Estimating learning models from experimental data*

Antonio Cabrales† **Walter García-Fontes**‡

September 2000

*We thank Xiaohong Chen, Jaume García-Villar, José García-Montalvo, Margarita Genius, Brit Grosskopf, Arturo Kohatsu, Rosemarie Nagel, Albert Satorra and seminar audiences at at CEMFI, Universitat de les Illes Balears, Universidad Pablo de Olavide, University of Rotterdam, University of Amsterdam, Universitat Pompeu Fabra and at the XVII Latin American Meeting of the Econometric Society for helpful comments. The financial support of the Spanish Ministry of Education through grants DGYCIT PB96-0302 (Cabrales) and DGES PB8-1058-C03-01 (García-Fontes), Generalitat de Catalunya through grant 1999SGR-00157 (Cabrales) and 1998SGR0084 (García-Fontes), and Universitat Pompeu Fabra COFREA99.006 and COFREA99.003 are gratefully acknowledged.

†Department of Economics, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona (Spain). E-mail: antonio.cabrales@econ.upf.es. http://www.econ.upf.es/%7Ecabrales

‡Department of Economics, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona (Spain). E-mail: wgarcia@upf.es. http://puna.upf.es.

Abstract

We study the statistical properties of three estimation methods for a model of learning that is often fitted to experimental data: quadratic deviation measures without unobserved heterogeneity, and maximum likelihood with and without unobserved heterogeneity. After discussing identification issues, we show that the estimators are consistent and provide their asymptotic distribution. Using Monte Carlo simulations, we show that ignoring unobserved heterogeneity can lead to seriously biased estimations in samples which have the typical length of actual experiments. Better small sample properties are obtained if unobserved heterogeneity is introduced. That is, rather than estimating the parameters for each individual, the individual parameters are considered random variables, and the distribution of those random variables is estimated.

1 Introduction

A traditional view of equilibrium behavior is that it results from the introspection of rational players. As Fudenberg and Levine (1998) point out, this view has "nontrivial conceptual and empirical problems". Learning models are increasingly becoming an alternative foundation for equilibrium theory. A problem with this approach is that the predictions are very sensitive to the precise description of the model[1]. However, once (and if) the learning behavior of real agents is pinned down, the predictions of these models need not be less sharp than the full rationality alternatives. It is therefore important to find which learning model is empirically better founded.

One natural way to start this search is to estimate the models from field data. However, the structure of real life games is typically not well known, and there are many sources of dynamics besides learning. Thus, the use of field data requires to make a number of identifying assumptions that could confound the estimation of learning parameters. One useful alternative is to estimate learning models from experimental data. For one thing, the games played by laboratory agents are quite tightly controlled. Also, the subjects are often required to play the same exact game with different opponents over time, so the main (or only) source of dynamics is the learning process.

The purpose of this paper is to provide tools for the estimation of the form of learning used by laboratory agents. We use a quite general specification for learning, taken from Camerer and Ho (1999) (henceforth CH). We first give conditions which guarantee that the parameters are identified, that is, two different parameter vectors lead to different distributions for the data. Even when the parameters are *theoretically* identified, there may be practical identification problems. We explain these problems, as well as the conditions for identification, in terms of the underlying learning model. Basically, if the game is simple enough, or the agents are very sharp, they will realize early on what is the optimal way to play. Then the data will exhibit little or no variation over time. It is not feasible to extract parameters precisely with so little information.

We also provide conditions for consistency of the estimators, and characterize their limiting distribution, for different estimation methods (maximum likelihood and quadratic deviation)[2]. This is not trivial because the model implies

[1] See, e.g. Fudenberg and Levine (1998).

[2] Previous work in the area (CH, Tang 1996, Chen and Tang 1998, Nagel and Tang 1998, Cabrales, García-Fontes and Motta 2000) has either failed to provide the properties

that the distributions of the data from experiments will sometimes exhibit substantial heterogeneity and historical dependence. The players can end up locked into different equilibria under different realizations of the same underlying stochastic process (that is, for the same learning parameters). The different equilibria are more likely to be attractors *a priori* under different parameter vectors. This implies that with positive probability the parameter estimates will be biased even asymptotically. While the reason for this is formally similar to the familiar unit-root problem of economic time series[3], the particularities of the data are such that the usual solutions for the unit-root problem will not help here[4].

We also study the properties of the estimators in small samples using Monte Carlo simulations. When the length of the sample is from 30 to 50 periods (a usual length of actual experiments) the estimators are seriously biased. Given this problem, we propose and study an alternative method for understanding the learning behavior of experimental subjects. Rather than trying to estimate the learning coefficients of each individual, one can assume that these learning parameters are themselves random variables and estimate their distribution, that is introduce unobserved heterogeneity in the model. Unlike the individual learning coefficients, the parameters for their distribution can be estimated with small or no biases, even in small samples. However, we have assumed a particular distribution for the learning coefficients, and we have not explored exhaustively the parameter spaces for those distributions.

The structure of the paper is as follows. We first give some background on learning models in section 2. Then we describe the experience-weighted attraction learning model (CH model) in section 3. The identification conditions and the asymptotic properties of this model are described in section 4. The small sample analysis is performed in section 5. The paper ends with some conjectures on directions for further research.

---

of the estimators or has not done any statistical inference on the estimates provided.

[3]The problems caused by unit roots in standard economic models are not related to consistency, but rather to non-standard asymptotic distributions of the estimates of other parameters.

[4]We propose as an alternative to make the parameters of the model random, and estimate their distribution. This is the same solution we use to deal with small sample problems (as the problem is conceptually similar).

2 Some learning models and experimental evidence

Most of the learning models that have been considered in the literature assume that strategies (actions) that have given the agents higher payoffs in the past are used more frequently in the future. Within this general perspective, two strands of models stand out particularly. In the first type of model the agents choose the best strategy, given some average of past behavior of the other players. This includes models like fictitious play (Brown 1951, Robinson 1951) or best response dynamics (Cournot 1971, Matsui 1991). The alternative assumption, used for example in the learning by reinforcement model (Bush and Mosteller 1951, Cross 1973, Roth and Erev 1995), is that agents choose their strategies with probabilities that are roughly in proportion to some average of past payoffs. This model in turn is related to the replicator dynamics of evolutionary game theory (Taylor and Jonker 1978), as Börgers and Sarin (1997) have shown.

The model of CH unifies the two types of models with a relatively parsimonious specification[5]. For this reason we use the CH model to showcase the econometric methods. Strictly speaking the "best responsive" models are only limiting cases (when one of the parameters goes to infinity) of CH. On the other hand, the pure version of learning by reinforcement is in the strict interior of the parameter set. We will find, however, that the degree of dependence of the stochastic process of (the pure version of) learning by reinforcement or fictitious play is very high. So much, in fact, that the asymptotic properties of the estimators in those cases are poor.

We will now formally describe the models. We have an I-player game, where each player i has N_i strategies. Let s_i^j be the strategy j of player i, $s_i(t)$ denotes the strategy played by player i at time t. $s(t) = \{s_1(t), s_2(t), \ldots, s_N(t)\}$, is the strategy profile of all the players at time t, and $s_{-i}(t) = \{s_1(t), \ldots, s_{i-1}(t), s_{i+1}(t) \ldots, s_N(t)\}$ is the strategy profile of players other than i at time t. The payoff function for a player i using strategy j against a strategy profile s_{-i} for the other players is denoted $\pi_i(s_i^j, s_{-i})$.

The probability that strategy $s_i^j \in S_i$ is chosen by agent i at time $t+1$, conditional on $s(t), \ldots, s(0)$ is denoted by $P_i^j(t+1)$, that is, $P_i^j(t+1) \equiv$

[5] Cheung and Friedman (1997), Mookherjee and Sopher (1997) and Blume, DeJong, Neumann and Savin (2000) also encompass belief learning with reinforcement learning models. Although their models are not identical to CH, they are close enough to conjecture that our results should mostly apply to these models as well

$P(s_i(t+1) = s_i^j | s(t), \ldots, s(0))$.

One popular learning model that has been fitted to the data is called *learning by reinforcement* (LR) model (see Roth and Erev 1995). Let $I(s_i^j, s_i(t))$ be an indicator function that is 1 if strategy j was used by agent i at time t and 0 otherwise and $A_i^j(t)$ is the "propensity" to play strategy j.

$$A_i^j(t) = \phi_i A_i^j(t-1) + I(s_i^j, s_i(t))\pi_i(s_i^j, s_{-i}(t))$$

where ϕ_i is a parameter that measures how past information is discounted.

Notice that a strategy which is not played does not increase its "propensity". The learning model predicts that strategy j for player i will be played with probability

$$P_i^j(t+1) = \frac{A_i^j(t)}{\sum_{k=1}^{N_i} A_i^k(t)}.$$

The parameter ϕ_i (as well as the initial propensities $A_i^j(0)$) has been estimated (see for example Cabrales, García-Fontes and Motta 2000, or Chen and Tang 1998) by minimizing on a grid the quadratic deviation measure

$$QDM = \sum_{i=1}^{I}\sum_{j=1}^{N_i}\sum_{t=1}^{T} \frac{(I(s_i^j, s_i(t)) - P_i^j(t))^2}{IT\sum_{i=1}^{I} N_i}.$$

The higher the initial propensities, the more important initial play in the resulting outcomes.

Another quite popular model in the literature is the model of fictitious play (Brown 1951, Robinson 1951). In fictitious play agents best respond to their beliefs, which are formed by doing a simple average of past behavior by the opponents. A related model is the best-response dynamics, in which agents best respond to the behavior of the opponent in the last period. Cournot (1971) already has a description of oligopoly behavior in those terms. For more recent uses of the dynamics in game theory see Matsui (1991). More formally, assume that each agent i forms beliefs about the probability that her opponents k will use strategy s_k^j, which we denote by $\hat{s}_{ik}^j(t)$,

$$\hat{s}_{ik}^j(t) = (1 - \lambda_i(t))\hat{s}_{ik}^j(t-1) + \lambda_i(t)I(s_k^j, s_k(t))$$

where $\lambda_i(t) = \phi_i + \rho_i/t$. Denote by $\hat{s}_{ik}$ the vector of beliefs of player i about the strategies used by player k, and $\hat{s}_{-i}$ the vector of beliefs of player i about all other players. Let $BR_i(\hat{s}_{-i}(t))$ be the set of best responses of agent i to $\hat{s}_{-i}(t)$. Then

$$P_i^j(t+1) = \begin{cases} 1 & \text{if } \ j = BR(\hat{s}_{-i}(t)) \\ 0 & \text{if } \ j \notin BR(\hat{s}_{-i}(t)) \\ \gamma_i^j & \text{if } \ j \in BR(\hat{s}_{-i}(t)) \text{ and } j \neq BR(\hat{s}_{-i}(t)) \end{cases}$$

so that we have fictitious play strictly speaking when $\phi_i = 0$, $\rho_i = 1$ for all i; and best response dynamics when $\phi_i = 1$, $\rho_i = 0$.

The parameters ϕ_i and ρ_i as well as the initial beliefs $\hat{s}_{-i}(0)$, are also typically estimated with a grid from 0 to 1[6][7]

Two things should be noted about the results obtained so far in the two kinds of studies. The first thing is that in general, the "best responsive" models have a quite higher QDM than the reinforcement models (see Chen and Tang 1998 and Tang 1996, Cabrales, García-Fontes and Motta 2000) which warrants the conclusion that fictitious play is not a good representation in general for the way in which real agents learn and that the reinforcement model seems more adequate.[8]

The other observation is that in general the forgetting parameter ϕ_i has an intermediate value (neither 0 nor 1). This is important since the "pure" models with $\phi_i = 1$ are very popular, but we cannot provide asymptotic results for them. The parametric restrictions we use to study the properties of the estimators does not seem then to be a problem given the real behavior of agents.

Although the analysis in CH is quite general, as it encompasses the models we just described, not all learning models can be incorporated in their framework. A particularly important exception is that of games with a continuum of actions. One could in principle extend mechanically the model by discretizing the strategy space and then applying CH. But this would miss important effects produced by the structure of some games. For example, CH (like learning by reinforcement) does not reinforce strategies that have

[6]The players are typically never indifferent between actions 0 or 1, so γ does not have to be estimated.

[7]For papers where this model is estimated, see Tang (1996), Cabrales, García-Fontes and Motta (2000), Sefton (1999). Boylan and El-Gamal (1993) using a Bayesian approach try to discriminate between best response dynamics and fictitious play.

[8]Sefton (1998) estimates another variant of fictitious play with mistakes and he obtains adequate results to explain the experimental data of Clark, Kay and Sefton (1999). He does not compare his results with those for other learning models, however.

not been used in the last period. But strategies that are "close" to the one chosen should probably not be treated equally, and this is specially important in games with a continuum of strategies. Armantier (1999) generalizes learning by reinforcement for auctions and then proposes a robust estimation method. Crawford (1995) and Broseta (2000) propose and estimate adaptive learning models for coordination games in which the best response of a player only depends on a summary statistic of other players' behavior[9].

3 The experience-weighted attraction learning model

In the model of Camerer and Ho (1997), the probabilities that strategy $s_i^j \in S_i$ is chosen by agent i at time $t+1$, conditional on $s(t), \ldots, s(0)$ is now given by[10]:

$$P_i^j(t+1) = \frac{e^{\lambda_i A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda_i A_i^k(t)}}$$

where $A_i^j(t)$ is the "attraction" of strategy j for agent i at time t, which is given by:

$$A_i^j(t) = \frac{\phi_i N_i(t-1) A_i^j(t-1) + [\delta_i + (1-\delta_i) I(s_i^j, s_i(t))] \pi_i(s_i^j, s_{-i}(t))}{N_i(t)}.$$

In this formula $I(x,y)$ denotes the indicator function which is 0 if $x \neq y$ and is 1 if $x = y$, and $N_i(t)$ is a variable that is used to express the importance of past experience and is recursively defined by

$$N_i(t) = \rho_i N_i(t-1) + 1, \quad t \geq 1.$$

the variables $N_i(t)$ and $A_i^j(t)$ are started with some initial values $N_i(0)$, and $A_i^j(0)$.

Let $F_i^j(s(t)) = [\delta_i + (1-\delta_i) I(s_i^j, s_i(t))] \pi_j(s_i^j, s_{-i}(t))$, then one can also write:

[9] We have concentrated on the econometric issues of learning dynamics. A complementary strategy is followed by Haruvy and Stahl (1998). They have studied the way in which experimental subjects make their decisions on the first round of play.

[10] This expression makes the model similar to a multinomial logit or probit model. However, notice that unlike in those models, these probabilities are not derived from the distribution of some noise term in a more fundamental behavioral equation.

$$A_i^j(t) = \frac{\sum_{k=0}^{t-1} \phi_i^k F_i^j(s(t-k)) + \phi_i^t N(0) A_i^j(0)}{\rho_i^t N(0) + \sum_{k=0}^{t-1} \rho_i^k}.$$

In order to understand the meaning of the different elements of the model and the coefficients, notice first that the attractions are related to the payoffs. This means that when λ_i is large, small differences in (expected) payoffs lead to large differences in the probabilities that strategies are played. For this reason, large values of λ_i will make the model be more like fictitious play or best response, which put weight only on the "best" strategy, rather than learning by reinforcement, which put weight roughly in proportion to payoffs. Strictly speaking, the model is exactly like fictitious play only if $\lambda_i = \infty$, but for practical purposes there will not be much difference in behavior for high values of λ_i.

The parameter δ_i regulates how much more the payoff of the strategy that has actually been played in a given period gets incorporated in the attraction with respect to strategies that have not been played. With $\delta_i = 0$ only the strategy that has actually been played gets its payoff incorporated in the attractions (as it happens, for example, in learning by reinforcement) . With $\delta_i = 1$ (as in best reply or fictitious play) the payoffs to all strategies (given the strategy played by the opponent in the present period) get incorporated in the attractions.

The parameters ρ_i and ϕ_i tell us how much the past is discounted when updating attractions. When $\rho_i = 1$ and $\phi_i = 1$, we have a model like fictitious play which gives all periods the same weight, whereas $\rho_i = 0$ and $\phi_i = 0$ is more like best reply which reacts only to last period's experience.

Summarizing $\rho_i = \phi_i = 1$, $\delta_i = 1$, $\lambda_i = \infty$ for all i gives us fictitious play $\rho_i = \phi_i = 0$, $\delta_i = 1$, $\lambda_i = \infty$ gives us best reply, and $\rho_i = \phi_i = 0$, $\delta_i = 0$, $\lambda_i = 1$ gives us learning by reinforcement. In general $\rho_i = \phi_i$, and $\delta_i = 1$ gives us some kind of geometric-weighted belief model.

Let $\beta_i = (\rho_i, \phi_i, \delta_i, \lambda_i, N_i(0), A_i^j(0))$, the vector of individual parameters of this model and $\beta = (\beta_1, \beta_2, \ldots, \beta_I)$. The vector β can be estimated by minimizing a function

$$Q_n(s(1), \ldots, s(n), \beta) = n^{-1} \sum_{t=1}^{n} q(s(t), A(t-1), \beta)$$

where the function

$$q(s(t), A(t-1), \beta) = \sum_{j=1}^{J} \sum_{i=1}^{m_i} [I(s_i^j, s_i(t)) - P_i^j(t)]^2$$

in the case of minimum quadratic deviations or

$$q(s(t), A(t-1), \beta) = -\sum_{j=1}^{J} \sum_{i=1}^{m_j} [I(s_i^j, s_i(t)) \log P_i^j(t)])]$$

in the case of maximum likelihood.

In order to distinguish this two cases we will call $Q_n^Q(s(1), \ldots, s(n), \beta)$ the function that uses the minimum quadratic deviation, and $Q_n^L(s(1), \ldots, s(n), \beta)$ the function that uses the maximum likelihood. In the remainder we will assume that the parameter space B is restricted so that B is a compact set. This is true for example if there is a value $\overline{\lambda} < \infty$, such that $\overline{\lambda} \geq \lambda_i \geq 0$ for all i. The other parameters have natural bounds, so that $1 \geq \delta_i \geq 0, 1 \geq \rho_i \geq 0$ and $1 \geq \phi_i \geq 0$. But we also assume (as we will see, this is also important for our results) that there exists a $\overline{\phi} < 1$, such that $\overline{\phi} \geq \phi_i \geq 0$.

4 Asymptotic properties of the estimators

4.1 Identification

One problem with nonlinear models is that it is not always clear if the parameters are identified. We say that the parameter vector is identified, if for any two $\beta \neq \beta'$ there are realizations of the stochastic process $\{s(i)\}_{i=0}^{\infty}$ which have positive and distinct probability under β and β'. We will now show that under some mild restrictions on the parameters and the game, the model is identified. However, we will see that in some circumstances where the restrictions are satisfied, there can be a *practical* identification problem.

Proposition 1. Let β^* be the true parameter vector. Then β^* is identified, provided that for all i, $\delta_i^* \neq 1$, $N_i^*(0) \neq 1/(1-\rho_i^*)$, and that there exists a strategy profile s_{-i}^*, such that for some agent i and strategies j^*, k^*, $\pi_i(s_i^{j^*}, s_{-i}^*) \neq \pi_i(s_i^{k^*}, s_{-i}^*)$.

Proof: See appendix A.

The reasons why the assumptions are needed are easy to see. With $\delta_i^* = 1$, we have a "belief"-based model. Suppose that all strategies for some player are such that the payoff is independent of the other player's behavior. Then the belief about its payoff is independent of the "discount"-factor ϕ_i, which would not be identified[11]. If $N_i^*(0) = 1/(1-\rho_i^*)$, then $N_i^*(t)$ is a constant and therefore only $\lambda_i^*/N_i^*(t)$ is identified. Identical payoff functions for all strategies $(\pi_i(s_i^j, s_{-i}^*) = \pi_i(s_i^k, s_{-i}^*)$. for all $i, j, k)$ also create some identification problems, but this is not a very interesting case to study anyway.

But near non-identification problems can arise in other cases. When $\rho_i^* \neq 1$, $N_i^*(t)$ tends to a constant as t goes to infinity, which then multiplies λ_i^*, so that only the product $\lambda_i^*/N_i(t)^*$ can be identified in the limit. If ρ_i^* is close to 0, this convergence is very fast, which might give numerical identification problems even in small samples[12].

Another problem that is not addressed in the theoretical result is that "best responsive" dynamics can also lead to identification problems. Suppose that one strategy is strictly dominant (or is a unique best response given initial attractions). Then, that is the only strategy that will be played, irrespective of the values of ϕ_i^*, or δ_i^* which would not be identified. This problem does not arise in Proposition 1 because we have already assumed that $\lambda_i^* \neq \infty$, and thus excluded the case. However, for practical matters a high value of λ_i^* can also create serious identification problems (which typically show up in slow convergence of the estimation).

4.2 Consistency

The estimators $\hat{\beta}_n^Q$ and $\hat{\beta}_n^L$ are consistent estimators of β. To see this we first have to show that a sequence of minimizers of the expectation of the $Q_n^k(s(1), \ldots, s(n), \beta)$ function converges to the true vector of parameters, β^*[13].

[11] This also requires that $\phi_i^* = \rho_i^*$ and $(A_i^j(0))^* = \pi_i(s_i^j, s_{-i})$, for exact nonidentification, but in practice, $\delta_i^* = 1$ is enough as the effect of $N_i^*(t)$ goes away very fast when $\rho_i^* \neq 1$.

[12] In our Monte Carlo simulations we only estimate ϕ_i, δ_i and λ_i, since we were unable to estimate ρ_i or $N_i(0)$ with any reasonable precision. For practical purposes this is not a very serious shortcoming of this model, since only the first three parameters have a clear economic interpretation.

[13] This result assumes, as we always do in this paper that the CH model is the true data generating process. Nevertheless the results in this section go through even if the model is misspecified, provided that one interprets then β^* as the value of β which minimizes the KLIC (Kullback-Leibler Information Criterion). See e.g. White (1996).

Lemma 1. Denote by β^* the true vector of parameters, then

$\beta^* = \text{argmin}_\beta E(Q_n^k(s(1), \ldots, s(n), \beta)$, for all $k \in \{Q, L\}$.

Proof: See appendix A.

Another necessary intermediate step is to show that the stochastic process $s(t)$ is α-mixing.

Definition 1. Let $(s(t))_{t\in R}$ be a stochastic process on $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_1^i$ be the σ-field generated by sequences $(s(1), \ldots, s(i))$ and $\mathcal{F}_k^\infty$ be the σ-field generated by sequences $(s(k), s(k+1), \ldots)$. Define

$$\alpha(l) = \sup_{k\in\mathfrak{N}} \sup\{|P(F \cap G) - P(F)P(G)| : G \in \mathcal{F}_1^k, F \in \mathcal{F}_{k+l}^\infty\},$$

$$\phi(l) = \sup_{k\in\mathfrak{N}} \sup\{|P(F|G) - P(F)| : G \in \mathcal{F}_1^k, F \in \mathcal{F}_{k+l}^\infty, P(G) > 0\}.$$

If $\alpha(l)$ ($\phi(l)$) goes to zero as l approaches infinity, we call the process α-mixing (ϕ-mixing). Notice that every ϕ-mixing process is α-mixing.

Lemma 2. The process $(s(t))_{t\in R}$ is ϕ-mixing of size $-r/(r-1)$, for $r > 2$. That is, the sequence $\phi(j)$ is $O(m^{-\lambda})$ for some $\lambda > r/(r-1)$.

Proof: See appendix A.

We use in this proof the fact that $\overline{\phi} < 1$. Without this condition the $s(t)$ process need not be mixing. To see why, consider a one-player, two-strategy game with $\pi_i(s_i^j, s_{-i}(t)) = 1$, for the two strategies. Let $s(t)$ be defined by the limiting case that $\phi_1 = 1$ and also that $\lambda_1 = 1, \delta_1 = 0, \rho_1 = 1$.

Define $z_{u,t} = (s(u), s(u+1), ..., s(t))$, and let $z_{1,k}$ such that $s(1) = ... = s(k) = 1$, and $z'_{1,k}$ such that $s(1) = ... = s(k) = 0$. Let also $N_{z_{k+1,l}}$ be the number of times that strategy 0 is used between times k and $k + l$.

Then we have that

$$P(s(k+l) = 0|z_{1,k}, z_{k+1,l}) = \exp(\frac{N_{z_{k+1,l}} + k}{k+l} - 1)$$

and

$$P(s(k+l) = 0|z'_{1,k}, z_{k+1,l}) = \exp(\frac{N_{z_{k+1,l}}}{k+l} - 1).$$

Thus

$$\frac{P(s(k+l) = 0|z_{1,k}, z_{k+1,l})}{P(s(k+l) = 0|z'_{1,k}, z_{k+1,l})} = \exp(\frac{k}{k+l})$$

Therefore since we can make $k = l$ we have sets F and G, such that for some constant K

$$\sup_{k \in \mathfrak{N}} \sup\{|P(F \cap G) - P(F)P(G)| : G \in \mathcal{F}_1^k, F \in \mathcal{F}_{k+l}^{\infty}\} > K$$

and the process is not mixing.

As one can easily see the problem is that the definition of mixing requires that the weight of observations of the process that are more than l periods in the past have to go to zero as l goes to infinity *independently of the total amount of time that the stochastic process has been running ($k+l$)*. But when $\phi_i = 1$, then the weight of *any* l observations in the conditional probabilities is precisely their weight in the total length of time that the game has been played.

Proposition 2. The estimators $\hat{\beta}_n^Q$ and $\hat{\beta}_n^L$ are consistent estimators of β^*[14]. That is

$$|\hat{\beta}_n^Q - \beta^*| \to 0 \text{ } i.p. \text{ } as \text{ } n \to \infty$$

$$|\hat{\beta}_n^L - \beta^*| \to 0 \text{ } i.p. \text{ } as \text{ } n \to \infty$$

Proof: Lemma 1 shows that $\beta^* = \text{argmin}_\beta E(Q_n^k(s(1), \ldots, s(n), \beta)$, for all $k \in \{Q, L\}$, so by theorem 7.1 (p. 81) in Pötscher and Prucha 1997, the result follows provided the following conditions hold:

1. The process $s(t)$ is defined in a subset of $\mathfrak{R}^p$, and the parameter space B is a compact metric space.
2. The process $s(t)$ is α-mixing.
3. The function $q^k(s(t), A(t-1), \beta)$ is continuous.
4. $\sup_n n^{-1} \sum_{t=1}^n E|s(t)| < \infty$.

Conditions 1 and 4 are satisfied by the finiteness of the strategy spaces we consider and the limits we impose on the parameters. Condition 2 is verified in Lemma 2. Condition 3 is easily verified by inspection of function $q(s(t), A(t-1), \beta)$.□

[14] See footnote 13.

The problem with $\phi_i = 1$ is akin to the one of a time series having a unit root. The standard approach to solving a unit root problem is (after testing for its existence) to do some transformation to the series, such as differencing it, to estimate other parameters of the model. The type of data that we are dealing with may create problems that cannot be solved in this way.

To see why, notice that with $\phi_i = 1$ Posch (1996) shows[15] that the $P_i^j(t)$ process converges almost surely to one of the strict equilibria in a 2×2 coordination game. The likelihood of convergence to the different equilibria (for given starting values) is different for different parameters. For example, with very high values of λ_i the process is very likely to converge to the equilibrium in whose basin of attraction it starts, whereas the likelihood of convergence to that equilibrium is less strong (but still positive) for smaller values of λ_i. So observing that the process converges to that equilibrium will yield a high estimate of λ_i, even if the real value is a low one (which can be true with positive probability). Consequently the estimate of β cannot converge in probability to its true value, and Proposition 2 would not hold.

One solution for this problem could be to use the information available from other individuals or sessions. This is the approach we propose to deal with the small sample problems that we uncover in section 5.

4.3 Asymptotic normality

For every $k \in \{Q, L\}$, let

$$C_n^k = E(\nabla_{\beta\beta}(Q_n^k(s(1), \ldots, s(n), \beta)))|_{\beta^*}$$

and

$$D_n^k = (nE(\nabla_{\beta'}(Q_n^k(s(1), \ldots, s(n), \beta)\nabla_{\beta}(Q_n^i(s(1), \ldots, s(n), \beta)))|_{\beta^*})^{1/2}.$$

Let $\lambda_{\min}^{kn}$ be the minimum eigenvalue of the matrix $E(\nabla_{\beta\beta}(Q_n^k(s(1), \ldots, s(n), \beta)))$.

Proposition 3. Assume that for all k, $\liminf_{n\to\infty} \lambda_{\min}^{kn} > 0$. Then the estimators $\hat{\beta}_n^Q$ and $\hat{\beta}_n^L$ are asymptotically normal estimators of β^*. That is, for every $k \in \{Q, L\}$,

[15] Using stochastic approximation theory.

$$n^{1/2}(\hat{\beta}_n^k - \beta^*) = (C_n^k)^{-1} D_n^k \zeta_n + o_p(1)$$

with

$$\zeta_n \to N(0, I)$$

and

$$n^{1/2}(D_n^k)^{-1} C_n^k (\hat{\beta}_n^k - \beta^*) \to N(0, I)$$

Proof: By theorem 11.2 (p. 108) in Pötscher and Prucha 1997 the result follows provided the following conditions hold:

1. The process $s(t)$ is defined in a subset of $\mathfrak{R}^p$, and the parameter space B is a compact metric space.
2. The functions $q^k(s(t), .A(t-1), .)$ are twice continuously partially differentiable at every point in B, and the function $q^k(.,., \beta)$ is measurable in S.
3. The sequence of estimators $\hat{\beta}_n^k$ satisfies $|\hat{\beta}_n^k - \beta^*| = o_p(1)$.
4. $\sup_n n^{-1} \sum_{t=1}^n E|s(t)| < \infty$.
5. The function $\nabla_{\beta\beta}(q^k(s(t), A(t-1), \hat{\beta}_n^k))$ is continuous on S, B, and
$$\sup_n n^{-1} \sum_{t=1}^n E[\sup |\nabla_{\beta\beta}(q^k(s(t), A(t-1), \hat{\beta}_n^k))|^2] < \infty$$
6.
$$E(\nabla_\beta(Q_n^k(s(1), \ldots, s(n), \beta^*)) = 0$$
7.
$$\liminf_{n\to\infty} \lambda_{\min}^{kn} E(\nabla_{\beta\beta}(Q_n^k(s(1), \ldots, s(n), \beta))) > 0$$
8. The process $(s(t))_{t\in R}$ is ϕ-mixing of size $-r/(r-1)$, for $r > 2$.

Proof: Conditions 1 and 4 are satisfied by the finiteness of the strategy spaces we consider and the limits we impose on the parameters. Condition 8 is verified in Lemma 2. Conditions 2 and 5 are easily verified by inspection of function $q^k(s(t), A(t-1), \beta)$. Condition 3 follows from Proposition 2. Condition 6 follows from the way in which estimators are obtained. Condition 7 is an assumption of the proposition.□

Remark. It is possible that the assumption of the theorem is satisfied trivially, as we have not verified it. In our Monte Carlo simulations we have never encountered any invertibility problems of that matrix.

5 Monte Carlo simulations

In this section, we perform Monte Carlo simulations using the Quadratic Deviation and Maximum Likelihood estimators. We consider two games, the 2×2 coordination game studied in Cabrales, García-Fontes and Motta (2000), based on the following payoff matrix:

		Choice of B	
		0	1
Choice of A	0	2.92, 1.64	6.12, 3.62
	1	3.64, 5.09	3.67, 3.46

and a dominance solvable game:

		Choice of B	
		0	1
Choice of A	0	8, 1	6, 3
	1	5, 5	3, 6

We simulate the data for 14 individuals, 7 of type A and 7 of type B. Each period each player of type $i, i \in \{A, B\}$ is randomly matched with a player of type $i', i' \in \{A, B\}, i' \neq i$. Initial values for the attractions are fixed as as an average of the possible payoffs. For instance for the coordination game, the attractions for player $i \in \{A, B\}$ for strategy $j \in \{0, 1\}$ are:

$$A_A^0 = (2.92 + 6.12)/2$$

$$A_A^1 = (3.64 + 3.67)/2$$

$$A_B^0 = (1.64 + 5.09)/2$$
$$A_B^1 = (3.62 + 3.46)/2$$

We choose the attractions for the dominance solvable game in a similar way.

We will only estimate ϕ_i, δ_i and λ_i for the learning models (we were not able to solve the identification problems for ρ_i and the initial value for N_i in the denominator of the attractions). We fix the values $\rho_i = 0.5$ and $N_i = 1$. We consider different combinations of values for ϕ_i, δ_i and λ_i for the different individuals, which are given in tables 1 and 2 for the coordination game, and tables 3 and 4 for the dominance solvable game.

The recursive procedure for generating the data works as follows: starting from the initial values of the attractions, we generate the probabilities of playing strategies 0 and 1 for each player for period 1. Conditional on these probabilities we draw from a binomial distribution to generate the actual choice of strategy for period 1[16].

Tables 1, 2, 3 and 4 report the results from 200 simulations for each individual using sample sizes of $T = 30$ and $T = 1000$. Initial values for the parameters to be estimated were obtained by a grid search in the interval $[0, 1]$ of size 0.1. We report the mean of the estimated values over the 200 simulations, the median, the Root Mean Squared Error (RMSE) and the asymptotic coverage, that is, the percentage of observations that fall outside 2/1.645/1 asymptotic standard errors of the true parameter values, as computed from the asymptotic normal distribution obtained in the previous sections. The values reported in the tables should be compared to the percentages 5/10/32.

The performance of the estimators in the small sample ($T = 30$) is poor, as can be seen in tables 1 and 3[17]. There are also numerical problems, as shown by the high number of cases where the optimization procedures do not find the minimizing vectors. Not only the point estimates are not very precise,

[16] The actual algorithm that we use is the following: we draw a random number u^0 from a uniform distribution in the interval $[0, 1]$. If p^0 is the probability of playing strategy 0 and if $u^0 \leq p^0$ we choose strategy 0, otherwise we choose strategy 1. Given the choices for period 1 we generate the attractions for period 1 and the probabilities for period 2, and we continue recursively until the last period. The random number generator for both the actual strategy choice, as well as the random matching of players, is started with the same initial values so that the data generated for testing the different estimating procedures are the same.

[17] We also tried with samples of 50 and 100 periods, and the results did not improve substantially. The precision of the estimates started to look reasonable only for samples of around 500 periods.

Table 1: Simulation results: Coordination game ($T = 30$)

	QDM					MLE				
	Mean	Median	RMSE	Asympt. coverage	Did not converge	Mean	Median	RMSE	Asympt. coverage	Did not converge
Individual 1 (A)					71					65
$\phi = 0.5$	0.32	0.39	0.21	29/32/46		0.33	0.42	0.19	22/25/42	
$\delta = 0.5$	0.54	0.61	0.19	32/38/50		0.54	0.60	0.21	27/33/47	
$\lambda = 0.5$	0.89	0.73	0.57	31/38/53		0.89	0.69	0.80	21/29/50	
Individual 2 (B)					101					101
$\phi = 0.5$	0.28	0.41	0.28	22/25/43		0.29	0.42	0.25	19/24/40	
$\delta = 0.5$	0.45	0.63	0.46	29/32/45		0.39	0.60	0.44	32/34/44	
$\lambda = 0.5$	0.72	0.65	0.32	22/29/41		0.74	0.66	0.30	18/23/38	
Individual 3 (A)					88					80
$\phi = 0.2$	0.14	0.14	0.11	14/21/47		0.17	0.20	0.10	21/25/44	
$\delta = 0.2$	0.27	0.43	0.27	32/37/53		0.28	0.44	0.24	23/31/45	
$\lambda = 0.6$	1.13	0.84	1.19	37/43/60		1.00	0.78	0.84	35/40/54	
Individual 4 (B)					126					131
$\phi = 0.7$	0.28	0.49	0.45	27/31/44		0.28	0.50	0.46	27/35/51	
$\delta = 0.7$	0.66	0.83	0.32	20/25/33		0.60	0.87	0.45	17/25/36	
$\lambda = 0.4$	0.71	0.57	0.35	19/24/37		0.68	0.60	0.34	11/19/35	
Individual 5 (A)					102					89
$\phi = 0.2$	0.15	0.21	0.15	16/23/45		0.12	0.20	0.16	14/17/41	
$\delta = 0.2$	0.20	0.33	0.28	38/45/57		0.18	0.34	0.32	35/41/50	
$\lambda = 0.6$	0.97	0.71	0.71	38/40/56		0.97	0.76	0.80	32/41/53	
Individual 6 (B)					118					122
$\phi = 0.7$	0.25	0.48	0.50	28/29/48		0.32	0.49	0.41	31/33/43	
$\delta = 0.7$	0.69	0.77	0.26	23/28/42		0.64	0.76	0.29	16/24/33	
$\lambda = 0.4$	0.97	0.68	1.33	27/34/53		0.88	0.68	0.57	24/31/45	
Individual 7 (A)					98					80
$\phi = 0.2$	0.15	0.22	0.14	22/27/49		0.15	0.17	0.14	21/27/44	
$\delta = 0.2$	0.31	0.37	0.25	39/42/54		0.25	0.34	0.33	33/39/50	
$\lambda = 0.6$	1.09	0.83	1.02	40/47/65		0.93	0.82	0.50	37/44/61	
Individual 8 (B)					136					122
$\phi = 0.7$	0.38	0.57	0.36	30/33/48		0.44	0.58	0.31	22/26/47	
$\delta = 0.7$	0.63	0.80	0.40	28/29/41		0.68	0.78	0.36	26/31/44	
$\lambda = 0.4$	0.60	0.62	0.24	25/30/47		0.60	0.63	0.22	17/26/47	
Individual 9 (A)					112					106
$\phi = 0.7$	0.45	0.55	0.30	25/31/50		0.47	0.54	0.28	25/27/47	
$\delta = 0.7$	0.82	0.81	0.19	16/20/35		0.80	0.81	0.18	16/20/30	
$\lambda = 0.4$	1.04	0.67	2.17	31/34/53		0.80	0.64	0.61	20/30/49	
Individual 10 (B)					104					97
$\phi = 0.2$	0.19	0.26	0.11	23/24/45		0.14	0.27	0.14	16/19/40	
$\delta = 0.2$	0.20	0.35	0.34	27/33/45		0.17	0.35	0.38	31/37/45	
$\lambda = 0.6$	1.06	0.75	1.22	30/38/51		0.94	0.82	0.53	23/28/50	
Individual 11 (A)					120					117
$\phi = 0.7$	0.44	0.55	0.29	29/33/48		0.43	0.54	0.30	36/41/54	
$\delta = 0.7$	0.78	0.78	0.24	26/28/41		0.75	0.80	0.29	27/29/42	
$\lambda = 0.4$	0.93	0.64	1.24	23/29/51		0.71	0.63	0.38	23/28/55	
Individual 12 (B)					94					89
$\phi = 0.2$	0.16	0.22	0.13	8/12/26		0.11	0.18	0.17	13/14/28	
$\delta = 0.2$	0.15	0.29	0.32	34/37/48		0.15	0.27	0.31	28/34/43	
$\lambda = 0.6$	0.84	0.70	0.50	26/31/52		0.88	0.70	0.43	20/28/45	
Individual 13 (A)					91					83
$\phi = 0.5$	0.34	0.40	0.20	17/18/29		0.32	0.40	0.23	21/21/38	
$\delta = 0.5$	0.58	0.61	0.21	23/31/42		0.58	0.62	0.22	23/29/37	
$\lambda = 0.5$	1.15	0.72	2.60	24/31/44		0.89	0.74	0.61	21/27/42	
Individual 14 (B)					109					100
$\phi = 0.5$	0.30	0.40	0.23	19/19/34		0.30	0.42	0.24	19/20/37	
$\delta = 0.5$	0.43	0.57	0.41	21/28/39		0.44	0.55	0.32	22/26/42	
$\lambda = 0.5$	0.90	0.63	1.16	19/22/41		0.73	0.62	0.26	10/14/34	

Table 2: Simulation results: Coordination game ($T = 1000$)

	QDM					MLE				
	Mean	Median	RMSE	Asympt. coverage	Did not converge	Mean	Median	RMSE	Asympt. coverage	Did not converge
Individual 1 (A)					0					0
$\phi = 0.5$	0.49	0.50	0.007	5/9/42		0.49	0.50	0.008	5/8/43	
$\delta = 0.5$	0.49	0.49	0.010	3/7/50		0.49	0.50	0.009	3/4/45	
$\lambda = 0.5$	0.50	0.49	0.006	7/10/40		0.50	0.49	0.006	6/8/42	
Individual 2 (B)					2					5
$\phi = 0.5$	0.50	0.50	0.006	7/10/45		0.50	0.50	0.006	8/11/43	
$\delta = 0.5$	0.50	0.51	0.006	6/8/48		0.50	0.51	0.006	4/6/45	
$\lambda = 0.5$	0.50	0.49	0.007	9/11/42		0.50	0.49	0.006	6/8/42	
Individual 3 (A)					9					12
$\phi = 0.2$	0.20	0.20	0.004	6/10/42		0.20	0.20	0.004	4/7/38	
$\delta = 0.2$	0.20	0.21	0.010	6/16/46		0.20	0.21	0.009	2/9/46	
$\lambda = 0.6$	0.62	0.61	0.017	12/15/51		0.61	0.61	0.015	5/8/53	
Individual 4 (B)					29					45
$\phi = 0.7$	0.69	0.70	0.007	9/10/48		0.69	0.70	0.010	11/13/49	
$\delta = 0.7$	0.70	0.71	0.006	5/16/44		0.69	0.70	0.009	4/5/45	
$\lambda = 0.4$	0.41	0.40	0.008	7/15/46		0.41	0.40	0.011	10/10/45	
Individual 5 (A)					10					10
$\phi = 0.2$	0.19	0.20	0.009	5/7/40		0.19	0.19	0.008	5/7/42	
$\delta = 0.2$	0.20	0.20	0.013	10/15/45		0.20	0.21	0.012	7/13/45	
$\lambda = 0.6$	0.62	0.60	0.017	12/16/49		0.61	0.61	0.016	7/16/51	
Individual 6 (B)					26					40
$\phi = 0.7$	0.69	0.69	0.008	10/12/48		0.69	0.69	0.009	10/12/49	
$\delta = 0.7$	0.69	0.70	0.008	3/6/41		0.69	0.70	0.010	3/5/40	
$\lambda = 0.4$	0.41	0.41	0.010	7/8/45		0.41	0.41	0.008	6/9/43	
Individual 7 (A)					16					10
$\phi = 0.2$	0.20	0.20	0.005	4/7/38		0.20	0.20	0.004	4/7/35	
$\delta = 0.2$	0.19	0.20	0.013	7/15/41		0.19	0.19	0.019	4/10/47	
$\lambda = 0.6$	0.60	0.60	0.008	14/18/46		0.60	0.59	0.007	12/17/47	
Individual 8 (B)					30					45
$\phi = 0.7$	0.69	0.70	0.008	6/8/40		0.69	0.70	0.008	8/11/41	
$\delta = 0.7$	0.70	0.70	0.006	6/9/45		0.69	0.70	0.009	6/9/46	
$\lambda = 0.4$	0.41	0.41	0.009	6/8/41		0.40	0.41	0.006	6/8/40	
Individual 9 (A)					0					0
$\phi = 0.7$	0.71	0.71	0.008	4/6/37		0.71	0.71	0.008	6/9/38	
$\delta = 0.7$	0.70	0.70	0.005	7/11/47		0.70	0.70	0.005	9/12/47	
$\lambda = 0.4$	0.39	0.40	0.008	5/8/42		0.39	0.40	0.009	4/8/42	
Individual 10 (B)					12					16
$\phi = 0.2$	0.20	0.19	0.006	3/4/41		0.20	0.19	0.006	2/3/38	
$\delta = 0.2$	0.20	0.21	0.009	3/7/46		0.20	0.21	0.009	7/8/45	
$\lambda = 0.6$	0.60	0.60	0.008	9/11/46		0.61	0.60	0.009	6/8/46	
Individual 11 (A)					0					0
$\phi = 0.7$	0.70	0.70	0.003	8/12/41		0.70	0.70	0.002	10/13/43	
$\delta = 0.7$	0.71	0.71	0.009	10/14/47		0.71	0.71	0.009	9/12/44	
$\lambda = 0.4$	0.40	0.39	0.006	7/12/40		0.40	0.40	0.006	10/14/41	
Individual 12 (B)					6					8
$\phi = 0.2$	0.19	0.19	0.011	4/5/46		0.19	0.19	0.014	5/8/43	
$\delta = 0.2$	0.20	0.20	0.009	5/10/47		0.20	0.21	0.009	6/9/44	
$\lambda = 0.6$	0.61	0.60	0.010	5/6/50		0.61	0.61	0.011	6/7/50	
Individual 13 (A)					0					0
$\phi = 0.5$	0.50	0.50	0.004	7/8/44		0.50	0.50	0.004	8/10/41	
$\delta = 0.5$	0.50	0.51	0.006	6/9/46		0.50	0.51	0.006	4/7/43	
$\lambda = 0.5$	0.51	0.50	0.007	8/10/44		0.50	0.50	0.007	8/10/46	
Individual 14 (B)					1					2
$\phi = 0.5$	0.50	0.50	0.005	7/9/45		0.50	0.50	0.005	10/12/46	
$\delta = 0.5$	0.50	0.51	0.007	6/10/48		0.50	0.51	0.007	6/9/52	
$\lambda = 0.5$	0.50	0.49	0.006	7/8/47		0.50	0.49	0.005	8/11/44	

Table 3: Simulation results: Dominance solvable game ($T = 30$)

	QDM					MLE				
	Mean	Median	RMSE	Asympt. coverage	Did not converge	Mean	Median	RMSE	Asympt. coverage	Did not converge
Individual 1 (A)					130					129
$\phi = 0.5$	0.11	0.14	0.41	34/36/60		0.12	0.16	0.41	32/35/61	
$\delta = 0.5$	0.68	0.74	0.30	31/39/64		0.65	0.68	0.31	24/30/50	
$\lambda = 0.5$	1.06	0.94	0.75	40/43/59		1.12	0.90	0.90	41/48/56	
Individual 2 (B)					98					93
$\phi = 0.5$	0.25	0.28	0.28	11/15/32		0.21	0.20	0.32	17/22/40	
$\delta = 0.5$	0.48	0.59	0.40	19/24/41		0.61	0.63	0.55	21/28/43	
$\lambda = 0.5$	0.84	0.74	0.51	16/22/32		0.84	0.72	0.43	20/26/34	
Individual 3 (A)					74					67
$\phi = 0.2$	0.04	0.06	0.19	15/20/47		0.03	0.04	0.20	21/26/48	
$\delta = 0.2$	0.17	0.22	0.22	26/29/46		0.20	0.23	0.28	25/29/44	
$\lambda = 0.6$	0.75	0.63	0.26	23/28/47		0.78	0.61	0.32	25/32/47	
Individual 4 (B)					121					124
$\phi = 0.7$	0.30	0.31	0.44	34/41/51		0.31	0.28	0.41	41/43/54	
$\delta = 0.7$	0.84	0.82	0.28	18/19/28		0.78	0.75	0.19	17/17/29	
$\lambda = 0.4$	1.04	0.92	0.84	44/49/58		0.99	0.86	0.73	41/47/57	
Individual 5 (A)					96					76
$\phi = 0.2$	0.02	-0.01	0.20	16/21/43		-0.03	-0.05	0.25	19/23/53	
$\delta = 0.2$	0.22	0.31	0.33	37/45/60		0.16	0.27	0.32	32/38/52	
$\lambda = 0.6$	0.83	0.69	0.31	24/30/41		0.87	0.71	0.40	30/36/54	
Individual 6 (B)					114					116
$\phi = 0.7$	0.22	0.28	0.52	43/48/59		0.22	0.26	0.52	43/49/58	
$\delta = 0.7$	0.78	0.83	0.21	19/24/35		0.75	0.78	0.21	18/20/35	
$\lambda = 0.4$	1.06	0.92	0.77	48/50/64		1.14	0.89	0.93	43/50/65	
Individual 7 (A)					80					72
$\phi = 0.2$	0.02	0.01	0.20	13/16/41		-0.02	-0.02	0.23	18/23/52	
$\delta = 0.2$	0.11	0.20	0.35	33/34/51		0.01	0.09	0.44	36/41/57	
$\lambda = 0.6$	0.77	0.72	0.26	23/32/46		0.78	0.64	0.37	29/34/50	
Individual 8 (B)					120					122
$\phi = 0.7$	0.30	0.32	0.42	45/50/61		0.30	0.37	0.44	37/42/50	
$\delta = 0.7$	0.82	0.86	0.49	20/23/43		0.89	0.87	0.36	27/29/38	
$\lambda = 0.4$	1.00	0.89	0.74	46/51/61		1.01	0.76	0.75	38/44/56	
Individual 9 (A)					168					165
$\phi = 0.7$	0.18	0.28	0.54	56/72/88		0.14	0.13	0.59	57/63/74	
$\delta = 0.7$	1.00	1.02	0.31	16/19/34		0.95	0.92	0.29	14/17/31	
$\lambda = 0.4$	1.32	1.24	1.03	63/66/78		1.35	1.19	1.19	57/57/63	
Individual 10 (B)					78					84
$\phi = 0.2$	0.05	0.08	0.20	14/16/30		0.05	0.07	0.21	13/15/28	
$\delta = 0.2$	0.13	0.27	0.57	21/26/41		0.20	0.29	0.41	23/29/37	
$\lambda = 0.6$	0.92	0.76	0.53	25/30/43		0.89	0.77	0.45	25/32/48	
Individual 11 (A)					163					160
$\phi = 0.7$	0.09	0.10	0.63	68/76/89		0.07	0.06	0.67	68/70/83	
$\delta = 0.7$	0.88	0.94	0.30	22/30/43		0.84	0.85	0.44	18/20/28	
$\lambda = 0.4$	1.37	1.36	1.08	54/62/70		1.33	1.39	1.02	73/73/78	
Individual 12 (B)					76					82
$\phi = 0.2$	0.11	0.12	0.15	13/17/31		0.11	0.12	0.14	6/8/22	
$\delta = 0.2$	-0.01	0.19	0.66	27/31/43		-0.05	0.09	0.55	15/21/34	
$\lambda = 0.6$	0.80	0.67	0.42	23/31/53		0.70	0.64	0.21	15/21/45	
Individual 13 (A)					126					128
$\phi = 0.5$	0.15	0.20	0.38	30/31/54		0.10	0.19	0.44	33/40/67	
$\delta = 0.5$	0.53	0.63	0.37	34/38/54		0.53	0.61	0.26	29/33/56	
$\lambda = 0.5$	0.93	0.72	0.60	30/35/50		0.94	0.71	0.59	40/46/54	
Individual 14 (B)					100					100
$\phi = 0.5$	0.22	0.31	0.32	22/25/38		0.20	0.29	0.37	25/30/44	
$\delta = 0.5$	0.44	0.49	0.42	21/28/38		0.41	0.41	0.40	25/27/43	
$\lambda = 0.5$	0.91	0.71	0.65	32/35/53		0.87	0.69	0.53	26/30/45	

Table 4: Simulation results: Dominance solvable game ($T = 1000$)

	QDM					MLE				
	Mean	Median	RMSE	Asympt. coverage	Did not converge	Mean	Median	RMSE	Asympt. coverage	Did not converge
Individual 1 (A)					0					0
$\phi = 0.5$	0.47	0.48	0.04	13/16/42		0.48	0.49	0.03	11/15/36	
$\delta = 0.5$	0.52	0.52	0.03	11/15/40		0.52	0.51	0.02	15/18/46	
$\lambda = 0.5$	0.55	0.53	0.06	11/18/45		0.54	0.52	0.04	15/22/45	
Individual 2 (B)					2					5
$\phi = 0.5$	0.49	0.50	0.02	8/11/40		0.49	0.50	0.02	11/14/39	
$\delta = 0.5$	0.49	0.49	0.02	6/9/37		0.49	0.49	0.02	7/9/31	
$\lambda = 0.5$	0.51	0.50	0.02	6/11/40		0.52	0.50	0.03	7/10/37	
Individual 3 (A)					9					12
$\phi = 0.2$	0.20	0.19	0.01	4/6/38		0.20	0.19	0.01	6/10/44	
$\delta = 0.2$	0.20	0.20	0.01	6/10/44		0.20	0.20	0.01	7/10/45	
$\lambda = 0.6$	0.61	0.61	0.01	7/9/49		0.61	0.61	0.01	6/11/48	
Individual 4 (B)					29					45
$\phi = 0.7$	0.61	0.64	0.09	13/18/37		0.63	0.65	0.07	13/17/37	
$\delta = 0.7$	0.72	0.73	0.02	14/17/36		0.71	0.73	0.02	12/15/36	
$\lambda = 0.4$	0.52	0.48	0.13	14/19/38		0.50	0.48	0.11	14/18/38	
Individual 5 (A)					10					10
$\phi = 0.2$	0.18	0.18	0.02	3/4/48		0.19	0.19	0.02	6/8/47	
$\delta = 0.2$	0.20	0.20	0.01	4/6/44		0.20	0.19	0.01	6/9/52	
$\lambda = 0.6$	0.62	0.62	0.02	7/11/50		0.62	0.61	0.02	5/9/51	
Individual 6 (B)					26					40
$\phi = 0.7$	0.62	0.64	0.10	17/22/43		0.62	0.64	0.08	19/21/42	
$\delta = 0.7$	0.73	0.73	0.03	14/19/43		0.72	0.73	0.02	14/18/41	
$\lambda = 0.4$	0.52	0.48	0.13	16/23/47		0.51	0.47	0.12	19/22/43	
Individual 7 (A)					16					10
$\phi = 0.2$	0.19	0.20	0.01	6/7/44		0.19	0.19	0.01	6/9/45	
$\delta = 0.2$	0.20	0.20	0.01	3/6/43		0.21	0.20	0.01	6/7/44	
$\lambda = 0.6$	0.61	0.60	0.01	9/12/43		0.61	0.61	0.01	7/9/41	
Individual 8 (B)					30					45
$\phi = 0.7$	0.65	0.68	0.05	15/21/44		0.64	0.67	0.06	14/18/37	
$\delta = 0.7$	0.71	0.72	0.02	15/19/36		0.72	0.72	0.02	12/16/37	
$\lambda = 0.4$	0.47	0.43	0.08	15/20/43		0.48	0.44	0.09	12/19/39	
Individual 9 (A)					0					0
$\phi = 0.7$	0.63	0.68	0.09	22/27/48		0.66	0.70	0.06	22/25/40	
$\delta = 0.7$	0.72	0.73	0.03	22/29/51		0.68	0.68	0.03	19/23/46	
$\lambda = 0.4$	0.52	0.42	0.17	25/36/48		0.47	0.40	0.11	24/29/46	
Individual 10 (B)					12					16
$\phi = 0.2$	0.20	0.20	0.01	5/6/40		0.20	0.21	0.01	4/7/37	
$\delta = 0.2$	0.19	0.20	0.01	4/7/46		0.20	0.20	0.01	4/5/42	
$\lambda = 0.6$	0.60	0.60	0.01	6/7/47		0.60	0.60	0.01	4/6/45	
Individual 11 (A)					0					0
$\phi = 0.7$	0.60	0.65	0.12	28/36/51		0.62	0.67	0.09	32/35/47	
$\delta = 0.7$	0.73	0.75	0.04	22/30/47		0.70	0.72	0.03	19/22/42	
$\lambda = 0.4$	0.57	0.48	0.23	31/38/55		0.53	0.43	0.17	33/39/52	
Individual 12 (B)					6					8
$\phi = 0.2$	0.19	0.20	0.01	4/7/41		0.19	0.20	0.01	6/9/38	
$\delta = 0.2$	0.20	0.21	0.01	4/6/42		0.20	0.20	0.01	4/8/42	
$\lambda = 0.6$	0.61	0.61	0.02	3/5/42		0.61	0.62	0.01	5/9/42	
Individual 13 (A)					0					0
$\phi = 0.5$	0.48	0.50	0.02	7/10/38		0.49	0.51	0.01	11/15/39	
$\delta = 0.5$	0.50	0.50	0.01	13/15/47		0.50	0.50	0.01	17/21/51	
$\lambda = 0.5$	0.53	0.50	0.04	8/13/41		0.52	0.50	0.02	13/16/44	
Individual 14 (B)					1					2
$\phi = 0.5$	0.47	0.50	0.03	9/14/44		0.48	0.49	0.03	10/14/40	
$\delta = 0.5$	0.51	0.51	0.01	10/14/41		0.51	0.52	0.01	9/11/39	
$\lambda = 0.5$	0.53	0.51	0.04	10/14/41		0.53	0.51	0.03	9/13/41	

as shown by the high values of the RMSEs, but they also fall outside the asymptotic confidence bands. The estimators seem to perform slightly worse for the dominance solvable game, both with respect to numerical problems of the optimization procedures and with respect to the statistical properties of the estimators, as can be seen by higher RMSEs and larger percentages of cases falling outside the asymptotic confidence bands.

Given the high nonlinearity of the problem it is hard to guess the reason for (and therefore solve) the small sample problems detected in the simulations. The lack of convergence results from identification problems. In short samples the process generates often individual histories with so little variation that many combinations of the parameters can explain well the data. At the end of Appendix B we provide a formal discussion of a similar issue in the context of the model with random learning parameters.

As for the bias, notice that the estimate of δ does not differ in a systematic way from the true value (and the difference is not large). Notice also (see tables 1 and 2) that the product $\phi \times \lambda$ does not appear to have a large systematic bias either. One can conjecture that the problem with ϕ comes from the estimation of λ. Let us look at a simplified model to see why is λ a problem.

Assume that the players are playing a one-person game with only two strategies (s_1, s_2), whose payoffs are respectively π_1, π_2 with $\pi_1 > \pi_2$. Assume also that $\phi = 0, \delta = 1, \rho = 0$, and $N(0) = 1$. We only have to estimate λ. The probability that strategy s_1 is played is therefore $p_1 = \frac{1}{e^{\lambda(\pi_2 - \pi_1)+1}}$. The model then reduces to the choice of s_1 following a binomial distribution with probabilities given by p_1. Let K be the number of times that strategy s_1 is played in a given sample of size T. The likelihood function of that sample is

$$\left(\frac{1}{e^{\lambda(\pi_2-\pi_1)}+1}\right)^K \left(\frac{1}{e^{\lambda(\pi_1-\pi_2)}+1}\right)^{T-K}.$$

This likelihood is maximized for

$$\hat{\lambda} = (\pi_2 - \pi_1) \ln\left(\frac{T}{K} - 1\right).$$

But we know that in any reasonable learning models $\lambda \geq 0$, so we have to make

$$\hat{\lambda} = \max\left\{0, (\pi_2 - \pi_1) \ln\left(\frac{T}{K} - 1\right)\right\}.$$

This makes the estimator a convex function of K. Since K follows a binomial its expectation is simply $p_1 T$. So by Jensen's inequality we have that

$$\begin{aligned} E(\hat{\lambda}) &> (\pi_2 - \pi_1) \ln \left(\frac{T}{E(K)} - 1 \right) \\ &= (\pi_2 - \pi_1) \ln \left(\frac{T}{p_1} - 1 \right) \\ &= \lambda \end{aligned}$$

So the estimator is biased upwards as we observe in the simulations. In the example this could in principle be solved by choosing carefully an upper bound for $\hat{\lambda}$, but in the general case, the choice of an upper bound that gets rid of the bias depends on knowledge of the other parameters, which are not available. For this reason this does not seem like a workable approach in general.

6 Random learning parameters

We have so far assumed that the learning parameters are estimated individual by individual. As we showed in section 4, this can be done in a consistent way, as long the experiment lasts for a long enough period of time. Section 5 shows that this can be very long for reasonable games. What can be done about that? Lengthening the duration of an experiment to the extent suggested by our simulations would not be practical. Something that can often be done is to run more sessions of the same experiment. This would not help if we insisted in obtaining the parameters for each individual, but it can be useful if we are interested in the *distribution* of learning parameters.

We need to make some adjustments in notation for the new framework. As before, we have an I-player game, where each player i has N_i strategies. Suppose that each individual i has a vector of learning parameters β_i that is randomly (and independently across players) drawn from some distribution $F(\beta_i, \alpha)$, where α is the vector of parameters which characterizes this distribution. This distribution is the same across individuals and sessions. The parameter β_i is fixed for the duration of play, which consists of a number T of repetitions of the game, for all individuals. We call this T-fold repetition of the game a *session*. An *experiment* consists of some number n of *sessions*, in each one of which a new set of I players draw (independently from the others) their learning parameter from the distribution $F(\beta_i, \alpha)$.

Let us denote by $s_i^l(t)$ the strategy played by player i at time t in session l. Let $s^l(t) = \{s_1^l(t), s_2^l(t), \ldots, s_N^l(t)\}$, be the strategy profile of all the players at time t in session l, and $s_{-i}^l(t) = \{s_1^l(t), \ldots, s_{i-1}^l(t), s_{i+1}^l(t) \ldots, s_N^l(t)\}$ the strategy profile of players other than i at time t in session l. Let $S(l) = \{s^l(1), s^l(2), \ldots, s^l(T)\}$ be the sequence of observations of strategy choices for all players in session l. Let S^T be the set of all possible $S(l)$ sequences, and s^T a generic element of S^T.

We can now define the probability of observing a particular sequence s^T, given a vector of parameters α is thus:

$$P(s^T, \alpha) = \prod_{i=1}^{I} \int_{\beta_i \in B} \prod_{j=1}^{N_i} \prod_{t=1}^{T} [P_i^j(t)^{I(s_i^j, s_i^l(t))}] dF(\beta_i, \alpha).$$

We denote by $I(s^T, S(l))$ the indicator function that is one if $s^T = S(l)$, and zero otherwise. We define now

$$q(S(l), \alpha) = - \sum_{s^T \in S^T} [I(s^T, S(l)) \log P(s^T, \alpha)])]$$

and

$$Q_n(S(1), \ldots, S(n), \alpha) = n^{-1} \sum_{l=1}^{n} q(S(l), \alpha)$$

The model can now be estimated by maximizing $Q_n(S(1), \ldots, S(n), \alpha)$ over the vector of parameters α. When the parameter vector α is identified[18], its estimator $\hat{\alpha}$, is consistent (see appendix B, where we also characterize the asymptotic distribution) under mild assumptions about the function $F(.,.)$.

Notice that the results hold because behavior between sessions is independent, so that we can apply appropriate law of large numbers[19]. But the number of sessions will, in practice, have to be not much more than 10 to 20. For this reason the large sample results we prove in the appendix are of limited usefulness, without having a better idea of what is a "large" sample in practice. Thus, we focus here on the small sample properties of the estimators.

[18] Identification can be a serious issue in this context, as we show in appendix B.

[19] For this reason, it is not equivalent to have n sessions with I players each, than one session with $n \times I$ players. With just one session, play can easily become correlated.

One of the biggest problem that one finds in this context is the specification of the distribution of the parameter vectors β_i. Economic theory can be of little help in proposing a distribution. But we know from the work on "random-effects" models in panel data (see e.g. Heckman and Singer 1984) that even mild incorrect specifications in the distribution of β_i, can lead to serious biases in the estimation process.

To illustrate this estimation procedure, we use again the 2×2 coordination game of section 5. As before, we are interested in ϕ_i, δ_i and λ_i. To generate the data for our estimations, we assume that ϕ_i is drawn at the beginning of each experiment from a beta distribution with support in the interval $(0, 1)$:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(x)^{a-1}(1-x)^{b-1},$$

where Γ is the gamma function, and we take $a = 2$ and $b = 2$. We also assume that the other coefficients to be estimated, δ_i and λ_i, are fixed and equal to 0.5 for all individuals. We estimate the three coefficients jointly by maximum likelihood. Each simulation pools the data from 10 experiments with 30 periods each, thus replicating the estimating procedure that we suggested above.

The results can be found in table 5. The assumption of random coefficients, assuming that the specification of the model is correct, reduces the biases discussed in the previous sections but the precision is still poor, as illustrated by the high root mean squared errors for a and b.

7 Further research

In this paper we have studied the properties of parameter estimators of a model that encompasses a wide variety of learning patterns. We hope that this will help to understand better the behavior of laboratory agents. The model has some important deficiencies that further research should take into account. One problem is that it assumes that agents have feedback on the "strategies" that other agents use (that is, complete contingent plans of action), but in nontrivial extensive games the agents may only know the "actions" that the other agents take given the path of the game actually taken. For example, in an ultimatum game a proposer chooses a split of a "pie" from a (possibly large) set of splits. Then a responder accepts of rejects the proposed split. In the usual version of the game, the proposer will only know what happens with the actually proposed split, and not with some others.

Table 5: Simulation results: random ϕ_i with beta distribution ($a = 2$, $b = 2$), $\delta_i = 0.5$ and $\lambda_i = 0.5$ for all individuals (200 simulations with 10 experiments each, $T = 30$).

	a	b	δ	ρ
Population value	2	2	0.5	0.5
Means	2.01	2.02	0.50	0.50
Medians	1.90	1.99	0.50	0.50
Root Mean Squared Error	0.52	0.30	0.00	0.00
Asymptotic Coverage				
5 %	16.66	10.14	5.07	7.97
10 %	24.64	18.84	7.25	9.42
32 %	42.75	39.13	36.23	38.41
Number of simulations that did not converge	62			

A more satisfactory version of the learning model should be able to describe how the player uses the available information on “actions” to adjust the belief on “strategies.” Some efforts on this direction are done by Anderson and Camerer (1999), Camerer, Ho and Wang (1999) and Grosskopf (1999).

Another missing factor is active learning with experimentation. A real agent may choose a strategy she currently believes suboptimal in order to gather more precise information about its performance. There are many models in the theoretical literature with this feature, but to our knowledge none has been fitted to experimental data. This is particularly important when, as we discussed in the previous paragraph, there is only information on “actions” rather than “strategies”.

So far we have talked about improvements of the underlying theoretical model (CH). But even within the CH model, there are econometric aspects that need to be dealt with. To the extent that the pure fictitious play or learning by reinforcement models are true (that is, when $\overline{\phi} = 1$), we need to establish the properties of the estimators for other parameters, which we could not provide with our theorems. In the same vein, it would be useful to know how general are the small sample properties of the estimators that we have partially analyzed with Monte Carlo methods.

We have taken as given the game that subjects are playing. But, as the results of Feltovich (2000) make clear, one can design the game to obtain more clear separation between the predictions of different learning models. Our results also underscore the fact that the properties of the estimates are going to depend very substantially on the game that is played.

Erev and Haruvy (2000) have shown that, depending on the model comparison criteria that is used, different learning models can appear to account best for the data. An important issue in this context is whether parameters are allowed to vary across games or not. It would be interesting to check if these discrepancies are as large once one takes into account the possibilities for inconsistent estimates that we have uncovered. One could, for example, use the model with random parameters to see if the distributions of parameters are statistically different between games.

One obvious last question is whether this sort of analysis will give substantial clues as to how do people learn outside the laboratory. For example, in real life one can count on the experience of other players in the same or similar games to guide one's own behavior to an extent that is typically precluded in the laboratory. Although this is true, one can adjust the experimental environment to permit a degree of controlled learning along some of those dimensions (see, for example, Bosch and Saez-Martí 1999 for an experiment in which controlled learning from others is permitted).

8 Appendix A

Proposition 1. Let β^* be the true parameter vector. Then β^* is identified, provided that for all i, $\delta_i^* \neq 1$, $N_i^*(0) \neq 1/(1-\rho_i^*)$, and that there exists a strategy profile s_{-i}^*, such that for some agent i and strategies j^*, k^*, $\pi_i(s_i^{j^*}, s_{-i}^*) \neq \pi_i(s_i^{k^*}, s_{-i}^*)$.

Proof:

Case 1. Let $\alpha_i^*(t) = (\lambda_i^* \sum_{s=0}^{t-1}(\phi_i^*)^s)/N_i^*(t)$ and let $\alpha_i^{'}(t)$ be defined analogously. Let $\alpha_i^*(\infty) = \lim_{t\to\infty}\, \alpha^*(t)$. Assume that for all t, there is a t' such that $\alpha_i^*(t') \neq \alpha_i^{'}(t')$, and let the sequence where this is true $\{t\}^*$. Assume also that $\alpha^*(\infty) \neq \alpha^{'}(\infty)$

Let $z_{u,t} = (s(u), s(u+1), ..., s(t))$, and let a realization of the stochastic process $z_{1,t-1}$ such that $s_{-i}(1) = ... = s_{-i}(t-1) = s_{-i}^*$ and that $s_i(1) = ... = s_i(t-1) = s_i^{j^*}$.

Denote $\Pi_{ij^*} = \pi_i(s_i^{j^*}, s_{-i}^*)$ and $\Pi_{ik^*} = \pi_j(s_i^{k^*}, s_{-i}^*)$. Then

$$\begin{aligned} & \frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1},\beta^*)}{P(s_i(t) = s_i^{k^*}|z_{1,t-1},\beta^*)} = \\ = \quad & \exp(\alpha_i^*(t)(\Pi_{ij^*} - \delta_i^*\Pi_{ik^*} + \frac{(\phi_i^*)^t}{\sum_{s=0}^{t-1}(\phi_i^*)^s} N_i^*(0)(A_i^{j^*}(0)^* - A_i^{k^*}(0)^*))) \end{aligned}$$

Suppose there is a $t \in \{t\}^*$ such that

$$\frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1},\beta^*)}{P(s_i(t) = s_i^{k^*}|z_{1,t-1},\beta^*)} \neq \frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1},\beta^{'})}{P(s_i(t) = s_i^{k^*}|z_{1,t-1},\beta^{'})},$$

then the result follows for this case. Suppose to the contrary that for all $t \in \{t\}^*$

$$\frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1},\beta^*)}{P(s_i(t) = s_i^{k^*}|z_{1,t-1},\beta^*)} = \frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1},\beta^{'})}{P(s_i(t) = s_i^{k^*}|z_{1,t-1},\beta^{'})},$$

Then if we choose $z'_{1,t-1}$ such that $s_{-i}(1) = ... = s_{-i}(t-1) = s^*_{-i}$ and that $s_i(1) = ... = s_i(t-1) = s_i^{k^*}$ we have that for some $t \in \{t\}^*$

$$\frac{P(s_i(t) = s_i^{j^*}|z'_{1,t-1}, \beta^*)}{P(s_i(t) = s_i^{k^*}|z'_{1,t-1}, \beta^*)} \neq \frac{P(s_i(t) = s_i^{j^*}|z'_{1,t-1}, \beta')}{P(s_i(t) = s_i^{k^*}|z'_{1,t-1}, \beta')}$$

and the result would follow, because otherwise we would have that for all $t \in \{t\}^*$

$$\frac{(\Pi_{ij^*} - \delta_i^*\Pi_{ik^*}) + \frac{(\phi_i^*)^t}{\sum_{s=0}^{t-1}(\phi_i^*)^s} N_i^*(0)(A_i^{j^*}(0)^* - A_i^{k^*}(0)^*)}{(\delta_i^*\Pi_{ij^*} - \Pi_{ik^*}) + \frac{(\phi_i^*)^t}{\sum_{s=0}^{t-1}(\phi_i^*)^s} N_i^*(0)(A_i^{j^*}(0)^* - A_i^{k^*}(0)^*)}$$

$$= \frac{(\Pi_{ij^*} - \delta'_i\Pi_{ik^*}) + \frac{(\phi'_i)^t}{\sum_{s=0}^{t-1}(\phi'_i)^s} N'_i(0)(A_i^{j^*}(0)' - A_i^{k^*}(0)')}{(\delta'_i\Pi_{ij^*} - \Pi_{ik^*}) + \frac{(\phi'_i)^t}{\sum_{s=0}^{t-1}(\phi'_i)^s} N'_i(0)(A_i^{j^*}(0)' - A_i^{k^*}(0)')}$$

but then taking t to infinity we would have that

$$\frac{\Pi_{ij^*} - \delta_i^*\Pi_{ik^*}}{\delta_i^*\Pi_{ij^*} - \Pi_{ik^*}} = \frac{\Pi_{ij^*} - \delta'_i\Pi_{ik^*}}{\delta'_i\Pi_{ij^*} - \Pi_{ik^*}}.$$

This, in turn, would imply that either $\Pi_{ij^*} = \Pi_{ik^*}$, which is a contradiction, or $\delta_i^* = \delta'_i$ which is again a contradiction with the fact that $\alpha_i^*(\infty) \neq \alpha'_i(\infty)$, and that for all $t \in \{t\}^*$

$$\frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1}, \beta^*)}{P(s_i(t) = s_i^{k^*}|z_{1,t-1}, \beta^*)} = \frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1}, \beta')}{P(s_i(t) = s_i^{k^*}|z_{1,t-1}, \beta')},$$

since $\alpha_i^*(\infty) \neq \alpha'_i(\infty)$ implies that

$$\lim_{t\to\infty} \frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1}, \beta^*)}{P(s_i(t) = s_i^{k^*}|z_{1,t-1}, \beta^*)} = \exp(\alpha_i^*(\infty)(\Pi_{ij^*} - \delta_i^*\Pi_{ik^*}))$$

$$\neq \exp(\alpha'_i(\infty)(\Pi_{ij^*} - \delta_i^*\Pi_{ik^*})) = \lim_{t\to\infty} \frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1}, \beta')}{P(s_i(t) = s_i^{k^*}|z_{1,t-1}, \beta')}.$$

Case 2. Let $\alpha_i^*(t) = \alpha'_i(t)$, for all $t \geq t^*$ Assume also $\delta_i^* \neq \delta'_i$. Then we must have that $\Pi_{ij^*} - \delta_i^*\Pi_{ik^*} \neq \Pi_{ij^*} - \delta'_i\Pi_{ik^*}$, so that for a value of t high enough:

$$\begin{aligned}
&\Pi_{ij^*} - \delta_i^* \Pi_{ik^*} + \frac{(\phi_i^*)^t}{\sum_{s=0}^{t-1}(\phi_i^*)^s} N_i^*(0)(A_i^{j^*}(0)^* - A_i^{k^*}(0)^0)) \\
\neq\ & (\Pi_{ij^*} - \delta_i^{'} \Pi_{ik^*}) + \frac{(\phi_i^{'})^t}{\sum_{s=0}^{t-1}(\phi_i^{'})^s} N_i^{\prime}(0)(A_i^{j^*}(0)' - A_i^{k^*}(0)')
\end{aligned}$$

which implies, since $\alpha_i^*(t) = \alpha_i^{'}(t)$ for all $t \geq t^*$, that for t high enough,

$$\frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1}, \beta^*)}{P(s_i(t) = s_i^{k^*}|z_{1,t-1}, \beta^*)} \neq \frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1}, \beta^{'})}{P(s_i(t) = s_i^{k^*}|z_{1,t-1}, \beta^{'})}.$$

Case 3. Let $\alpha_i^*(t) = \alpha_i^{'}(t)$, for all $t \geq t^*$, $\delta_i^* = \delta_i^{\prime}$ and $\phi_i^* \neq \phi_i^{\prime}$.

First note that in this case

$$\begin{aligned}
&\frac{\sum_{s=0}^{t-1}(\phi_i^*)^s - 1}{\sum_{s=0}^{t-1}(\phi_i^*)^s}(\Pi_{ij^*} - \delta_i^* \Pi_{ik^*})) + \frac{1}{\sum_{s=0}^{t-1}(\phi_i^*)^s}(\delta_i^* \Pi_{ij^*} - \Pi_{ik^*}) \quad (1) \\
\neq\ & \frac{\sum_{s=0}^{t-1}(\phi_i^{\prime})^s - 1}{\sum_{s=0}^{t-1}(\phi_i^{\prime})^s}(\Pi_{ij^*} - \delta_i^* \Pi_{ik^*})) + \frac{1}{\sum_{s=0}^{t-1}(\phi_i^{\prime})^s}(\delta_i^* \Pi_{ij^*} - \Pi_{ik^*}) \quad (2)
\end{aligned}$$

as otherwise we would have $\Pi_{ij^*} = -\Pi_{ik^*}$ and we have defined payoffs to be positive. This implies that when t is high enough, if we choose $z_{1,t-1}^{\prime\prime}$ such that $s_{-i}(1) = ... = s_{-i}(t-1) = s_{-i}^*$ and that $s_i(1) = ... = s_i(t-2) = s_i^{k^*}$, and $s_i(t-1) = s_i^{j^*}$ we have that

$$\begin{aligned}
&\frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1}^{\prime\prime}, \beta^*)}{P(s_i(t) = s_i^{k^*}|z_{1,t-1}^{\prime\prime}, \beta^*)} = \\
=\ & \exp(\alpha_i^*(t)(\frac{\sum_{s=0}^{t-1}(\phi_i^*)^s - 1}{\sum_{s=0}^{t-1}(\phi_i^*)^s}(\Pi_{ij^*} - \delta_i^* \Pi_{ik^*})) + \frac{1}{\sum_{s=0}^{t-1}(\phi_i^*)^s}(\delta_i^* \Pi_{ij^*} - \Pi_{ik^*}) + \\
&\frac{(\phi_i^*)^t}{\sum_{s=0}^{t-1}(\phi_i^*)^s} N_i^*(0)(A_i^{j^*}(0)^* - A_i^{k^*}(0)^*)))
\end{aligned}$$

so that for t high enough and given equation 1 and $\alpha_i^*(t) = \alpha_i^{'}(t)$,

$$\frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1}^{\prime\prime}, \beta^*)}{P(s_i(t) = s_i^{k^*}|z_{1,t-1}^{\prime\prime}, \beta^*)} \neq \frac{P(s_i(t) = s_i^{j^*}|z_{1,t-1}^{\prime\prime}, \beta^{'})}{P(s_i(t) = s_i^{k^*}|z_{1,t-1}^{\prime\prime}, \beta^{'})}$$

Case 4. Let $\alpha_i^*(t) = \alpha_i^{'}(t)$, for all $t \geq t^*$, $\delta_i^* = \delta_i', \phi_i^* = \phi_i'$, and $A_i^{j^*}(0)^* \neq A_i^{j^*}(0)'$ for some $s_i^{j^*}$.

In this case, trivially $P(s_i(1) = s_i^{j^*}|\beta^*) \neq P(s_i(1) = s_i^{j^*}|\beta')$.

The only possible case that remains is that $\alpha_i^*(t) = \alpha_i^{'}(t)$, for all $t \geq t^*$, $\delta_i^* = \delta_i', \phi_i^* = \phi_i'$, and $A_i^j(0)^* = A_i^j(0)'$ for all $s_i^{j^*}$. We will now show that this in fact implies that $\beta^* = \beta'$ so that cases 1 through 4 exhaust all the possible cases that have to be proved.

$\alpha_i^*(t) = \alpha_i^{'}(t)$, for all $t \geq t^*$ implies that $N_i^*(t)/\lambda_i^* = N_i^{'}(t)/\lambda_i^{'}$ for all $t \geq t^*$. For this reason we must have that

$$\frac{\rho_i^* N_i^*(t)}{\lambda_i^*} + \frac{1}{\lambda_i^*} = \frac{N_i^*(t+1)}{\lambda_i^*} = \frac{N_i^{'}(t+1)}{\lambda_i^{'}} = \frac{\rho_i^{'} N_i^{'}(t)}{\lambda_i^{'}} + \frac{1}{\lambda_i^{'}}$$

so that given $N_i^*(t)/\lambda_i^* = N_i^{'}(t)/\lambda_i^{'}$

$$(\rho_i^* - \rho_i^{'})\frac{N_i^*(t)}{\lambda_i^*} = \frac{1}{\lambda_i'} - \frac{1}{\lambda_i^*}$$

but given that $N_i^*(0) \neq 1/(1-\rho_i^*)$, we must have that $N_i^*(t)$, is not a constant, therefore the previous expression can only hold if $\rho_i^* = \rho_i^{'}$ and $\lambda_i' = \lambda_i^*$. But then we must also have that $N_i^{'}(0) = N_i^*(0)$, so that $N_i^*(t)/\lambda_i^* = N_i^{'}(t)/\lambda_i^{'}$ for all $t \geq t^*$. Since we already know that $\delta_i^* = \delta_i', \phi_i^* = \phi_i'$, and $A_i^{j^*}(0)^* = A_i^{j^*}(0)'$ for all s_i^{j*}, then $\beta^* = \beta'$ so that cases 1 through 4 exhaust all the possible cases, and the result follows.□

Lemma 1. Denote by β^* the true vector of parameters, then

$\beta^* = \text{argmin}_\beta E(Q_n^k(s(1), \ldots, s(n), \beta)$, for all $k \in \{Q, L\}$.

Proof:

Let first $k = Q$. Let $P_i^j(t, \beta)$ be the probability that agent i uses strategy j at time t, conditional on $s(t-1), \ldots, s(0)$, for a vector of parameters β. Then

$$E([I(s_i^j, s_i(t)) - P_i^j(t,\beta)]^2 | s(t-1), \ldots, s(0)) =$$
$$= E([I(s_i^j, s_i(t))]^2 - 2I(s_i^j, s_i(t))P_i^j(t,\beta) + [P_i^j(t,\beta)]^2 | s(t), \ldots, s(0))$$
$$= E([I(s_i^j, s_i(t))] - 2I(s_i^j, s_i(t))P_i^j(t,\beta) + [P_i^j(t,\beta)]^2 | s(t), \ldots, s(0))$$
$$= P_i^j(t,\beta^*) - 2P_i^j(t,\beta^*)P_i^j(t,\beta) + [P_i^j(t,\beta)]^2$$

$$= P_i^j(t,\beta^*) - [P_i^j(t,\beta^*)]^2 + [P_i^j(t,\beta^*)]^2 - 2P_i^j(t,\beta^*)P_i^j(t,\beta) + [P_i^j(t,\beta)]^2$$
$$= P_i^j(t,\beta^*) - [P_i^j(t,\beta^*)]^2 + [P_i^j(t,\beta^*)]^2 - P_i^j(t,\beta)]^2$$

and this expression is minimized when $P_i^j(t,\beta) = P_i^j(t,\beta^*)$, which happens uniquely when $\beta = \beta^*$, provided the parameter β^* is identified.

Let now $k = L$,

$$E(\sum_{i=1}^{m_j}[-I(s_i^j, s_i(t)) \log P_i^j(t)]|s(t),\ldots,s(0)) = \sum_{i=1}^{m_j} R_i^j(t) \log P_i^j(t).$$

We will now show that this expression is minimized when $R_i^j(t) = P_i^j(t)$. The first order conditions of the problem

$$\max_{P_i^j(t)} \sum_{i=1}^{m_j} R_i^j(t) \log P_i^j(t)$$

$$\text{subject to} \sum_{i=1}^{m_j} P_i^j(t) = 1$$

are $\frac{R_i^j(t)}{P_i^j(t)} = \lambda$ or $R_i^j(t) = P_i^j(t)\lambda$. Adding over j we get $\lambda = 1$, which gives $R_i^j(t) = P_i^j(t)$ as we wanted. □

Lemma 2. The process $(s(t))_{t\in R}$ is ϕ-mixing of size $-r/(r-1)$, for $r > 2$. That is, the sequence $\phi(j)$ is $O(m^{-\lambda})$ for some $\lambda > r/(r-1)$.

Proof: Let $z_{t,u} = (s(u), s(u+1), ..., s(t))$. For $G \in \mathcal{F}_1^k, F \in \mathcal{F}_{k+j}^\infty$ we have that

$$|P(F|G) - P(F)| \leq \max_{\substack{z_{1,k} \in G, z'_{1,k} \in \mathcal{F}_1^k, \\ z_{k+1,k+j-1} \in \mathcal{F}_{k+1}^{k+j-1}}} |P(F|z_{1,k}, z_{k+1,k+j-1}) - P(F|z'_{1,k}, z_{k+1,k+j-1})|$$

$$= \max_{\substack{z_{1,k} \in G, z'_{1,k} \in \mathcal{F}_1^k, \\ z_{k+1,k+j-1} \in \mathcal{F}_{k+1}^{k+j-1}}} \left| \sum_{z_{k+j}^\infty \in F} (P(z_{k+j,\infty}|z_{1,k}, z_{k+1,k+j-1}) - P(z_{k+j,\infty}|z'_{1,k}, z_{k+1,k+j-1})) \right|$$

$$\leq \max_{\substack{z_{1,k} \in G, z'_{1,k} \in \mathcal{F}_1^k, \\ z_{k+1,k+j-1} \in \mathcal{F}_{k+1}^{k+j-1}}} \sum_{z_{k+j}^\infty \in F} \frac{|P(z_{k+j,\infty}|z_{1,k}, z_{k+1,k+j-1}) - P(z_{k+j,\infty}|z'_{1,k}, z_{k+1,k+j-1})|}{P(z_{k+j,\infty}|z'_{1,k}, z_{k+1,k+j-1})}$$
$$\times P(z_{k+j,\infty}|z'_{1,k}, z_{k+1,k+j-1})$$

$$
\begin{aligned}
&\leq \max_{\substack{z_{1,k} \in G, z'_{1,k} \in \mathcal{F}_1^k, \\ z_{k+1,k+j-1} \in \mathcal{F}_{k+1}^{k+j-1}}} \sup_{z_{k+j}^{\infty} \in F} \frac{|P(z_{k+j,\infty}|z_{1,k}, z_{k+1,k+j-1}) - P(z_{k+j,\infty}|z'_{1,k}, z_{k+1,k+j-1})|}{P(z_{k+j,\infty}|z'_{1,k}, z_{k+1,k+j-1})} \\
&= \max_{\substack{z_{1,k} \in G, z'_{1,k} \in \mathcal{F}_1^k, \\ z_{k+1,k+j-1} \in \mathcal{F}_{k+1}^{k+j-1}}} \sup_{z_{k+j}^{\infty} \in F} \left| \frac{P(z_{k+j,\infty}|z_{1,k}, z_{k+1,k+j-1})}{P(z_{k+j,\infty}|z'_{1,k}, z_{k+1,k+j-1})} - 1 \right|
\end{aligned}
$$

Let $s^{z_{t,u}}(r)$ be such that $z_{t,u} = (s(u), s(u+1), ..., s^{z_{t,u}}(r), ..., s(t))$. Also, let $\Pi^M = |\max_{i,j,s_i^j,s_{-j}} \pi_i(s_i^j, s_{-i})|$ and $\Pi^m = |\min_{i,j,s_i^j,s_{-j}} \pi_i(s_i^j, s_{-i})|$. Then,

$$
\begin{aligned}
&\frac{P(s_i^{z_{k+j,\infty}}(t)|z_{1,k}, z_{k+1,t-1})}{P(s_i^{z_{k+j,\infty}}(t)|z'_{1,k}, z_{k+1,t-1})} \\
&= \frac{\dfrac{\exp(\frac{\sum_{n=0}^{t-k-1}(\phi_i)^n F_i^j(s^{z_{k+1,t-1}}(t-n)) + \sum_{n=t-k}^{t-1}(\phi_i)^n F_i^j(s^{z_{1,k}}(t-n)) + (\phi_i)^t N_i(0) A_i^j(0)}{(\rho_i)^t N_i(0) + \sum_{n=0}^{t-1}(\rho_i)^n})}{\sum_{m=1}^{N_m} \exp(\frac{\sum_{n=0}^{t-k-1}(\phi_i)^n F_i^m(s^{z_{k+1,t-1}}(t-n)) + \sum_{n=t-k}^{t-1}(\phi_i)^n F_i^m(s^{z_{1,k}}(t-n)) + (\phi_i)^t N_i(0) A_i^j(0)}{(\rho_i)^t N_i(0) + \sum_{n=0}^{t-1}(\rho_i)^n})}}{\dfrac{\exp(\frac{\sum_{n=0}^{t-k-1}(\phi_i)^n F_i^j(s^{z_{k+1,t-1}}(t-n)) + \sum_{n=t-k}^{t-1}(\phi_i^n F_i^j(s^{z'_{1,k}}(t-n)) + (\phi_i)^t N_i(0) A_i^j(0)}{(\rho_i)^t N_i(0) + \sum_{n=0}^{t-1}(\rho_i)^n})}{\sum_{m=1}^{N_m} \exp(\frac{\sum_{n=0}^{t-k-1}(\phi_i)^n F_i^m(s^{z_{k+1,t-1}}(t-n)) + \sum_{n=t-k}^{t-1}(\phi_i)^n F_i^m(s^{z'_{1,k}}(t-n)) + (\phi_i)^t N_i(0) A_i^j(0)}{(\rho_i)^t N_i(0) + \sum_{n=0}^{t-1}(\rho_i)^n})}} \\
&\leq \frac{\dfrac{\exp(\frac{\sum_{n=t-k}^{t-1}(\phi_i)^n \Pi^M)}{(\rho_i)^t N_i(0) + \sum_{n=0}^{t-1}\rho_i^n})}{\exp(\frac{\sum_{n=t-k}^{t-1}(\phi_i^n(-\Pi^m)}{(\rho_i)^t N_i(0) + \sum_{n=0}^{t-1}(\rho_i)^n}) \sum_{m=1}^{N_m} \exp(\frac{\sum_{n=0}^{t-k-1}(\phi_i)^n F_i^m(s^{z_{k+1,t-1}}(t-n)) + (\phi_i)^t N_i(0) A_i^j(0)}{(\rho_i)^t N_i(0) + \sum_{n=0}^{t-1}(\rho_i)^n})}}{\dfrac{\exp(\frac{\sum_{n=t-k}^{t-1}(\phi_i)^n(-\Pi^m)}{(\rho_i)^t N_i(0) + \sum_{n=0}^{t-1}(\rho_i)^n})}{\exp(\frac{\sum_{n=t-k}^{t-1}(\phi_i)^n \Pi^M)}{(\rho_i)^t N_i(0) + \sum_{n=0}^{t-1}(\rho_i)^n}) \sum_{m=1}^{N_m} \exp(\frac{\sum_{n=0}^{t-k-1}(\phi_i)^n F_i^m(s^{z_{k+1,t-1}}(t-n)) + (\phi_i)^t N_i(0) A_i^j(0)}{(\rho_i)^t N_i(0) + \sum_{n=0}^{t-1}(\rho_i)^n})}} \\
&= \exp 2(\frac{\sum_{n=t-k}^{t-1}(\phi_i)^n(\Pi^M + \Pi^m))}{(\rho_i)^t N_i(0) + \sum_{n=0}^{t-1}(\rho_i)^n}) \leq \exp 2(\frac{(\phi_i)^{t-k}(\Pi^M + \Pi^m))}{1 - \phi_i})
\end{aligned}
$$

where for the last inequality we use the fact that $\phi_i < 1$.

Therefore,

$$
\begin{aligned}
\frac{P(z_{k+j,\infty}|z_{1,k}, z_{k+1,k+j-1})}{P(z_{k+j,\infty}|z'_{1,k}, z_{k+1,k+j-1})} &= \prod_{t=k+j}^{\infty} \prod_{i=1}^{I} \frac{P(s_i^{z_{k+j,\infty}}(t)|z_{1,k}, z_{k+1,t-1})}{P(s_i^{z_{k+j,\infty}}(t)|z'_{1,k}, z_{k+1,t-1})} \\
&\leq \prod_{i=1}^{I} \prod_{t=k+j}^{\infty} \exp 2(\frac{(\phi_i)^{t-k}(\Pi^M + \Pi^m))}{1 - \phi_i}) = \prod_{i=1}^{I} \exp 2(\frac{(\phi_i)^j(\Pi^M + \Pi^m))}{(1 - \phi_i)^2})
\end{aligned}
$$

Therefore $\sup_{k \in R} \sup\{|P(F|G) - P(F)| : G \in \mathcal{F}_1^k, F \in \mathcal{F}_{k+j}^{\infty}, P(G) > 0\}$.

$$\begin{aligned} |P(F|G) - P(F)| &\leq \max_{z_{1,k} \in G, z'_{1,k} \in \mathcal{F}_1^k, z_{k+1,k+j-1} \in \mathcal{F}_{k+1}^{k+j-1}} \sup_{z_{k+j}^{\infty} \in F} \left| \frac{P(z_{k+j,\infty} | z_{1,k}, z_{k+1,k+j-1})}{P(z_{k+j,\infty} | z'_{1,k}, z_{k+1,k+j-1})} - 1 \right| \\ &\leq \prod_{i=1}^{I} \exp 2 \left(\frac{(\phi_i)^j (\Pi^M + \Pi^m))}{(1 - \phi_i)^2} \right) - 1 \end{aligned}$$

Since this is true for any $k \in \mathfrak{N}$ and any $G \in \mathcal{F}_1^k, F \in \mathcal{F}_{k+j}^{\infty}, P(G) > 0$, the result follows.$\square$

9 Appendix B

We investigate here the asymptotic properties of estimators of the α parameter vector which controls the distribution of individual learning parameters in the population. We first have the analog of Lemma 1

Lemma 3. Denote by α^* the true vector of parameters, then

$\alpha^* \in \text{argmin}_\alpha E(Q_n(S(1), \ldots, S(n), \alpha))$.

Proof:

$$E(-\sum_{s^T \in S^T} [I(s^T, S(l)) \log P(s^T, \alpha)]) = -\sum_{s^T \in S^T}^{m_j} P(s^T, \alpha^*) \log P(s^T, \alpha).$$

We will now show that this expression is minimized when $\alpha = \alpha^*$. The first order conditions of the problem

$$\max_\alpha \sum_{s^T \in S^T}^{m_j} P(s^T, \alpha^*) \log P(s^T, \alpha)$$

$$\text{subject to} \sum_{s^T \in S^T}^{m_j} P(s^T, \alpha) = 1$$

are $\frac{P(s^T,\alpha^*)}{P(s^T,\alpha)} = \lambda$ or $P(s^T, \alpha^*) = P(s^T, \alpha)\lambda$. Adding over $s^T \in S^T$ we get $\lambda = 1$, which gives $P(s^T, \alpha^*) = P(s^T, \alpha)$ so that a minimizer is α^*as we wanted. □

We can now prove the analog of Propositions 1 and 2.

Proposition 4. If α^* is identified and $F(.,.)$ is a continuous function of α, then $\hat{\alpha}$ is a consistent estimator of α^*. That is

$$|\hat{\alpha} - \alpha^*| \to 0 \; i.p. \; as \; n \to \infty$$

Proof: Lemma 3 shows that $\alpha^* \in \text{argmin}_\alpha E(Q_n(S(1), \ldots, S(n), \alpha))$, and since the parameter vector is identified that requirement is satisfied with equality. So by theorem 7.1 (p. 81) in Pötscher and Prucha 1997, the result follows provided the following conditions hold:

1. The process $S(l)$ is defined in a subset of $\Re^p$, and the parameter space A is a compact metric space.
2. The process $S(l)$ is α-mixing.
3. The function $q(S(l), \alpha)$ is continuous.
4. $\sup_n n^{-1} \sum_{l=1}^{n} E|S(l)| < \infty$.

Conditions 1 and 4 are satisfied by the finiteness of the strategy spaces we consider and the limits we impose on the parameters. Condition 2 is satisfied as the behavior between sessions as well as the learning parameters draws between individuals is independent. Condition 3 is easily verified by inspection of the function $q(S(l), \alpha)$.□.

Let

$$C_n = E(\nabla_{\alpha\alpha}(Q_n(S(1), \ldots, S(n), \alpha)))$$

and

$$D_n = (nE(\nabla_{\alpha'}(Q_n(S(1), \ldots, S(n), \alpha)\nabla_{\alpha}(Q_n(S(1), \ldots, S(n), \alpha))))^{1/2}.$$

Proposition 5. Assume that $\liminf_{n\to\infty} \lambda_{\min} E(\nabla_{\alpha\alpha}(Q_n(S(1), \ldots, S(n), \alpha))) > 0$ and that $F(.,.)$ is a twice continuously differentiable function. Then $\hat{\alpha}$ is an asymptotically normal estimator of α^*. That is,

$$n^{1/2}(\hat{\alpha}_n - \alpha^*) = (C_n)^{-1} D_n \zeta_n + o_p(1)$$

with

$$\zeta_n \to N(0, I)$$

and

$$n^{1/2}(D_n)^{-1} C_n(\hat{\alpha}_n - \alpha^*) \to N(0, I)$$

Proof: By theorem 11.2 (p. 108) in Pötscher and Prucha 1997 the result follows provided the following conditions hold:

1. The process $S(l)$ is defined in a subset of $\Re^p$, and the parameter space A is a compact metric space.

2. The functions $q(S(l), .)$ is twice continuously partially differentiable at every point in A, and the function $q(., \alpha)$ is measurable in S.

3. The sequence of estimators $\hat{\alpha}_n$ satisfies $|\hat{a}_n - \alpha^*| = o_p(1)$.

4. $\sup_n n^{-1} \sum_{l=1}^n E|S(l)| < \infty$.

5. The function $\nabla_{\alpha\alpha}(q(S(l), \hat{\alpha}))$ is continuous on S, A, and

$$\sup_n n^{-1} \sum_{l=1}^n E[\sup |\nabla_{\alpha\alpha}(q(S(l), \hat{\alpha}))|^2] < \infty$$

6.

$$E(\nabla_\alpha(Q_n(S(1), \ldots, S(n), \alpha)) = 0$$

7.

$$\liminf_{n\to\infty} \lambda_{\min} E(\nabla_{\alpha\alpha}(Q_n(S(1), \ldots, S(n), \alpha))) > 0.$$

8. The process $(S(l))_{l\in N}$ is ϕ-mixing of size $-r/(r-1)$, for $r > 2$.

Proof: Conditions 1 and 4 are satisfied by the finiteness of the strategy spaces we consider and the limits we impose on the parameters. Condition 8 is satisfied as the behavior between sessions as well as the learning parameters draws between individuals is independent. Condition 2 and 5 is easily verified by inspection of function $q(S(l), \alpha)$ and by the continuous differentiability of $F(.,.)$. Condition 3 follows from Proposition 4. Condition 6 follows from the way in which estimators are obtained. Condition 7 is an assumption of the proposition.□

We have not been able to find meaningful sufficient conditions for identification of the parameter vector α. In fact, as we now show with an example, this is not a trivial issue in general.

Assume that the players are playing a one-person game with only two strategies (s_1, s_2), whose payoffs are respectively π_1, π_2 with $\pi_1 > \pi_2$. Assume also that $\phi = 0, \delta = 1, \rho = 0$, and $N(0) = 1$. We only have to estimate the distribution of λ. Assume that this game is played only once in every session, so that $T = 1$. The probability distribution of λ is such that there are only three possible values of λ, namely $\lambda_1, \lambda_2, \lambda_3$ whose respective probabilities are $p_1, p_2, (1 - p_1 - p_2)$. In this case, the vector $\alpha = (p_1, p_2)$. The probability that strategy s_1 is played is therefore $p_{s_1} = \frac{1}{e^{\lambda_1(\pi_2 - \pi_1)}+1}p_1 + \frac{1}{e^{\lambda_2(\pi_2 - \pi_1)}+1}p_2 + \frac{1}{e^{\lambda_3(\pi_2 - \pi_1)}+1}(1 - p_1 - p_2)$. The model then reduces to the choice of s_1 following a binomial distribution with probabilities given by p_1. Let K be the number of sessions (thus, the number of times) that strategy s_1 is played in a given sample of size n. The likelihood function of that sample is

$$(p_{s_1})^K (1 - p_{s_1})^{n-K}.$$

This likelihood is maximized for any $\hat{\alpha}$, such that

$$p_{s_1} = \frac{K}{n}.$$

But this is a linear equation with two unknowns, so we will have infinitely many solutions for the problem. In this particular case the difficulty will be resolved if each session has two periods, but in general notice that there will be a tension between the number of periods in each session and the complexity of the distribution function that needs to be estimated.

References

O. Armantier (2000), "The Influence of Observation on Economic Behavior: Experimental Evidence," mimeo.

C.M. Anderson and C.F. Camerer (1999), "Experience Weighted Attraction Learning in Sender-Receiver Games," mimeo.

A. Blume, D.V. DeJong, G.R. Neumann and N.E. Savin (2000), "Learning and Communication in Sender-Receiver Games: An Econometric Investigation," CentER WP 2000-09.

T. Börgers and R. Sarin (1997), "Learning through Reinforcement and Replicator Dynamics," *Journal of Economic Theory*, 77:1-14.

A. Bosch and M. Sáez-Martí (1999), "Credit Cycles in Theory and Experiment," WP 351 Universitat Pompeu Fabra.

R. Boylan and M. El-Gamal (1993), "Fictitious Play: A Statistical Study of Multiple Economic Experiments," *Games and Economic Behavior*, 5:205-222.

B. Broseta (2000), "Adaptive Learning and Equilibrium Selection in Experimental Coordination Games: An ARCH(1) Approach," forthcoming *Games and Economic Behavior*.

G.W. Brown (1951), "Iterative Solution of Games by Fictitious Play," in T.C. Koopmans, ed., *Activity Analysis of Production and Allocation*, New York: John Wiley and Sons.

R.R. Bush and F. Mosteller (1951), "A Mathematical Model for Simple Learning," *Psychological Review*, 58:313-323.

R.R. Bush and F. Mosteller (1955), *Stochastic Models for Learning* New York: John Wiley and Sons.

A. Cabrales, W. García-Fontes and M. Motta (2000), "Risk Dominance Selects the Leader. An Experimental Analysis," *International Journal of Industrial Organization*, 18:137-162.

C.F. Camerer, T.H. Ho (1999), "Experience-Weighted Attraction Learning in Games: A Unifying Approach," *Econometrica*, 67:827-874.

C.F. Camerer, T.H. Ho and X. Wang (1999), "Individual Differences in Experience Weighted Attraction Learning with Partial Payoff Information," mimeo.

Y.-W. Cheung and D. Friedman (1997), "Individual Learning in Normal

Form Games: Some Laboratory Results," *Games and Economic Behavior*, 19:46-76.

V.P. Crawford (1995), "Adaptive Dynamics in Coordination Games," *Econometrica*, 63:103-144.

Y. Chen and F. F. Tang, (1998) "Learning and Incentive Compatible Mechanisms for Public Goods Provision: An Experimental Study," *Journal of Political Economy*, 106:633-662.

K. Clark, S. Kay and M. Sefton (1999), "When are Nash Equilibria Self-Enforcing? An Experimental Analysis," *International Journal of Game Theory*, forthcoming.

A.A. Cournot (1971), *Researches into the Mathematical Principles of the Theory of Wealth*, New York: A.M. Kelley.

J.G. Cross (1973), "A Stochastic Learning Model of Economic Behavior," *Quarterly Journal of Economics* 87:239-266.

I. Erev and E. Haruvy (2000), "On the Potential Uses and Current Limitations of Data Driven Learning Models", mimeo.

N. Feltovich (2000), "Reinforcement-Based vs. Belief-Based Learning Models in Experimental Asymmetric-Information Games," *Econometrica* 68:605-642.

D. Fudenberg and D.K. Levine (1998), *The Theory of Learning in Games*, Boston MA: MIT Press.

W. Güth, R. Schmittberger and B. Schwarze, "An Experimental Analysis of Ultimatum Bargaining", *Journal of Economic Behavior and Organization*, 3, 367-388.

E. Haruvy and D.O. Stahl (1998), "Empirical Tests of Equilibrium Selection Principles in Symmetric Normal-Form Games," mimeo.

A. Matsui (1991), "Best Response Dynamics and Socially Stable Strategies," *Journal of Economic Theory*, 57: 343-63.

D. Mookherjee and B. Sopher (1997), "Learning and Decision Costs in Experimental Constant Sum Games," *Games and Economic Behavior*, 19:97-132.

R. Nagel, F.F. Tang (1998), "Experimental Results on the Centipede Learning in Normal Form: An Investigation on Learning", *Journal of Mathematical Psychology*, 42:356-384.

M. Posch (1996), "Cycling in a Stochastic Learning Algorithm for Normal-Form Games", mimeo, Institut Für Mathematik der Universität Wien.

B.M. Pötscher and I.R. Prucha (1997), *Dynamic Nonlinear Econometric Models*, Springer Verlag, Berlin.

J. Robinson (1951), "An Iterative Method of Solving a Game," *Annals of Mathematics*, 54:296-301.

A. Roth and I. Erev (1995), "Learning in Extensive Games: Experimental Data and Simple Dynamic Models in the Intermediate Term", *Games and Economic Behavior*, 8, 164-212.

M. Sefton (1998), "A Model of Behavior in Coordination Game Experiments," Technical Report, University of Manchester, mimeo.

F. F. Tang (1996), "Anticipatory Learning in Two-Person Games: An Experimental Study; Part II: Learning", SFB 303, Universität Bonn, Discussion Paper No. B-362 (Berlin: Springer Verlag, forthcoming).

P.D. Taylor and L.B. Jonker (1978), "Evolutionarily Stable Strategies and Game Dynamics," *Mathematical Biosciences*, 40:145-156.

H. White (1996), *Estimation, Inference and Specification Analysis*, Economic Society Monographs, Cambridge: Cambridge University Press.