Today

- **E**xperience-**W**eighted **A**ttraction learning
  - Model
  - Comparison

EWA

- Paper: Camerer & Ho (1999)
- Broad model with several parameters
- Contains Fictitious Play (FP) and Reinforcement Learning (RL) as special cases
  - FP and RL can be tested withhin the model
  - When fitted, EWA will always be as least as good as FP and RL
  - Similarities and differences between the two can be shown in the model

EWA: model

- n players, indexed by i
- $s_i^j \in S_i$ strategy j of player i, finite strategy space
- $\pi_i(s_i(t), s_{-i}(t))$ payoff of player i in round t
- In RL, the state was determined by, for each player, a vector which included one entry per strategy, called *propensities*
- In EWA, each player has a similar vector with one entry per strategy, $A(t)$, called *attractions*
- In addition to that, the state also includes a number $N(t)$, which is termed *observation-equivalents*
  - Note: N is not the number of players
- A and N are updated each round to form the learning theory

EWA: N

- $N(t)$ "Observation-equivalents": Roughly the same as previous rounds of experience
- $N(0)$ Value of N at the start of the game, one of the parameters of the theory
- Updating rule for N:

$$N(t) = \rho N(t-1) + 1,\ t \geq 1$$

- Set value of 1 is added each round
- Previous experience is depreciated with a discount factor ρ

EWA: A

- $A_i^j(t)$ “Attraction” of player i’s strategy j in round t
- $A_i^j(0)$ Initial attractions at the start of the game, parameters of the theory
- Updating rule for A consists of 3 parts:
  1. Keeping (discounted) old attraction: First entry in the numerator
  2. Adding value given the outcome of the current round: Second entry in the numerator
  3. Normalization with experience: Denominator

EWA: A

- 1. Keeping discounted old attractions:

$$\phi N(t-1)A_i^j(t-1)$$

- ϕ: Second discount rate, for old attractions, can be different from ρ

EWA: A

- 2. Adding value for current round

$$\phi N(t-1)A_i^j(t-1) + [\delta + (1-\delta)I(s_i^j, s_i(t))]\pi_i(s_i^j, s_{-i}(t))$$

- I is the indicator function, which is $= 1$ if the strategy was used this round, and $= 0$ if the strategy was not used
- δ determines weight added for used vs unused strategies:
  - $\delta = 0$ Only used strategy is updated (as in RL)
  - $\delta = 1$ Used and unused strategies treated the same (as in FP)
  - δ is another parameter of the theory

EWA: A

- 3. Normalization with experience

$$\frac{\phi N(t-1)A_i^j(t-1) + [\delta + (1-\delta)I(s_i^j, s_i(t))]\pi_i(s_i^j, s_{-i}(t))}{N(t)}$$

$$N(t) = \rho N(t-1) + 1$$

- $N(0)$ is pregame experience, further N ingame experience

EWA: Parameters

$$A_i^j(t) = \frac{\phi N(t-1) A_i^j(t-1) + [\delta + (1-\delta) I(s_i^j, s_i(t))] \pi_i(s_i^j, s_{-i}(t))}{\rho N(t-1) + 1}$$

- In FP, the initial weights determine, how strongly the “prior” beliefs about the game influence gameplay
- Similar here, $N(0)$ determines the strengths of pregame influence. Someone with extremely high $N(0)$ will not change his play according to actual experience, but stick to his pre-conceived knowledge of the game

EWA: Parameters

$$A_i^j(t) = \frac{\phi N(t-1) A_i^j(t-1) + [\delta + (1-\delta) I(s_i^j, s_i(t))] \pi_i(s_i^j, s_{-i}(t))}{N(t) = \rho N(t-1) + 1}$$

- The initial attractions need not be distributed evenly. Just as $N(0)$ reflects the strength of pre-game knowledge, $A_i^j(0)$ reflects what exactly you know prior to the game
- Note the similarity with RL's propensities and initial propensities

EWA: Parameters

$$A_i^j(t) = \frac{\phi N(t-1)A_i^j(t-1) + [\delta + (1-\delta)I(s_i^j, s_i(t))]\pi_i(s_i^j, s_{-i}(t))}{\rho N(t-1) + 1}$$

- Two discount factors. Camerer and Ho say:
  - "*These factors combine cognitive phenomena like forgetting with a deliberate tendency to discount old experience when the environment is changing*"
  - ... I don't see why you need 2 different discount factors for that ...
- Both parameters together govern the steady state attraction level: $(1-\rho)/(1-\phi)$ times the steady-state average payoff

EWA: Parameters

$$A_i^j(t) = \frac{\phi N(t-1)A_i^j(t-1) + [\delta + (1-\delta)I(s_i^j, s_i(t))]\pi_i(s_i^j, s_{-i}(t))}{\rho N(t-1)+1}$$

- δ determines what strategies players consider to base their learning on
- The bigger δ, the stronger are hypothetical payoffs used to base learning on
  - With $\delta = 0$, only actual payoffs are used
  - With $\delta = 1$, the learner does not differentiate between hypothetical and actual payoffs
- Easy to interpret δ in terms of FP and RL

EWA: Choice

$$A_i^j(t) = \frac{\phi N(t-1) A_i^j(t-1) + [\delta + (1-\delta) I(s_i^j, s_i(t))] \pi_i(s_i^j, s_{-i}(t))}{\rho N(t-1) + 1}$$

- Question for the audience: Which strategies are played if EWA is used?

EWA: Choice

$$A_i^j(t) = \frac{\phi N(t-1)A_i^j(t-1) + [\delta + (1-\delta)I(s_i^j, s_i(t))]\pi_i(s_i^j, s_{-i}(t))}{\rho N(t-1) + 1}$$

- Question for the audience: Which strategies are played if EWA is used?
- Hint: It is a trick question

EWA: Choice

$$A_i^j(t) = \frac{\phi N(t-1)A_i^j(t-1) + [\delta + (1-\delta)I(s_i^j, s_i(t))]\pi_i(s_i^j, s_{-i}(t))}{\rho N(t-1) + 1}$$

- Question for the audience: Which strategies are played if EWA is used?
- Hint: It is a trick question
- I did not tell you yet!

EWA: Choice

- So far, we only saw how attractions are generated from parameters and past play
- How do attractions map into behavior? Camerer and Ho propose 3 forms (logit, probit, power) and use logit

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

- Where $P_i^j(t+1)$ is the probability of player i to play strategy j in round t and m_i is the number of strategies player i has
- λ is then another parameter of the model

EWA: Parameters

- Before we continue, a quick parameter count:
  1. ρ discounting
  2. ϕ discounting
  3. $N(0)$ strength of previous experience
  4. $A_i^j(0)$ “shape” of previous experience (actually up to i times j parameters)
  5. δ weight on hypothetical payoffs
  6. λ sensitivity to attractions

- Model will obviously fit very well, due to many parameters used
- Questionable whether it can be used to predict behavior
- Good in organising behavior, separating different learning behaviors, all in one model, while keeping (some) psychological base

EWA: Special cases

- The model collapses to Reinforcement Learning if:
  - $\rho = 0$
  - $N(0) = 1$
  - $\delta = 0$
- The model is equivalent to Fictitious Play/Cournot adjustment if:
  - $\rho = \phi$ ($= 1$ for FP, $= 0$ for Cournot adj.)
  - $\delta = 1$
  - and a proper choice rule is used (max)
- Because of the two discounting parameters, EWA is not just a convex combination of RL and FP, but broader

$$(2.6) \qquad B^k_{-i}(0) = \frac{N^k_{-i}(0)}{N(0)},$$

with $N^k_{-i}(0) \geq 0$ and $N(0) > 0$. Beliefs are updated by depreciating the previous counts by ρ, and adding one for the strategy combination actually chosen by the other players. That is,

$$(2.7) \qquad B^k_{-i}(t) = \frac{\rho \cdot N^k_{-i}(t-1) + I(s^k_{-i}, s_{-i}(t))}{\sum_{h=1}^{m_{-i}} \left[\rho \cdot N^h_{-i}(t-1) + I(s^h_{-i}, s_{-i}(t)) \right]}.$$

Expressing beliefs in terms of previous-period beliefs,

$$B_{-i}^{k}(t) = \frac{\rho \cdot B_{-i}^{k}(t-1) + \dfrac{I(s_{-i}^{k}, s_{-i}(t))}{N(t-1)}}{\rho + \dfrac{1}{N(t-1)}} \tag{2.8}$$

$$= \frac{\rho \cdot N(t-1) \cdot B_{-i}^{k}(t-1) + I(s_{-i}^{k}, s_{-i}(t))}{\rho \cdot N(t-1) + 1}.$$

This form of belief updating weights observations from one period ago ρ times as much as the most recent observation. This includes Cournot dynamics ($\rho = 0$; only the most recent observation counts) and fictitious play ($\rho = 1$; all observations count equally) as special cases. The general case $0 \leq \rho \leq 1$ is a compromise in which all observations count but more recent observations count more.

Expected payoffs in period t, $E_i^j(t)$, are taken over beliefs according to

$$E_i^j(t) = \sum_{k=1}^{m_{-i}} \pi_i(s_i^j, s_{-i}^k) \cdot B_{-i}^k(t). \tag{2.9}$$

The crucial step is to express period t expected payoffs as a function of period $t-1$ expected payoffs. Substituting equation (2.8) into (2.9) and rearranging yields:

$$E_i^j(t) = \frac{\rho \cdot N(t-1) \cdot E_i^j(t-1) + \pi(s_i^j, s_{-i}(t))}{\rho \cdot N(t-1) + 1}. \tag{2.10}$$

This equation makes the kinship between the EWA and belief approaches transparent. Formally, suppose initial attractions are equal to expected payoffs

EWA: Experimental fit

- Camerer and Ho test EWA and its special cases extensively, using data from several games
- EWA fits better, even when penalized for additional parameters (but: RL and FP used with non-uniform initial weights/propensities, so more parameters here as well)
- However: Many technical decisions must be made when trying to design “fair” test, so test by the authors not entirely convincing
- Estimated parameters of EWA differ, sometimes widely, across different types of games

EWA lite

- Updated version of EWA, which tries to address the critique that EWA has to many parameters
- Replaces all EWA parameters with functions, determined by the game structure and one addition parameter (so EWA lite has just one parameter)
- Good inner compromise between edge cases of FP, RL, etc??