in games with partial information and large strategy spaces can be taken into account. Our approach emphasizes the importance of incorporating players' ex-post inference and beliefs in partial-information settings, where the players know the strategic environment but do not observe past actions of others. The superior performance of our model in the specific game we tested it in, suggests that applying it to other contexts where the strategic environment is known but feedback is sparse, could yield considerable predictive improvements. Whether this suspicion is true remains an open question, which we believe is worth investigating.

Acknowledgements

We would like to acknowledge the financial support of the Australian Research Council under DP120101831. We thank Thorsten Chmura, Dmitriy Kvasov, Changxia Ke, Paul Pezanis-Christou, and Jason Shachat for helpful comments. Mickey Chan's invaluable research assistance is gratefully acknowledged.

References

- Akiva, M. E. B. & Lerman, S. R. (1985), *Discrete choice analysis: theory and application to predict travel demand*, Vol. 9, MIT press.
- Bayer, R. C. & Ke, C. (2011), 'Are rockets and feathers caused by search or informational frictions?', *School of Economics Working Papers, University of Adelaide* .
- Bayer, R.-C., Renner, E. & Sausgruber, R. (2013), 'Confusion and learning in the voluntary contributions game', *Experimental Economics* pp. 1–19.
- Bruttel, L. V. (2009), 'Group dynamics in experimental studies: the bertrand paradox revisited', *Journal of Economic Behavior & Organization* **69**(1), 51–63.
- Camerer, C. F. & Ho, T. H. (1999), 'Experience-weighted attraction learning in normal form games', *Econometrica* **67**(4), 827–874.
- Camerer, C., Ho, T. & Chong, J. (2004), 'A cognitive hierarchy model of games', *The Quarterly Journal of Economics* **119**(3), 861–898.
- Cason, T. N. & Friedman, D. (1997), 'Price formation in single call markets', *Econometrica: Journal of the Econometric Society* pp. 311–345.

- Chen, Y. & Khoroshilov, Y. (2003), 'Learning under limited information', *Games and Economic Behavior* **44**(1), 1–25.
- Chen, Y. & Tang, F. (1998), 'Learning and incentive-compatible mechanisms for public goods provision: An experimental study', *Journal of Political Economy* **106**(3), 633–662.
- Cheung, Y. & Friedman, D. (1997), 'Individual learning in normal form games: Some laboratory results', *Games and Economic Behavior* **19**(1), 46–76.
- Dufwenberg, M., Gneezy, U., Goeree, J. & Nagel, R. (2007), 'Price floors and competition', *Economic Theory* **33**(1), 211–224.
- Erev, I. & Roth, A. (1998), 'Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria', *American economic review* pp. 848–881.
- Fudenberg, D. & Levine, D. K. (1995), 'Consistency and cautious fictitious play', *Journal of Economic Dynamics and Control* **19**(5), 1065–1089.
- Grosskopf, B. (2003), 'Reinforcement and directional learning in the ultimatum game with responder competition', *Experimental Economics* **6**(2), 141–158.
- Ho, T., Camerer, C. & Chong, J. (2007), 'Self-tuning experience weighted attraction learning in games', *Journal of Economic Theory* **133**(1), 177–198.
- McKelvey, R. D. & Palfrey, T. R. (1995), 'Quantal response equilibria for normal form games', *Games and Economic Behavior* **10**(1), 6–38.
- Sarin, R. & Vahid, F. (1999), 'Payoff assessments without probabilities: a simple dynamic model of choice', *Games and Economic Behavior* **28**(2), 294–309.

- Sarin, R. & Vahid, F. (2004), 'Strategy similarity and coordination', *The Economic Journal* **114**(497), 506–527.
- Selten, R., Abbink, K. & Cox, R. (2005), 'Learning direction theory and the winner's curse', *Experimental Economics* **8**(1), 5–20.
- Selten, R. & Buchta, J. (1999), 'Experimental sealed bid first price auctions with directly observed bid functions', *Games and human behavior: Essays in honor of Amnon Rapoport* pp. 101–116.
- Selten, R. & Stoecker, R. (1986), 'End behavior in sequences of finite prisoner's dilemma supergames a learning theory approach', *Journal of Economic Behavior & Organization* **7**(1), 47–70.
- Wilcox, N. (2006), 'Theories of learning in games and heterogeneity bias', *Econometrica* **74**(5), 1271–1292.

Appendix

A Experimental Instructions

Thank you for participating in this experiment. Please read the instructions carefully. This is important, as understanding the instructions is crucial for earning money. Please note that you are not allowed to communicate with other participants during the experiment. If you do not obey to this rule we may exclude you from the experiment. If you have any questions, please raise your hand. We will come to answer your questions individually. The currency in this game is called E-Dollars. At the end of the game we will convert the E-Dollars you have earned into real money. The exchange rate is $100 \text{ E-Dollars} = 2 \text{ Australian Dollars}$. You will also be paid an appropriate base payment for your participation.

A.1 Your task

You will play a market game in this experiment. There are two types of players in the game: sellers and the buyers. You will be randomly assigned your role (either as a seller or a buyer) at the beginning of the experiment. Your role will be announced to you and fixed for the whole duration of the experiment. In each round we will randomly pair two sellers with one buyer. Each of the two sellers wants to sell one unit of a good which will cost the seller $MC = 30 \text{ E-Dollars}$ to produce and sell. The buyer can buy one unit of the good, which he values at $V = 100 \text{ E-Dollars}$. The profits for seller will be the selling price minus the cost MC if a sale takes place and zero otherwise. The profit of the buyer will be the valuation V minus the selling price, if a purchase takes place and zero otherwise. The higher your profit is the higher is the Australian dollars you could gain from this experiment.

A.2 The trading environment

The game is composed of two decision-making stages: the sellers' stage and the buyer's stage. In the sellers' stage, the two sellers in the same group simultaneously set the prices in E-Dollars at which they want to sell. After both sellers have entered their selling prices, the buyers enter the game. In the buyer's stage, the buyer will be randomly given one out of the two prices offered by the two sellers in the group. Then the buyer can decide if he a) wants to immediately accept the offer, or b) to check the price of the other seller, or c) to exit the market. In case the buyer sees both prices, he can either choose to accept one of the two offers or to exit the market. Exit brings zero profit to the buyer and also to each of the sellers.

A.3 Your Profit

The round profits will depend on the prices set by the sellers and the decision of the buyers. Depending on the type (seller or buyer) the profits will be given as follows:

a) Sellers:

- *Price (P) – cost (MC = 30, initially)* if the unit was traded
- *zero* if the unit was not traded

Note that the production cost MC is only incurred if the unit is actually traded. Furthermore, the production cost is initially fixed at 30 but may change during the game (see below).

b) **Buyers:**

- *Valuation* ($V=100$) – *Price* (P) if the unit was purchased
- *zero* if the unit was not purchased

A.4 Summary

In this market game you will be a buyer or a seller. Your role will be fixed through the entire session. There are always two sellers and one buyer in a trading group. However, the members of each group will be randomly replaced in each round. If you are a seller you want to sell a unit of a good, if you are a buyer you can buy a unit of the good from the seller you choose.

Again, please make sure that you understand the instructions clearly, as this is crucial for your earnings in this experiment. If you have any questions please raise your hand. We will come and answer your question. Once you are ready, we will play a trial period, which is of no consequence for your payoff. After that you can raise your hand again and ask clarifying questions before we start with the 30 rounds (which determine your earnings).