ESTIMATING LEARNING RULES: STRATEGIES VERSUS ACTIONS

Aaron Lowen

Seidman College of Business

Grand Valley State University

Grand Rapids, MI 49506

(Lowena@gvsu.edu)

616-331-7441

July 14, 2005

Work in progress. Do not quote or cite without permission of the author.

Abstract

Recent work in learning theory has focused on estimating parameters of, and testing among, opposing rules. Unfortunately, there are alternative formulations of each learning rule. An open question is whether players use information gained along the path of play to update off-path propensities. Using a sender-receiver game, the stimulus-response learning rule is expressed under each formulation. Maximum likelihood estimation is applied to both simulated and experimental data. Identification issues play a significant role in estimation and testing. Model selection tests are performed and initial results indicate formulation choice matters. This has important implications for learning research, including compound rules that nest simpler rules where formulation matters.

I. Introduction

Recently, researchers have attempted to estimate learning rule parameters and test among these models using experimental data (Roth and Erev, 1995; Erev and Roth, 1995; Camerer and Ho, 1999; Cabrales and García-Fontes, 2000; Salmon, 2001; Blume, DeJong, Neumann, and Savin (BDNS), 2002). These researchers have, implicitly or explicitly, constrained themselves to using only passively observable data by not eliciting intended play in unrealized conditions. Others (Selten, 1967; Mitzkewitz and Nagel, 1993; Selten, Mitzkewitz, and Uhlich, 1997) have emphasized an active method of data collection, the "strategy method"[1], which yields more information about player behavior.

Suppose a player gains information when type θ_1 in a repeated game. In a subsequent period that same player becomes type θ_2 for the first time. A relevant issue is whether the player has updated propensities for behavior when θ_2 based on θ_1 experiences. In other words, whether players only update their propensities for behavior for the type they experienced or update across all possible types. The case where a player updates propensities for behavior for only one type at a time corresponds to learning in an information node in an extensive form representation of the stage game. This type of learning will be referred to as "actions" learning, and is well served by those who passively observe data. The case where information is used to update all types simultaneously can be thought of as learning in the strategic form of the stage game, or "strategy" learning, and Selten's method could be used to elicit underlying strategies. Note that we consider how boundedly rational players update their propensities for behavior in a repeated environment, *not* the equivalence between the extensive and strategic form representations of a game.[2] Instead, we ask whether players allow information to cross from one contingency to another in their learning rules. Here, we form different specifications of the Stimulus-Response (SR) learning rule corresponding to this distinction and examine their empirical relevance.

---

[1] To prevent confusion with similar terms this will be referred to as "Selten's method".

[2] The game theory literature has articulated the equivalence conditions and similarities between extensive and strategic forms (Kuhn (1953), Thompson (1952a and 1952b), Dalkey (1953), Kohlberg and Mertens (1986), van Damme (1984), and Mailath, Samuelson, and Swinkels (1993), among others). However, other work has shown that learning effects permit different equilibria in different extensive form games with the same strategic form (Fudenberg and Levine (1993a, 1993b) and Kalai and Lehrer (1995)).

Distinguishing between the action and strategy representations is important since the researcher may be omitting a relevant model that is similar to the one under consideration. Furthermore, this alternative may not nest nor be nested by the original model. Comparing a simple model with a more sophisticated model may be misleading if the simple model is incorrectly specified. Similarly, a more sophisticated model may nest the miss-specified simple model.

Two formulations of the stimulus response (SR)[3] learning model are estimated with maximum likelihood using experimental data from a sender-receiver game. A classical non-nested test (the J-test) is used for model selection, and a conditional means test is proposed to weaken the assumptions necessary for identification. Under the assumption of the SR learning rule, the strategy formulation is preferred to actions in at least one of the three data sets. These results were supported by the classical J-test and[4] the conditional means test.

In the second section we consider a sender-receiver game, the SR rule, and a description of how the different formulations (actions and strategies) are realized. The next section contains estimation and testing results that includes a description of how the game and rules were simulated and estimated with both real and simulated data. The fourth section concludes.

II. Data Generation Process

First a sender-receiver game and its strategic and extensive form representations is described. Then the SR model and the implementations of action and strategy learning is presented.

A. Sender-Receiver Game

Consider a repeated sender-receiver game with perfectly aligned preferences, and a message space exactly rich enough to allow type separation (M and Θ have the same cardinality). In this section senders are considered, with receivers having a similar treatment.

---

[3] A similar analysis for a belief-based (BBL) learning model and cross-model comparisons is available in a longer version of the paper. We find that SR dominates BBL in the experimental data sets used and when considering the BBL model, all data sets support the strategy form over actions.

[4] Also confirmed using Bayesian decision making techniques; available from the author.

There are separate populations of senders and receivers, each of size N. At the beginning of each round all players are paired, one sender with one receiver. Senders receive private information in the form of their type for that round[5], $\Theta = \{\theta_1, \theta_2\}$, with each type equally likely. Each sender then transmits one of two messages, $M = \{m_1, m_2\}$, which is privately observed by their partner. Messages are costless to send and are *a priori* meaningless. Receivers respond to their message with one of two actions, $A = \{a_1, a_2\}$. Both players in the pair receive positive payoff x is received if the $[\Theta, A]$ combination is $[\theta_1, a_2]$ or $[\theta_2, a_1]$ and zero otherwise. Positive payoffs are off-diagonal to create an initial ambiguity in messages and eliminate focal points[6] (e.g., with this design it is not clear to players whether message m_1 communicates sender type or requested receiver action).

Each player in a sender-receiver pair is informed of (or can deduce) the type, message, action, and payoff at the end of the round. No player observes this information for any pair other than their own, nor do they observe what their partner would have done if their information set had been different. In other words, the receiver does not know what message the sender would have sent as the other type, and the sender does not know what action the receiver would have taken given the other message. This is the end of a round, and the game repeats T times.

Extensive form or "actions" updating has an interpretation of the stage game based on the tree in Figure 1. In this representation, it is as if each stage game information set was run by a different representative of each player. For example, a sender uses one representative when type θ_1, and another when θ_2. The player does not coordinate information between the two, and they do not interact. Any experience gained when θ_1 does not influence behavior when θ_2. The extensive form game may be formally defined as $\Gamma_E = \{\aleph, M, A, \mathrm{H}, \{u_i(\cdot)\}\}$ that specifies players, the timing of each player's moves, actions available at each node, what is known at each node, and outcomes.

The strategic form is an alternative to the extensive form representation. There are only four stage-game strategies for each sender and each receiver. These are given in Table 1 for senders, with receiver strategies similarly constructed. A strategy for the sender maps types into

[5] A new type draw is made every round for every sender.

[6] This effect was created through design in the experimental data, and simulated agents were not programmed to have focal points.

messages, s: $\Theta \rightarrow M$; for the receiver, messages into actions, r: $M \rightarrow A$. A formal definition of the strategic game is $\Gamma_S = \{\aleph, \{S_i\}, \{u_i(\cdot)\}\}$, specifying the players, strategies, and outcomes.

This simple game is ideal for analyzing the issue of cross-node updating. First, it has finite, discrete action spaces. This avoids the trouble of defining "nearby" strategies which would be required in a game with a continuous (or infinite) action space. This is also advantageous because the strategies in the stage-game are easily enumerated and there are few type-conditioned behaviors. In essence, the learning environment is simple in that there are only two choices available to each player in each round (conditional on the realization of one of two possible types each round). Next, the stage-game equilibria are well understood and trivial to compute. Third, payoffs to each player in each of the separating equilibria are made to be equal, eliminating the focal outcome of a higher-payoff (Pareto dominant) equilibrium. This symmetry suggests natural assumptions on prior beliefs, even for experimentally generated data. Finally, the game is sufficiently simple to be implemented in an experimental lab.

The equilibrium concept for this dynamic game of incomplete information is typically perfect Bayesian equilibrium (PBE). In a PBE, each player has a belief about which node they have reached in their information set. Such a belief is a probability distribution over nodes in the information set. Player behavior at each information set must be optimal given these beliefs on player strategies.

B. Stimulus-Response Learning Rule (SR)

To explain the underlying decision-making process the focus is on senders, with receivers using an equivalent procedure. Following Roth and Erev (1995), define the propensity, $Q_{ij}(t)$[7], of player i to make choice j at time t as:

$$Q_{ij}(t) = \varphi_0 Q_{ij}(t-1) + \varphi_1 X_i(t-1)\, I_{ij}(t-1) \qquad (1)$$

where $X_i(t-1)$ is the reward player i received in period t-1. The indicator variable I_{ij} is one if the player made choice j at time t-1, and zero for all other choices in the information set. Parameter

[7] Sender notation will be repressed when we are considering a representative sender.

φ_0 measures the importance of previous observations, and φ_1 is the weight given to current rewards.[8] The probability player *i* chooses *j* from the available behaviors (*j'*) at the current information set at time *t* comes from the logit-like function:

$$P_{ij} = \frac{\exp(Q_{ij}(t))}{\sum_{j'} \exp(Q_{ij'}(t))} \tag{2}$$

The computation of propensities and probabilities is applied to the action and strategy formulations in similar ways, and is discussed in the next two sub-sections.

Specification of the SR model also requires initial propensities. Values chosen for $Q_{ij}(t=1)$, relative to payoff magnitude, affect the speed with which rewards change propensities and hence probabilities of making a particular choice. For example, setting prior propensities to be large compared to payoffs means that the probabilities for play will evolve slowly over time, with any single payoff only slightly changing the probabilities over behaviors.

Let the indicator function $I_i(t)$ equal one if the sender used m_2 and zero otherwise, and $P_{i2}(t)$ be the probability player *i* sent m_2. The maximized likelihood function, then, finds the parameters which make the predicted probability for each message, given type, as close as possible to the frequency of observed values. The log likelihood function for sender data is:

$$\ln(l(\varphi_0, \varphi_1)) = \sum_{i=1}^{N} \sum_{t=2}^{T} [I_i(t) \ \ln(P_{i2}(t)) + (1 - I_i(t)) \ \ln(1 - P_{i2}(t))] \tag{3}$$

As is shown in Blume, DeJong, Lowen, Neumann, and Savin (2002), henceforth BDLNS, the likelihood depends on the parameters and the differences in initial propensities, $Q_{ij}(t=0)$. Since symmetry in equilibria and information has been imposed, the differences in initial propensities vanish so long as they are equal. This symmetry of priors assumption will be imposed in the following simulations, and was created in the real data through experimental

---

[8] Typically $\varphi_0 \in (0, 1)$ and is referred to as the "memory" parameter; φ_1 is positive and referred to as the "updating" parameter.

design (see Blume, DeJong, Kim, and Sprinkle (1998) for details). Thus, differences in initial propensities drop out of the likelihood function, and it depends only on parameter values.
The receiver likelihood function is similarly constructed, with action given message the relevant conditional choice instead of message given type. As was also shown in BDLNS, the joint likelihood function for senders and receivers can be factored into the product of their separate likelihood functions so long as the parameter values are different for senders and receivers. For now we restrict our attention to sender data, and consider pooling in a later section.

Under the SR rule, propensities for behavior are only affected by rewards from the player's own experience. These rewards are incorporated so that stochastic choice is an important part of the decision process. SR is a "low rationality" model, where players are unsophisticated in the analysis of their partner's behavior, and players need not know opponent payoffs, make assumptions about opponent rationality, nor know they are playing a game.[9] Furthermore, memory requirements are low; players need only remember their propensities and payoffs from the previous period, not their entire history.[10]

The SR rule was selected for a number of reasons. First, it is simple in that it has only two parameters to be estimated. Next, it imposes weak requirements on player rationality, memory, and computational skill. Third, the SR learning rule results in behavior which appears to be consistent with existing experimental evidence: the Law of Effect (positive payoffs increase the likelihood of a particular choice in the future), the Power Law of Practice (learning curves flatten over time), and stochastic choice (players do not always use best-response behavior). The SR learning model is commonly used in the literature, and is (arguably) considered to be adequate in approximating the learning process of real players. Finally, it is easy to implement the action and strategy formulations in simulation and estimation.

C. The Stimulus-Response (SR) Learning Rule Under Different Domains

Both the action and strategic updating regimes can be defined under the SR learning rule. The actions form is considered first, followed by the strategic form.

---

[9] This permits the simplification that players ignore partner identities.

[10] Players could use their propensities to back out their history of play if they knew their own learning rule, priors, and parameter values. There are some mild restrictions on parameter values and payoffs for this inference to be possible.

1. Action Updating in SR

To see how propensities evolve for actions in equation (1), consider a history for a sender where the pair [θ_1, m_1] was played in period t-1. Propensities $Q_{\theta, m}$ for player i evolve in the following way for period t:

$$
\begin{aligned}
Q_{1,1}(t) &= \varphi_0 Q_{1,1}(t-1) + \varphi_1 X(t-1)*1 \\
Q_{1,2}(t) &= \varphi_0 Q_{1,2}(t-1) + \varphi_1 X(t-1)*0 \\
Q_{2,1}(t) &= Q_{2,1}(t-1) \\
Q_{2,2}(t) &= Q_{2,2}(t-1)
\end{aligned}
\tag{4}
$$

The zero and one values come from the realization of the indicator variable, where $Q_{1,1}(t)$ has the potential to be increased because the previous behavior realization was [θ_1, m_1]. Propensity $Q_{1,2}(t)$ is included in the updating since behavior m_2 is available from the realized type node, θ_1; $Q_{2,1}(t)$ is not included since θ_2 was not realized. Under actions, the probability sender i sends m_1 when θ_1 in period t is:

$$
P_{1,1}(t) = \frac{\exp(Q_{1,1}(t))}{\exp(Q_{1,1}(t)) + \exp(Q_{1,2}(t))} \tag{5}
$$

The updating process for actions learning has a natural condition on priors. Since types and messages are symmetric prior beliefs should be neutral to equilibrium selection; priors should not favor one of the separating equilibria. This symmetry implies types and messages contingent on type should be treated identically in prior beliefs.

Assumption #1: *In actions learning initial probability weight on each action within and across types is identical. Suppressing sender notation, let $Q_{\theta,m}$ indicate the propensity for type θ to send message m. This implies $Q_{1,1} = Q_{1,2} = Q_{2,1} = Q_{2,2}$ at the beginning of the first period (t=1).*

2. Strategy Updating in SR

Under strategies, information crosses type nodes. Thus, experience influences propensities of all strategies every period. The updating equation system for period t strategies comes from equation (1), for convenience we suppress player subscript:

$$Q_k(t) \;=\; \varphi_0 Q_k(t\text{-}1) \;+\; \varphi_1 X(t\text{-}1) I_k(t\text{-}1) \tag{6}$$

where Q_{k} represents the propensity for strategy s_{k}, $X(t\text{-}1)$ indicates the reward received in period t-1, and I_{k} is the indicator function which is 1 if strategy s_{k} was played in period t-1 and zero otherwise.

For example, if a player implemented strategy 1 in period t-1, propensities over strategies in period t are:

$$\begin{aligned}
Q_1(t) \;&=\; \varphi_0 Q_1(t\text{-}1) \;+\; \varphi_1 X(t\text{-}1)*1 \\
Q_2(t) \;&=\; \varphi_0 Q_2(t\text{-}1) \;+\; \varphi_1 X(t\text{-}1)*0 \\
Q_3(t) \;&=\; \varphi_0 Q_3(t\text{-}1) \;+\; \varphi_1 X(t\text{-}1)*0 \\
Q_4(t) \;&=\; \varphi_0 Q_4(t\text{-}1) \;+\; \varphi_1 X(t\text{-}1)*0
\end{aligned} \tag{7}$$

Under strategies, the probability sender i uses strategy s_2 in period t is:

$$P_2(t) \;=\; \frac{\exp(Q_2(t))}{\sum_{k=1}^{4} \exp(Q_k(t))} \tag{8}$$

The primary difference between action and strategy learners is the relation between types. Strategy learners let information gained when θ_1 influence behavior when θ_2, and vice versa. This may be due to more sophisticated strategic behavior or a different perception of the game. A player who has restricted their choice set to separating strategies would seem to imply the first, while players who use pooling strategies with positive probability would seem to imply the second.

Strategic form learning has a natural assumption on priors similar to that for action learning. As with actions, strategy learners should have prior beliefs that are neutral to separating equilibria.

Assumption #2: *In strategic form learning initial probability weights on the pooling strategies (s_1 and s_4) should be equal, as should the weights on separating strategies (s_2 and s_3). Referring to Table 2, let Q_k indicate the propensity to use strategy s_k, so $Q_1 = Q_4$ and $Q_2 = Q_3$ at the beginning of the first period (t=1).*

III. Estimation and Testing

A. Identification and Optimization

The task of parameter estimation under strategies is more complicated than it first appears. Passively observable sender data such as $[\theta_1, m_1]$ could be generated by two different underlying strategies (s_1 or s_2). Estimation of parameters under strategies requires an approach that can infer intended behavior off the path of play. One option from the literature is that proposed by Selten (1967); players declare the strategy to be used for the upcoming round to the researcher who implements the declared strategy for the player. In terms of experiment or simulation design, this means the researcher would elicit a complete contingent plan at the beginning of each play of the stage game from each player. Without application of Selten's method only realized types, messages, actions, and rewards would be observed.

Unfortunately, Selten's method may produce undesirable framing effects. One complication is that players who do not update in the strategic form are being forced to do so, and it is unclear how they will respond. With real players such an intervention may change behavior but no consensus has been reached. For example, Schotter, Weigelt, and Wilson (1994) found significant framing effects, but Brandts and Charness (1998) did not. Framing effects, if they existed, would complicate learning rule estimation and player classification. Eliciting unrealized behavior after a round or session concludes would avoid framing effects, but would elicit cheap talk. Players would have no incentive to recall or report accurately. Further, it is not obvious that players choose their type-dependent behavior until called upon, particularly if decision making is costly. If framing effects are of concern, the econometrician must find an alternative to direct estimation.

The approach used here is to make assumptions on prior propensities that make the underlying strategies (and hence parameters) identifiable by setting either pooling or separating strategy prior propensities to -∞. The more reasonable assumption is to eliminate pooling strategies from play, as these are not consistent with the stable outcome of a separating equilibrium. In Assumption 2 this implies[11] $Q_1 = Q_4 = -\infty$. This allows identification of parameters without applying Selten's method or adding variables to the learning rule.

In addition to the generally unobservable nature of strategies, the memory parameter in the SR model is not identified when the updating parameter is equal to zero ($\varphi_1 = 0$). The identification problem arises because each behavior is used with a probability determined exclusively by priors, regardless of the history of payoffs or value of the memory parameter (variable φ_0 can be assigned any value). In an applied setting this causes complications in estimation, since the variance of the distribution of estimates grows as they approach the unidentified value. Lack of identification also complicates classical testing of the model; the natural null hypothesis is one that contains $\varphi_1 = 0$. If the model is not identified at the null we cannot find the distribution of a test statistic under that null. In short, if players put no weight on new information, and no learning occurs and learning parameter estimation and model testing cannot be meaningfully done.

Parameters were estimated by numerically maximizing the log likelihood function using the OPTMUM procedure in GAUSS. As is shown in BDNS (2002), the SR likelihood function is not globally concave. Hence a quasi-Newton search method, Broyden-Fletcher-Goldfarb-Shanno (BFGS), was used to ensure positive definite Hessian approximations and not use second derivative information. Standard error estimates used later in the paper come from approximations of the Hessian at the optimum, not the Hessian update from the algorithm.

There are two possible scenarios: either the researcher knows the underlying data generating process (DGP) or does not. If the true DGP is known[12] the task is to estimate the parameters of the process. However, knowing the DGP does not mean we can successfully

---

[11] One possible assumption that would justify this is allowing players an understanding of the environment that permits them to eliminate strategies inconsistent with the separating equilibria.

[12] A Monte Carlo study of the small sample properties shows that the empirical rejection properties are slightly greater than the nominal rejection probabilities. Again, the expanded version is available on request.

estimate the parameters from the small samples available in practice. If the DGP is not known, the task is to find which model best describes the data. Classical testing is non-trivial since the models in this application are not nested.

B. Results

We now apply the SR formulations to real (experimental) data, where the underlying data generating process cannot be known with certainty. Experimental data comes from Blume, DeJong, Kim, and Sprinkle (1998), henceforth BDKS. Their "no-history" treatments, session 2, closely fit the environment described above, with $x = 0.70$, $N = 6$, and $T = 20$. One feature of their work was that the players were not allowed to know the identity of their partners. Presently, we have assumed either that players do not know or do not keep track of identities; these experiments imposed this anonymity. Second, the mapping of messages was privatized in such a way that there were no focal points in either the type-message or message-action spaces. Further, the experimental design only revealed information from the player's own pair, so these own-information models are given the best chance of working.

Testing is first done with the J-test using asymptotic critical values. Then, a conditional means test which weakens the identifying assumptions on pooling strategies is applied.

1. Classical Testing: The J-Test

Since the models and formulations used are not nested, we cannot perform the classical model selection tests for nested models. One of the classical alternatives for non-nested models is the J-test, originally proposed by Davidson and MacKinnon (1981) in a linear regression setting. The goal is to test whether one model contributes to explaining the variation in the data given the other model is already included.

The two models considered have behavior options denoted by j and k, and coefficient vectors φ and β. We compare the two models as competing hypotheses trying to explain player i's choice of behavior l:

$$H_0 : P_{i,l} = \frac{\exp(\varphi_0 Q_{il}(t-1) + \varphi_1 X_i(t-1) I_{il}(t-1))}{\sum_j \exp(\varphi_0 Q_{ij}(t-1) + \varphi_1 X_i(t-1) I_{ij}(t-1))}$$

against

$$H_1: P_{i,l} = \frac{\exp(\beta_0 Q_{il}(t-1) + \beta_1 X_i(t-1) I_{il}(t-1))}{\sum_k \exp(\beta_0 Q_{ik}(t-1) + \beta_1 X_i(t-1) I_{ik}(t-1))}$$

For convenience, let the right hand side of the first equation be denoted P_j, the right hand side of the second equation P_k. The J-test procedure is to first obtain estimates of the parameters for model H_1 and generate predicted values of P_k, $\hat{P}_k$, then test $\alpha = 0$ in the equation:

$$P_{i,l} = (1-\alpha)P_j + \alpha(\hat{P}_k) \tag{14}$$

which, operationally, takes the form:

$$P_{i,l} = P_j + \alpha(\hat{P}_k - P_j) \tag{15}$$

The hypotheses (models) are then reversed and the procedure repeated.

In Davidson and MacKinnon's linear regression setting, they show that, under H_0, plim $\hat{\alpha} = 0$, and the t-ratio ($\hat{\alpha}$/se($\hat{\alpha}$)) is asymptotically distributed standard normal. Rejecting the hypothesis that $\alpha = 0$ is rejecting the first model (φ, j) in favor of the second (β, k). One hopes to reject the null for one model but not the other, but it is possible to reject both or neither in finite samples. Here α is estimated as a third variable in the maximum likelihood estimation process.

Results for the J-test on each of the three replications from the BDKS data are given in Table 2. The rows of the table contain "model 0", or that model which was being estimated; model 0 is the null hypothesis or model j in (14). Columns contain "model 1", the model for which predicted values $\hat{P}$ were obtained. Estimates of α are given for each of the data sets. The corresponding p-values from standard asymptotic theory are below in parentheses. Data points where a p-value is recorded by an asterisk (*) indicate a failure to converge after 200 iterations. For these data sets the $\hat{\alpha}$'s were recorded, but are difficult to interpret; we have no approximation of the inverse Hessian at an optimum and hence no estimate of the standard error. For convenience, the t-statistics[13] are provided below the p-values in brackets.

Consider the cell in the first row and second column. In this cell the model $SR_{actions}$ is being estimated with the predicted values from $SR_{strategies}$ included. The numbers on the top line are the $\hat{\alpha}$'s from estimating (14) from each of the three BDKS data sets. For the first two data sets we get $\hat{\alpha}$'s of 0.62 and 0.61, which may be considered to be significantly different from zero since they have p-values of about zero. When we interpret the results as a test between the models, we reject the $SR_{actions}$ model in favor of $SR_{strategies}$. The third predicted value was 0.82, but is difficult to interpret since we have no estimate of the standard error.

Next we reverse the order of the models and estimate again. The cell in the second row and first column contains the results for estimation of $SR_{strategies}$ in the presence of predicted values from $SR_{actions}$. The first and third data sets have $\hat{\alpha}$'s of 0.52 and 0.27, both of which have p-values greater than 0.04. Depending on the choice of significance level, these may or may not be significantly different from zero. If this is the case, then the $SR_{actions}$ formulation does not contribute in the presence of $SR_{strategies}$. For the second data set we are caught in the troubling case where we reject each for the other.

2. Testing Forms Through Observation Only

The results of the previous section indicate that 1) we should consider the formulation of learning rules that permit differentiation between actions and strategies, and 2) the strategy formulation may outperform actions in this data. However, strategies are troublesome since we cannot estimate the strategies form without either observing underlying strategies or else making strong assumptions about prior propensities (such as in Assumption 2 above). Without knowing underlying strategies we cannot generate predicted values and parameter estimates. Given SR generates the data, however, it is possible to test which of the formulations underlies the data without knowing the underlying strategy choices, and without assuming pooling strategies are

---

[13] There is a concern that for relatively small experimental data sets such as those used here, first-order asymptotic theory provides a poor approximation to the finite sample distribution of the J-test. In an effort to address this concern, a parametric bootstrap (Fan and Li (1995) and Godfrey (1998)) was implemented. Incorporating bootstrap critical values did not change results qualitatively, and made virtually no change in the quantitative results.

never played. Below is proposed a test for distinguishing between the strategy and actions forms of the SR model without eliciting underlying strategies.

The null hypothesis of interest is that players update within type nodes (actions), or that payoffs received when a player is type θ do not affect behavior probabilities when type θ'. The alternative hypothesis is that players update across type nodes (strategies), or that payoffs received when a player is type θ do affect behavior probabilities when type θ'.

Suppose the game described above is played for two periods. Without loss of generality, take θ to be one of the two types and θ' the other. Thus, in the rest of this section we will discuss type realization [θ, θ], but not [θ', θ'] since θ is arbitrary. Similarly, let m be either of the two messages and m' the other. Using this notational convention, there are three distinct histories the sender may have experienced when choosing a message in the second round. These are given in Table 3. Case A is the history where the same type is realized in both periods. Cases B and C have different type realizations in each period. The difference between B and C is whether the type-message combination from period 1 received a positive payoff. Given symmetry over types and messages, Pr(Case A) = 0.5, Pr(Case B) = Pr(Case C) = 0.25. Case A occurs half the time since the probability a player is the same type in both periods is 0.5. Cases B and C occur with equal probability since there is a 0.5 probability of being different types in the two periods, and the likelihood of successfully matching in the first period is also 0.5.

When a player is the same type in both periods, we are not able to test whether they let information cross type nodes. Hence, Case A gives no information about play for the type that was not used in the first round, and does not allow a test of strategies versus actions. Thus data corresponding to Case A can be ignored when conducting the conditional means test.

In Case B both types are played, but the payoff is zero. Regardless of formulation, the SR rule does not change probabilities of behavior unless there is a positive payoff. This is the Law of Effect: only behaviors that receive positive payoff are played more frequently in the future. Hence, both updating regimes predict the same conditional probability of behavior:

$$\Pr(M_2 = m' \mid \Theta_1 = \theta, M_1 = m, X_1 = 0, \Theta_2 = \theta') = 0.5 \qquad (21)$$

where Θ_t, M_t, and X_t represent the type, message, and payoff in period t. Like Case A, data classified as Case B does not allow us to test the underlying formulation, and is also discarded.

For data that fits Case C, let subscripts indicate period (M_t is message in period t) and define:

Q_P = initial weight on each pooling strategy,

Q_S = initial weight on each separating strategy,

$\delta = Q_P - Q_S$,

$\Pi = \Pr(M_2 = m' \mid \Theta_1 = \theta, M_1 = m, X_1 = 0, \Theta_2 = \theta', \text{actions})$,

$\Pi_P = \Pr(M_2 = m' \mid \Theta_1 = \theta, M_1 = m, X_1 = 1, \Theta_2 = \theta', S_1 \in \text{pooling})$, and

$\Pi_S = \Pr(M_2 = m' \mid \Theta_1 = \theta, M_1 = m, X_1 = 1, \Theta_2 = \theta', S_1 \in \text{separating})$

where S_t is the strategy played in period t. Under the null of actions learning, $\Pi = 0.5$; information gained when type θ does not influence behavior when type θ'. Under the alternative of strategies probabilities are:

$$\Pi_P = \frac{\exp(\varphi_0 Q_P) + \exp(\varphi_0 Q_P + \varphi_1 X_1)}{\exp(\varphi_0 Q_P) + 2\exp(\varphi_0 Q_S) + \exp(\varphi_0 Q_P + \varphi_1 X_1)}$$

$$\Pi_S = \frac{\exp(\varphi_0 Q_S) + \exp(\varphi_0 Q_S + \varphi_1 X_1)}{2\exp(\varphi_0 Q_P) + \exp(\varphi_0 Q_S) + \exp(\varphi_0 Q_S + \varphi_1 X_1)} \tag{22}$$

Players know their strategy, but the researcher does not. The researcher only observes the empirical frequency of messages conditional on type in the population of senders. This empirical frequency among Case C senders in the second period is:

$$\frac{\exp(Q_P)}{\exp(Q_S) + \exp(Q_P)}\Pi_P + \frac{\exp(Q_S)}{\exp(Q_S) + \exp(Q_P)}\Pi_S \tag{23}$$

Consider the special case where $Q = Q_S = Q_P \in R$, $\varphi_0 > 0$, and $\varphi_1 > 0$. Equation (23) is then equal to 0.5. Under this condition, the researcher is unable to distinguish between the formulations due to the offsetting nature of pooling and separating strategies in the population. The identifying assumption consistent with the Law of Effect is therefore: $Q_S \neq Q_P$. As before, assume that separating strategies are used more often than pooling strategies, but with the weaker assumption that $Q_S > Q_P$ instead of $Q_P = -\infty$.

Despite the problems associated with Case C data, it is possible to test whether players update across or within type nodes. Consider the case where the data is generated under strategies with $Q_S = 0.5$, $Q_P = 0.2$, $\varphi_0 = \varphi_1 = 0.8$, and $x = 0.7$. Thus,

$$\Pi_S = \frac{\exp(\varphi_0 Q_S + \varphi_1 x_1) + \exp(\varphi_0 Q_S)}{2\exp(\varphi_0 Q_P) + \exp(\varphi_0 Q_S + \varphi_1 x_1) + \exp(\varphi_0 Q_S)} = 0.636$$

Given this probability we can determine the probability of a type I error (α) in a finite sample, and the power of the test. Table 4 gives the power of the conditional means test using this probability of message given type. As should be expected, the power of the test increases with the number of data points.

The consideration of conditional means is similar in longer games. Let t^* be the first period in which a player receives a positive payoff, and t^{**} the first period of positive payoff when the other type. For some histories it is possible to make predictions under the Law of Effect and SR rule even if all that is known of priors and parameters is: $Q_S \neq \pm\infty$, $Q_P \neq \infty$, $Q_S \neq Q_P$, and $\varphi_0, \varphi_1 \in \mathrm{R}^+$. We can strengthen the assumption $Q_S \neq Q_P$ to $Q_S > Q_P$.

Assume the above restrictions on parameters and priors, the Law of Effect, and the actions learning rule. The first time a player has received $x > 0$ for some $[\theta, m]$, $\Pr[m \mid \theta] > 0.5$ in future play, while $\Pr[m' \mid \theta'] = 0.5$. Under strategies, the same payoff also causes $\Pr[m \mid \theta] > 0.5$, but differs in that $\Pr[m' \mid \theta'] > 0.5$ in the population. Hence, by testing $\Pr[m' \mid \theta']$ we can infer whether the data was generated by actions or strategies.

For the three BDKS data sets the results of the conditional means test are given in Table 5. The number of data points is less than $T*N = 120$ because not all data points contribute to the conditional means test. Periods $t < t^*$ do not change the probabilities of behavior, due to the Law of Effect. Periods $t > t^{**}$ are confounded because both types have received a payoff, and the null and alternative hypotheses cannot be used to make predictions of behavior probabilities without observing the underlying strategies (the parameters are not identified). In either case, both groups must be discarded, and so we can use at most periods $t \in (t^*, t^{**}]$. This pool of data is further reduced when $[\theta, m']$ receives a payoff in the $(t^*, t^{**}]$ time interval. Such a payoff changes the probabilities for play, under strategies, in a way that cannot be known without observing underlying strategy choices.

For the SR model we reject the actions formulation in favor of strategies for the second two data sets, but fail to reject actions for the first. These results parallel those found by the previous decision making approach.

3. Combining Receivers and Senders

Until now only sender information has been considered, since the likelihood function factors senders and receivers so long as they have different parameter values. This assertion can be tested. First, we use the classical approach to test whether the data can be pooled, then repeat the exercises from the previous section on the full (including both senders and receivers) data sets.

a. Pooling Tests

Before combining the sender and receiver data within each data set we must test whether the senders and receivers have the same parameter values given the data set and the model. The classical test is to estimate each model m_i using the sender and receiver portion of each data set ds_j separately, retaining the maximum log likelihood function values $LLF\{s,m_i,ds_j\}$ and $LLF\{r,m_i,ds_j\}$. Then, the full data set is used to estimate each model, retaining $LLF\{f,m_i,ds_j\}$. Next we compute a test statistic by 2*[($LLF\{s,m_i,ds_j\}$ + $LLF\{r,m_i,ds_j\}$) - $LLF\{f,m_i,ds_j\}$], which is distributed chi-squared with two degrees of freedom. The null hypothesis is pooling, the alternative to reject pooling. P-values for the classical test are given in Table 6. For 3 of the 6 data set-model combinations we fail to reject pooling at the 10% level. The other 3 cases are rejected between the 5 and 1% levels. Results are reported below as though pooling were acceptable for all data set–model combinations.

b. Results from Pooled Data Sets

The full data sets have twice as many player-period observations, and the situation becomes very similar to analyzing sender-only data with $N = 12$. Each data set now has 12 players with data for 20 rounds each. The classical model test is again the J-test, computed as before, with results in Table 7.

Results are qualitatively similar to those from the sender only results. As before, we found that fitted values from $SR_{actions}$ contribute in the presence of $SR_{strategies}$ with p-values of less than 0.05. Further, the fitted values from $SR_{strategies}$ do contribute in the presence of $SR_{actions}$ with p-values of less than 0.01. Under SR we reject actions in favor of strategies, and strategies in favor of actions.

Finally, the conditional means test results on pooled data are in Table 8. There is some push toward the strategies framework. None are very convincing, as the test statistics are "close" to the critical values indicated. Using pooled data sets instead of just sender data has little impact on the qualitative results; formulation choice is mixed by data set under the SR learning rule.

IV. Conclusions and Further Research

This paper considers the distinction between "actions" and "strategies," and the relevance of how players represent new information in simple learning rules. In particular, the stimulus-response learning rule was stated in these forms for a sender-receiver game. We find that the number of players necessary for well-behaved maximum likelihood estimates is moderate, and within the reach of experimental labs. If a particular lab does not have the resources to achieve this group size, and rejects the hypothesis of pooling, then bootstrapping can be used to account for small sample sizes. In classical and conditional means testing of the two SR formulations for the three data sets considered, results were mixed. **Typically, two were borderline, or not clearly preferred to each other, while a third was strong support for making a distinction between the actions and strategies formulations[14].** Finally, the proposed conditional means test supports the earlier result that the strategy formulation outperforms actions in the last two data sets. **Adding receiver data moderates the numerical outcomes, but has little effect on any of the qualitative results.**

Consider a researcher testing the SR actions learning rule against another (possibly nested) learning rule, but, unaware of the strategies framework, estimated only the actions framework. For two of the three experimental data sets used here, we have evidence that the SR model would perform more poorly than if the strategies framework had been used. In these cases

[14] In comparing the strategy and actions formulations of each model, the strategy version is favored in two of the three data sets for SR, and for all three under BBL.

the SR model could be incorrectly rejected for another model. If that researcher used a model that nested SR, the parameter estimates and performance of the more complicated model would be compromised.

These results point to a number of avenues for further research. One is to investigate the relevance of domain specification in other learning models, particularly in more sophisticated models. For example, the experience weighted attraction model (Camerer and Ho, 1999) nests simple models for which domain specification may be a relevant concern. If comparisons are to be made with models that require population-level data, experiments that provided such information to players should be used.

Finally, the assumption here was that players are homogeneous within a repetition. One distinct possibility is that players are heterogeneous; the class of their learning rule, formulation of their particular rule, and parameter values within formulation may differ across individuals. Here we had too little data on individual players to reliably confront this potential heterogeneity. Conducting experiments that observe multiple plays of the repeated game for each player seems necessary to address potential differences among players; mixture models are one reasonable approach to analyzing this sort of data.

Figure 1. Extensive Form Representation of the Stage-Game

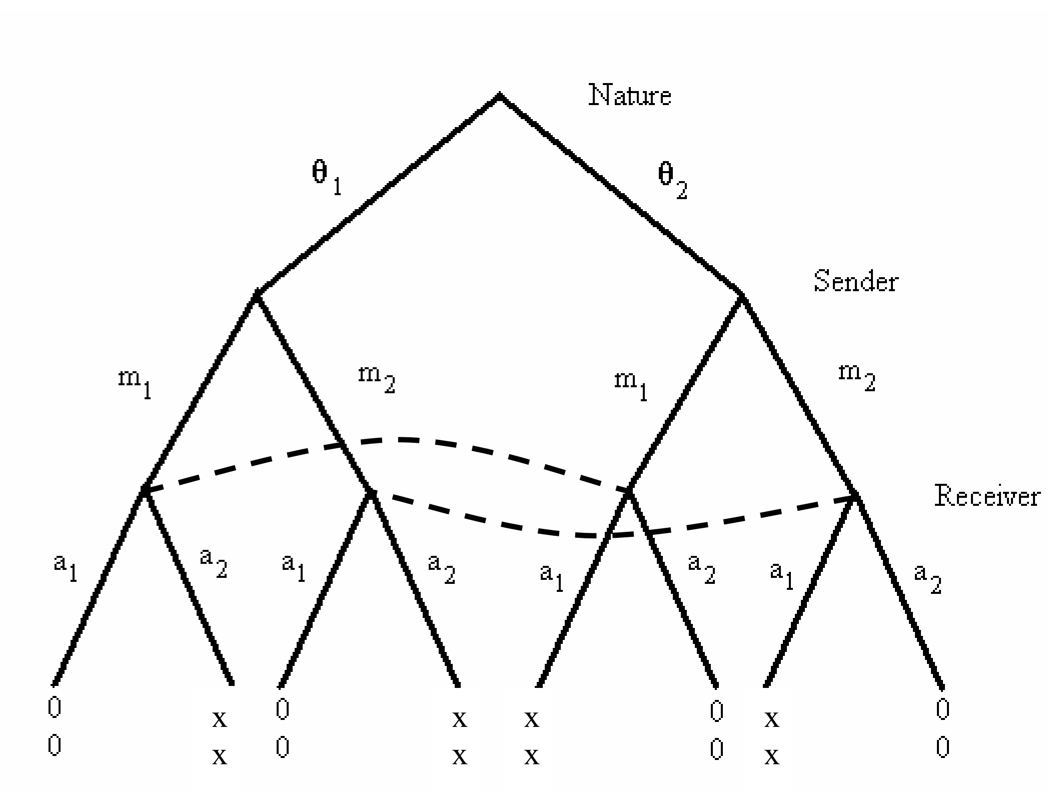

Table 1. Strategic Form Stage-Game Strategies for Senders Given Type Observation

Strategy	Variety	θ_1	θ_2
s_1	Pooling	m_1	m_1
s_2	Separating	m_1	m_2
s_3	Separating	m_2	m_1
s_4	Pooling	m_2	m_2

Table 2. J-Test Results for Experimental Data

Model 0 \ Model 1	$SR_{actions}$	$SR_{strategies}$
$SR_{actions}$	α-estimates p-estimates t-estimates	0.62,0.61,0.82 (0,0.004,*) [4.04,2.94,*]
$SR_{strategies}$	0.52,0.56,0.27 (0.04,0,0.047) [2.08,3.69,2.01]	- (-) [-]

Table 3. Outcomes in a Two Period Game

Case	Period	Type	Message	Payoff
A	1	θ	m	0 or 1
	2	θ		
B	1	θ	m	0
	2	θ'		
C	1	θ	m	1
	2	θ'		

Table 4. Power in the Conditional Means Test

N	Pr(success)	Critical Value	Pr(≤ CV)	α	Power
6	0.50	5	0.9844	0.0156	0.0662
12	0.50	9	0.9807	0.0193	0.1292
24	0.50	16	0.9680	0.0320	0.3056
400	0.50	216	0.9506	0.0494	0.9999

Table 5. Conditional Means Test Results for Experimental Data

Data Set	Data Points	Critical Value	$\#[m'\|\theta']$	Reject Actions?
1	7	6	5	No
2	18	12	13	Yes
3	16	11	13	Yes

Table 6. P-Values from the Classical Pooling Test

Data Set	$SR_{actions}$	$SR_{strategies}$
1	0.015	0.011
2	0.274	0.124
3	0.889	0.030

Table 7. J-Test Results for Experimental Data: Pooled Data Sets

Model 0 \ Model 1	$SR_{actions}$	$SR_{strategies}$
$SR_{actions}$	α-estimates p-estimates t-estimates	0.63,0.33,0.72 (0.005,0.01,0) [2.85,2.52,5.02]
$SR_{strategies}$	0.51,0.78,0.47 (0.04,0,0.006) [2.10,7.97,2.78]	- (-) [-]

Table 8. Conditional Means Test Results for Experimental Data: Pooled Data Sets

Data Set	Data Points	Critical Value	$\#[m'\|\theta']$	Reject Actions?
1	12	9	10	Yes
2	29	19	19	Yes
3	32	21	23	Yes

REFERENCES

Blume, A., D.V. DeJong, Y.G. Kim, and G.B. Sprinkle (1998): "Experimental evidence on the evolution of meaning of messages in sender-receiver games," *American Economic Review*, 88, 1323-1340.

Blume, A., D.V. DeJong, A.M. Lowen, G.R. Neumann, and N.E. Savin (2002): "Learning and experiments: the bootstrap to the rescue," Working Paper, University of Iowa and University of Pittsburgh.

Blume, A., D.V. DeJong, G.R. Neumann, and N.E. Savin (2002): "Learning and communication in sender-receiver games: An econometric investigation," *Journal of Applied Econometrics*, 17, 225-247.

Brandts, J. and G. Charness (1998): "Hot vs. cold: Sequential responses and preference stability in experimental games," Working Paper UPF 321, Departament d'Economia i Empresa.

Cabrales, A. and W. García-Fontes (2000): "Estimating learning models from experimental data," Working Paper UPF 501, Universitat Pompeu Fabra.

Camerer, C.F. and T. Ho (1999): "Experience-weighted attraction learning in normal form games," *Econometrica*, 67, 827-874.

Crawford, V.P. and J. Sobel (1982): "Strategic information transmission," *Econometrica*, 50, 1431-1451.

Dalkey, N. (1953): "Equivalence of information patterns and essentially determinate games," In *Contributions to the Theory of Games, vol. 2*, ed. by H.W. Kuhn and A.W. Tucker. Princeton: Princeton University Press, pp. 217-245.

Davidson, R. and J.G. MacKinnon (1981): "Several tests for model specification in the presence of alternative hypotheses," *Econometrica*, 49, 781-793.

Erev, I. and A.E. Roth (1998): "Predicting how people play games: Reinforcement learning in experimental games with unique mixed strategy equilibria," *American Economic Review*, 88, 320-881.

Fan, Y. and Q. Li (1995): "Bootstrapping J-type tests for non-nested regression models," *Economics Letters*, 48, 107-112.

Fudenberg, D. and D.M. Kreps (1993): "Learning mixed equilibria," *Games and Economic Behavior*, 5, 320-367.

Fudenberg, D. and D.M. Kreps (1995): "Learning in extensive-form games: I. self-confirming equilibria," *Games and Economic Behavior*, 8, 20-55.

Fudenberg, D. and D.K. Levine (1993a): "Self-confirming equilibrium," *Econometrica*, 61, 523-545.

Fudenberg, D. and D.K. Levine (1993b): "Steady state learning and Nash equilibrium," *Econometrica*, 61, 547-573.

Gibbons, R. (1992): *Game Theory for Applied Economists*. Princeton, New Jersey: Princeton University Press.

Godfrey, L.G. (1998): "Tests of non-nested regression models: Some results on small sample behavior and the bootstrap," *Journal of Econometrics*, 84, 59-74.

Greene, W.H. (1997): *Econometric Analysis*. Upper Saddle River, New Jersey: Prentice Hall.

Judd, K.L. (1998): *Numerical Methods in Economics*. Cambridge, Mass: MIT Press.

Kass, R.E. and A.E. Raftery (1995): "Bayes factors," *Journal of the American Statistical Association*, 90, 773-795.

Kohlberg, E. and J. Mertens (1986): "On the strategic stability of equilibria," *Econometrica*, 54, 1003-1037.

Kreps, D.M. and R. Wilson (1982): "Sequential equilibrium," *Econometrica*, 50, 863-894.

Kuhn, H.W. (1953): "Extensive games and the problem of information," in *Contributions to the Theory of Games, vol. 2*, ed. by H.W. Kuhn and A.W. Tucker. Princeton, N.J.: Princeton University Press, pp. 193-216.

Mailath, G.J., L. Samuelson and J.M. Swinkels (1993): "Extensive form reasoning in normal form games," *Econometrica*, 61, 273-302.

Mas-Colell, A., M.D. Whinston and J.R. Green (1995): *Microeconomic Theory*, New York: Oxford University Press.

Mitzkewitz, M. and R. Nagel (1993): "Experimental results on ultimatum games with incomplete information," *International Journal of Game Theory*, 22, 171-198.

Poirier, Dale J. (1995): *Intermediate Statistics and Econometrics*, Cambridge, Mass: MIT Press.

Roth, A.E. and I. Erev (1995): "Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term," *Games and Economic Behavior*, 8, 164-212.

Salmon, T. (2001): "An evaluation of econometric models of adaptive learning," *Econometrica*, 69, 1597-1628.

Schotter, A., K. Weigelt and C. Wilson (1994): "A laboratory investigation of multiperson rationality and presentation effects," *Games and Economic Behavior*, 6, 445-468.

Selten, R. (1967): "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes," in Heinz Sauermann, ed., *Beitreäge zur experimentellen Wirtschaftsforschung*. Tübingen: J.C.B. Mohr (Paul Siebeck), 136-168.

Selten, R., M. Mitzkewitz and G.R. Uhlich (1997): "Duopoly strategies programmed by experienced players," *Econometrica*, 65, 517-555.

Thompson, F. (1952a): *Equivalence of Games in Extensive Form*, RM 759, The Rand Corporation.

Thompson, F. (1952b): *Behavior Strategies in Finite Games*, RM 769, The Rand Corporation.

Van Damme, E.E.C. (1984): "A relation between perfect equilibria in extensive form games and proper equilibria in normal form games," *International Journal of Game Theory*, 13, 1-13.

Personal notes:

1. add new Camerer and / or Ho papers to page 1.
2. untrue claim in b. results from pooled data sets section. they are not qualitatively similar!
3. fix conclusion bolded sentence and accompanying paragraph