The University of Adelaide
School of Economics

Research Paper No. 2013-21
November 2013

Do We Learn from Our Own Experience or from Observing Others?

Ralph-C Bayer and Hang Wu

Do we learn from our own experience or from observing others?

Ralph-C Bayer* and Hang Wu†

Abstract

Learning in real life is based on different processes. Humans learn to a certain extent from their own experience but also learn by observing what non directly related others have done. In this paper we propose a generalized payoff assessment learning (GPAL) model which enables us to evaluate the relative influences of these two different models of learning. We apply GPAL to a homogeneous good Bertrand duopoly experiment with random matching and population pricing information. The model explains the observed pricing and learning behavior at least as well and often better than learning models from the literature but has the advantage that the relative influence of the learning models can be estimated. We find that the own experience overwhelmingly dominates learning, despite the useful information about behavior of potential future opponents contained in the population price distribution.

JEL Numbers: C91, D83, L13

Keywords: Learning, Information, Bertrand Duopoly, Experiment

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

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

1 Introduction

For many experimental studies where people engage repeatedly in a stage game, it is well known that the availability of information about other players' past actions can have substantial influence on an agent's strategic adjustment (e.g., Abbink et al. 2004; Abramson et al. 2005; Armantier 2004; Huck et al. 2000; and Mookherjee & Sopher 1994). In strategic settings where there are a large number of players and in each period a player plays against a randomly matched opponent, the information about the choice made by the matched player and the observation of the strategies of population members can have differing impact on subjects' learning. In this paper, we propose a generalized payoff assessment learning (GPAL) model, which can be used to evaluate the relative impact of the the opponent's behaviour and the population distribution over strategies. In GPAL subjects choose the strategy that they assess subjectively to lead to the highest payoff. Initial payoff assessments are assumed to lie outside the model. After a period, the payoff assessment for a certain strategy is updated as a convex combination of its previous payoff assessment, the foregone expected payoff of the strategy being played against the distribution of the population strategies, and the foregone payoff the subject would have obtained had she played the strategy against the action chosen by her matched opponent. The choice probability for a strategy is then determined by the updated payoff assessment of this strategy relative to the updated assessments of all other strategies.

We apply the model to explain subjects' pricing and learning behavior in a homogeneous good Bertrand duopoly experiment with random rematching. In the experiment, after each period the following set of information is revealed to a subject: the realized profit, the price set by the matched opponent, and the prices posted by all other sellers across all duopolies in the session. The two different types of information were presented to a subject separately using different ways: the matched opponent's price was given as a number, and the population members' prices were shown using markers in a diagram. The results of the model parameter estimation clearly demonstrate that the opponent's price is the main driver of learning. So the price that actually influenced the experienced payoff is strongly influencing how a subject adapts her strategy, while the much richer information on the price distribution does hardly matter. This is quite surprising as the random matching protocol means that the session price distribution contains the information needed to form an expectation about the price set by the next opponent.

Our learning model's main advantage over existing models in the literature is that it makes it possible to estimate the strength of the impact of the two components. Other models cannot distinguish between the impact of others behavior that was actually payoff relevant and others behavior that might allow the formation of beliefs about the action of future opponents. In order to show that this advantage of GPAL does not come at the expense of its explanatory power, we compare its fit with two well-established payoff-based models, experience weighted attraction learning (EWA; Camerer & Ho 1999) and weighted ficitious play (WFP; Cheung & Friedman 1997). GPAL manages to achieve a similar predictive power as EWA and proved to be much better than WFP at explaining observed behavior at both aggregate and individual levels.

The remainder of the paper is structured as follows. The next Section briefly describes the Bertrand duopoly experiment and lays out the most prominent features of the observed pricing behavior. Section 3 presents the generalized payoff assessment learning model. It also includes a brief description of experience weighted attraction learning and weighted fictitious play. In Section 4 we use the experimental data to structurally estimate the parameters of the different learning models. Then we explain the estimation results, compare the explanatory power of the models, and evaluate the goodness of fit for our model. We close with some concluding remarks.

2 Experiment

The experimental sessions were conducted at AdLab, the Adelaide University Laboratory for Experimental Economics. In total, 120 university students from various disciplines were recruited. They repeatedly engaged in a homogeneous good Bertrand duopoly game for fifteen periods.[1] At the beginning of the experiments the participants were randomly assigned roles as sellers or buyers at a fixed ratio of two to one. The roles were kept fixed throughout the experiment. In each period, duopoly markets were formed using random matching. Each market consisted of two sellers and a buyer and all subjects were assigned to participate in a market. In each market, two sellers simultaneously and independently set integer prices that could range from $\$E30$ (marginal production cost c for the sellers) to $\$E100$ (reservation price v for the buyer).[2] Afterwards, the buyer observes both prices costlessly and then chooses either to buy from one of the sellers or to leave without buying.[3] In each stage the payoff for a seller who managed to sell was her price less the cost. An unsuccessful seller earned a profit of zero. The buyers' payoffs were defined as their reservation value minus the price they paid if they bought and zero if they did not buy. After a period, each seller was given the following set of information: her realized profit, the price posted by her matched competitor, and the prices selected by all sellers in the session.[4] The two different types of

[1] The experiments were designed to investigate the responses of sellers' pricing behavior to various exogenous cost shocks. The shocks occured at the beginning of the 16th period. The cost shock was unanticipated such that it cannot have any impact on play in the first-phase periods. However, subjects knew that the experiment would run for 30 periods. Since the effects of exogenous cost shocks are not of interests for this study, we will focus on the first 15 periods of play.

[2] The currency was Experimental Dollars. In what follows we drop the currency symbols.

[3] In more than 99% cases the buyers bought from the seller with the lower price.

[4] There were between 12 and 18 sellers in a session.

information about others choices were presented to a subject separately in different ways: the matched opponent's price was given as a number, and the population members prices were shown as markers in a diagram. The random matching mechanism, the market rules and the payoff structures were thoroughly explained to subjects in written instructions before the experiment. At the end of their session, the subjects' accumulated payoffs were converted into Australian Dollars at a given exchange rate. On average the participants earned around 20 Australian Dollars for about one hour of their time.

In a Bertrand-Nash equilibrium of the game no seller ever sets a price that is higher than 32.[5] Observed experimental behavior substantively deviates from the Bertrand-Nash prediction. Prices posted by the sellers exhibit significant dispersion over the strategy space for all periods. Figure 1 shows the time series for the interquartile ranges of the prices (boxes) as well as the average prices (black marks). Red lines in the boxes represent the median price levels. We observe that the central 50 percent of prices exhibit substantial spreads for all periods. In terms of price dynamics, the average price started off at 59.8, with an interquartile range of 50 to 70 and a median price at 60. Then the prices declined quickly as the experiment proceeded. In period 15, the average price was 40.5, and median price was 38, with an interqartile range of 34 to 42.

Another striking feature of the prices is that the sellers' price adjustments are depending heavily on their sale results in the previous period. More specifically, the sellers' intertemporal price adjustments present strong evidence for the sellers being ex-post rational in the sence of directional learning theory (cf. Selten & Stoecker 1986, Selten & Buchta 1999), especially for the first few periods. Figure 2 shows the interquartile box plots of the sellers' intertemporal price changes in all periods, separately for the unsuccessful and the successful

[5] With continuous strategy space, the only Bertrand-Nash equilibrium is for all sellers setting price at the marginal cost.

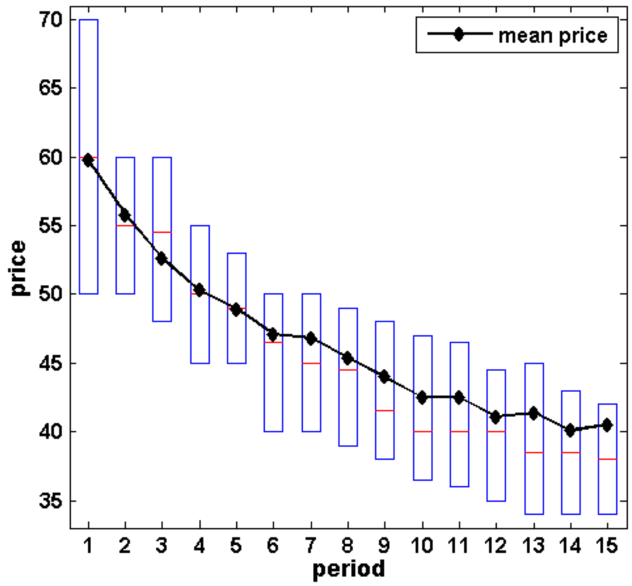

Figure 1: Time Series for Interquantile Ranges of Prices and Mean Prices

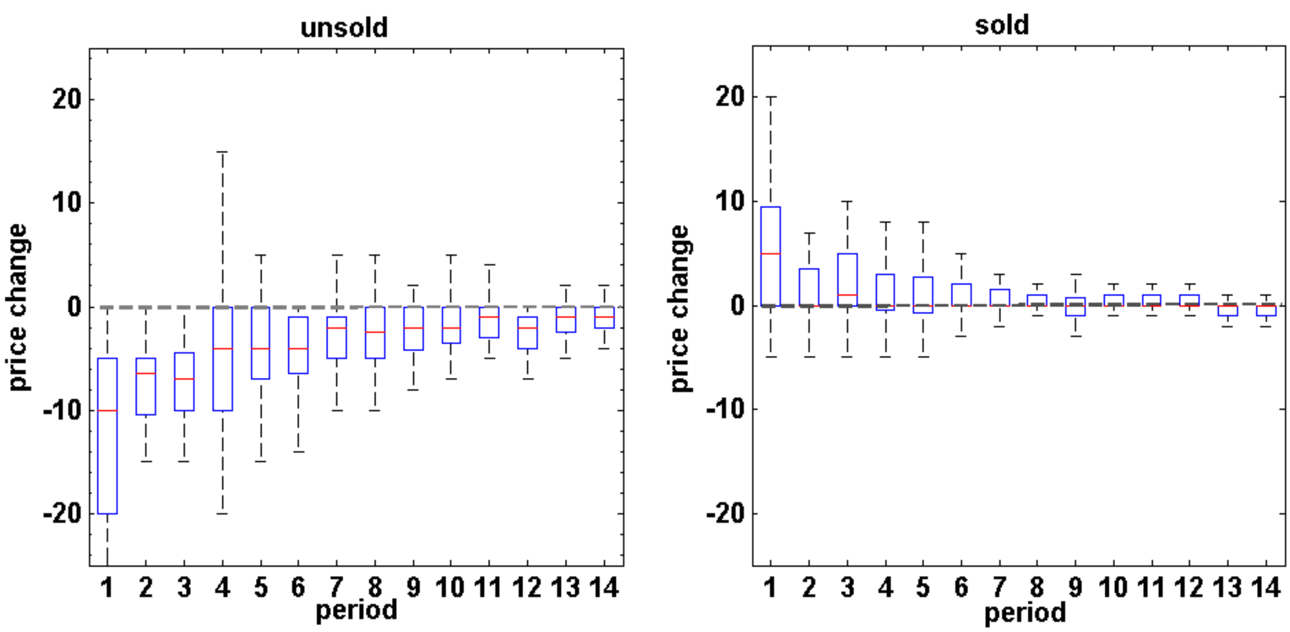

Figure 2: Time Series for Interquartile Plot of Price Adjustments by Sales Results

sellers. As can be seen from the plots, the sellers, who were not able to sell in the period before, usually reduced their prices. In contrast, in the first seven periods of the experiments those who experienced a sale success typically adjust prices upwards. From period eight to period fifteen, successful sellers' direction of price adjustment became less obvious. The observation that price adjustments depended on the selling success implies that the first component of learning (i.e., the experienced consequences of the direct opponent's price) clearly influences learning. We cannot say anything yet about the influence of information about the session price distribution.

3 Learning Models

In this Section we will first explain our proposed generalized payoff assessment learning (GPAL) model. Then the experience weighted attraction learning (EWA; Camerer & Ho 1999) model and the weighted fictitious play (WFP; Cheung & Friedman 1997) model are briefly discussed. We start with some notation. A set $I = \{1, 2, \cdots, n\}$ of sellers play repeatedly in the standard Bertrand duopoly game along the time horizon $T \equiv \{1, 2, \cdots, 15\}$. In each period $t \in T$, a seller is competing in price against a randomly determined competitor. Denote the price set as $\mathrm{P} \equiv \{c, c+1, \cdots, v\}$. The payoff for seller i is $p_i - c$ if her price is lower than that of her opponent, is $\frac{p_i - c}{2}$ in the case of a tie, and zero otherwise.[6] Denote the cumulative probability distribution function for price $p \in \mathrm{P}$ in period t as $F_t(p)$ and the associated density function as $f_t(p)$. After period t, a seller has access to the following set of information: her realized profit, her actual competitor's price, and the prices set by all N sellers.

[6] We assume a 50:50 tie-breaking rule for simplicity. Given that we have human buyers other rules are also possible.

3.1 Generalized Payoff Assessment Learning (GPAL)

GPAL extends the payoff assessment learning (PAL) model of Sarin & Vahid (1999), which assumes that subjects choose the strategy that they assess subjectively to lead to the highest payoff and update the payoff assessments upon the observation of new information. Initial assessments are assumed to be given and lie outside the model. As in most of reinforcement learning models (e.g., Roth & Erev 1995), the updating only occurs for the chosen strategy, so the only information that matters for a player's learning curve is that about the payoffs obtained from the chosen strategies. The PAL model works well only for games with small strategy sets or for limited information settings, where the players do not know the underlying game and observe only their own payoffs. In our context, the sellers had large strategy sets and were given more feedbacks than just their own payoffs. For this reason, the original PAL is not adequate for our purpose. GPAL extends the Sarin & Vahid (1999) model in two important ways so that it becomes applicable to our setting. First, GPAL takes into account the foregone payoffs for all unselected prices and updates the payoff assessments for those prices using foregone payoffs. Second, in GPAL, the payoff assessment for a strategy is defined as the weighted sum of the payoff assessments obtained using different factors that may affect learning. An important advantage of GPAL over other learning models is that GPAL allows us to evaluate the relative importance of different types of information in driving players learning behavior by including forgone profits in the actually played Bertrand game but also from a hypothetical game against the population price distribution..

According to GPAL, in each period $t \in T$, a seller aims to choose the price in P that she assesses to yield the highest payoff. Let $A_t^i(p)$ denote seller i's payoff assessment of choosing price p in period t. After having observed the information, the law of motion for seller i's payoff assessment of price p is governed by:

$$A_{t+1}^i(p) = \alpha A_t^i(p) + \beta U_t^i(p) + (1-\alpha-\beta)V_t^i(p);\ \forall i \in I,\ \forall p \in \mathrm{P}. \tag{3.1}$$

Here $A_t^i(p)$ is seller i's payoff assessment for price p brought forward from the previous period, which we refer to as the ***inertia factor*** of learning. The second component in Equation 3.1 $U_t^i(p)$, called the ***belief-learning factor***, is seller i's expected payoff if she had chosen p in response to the realized market price distribution. We can calculate the belief-learning factor as $U_t^i(p) \equiv (p-c)[1-F_t(p)+\frac{1}{2}f_t(p)]$. The remaining component $V_t^i(p)$ is the ***experiential-learning factor***. More specifically, $V_t^i(p)$ is the (forgone) payoff to seller i if she had posted p while facing the opponent she actually faced in the period before. Let p_t^j denote seller i's actual opponent's price in period t. Then the experiential-learning component $V_t^i(p)$ can be expressed as:

$$V_t^i(p) = \begin{cases} p-c & \text{if } p < p_t^j \\ \frac{1}{2}(p-c) & \text{if } p = p_t^j \\ 0 & \text{otherwise} \end{cases}. \tag{3.2}$$

By incorporating the experiential-learning factor, GPAL is able to account for the observed ex-post rational price adjustments. To see this, consider an expample where seller i posted p_t^i in period t and encountered seller j, whose price was p_t^j. Note that $A_t^i(p)$ is the payoff assessment for choosing price p, so we must have $0 \le A_t^i(p) \le p-c$. If $p_t^i < p_t^j$, seller i successfully sells and for all $p < p_t^j$ we have $V_t^i(p) = p-c$. Accordingly, the payoff assessments of the prices that are between p_t^i and p_t^j will be increased by more than those of the lower prices. A direct result is that we are more likely to observe prices between p_t^i and p_t^j. On the other hand, if $p_t^i > p_t^j$, seller i fails to sell. In this case, $V_t^i(p)$ is zero for $p > p_t^j$ and is $p-c$ for $p < p_t^j$. Consequently, $V_t^i(p)$ would have a positive (negative) effect on the choice probabilities for $p < p_t^j$ $(p > p_t^j)$. In other words, the experiential learning factor nicely captures the observed conditioning of adjustments on sale success.

The parameters $\alpha \in [0,1]$ and $\beta \in [0,1]$ measure the weights that the sellers allocate to the status quo $A_t^i(p)$ and belief learning $U_t^i(p)$. The term $1-\alpha-\beta$ is the weight being allocated to experiental learning $V_t^i(p)$. Two special cases of GPAL are noteworthy. First, when $\alpha+\beta = 1$ the model becomes a risk-neutral adaptive assessment model, where a player's payoff assessment of a price is equal to the weighted sum of all past period's (ex-post) expected payoffs playing against a random opponent from the pool of sellers. In this case, the learning process of a seller is driven only by the observation of the distribution of prices in the population. This has a flavor of weighted fictions play (Cheung & Friedman 1997) while still differing in one aspect: weighted fictitious play assumes that players play in response to the weighted sum of all past periods' distributions of opponent strategies. Second, with $\beta = 0$ a seller's learning behavior depends only on the prices set by the opponent she was matched with. The observation of the prices set by her potential future opponents does not matter then. Then the GPAL model becomes quite similar to the Sarin & Vahid (1999) PAL model. The only difference between the two is that GPAL with $\beta = 0$ uses the foregone payoffs of the unchosen strategies defined in (3.2) to update the payoff assessments of the unchosen strategies, while the Sarin & Vahid (1999) PAL model keeps the assessments of unchosen strategies unchanged.

3.2 Two alternative learning models

Next we describe two well-established alternative learning models: Experience Weighted Attraction (*EWA;* Camerer & Ho 1999) and Weighted Fictitous Play (*WFP*; Cheung & Friedman 1997). In Section 4, together with GPAL, we will estimate these models using the experimental data and evaluate the relative effectiveness of the different models in explaining observed behavior. A more important reason for our inclusion of the two alternative models is that, in modeling the updating rules of sellers, each of the two models uses only one different category of the revealed information. EWA only considers the informaton about a player's actually matched opponent's price. In contrast, in WFP all potential opponents choices are weighted equally in formulating a player's belief of her opponent's future strategy. In terms of the usage of information, EWA is identical to one special case of GPAL ($\beta = 0$), and WFP is in line with another special case of GPAL ($\alpha + \beta = 1$). Therefore, inculding EWA and WFP also serves our objective of evaluating the relative impacts of the two different types of information on learning behavior.

Experience Weighted Attraction (EWA) learning (Camerer & Ho 1999) is an attraction updating model of learning. It uses two updating variables to govern the learning process. The "observation-equivalents" variable, $N(t)$, with initial experience weight $N(1)$, is a measure of a player's past experience of playing the game. On the other hand, $A_t^i(p)$, with initial value $A_1^i(p)$, is a measure of the attraction of price p to seller i in period t.

The updating of $N(t)$ is governed by

$$N(t+1) = \rho \cdot N(t) + 1,\ t \geq 1, \tag{3.3}$$

where ρ is used to discount the influence of previous experiences. Denote $I(p,\, p_t^i)$ as the indicator variable for the chosen price p_t^i such that $I(p,\, p_t^i) = 1$ if $p = p_t^i$ and $I(p,\, p_t^i) = 0$ otherwise. The attraction variable for seller $i \in I$ is updated as

$$A_{t+1}^i(p) = \frac{\phi \cdot N(t) \cdot A_t^i(p) + [\delta + (1-\delta) \cdot I(p,\, p_t^i)]V_t^i(p)}{N(t+1)},\ t \geq 1. \tag{3.4}$$

where ϕ is a discount factor that depreciates previous attractions, δ is used to discount the effects of the foregone payoffs for the unchosen strategies, and $V_t^i(p)$ is the (foregone) payoff function defined in Equation (3.2). EWA is a hybridized learning model in the sense that it includes cumulative choice reinforcement learning ($\delta = 0$ and $\rho = 0$) and belief learning ($\delta = 1$ and $\rho = \phi$) as special cases. An important distinction of the belief learning special case of EWA and the WFP model we are using is that the belief learning case of EWA only counts a player's matched opponent's choices in formulating her beliefs, but in our WFP model we use the population members' past decisions. For simplicity, as in the GPAL model, we adopt the assumption that the foregone payoffs for unchosen strategies carry the same weight into the new attractions as the realized payoffs by setting $\delta = 1$. Another justification for doing this is that the observed price adjustments provide strong evidence for the sellers being ex-post rational in the sence of directional learning, which as argued by Camerer & Ho (1999), is essentially the same prediction as EWA with $\delta = 1$.

Weighted Fictitious Play (WFP) of Cheung & Friedman (1997) assumes that a player formulates beliefs about her opponents' future strategies and plays noisy responses to her beliefs. The belief of a player is assumed to be the weighted average of the strategies that she encountered in the past periods. Cournot learning and fictitious play learning are special cases of the model. While Cheung & Friedman (1997) assume that players' beliefs are formulated using the past strategies of their actual rivals, we assume that the sellers' current beliefs are the weighted average of all her potential opponents' past strategies. This makes sense as in any period any of the other sellers can be encountered. So this assumption basically adapts the original model for random matching.[7] Let $(F_1, \cdots, F_t)$ denote the vector of market price distributions observed from period 1 to period t, the belief firm i holds before period $t+1$ is:

$$B^i(F^j_{t+1}(p)) = \frac{F_t(p) + \sum_{\tau=1}^{t-1} \theta^\tau F_{t-\tau}(p)}{1 + \sum_{\tau=1}^{t-1} \theta^\tau}, \forall p \in \mathrm{P}. \tag{3.5}$$

Parameter θ captures the idea that different past histories enter with different weights into

[7] See Bayer et al. (2013) for an extension of WFP which incorporates risk aversion to explain sellers' behavior observed in the same Bertrand duopoly experiment.

the beliefs. When $0 < \theta < 1$ we have the typical case that recent histories carry more weight than older histories. Setting $\theta = 0$ yields the Cournot adjustment rule, where only the most recent period is relevant for the beliefs. Setting $\theta = 1$ yields standard fictitious play, where all past experiences are weighed evenly. Consequently, in period $t+1$ seller i's expected payoff for posting price p, given her belief $B^i_{t+1}(F^j_{t+1})$, can be written as

$$A^i_{t+1}(p) = (p - c)\left[1 - B^i(F^j_{t+1}(p)) + \frac{1}{2}B^i(f^j_{t+1}(p))\right]. \tag{3.6}$$

4 Analysis

In this Section we apply the learning models to the experimental data and evaluate their effectivenesses of predicting observed behavior. In total, we consider five models: GPAL, EWA, WFP, GPAL with $\beta = 0$, and GPAL with $\alpha + \beta = 1$. We include the two special cases of the generalized payoff assessment learning model to evaluate how the exclusion of one category of information would affect the performance of the model. First, we describe our estimation method. Then we evaluate the performance of the different learning models in explaining the observed behavior at both aggregate and individual levels.

4.1 Parameter Estimation

For parameter estimation, we use maximum likelihood estimation (MLE). That is, for each model, we search for the values of parameters that maximize the log-likelihood of observing the prices posted by the sellers in the experiment. To do so, we first transform the $A^i_t(p)$'s defined in all the above learning models into choice probabilities using the following choice function:

$$f_t^i(p) = \frac{[A_t^i(p)]^\lambda}{\sum_{k=c}^{v}[A_t^i(k)]^\lambda},\ \forall t \in T,\ \forall p \in \text{P}. \tag{4.1}$$

Here $\lambda \in [0, \infty)$ can be interpreted as the "mood-shock" or "bounded-rationality" parameter which measures how sensitive the sellers are to the payoff differences between different prices. As $\lambda \to \infty$, a seller tends to choose the price with the highest propensity and becomes fully rational. On the other hand, as $\lambda \to 0$, the seller becomes fully ignorant or confused and randomizes over all prices with equal probabilities. An appealing property of the choice rule is that prices that with higher propensities are more likely to be posted when $\lambda > 0$.

Our last task before turning to the estimation is to pin down the initial values of $A_t^i(p)$ for all prices. To avoid the possibility that different choices of initial propensity values in different models produce bias that is in favor of one model, we adopt identical set of initial propensities for all of the five models which we define as:

$$A_1^i(p) = (p-c)[1 - F_1(p) + \frac{1}{2}f_1(p)],\ \forall p \in \text{P}. \tag{4.2}$$

So given the market price distribution of period 1, we set $A_1^i(p)$ equal to the expected payoff of p being played against the actual period one prices. Starting with $A_1^i(p)$, when new information is revealed, the propensity vectors are updated according to the propensity updating rules. And, as a result, the sellers' choice probabilities can be calculated by Equation (4.1).

Given the updating rules and the vector of intial propensities, the MLE algorithm then searches for the values of learning model parameters that maximize the log-likelihood of observing the prices posted in periods two to fifteen. Formally, the log-likelihood function of a model can be stated as

$$LL = \log\left[\prod_{t=2}^{15}\prod_{i=1}^{80} f_t^i(p_t^i)\right] = \sum_{t=2}^{15}\sum_{i=1}^{80}\log\left[f_t^i(p_t^i)\right]. \tag{4.3}$$

Table 1: Maximum Likelihood Estimates

$GPAL$		WFP		EWA		$GPAL\,(\beta=0)$		$GPAL\,(\alpha+\beta=1)$	
α	0.518(0.02)	θ	0.321(0.01)	ϕ	0.576(0.01)	α	0.583(0.04)	α	0.606 (0.01)
β	0.023(0.01)			ρ	0.502(0.01)				
λ	1.522(0.06)	λ	1.542(0.02)	λ	1.663(0.02)	λ	1.601(0.07)	λ	1.673(0.02)
				$N(1)$	3.69(0.03)				
LL	-3457	LL	-3883	LL	-3460	LL	-3462	LL	-3853
BIC	3467	BIC	3890	BIC	3474	BIC	3469	BIC	3860

Note: Numbers in parentheses refer to standard errors derived by numerical differentiation.

4.2 Results

Table 1 reports the results of the estimation. In parentheses we report standard errors calculated using numerical differentiation. The table also includes the log-likelihood values (LL) and the corresponding Bayesian Information Criterion values (BIC) for the estimated models.[8] According to the estimates of the GPAL model, the belief-learning factor has a very small influence on learning ($\hat{\beta} = 0.023$) compared to the inertia factor ($\hat{\alpha} = 0.518$) and the experiential-learning factor ($1-\hat{\alpha}-\hat{\beta} = 0.459$). This is a suprising result becasue the random rematching protocol suggests that the observation of population members' actions should play a dominant role in determining players learning behavior. Moreover, the reported BIC values of the different learning models implies that the models with experiential-learning factors (GPAL, $BIC = 3467$; GPAL with $\beta = 0$, $BIC = 3469$; EWA, $BIC = 3474$) substantially outperform the models without the experiental-learning factors (GPAL with $\alpha + \beta = 1$, $BIC = 3860$; WFP, $BIC = 3890$) at an aggregte level.

Similar conclusions can be obtained by considering only the GPAL models. If we remove

[8] BIC is used to compare the relative goodness of fitting of different models. It penalizes models with additional parameters. According to BIC, a model with lower BIC is preferred. In our analysis, BIC is defined as $BIC = \frac{k}{2}\ln(N * T) - LL$, where k is the number of parameters, N is the number of sellers, and T is the number of periods considerred.

the belief-learning factor from the model by setting $\beta = 0$, the value of the maximum log-likelihood is reduced by 5, from -3457 to -3462. In contrast, if we drop the experiential-learning factor from the model and set $1 - \alpha - \beta = 0$, the maximum log-likelihood falls to -3853 which is much lower than -3457, suggesting that the exclusion of the experential learning factor substantially weakens the model's explainatory power. To summarize, at an aggregate level, there is strong evidence for the experiential-learning component dominating over the belief-learning component.

Next we turn to an evaluation of the comparative fit of the models on an individual level. To do so, we adopt the the quadratic scoring rule (QSR).[9] Given the predicted probabilities and the actually observed price, in period t the quadratic score for seller i is defined as:

$$q(i,t) = 2f_t^i(p_i^t) - \sum_{p=30}^{100} \left[f_t^i(p)\right]^2. \tag{4.4}$$

Here $f_t^i(p)$ is the predicted probability for price $p \in \mathrm{P}$, and p_i^t is the price acutally posted by player i in period t. According to Equation (4.4), we have $-1 \leq q(i,t) \leq 1$. When $q(i,t) = 1$, the model's prediction is perfectly in line with seller i's decision in period t. In contrast, if the model associates probability one to a price that is different to the chosen price of seller i, then the quadratic score is $q(i,t) = -1$. For comparison, we take the average quadratic score $\overline{q}$ of $q(i,t)$ for all five learning models over all periods from two to fifteen and across all sellers. According to QSR, the model with a higher $\overline{q}$ is more precise in predicting observed individual decisions.

Figure 3 plots mean quadratic scores for the five different learning models. As can be seen from the plot *EWA* has the highest mean quadratic score ($\overline{q} = 0.0742$), suggesting that *EWA* is the best in tracking individual's round-by-round choices. GPAL and its special case with

[9] see Selten (1998) for an aximatic characterization of QSR properties, and Chmura et al. (2012) for QSR applications.

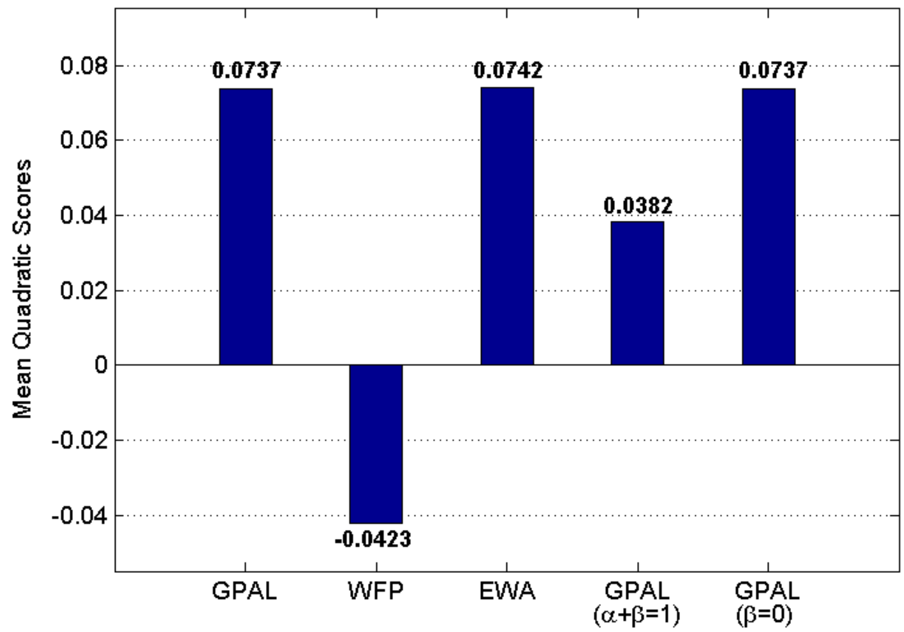

Figure 3: Mean quadratic scores

$\beta = 0$ have identical mean quadratic scores at $\overline{q} = 0.0737$ which are only marginally smaller than that of *EWA*. In contrast, dropping the experiential-learning factor from GPAL leads to a substantial reduction of $\overline{q}$ to 0.0382. *WFP* has the lowest score at -0.0423. Therefore, at the individual round-by-round level the three models incorporating the experiential-learning factor still substantially outperform the two pure belief-learning models, indicating that the observation of the matched player's choice has much more influence than the information about the population price distribution.

So far we have shown that GPAL can be used to evaluate the relative impact of different types of feedback on players' learning behavior. Moreover, GPAL fits the experimental data at least as well as other established experiential learning models that (such as EWA) and better than belief learning based models (such as WFP).

In what follows we provide a few more pieces of evidence for the good fit of GPAL. While we have established GPALs relative performance to other models it remains to show that the absolute fit it produces is also satisfactory. Figure 4 plots the time series for both

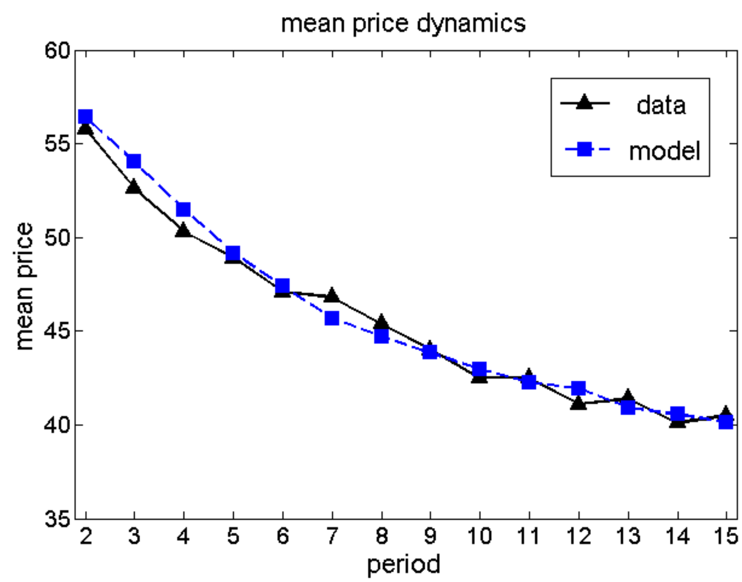

Figure 4: Average price dynamics: data observation and GPAL prediction

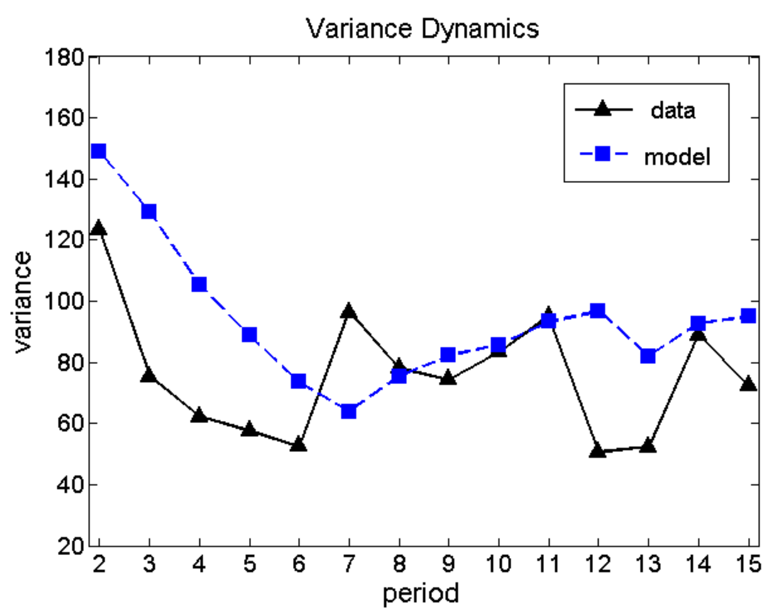

Figure 5: Price variance dynamics: data observation and GPAL prediction

the observed and predicted mean of the market prices. Figure 5 compares observed to predicted price variance. In addition, Figure 6 in Appendix A provides a comparison of the actual and predicted cumulative frequencies from period two to fifteen. The three plots further confirm that our model does an excellent job at capturing observed behavior. To assess the performance of our GPAL model in a more formal way, we conduct period-wise Kolmogorov-Smirnov(K-S) tests to test the null hypothesis of equality between the actual price distributions and those predicted by the model. For 10 out of 14 periods (71.4%) the test fails to reject the null hypothesis, which provides strong evidence for the good performance of GPAL.

5 Conclusion

In this paper we proposed a simple extension to the payoff assessment learning model of Sarin & Vahid (1999) to explain sellers' learning behavior in homogeneous-product Bertrand duopoly experiments with perfect information. The most prominent feature of the model is that it allows for different learning channels and therefore bridges the gap between existing models. Our model has got an experiential part and a belief-learning part. The experiential part uses actually experienced and forgone payoff to update attractions. The associated cognitive process can expressed by the question: "what would I have got if had chosen a different price?" Psychologically this component is quite salient. The second component is forward looking and hypothetical by using payoffs that can be expected to by achieved for certain prices playing against a random draw from the population price distribution. The associated question would be: "what can I expect in the next round if I set this price?" While the experiential component is salient the belief-learning component is appealing as it relates closely to how economists think about humans making decision, i.e. forming beliefs and conjectures and choosing (noisy) best responses. By combining both types learning from information, our model assumes that humans' learning is governed by a mixture of the two cognitive processes. However, we find in the context of a Bertrand duopoly that learning happens almost exclusively through the experiential channel. We believe that GPAL can be used and applied to data from other games in order to find out if our result is the expression of a regularity or only holds in our specific game.

Acknowledgements

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

References

Abbink, K., Sadrieh, A. & Zamir, S. (2004), 'Fairness, public good, and emotional aspects of punishment behavior', *Theory and decision* **57**(1), 25–57.

Abramson, C., Currim, I. S. & Sarin, R. (2005), 'An experimental investigation of the impact of information on competitive decision making', *Management Science* **51**(2), 195–207.

Armantier, O. (2004), 'Does observation influence learning?', *Games and Economic Behavior* **46**(2), 221–239.

Bayer, R. C., Wu, H. & Chan, M. (2013), 'Explaining price dispersion and dynamics in laboratory bertrand markets.', *Working Papers, University of Adelaide* .

Camerer, C. F. & Ho, T. H. (1999), 'Experience-weighted attraction learning in normal form games', *Econometrica* **67**(4), 827–874.

Cheung, Y. & Friedman, D. (1997), 'Individual learning in normal form games: Some laboratory results', *Games and Economic Behavior* **19**(1), 46–76.

Chmura, T., Goerg, S. J. & Selten, R. (2012), 'Learning in experimental 2× 2 games', *Games and Economic Behavior* .

Huck, S., Normann, H. & Oechssler, J. (2000), 'Does information about competitors' actions increase or decrease competition in experimental oligopoly markets?', *International Journal of Industrial Organization* **18**(1), 39–57.

Mookherjee, D. & Sopher, B. (1994), 'Learning behavior in an experimental matching pennies game', *Games and Economic Behavior* **7**(1), 62–91.

Roth, A. E. & Erev, I. (1995), 'Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term', *Games and economic behavior* **8**(1), 164–212.

Sarin, R. & Vahid, F. (1999), 'Payoff assessments without probabilities: a simple dynamic model of choice', *Games and Economic Behavior* **28**(2), 294–309.

Selten, R. (1998), 'Axiomatic characterization of the quadratic scoring rule', *Experimental Economics* **1**(1), 43–62.

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.

Appendix

A Cumulative Distributions of Prices - Data and Model

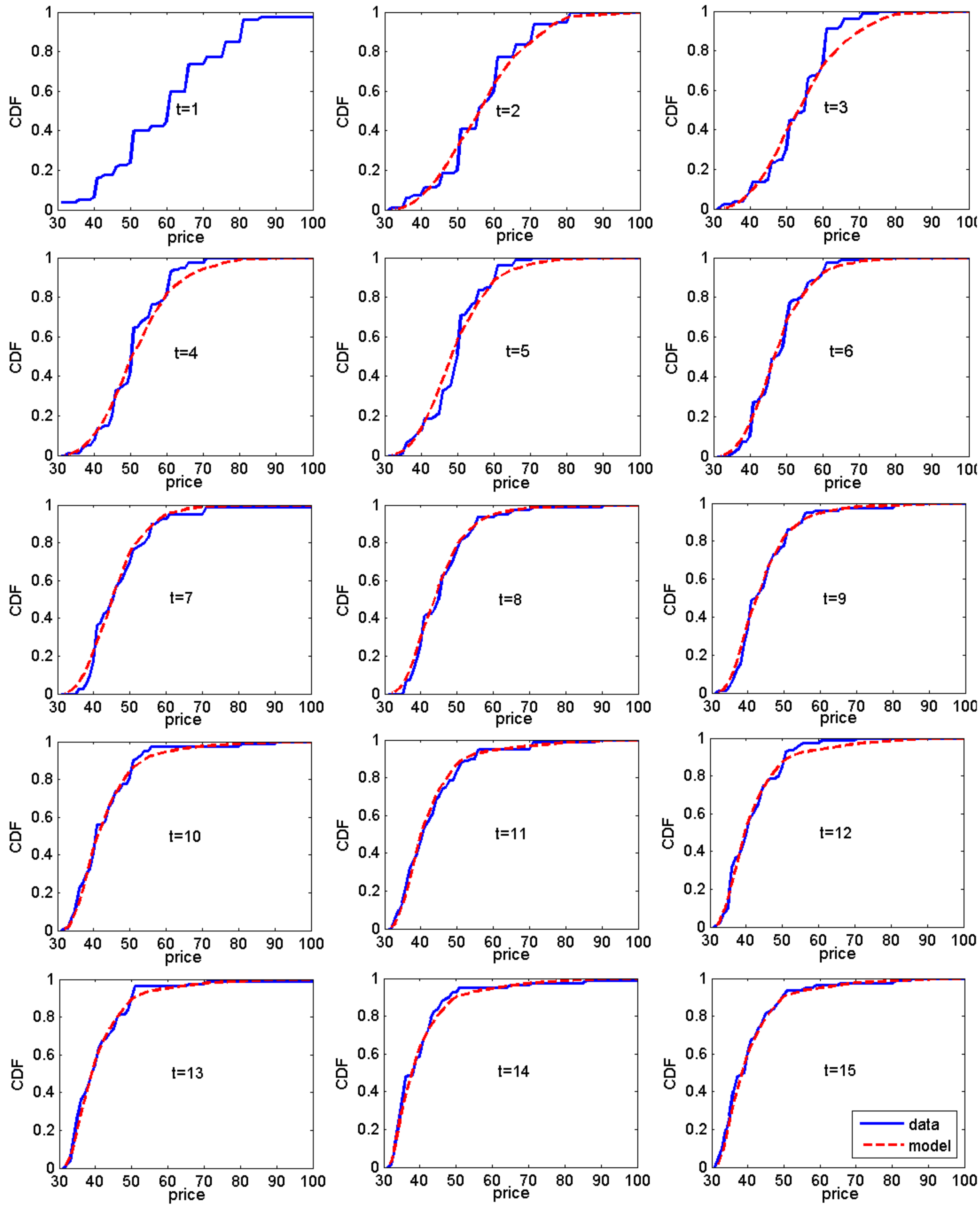

Figure 6: Cumulative Distributions of Prices - Data and Model

B 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 E-Dollars = 2 Australian Dollars. You will also be paid an appropriate base payment for your participation.

B.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 E-Dollars to produce and sell. The buyer can buy one unit of the good, which he values at V = 100 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.

B.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.

B.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

B.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).