Dynamic Games: Problems and Prospects.

Ariel Pakes (Harvard University).
DSE Summer-school 2025, UCL, London.

July 2, 2025

Background: Empirical Work in I.O.

• In developing the methodology needed to empirically analyze static questions in I.O. we took the framework for the analysis from our theory colleagues and used econometric methodology it to incorporate the richness of various market institutions.

• In particular we took

- from theory; the idea of using some form of Nash equilibrium as it allowed us to analyze counterfactuals and
- from econometrics; methodology for incorporating both observed and unobserved heterogeneity among both firms and consumers.

• This started in the analysis of full information take it or leave it games (retail pricing), but expanded to include Nash bargaining, models that allowed for asymmetric information, and non-price allocation (or matching) mechanisms.

Dynamic analysis started analogously.

• I.e. we took frameworks taken from our theory colleagues. They made assumptions which insured that the

1. state variables evolve as a Markov process
2. and the equilibrium is some form of Markov Perfection (no agent has an incentive to deviate at any value of the state variables).

• Star and Ho (1969) and more directly Maskin and Tirole (1988) who use full information environments.

• We started as in static analysis, trying to adapt the dynamic framework to the richness of different real world institutions.

• Ericson and Pakes (1995) develop a framework for doing this.

Basic Full Info Markov Perfect Game

- **States.**
  - $i \in \mathcal{Z}^+$ (could be multidimensional)
  - s_i will be the number of firms with efficiency level i,
  - $s = [s_i; \; i \in \mathcal{Z}^+]$ is the "industry structure" (a counting measure of the number of firms at each different efficiency level).
  - ζ aggregate state variable that evolves exogenously (perhaps as a Markov process). E.g. outside alternative.

Bellman Equation: "Capital Accumulation" Games

$$V(i,s) = \max\{\phi, \pi(i,s) +$$

$$\sup_{(x \geq 0)} [-cx + \beta \sum V(i',s')pr(i',s'|x,i,s,\zeta)p(\zeta)]\},$$

$$\text{where,} \quad pr(i',s'|x,i,s,\zeta) = pr(i'|i,x,\zeta)q(s-e(i)|i,x,s,\zeta).$$

- $pr(i'|i,x,\zeta)$. This is a game where my own investment only affects my own state variables.
- $q(s-e(i)|i,x,s,\zeta)$ perception of where my competitors will be.
- $\mathcal{P} = \{p(i'|x,\zeta); x \in \mathcal{R}^+\}$, stochastically increasing in x for every ζ.
- $q[\cdot|i,s,\zeta]$ is the equilibrium object. It embodies the incumbent's beliefs about entry and exit.

- Many possible entry models. Example:
  - Must pay x_e $(> \beta\phi)$ to enter,
  - enters one period later at state $\omega_e \in \Omega^e \subset \mathcal{Z}^+$ with probability $p^e(\cdot)$.
  - Only enters if the expected discounted value of future net cash flows from entering is greater than the cost of entry.)

- **Dynamic Equilibrium.**

1. Every agent chooses optimal policies given its perceptions on likely future industry structures
2. Those perceptions are consistent with the behavior of the agent's competitors everywhere, i.e. $\forall$ *possible* (i, s).

• Doraszelski and Satterwaite (2003) prove existence (to insure this we need random entry fees and exit costs).

• E-P have conditions which imply that any equilibria

1. Is "computable", i; never more than $\overline{n}$ firms active & Only observe "i" on $\Omega = \{1, ..., K\} \Rightarrow$ need only compute equilibria for $(i, s) \in \Omega \times S$

$$S \equiv \{s = [s_1, ..., s_k] : \sum s_j \leq \overline{n} < \infty\} \;\Rightarrow\; \#S \leq K^N$$

2. Generates a homogeneous Markov chain for industry structures [for $\{s_t\}$], i.e. $Q(s'|s)$

$$Pr[s_{t+1} = s'|s^t] = Pr[s_{t+1} = s'|s_t] \equiv Q[s'|s_t].$$

3. And provide conditions on the primitives such that insure that any equilibrium $Q[\cdot|\cdot]$ is *ergodic*. [Picture].
   - $\#R$ is frequently much smaller than $\#S$
   - Limiting probabilities converges to an invariant measure over points in R (often referred to as a "steady state" of the system, though "steady" seems to be a misnomer; as the state is not constant).

Why were we less successful in adapting these models to empirically analyzing dynamic issues?

The complexity issue. In empirical settings the model presumes agents

1. Can access a large amount of information and
2. Either compute or learn an unrealistic number of strategies (one for each information set).

- The complexity of Markov Perfection both
  - limited our ability to do dynamic analysis of market outcomes,
  - and led to a question of whether some other notion of equilibria will better approximate agents' behavior.

What would an agent (or a researcher) have to do to compute an equilibrium?

• Think of this as an iterative algorithm as in single agent problems, but when you calculate the continuation value you use the optimal policies of competitors in the last iteration to integrate where the algorithm expects them to be in the coming period. You then maximize current returns plus discounted continuation values and iterate until convergence.

• If we were to use a standard iterative computational tool (e.g. Pakes and McGuire, 1994), the burden would be a function of three factors

- the number of points evaluated at each iteration;
- the computational burden per point
- the number of iterations (value &/or policy function iterations).

Number of Points.

• Since each of the $\overline{n}$ active firms can only be at K distinct states, the number of points we need to evaluate at each iteration, or

$$\#S \leq K^{\overline{n}}.$$

Exchangeability of functions in the state variables of a firm's competitors: we do not need to differentiate between two vectors of competitors that are permutations of one another (see Pakes (1993) for details)

Burden per Point. Determined by

- the cost of calculating the expected value of future states the firm's own, as well as its competitors. If there is positive probability on each of κ points for each of the $m-1$ active competitors of a given firm. Then we need to sum over κ^m possible future states.
- The cost of obtaining the optimal polices and given the continuation value.

Conclusion

• The computational burden of the model grows quickly in both the number of firms ever simultaneously active (it grows geometrically in this dimension) and the number of state variables per firm (it grows exponentially in this dimension). "The Curse of Dimensionality" .

• Consequently though pointwise algorithm has been used both; (i) as a tool for investigating theoretical issues where analytic solutions were not possible, and (ii) as a framework fo controlling for dynamic selection in static problems, they have seen limited use in empirical work.

• There were different responses to this in the computational literature, and in the theory literature; both of which were well aware of the problem.

- We look first to the computational literature which turned to approximations.
- Come back to the theory literature: look for weaker equilibrium concepts.

Approximation Techniques.

• A number are available. Each has their problems, but they compute equilibria with much less of a computational burden than the standard algorithm. I leave a discussion of them to others at the conference. Early examples

- Deterministic approximation techniques. This starts in economics with the book by Judd(2004). It has now expanded with the use of various AI related tools.
- Stochastic Algorithm (Pakes and Mcguire, 2001)
- Continuous time algorithm. Doraszelski and Judd (2004).
- Oblivious equilibrium (mean field theory). Benkard, Van Roy, Weintraub, 2010. Approximate using moments of the distribution.

• An issue that arises with approximations is that when attempting to analyzing a counterfactual, you are trying to analyze an environment where the data used for approximations are not directly relevant.

Theory: First ask if assuming asymmetric information helps if we stay with perfection?

• It decrease the number of state variables by assuming agents only have access to a subset of the state variables.

- Since agents presumably know their own characteristics and these tend to be persistent, the "perfectness" notion would then lead us to a "Bayesian" Markov Perfect solution.

• **Is assuming "Bayesian MP" more realistic?** It decreases the information access and retention conditions but increases the burden of computing the policies significantly. The additional burden results from the need to compute posteriors, as well as optimal policies; and the requirement that they be consistent with one another.

• There is a question of whether agents can learn the optimal policies. I come back to this below.

Theory issues for the applied researcher.

- Premise: the complexity of Markov Perfection both
  - limits our ability to do dynamic analysis of market outcomes, but also
  - leads to a question of whether some other notion of equilibria will better approximate agents' behavior.

- What assumptions of MP might we relax? The initial frameworks made assumptions which insured that the
  1. state variables evolve as a Markov process
  2. and the equilibrium is some form of Markov Perfection (no agent has an incentive to deviate at any value of the state variables).

On the Markov Assumption.

• Except in situations involving active experimentation to learn (where policies are transient), we are likely to stick with the assumption that states evolve as a finite order Markov process. Reasons

- Convenience and fits the data well.
- Realism suggests information access and retention conditions limit the memory used.
- We can bound unilateral deviations (Ifrach and Weintraub, 2014), and have conditions which insure those deviations can be made arbitrarily small by letting the length of the kept history grow (White and Scherer, 1994).

• There are situations where we might not want to have the process be time homogenous (this often depends on the assumptions on "outside alternatives" in demand). I come back to this below.

Perfection.

• The fact that Markov Perfect framework becomes unwieldily when confronted by the complexity of real world institutions, raises the question of whether some other notion of equilibrium will better approximate agents' behavior.

• The complexity issue implies that agents

- have access to and can retain a large amount of information (all state variables), and
- can either compute or learn an unrealistic number of strategies (one for each information set).

How demanding is this? Durable goods example.

The response from theory.

•. Ttheory literature considered weaker equilibrium concepts. Examples

- Self-Confirming Equilibrium (off path beliefs which no longer need to be "subgame perfect", see Fudenberg and Levine 1993),
- Games with "procedurally rational" players. Each agent only knows its actions but the consequences of those actions are consistent with all players' behavior (Osborne and Rubinstein, 1998),
- Berk-Nash equilibrium: prior beliefs of agents may not put positive probability on the true model but a learning process leads them to converge to a close alternative (Esponda and Pouzo, 2016).

• These are all models of learning about a fixed state. Empirical work needs to allow states to evolve over time with the outcomes of investment processes. Still they are heading in a direction which we can mimic.

Abandoning Perfection In A Dynamic Game.

• I start with strategies that are "rest points" to a dynamical system. This makes the job of defining a rest point much easier because

- strategies at the rest point likely satisfy a Nash condition of some sort; else someone has an incentive to deviate.
- However it still leaves opens the question: What is the form of the Nash Condition?

• There remains the question of how might firm's get to it which becomes important for counterfactuals for two reasons

- We don't believe firms firms jump to an equilibrium instantaneously, and we might want to analyze the transition path.
- There are multiple equilibria in realistic dynamic games, and the transition path will pick out the relevant one.

I come back to these issues below.

What Conditions Can We Assume for the Rest Point at States that are Visited Repeatedly?

We expect (and I believe should integrate into our modelling) that

1. Agents perceive that they are doing the best they can at each of these points, and
2. These perceptions are at least consistent with what they observe.

Note. It might be reasonable to assume more than this: that agents (i) know and/or (ii) explore, properties of outcomes of states not visited repeatedly. I come back to this below.

Formalization of Assumptions.

• Denote the information set of firm i in period t by $J_{i,t}$. $J_{i,t}$ will contain both public (ξ_t) and private ($\omega_{i,t}$) information, so $J_{i,t} = \{\xi_t, \omega_{i,t}\}$.

• Assume $(J_{1,t}, \ldots J_{n_t,t})$ evolves as a finite state Markov process on $\mathcal{J}$ (or can be adequately approximated by one).

• Policies, say $m_{i,t} \in \mathcal{M}$, will be functions of $J_{i,t}$. For simplicity assume $\#\mathcal{M}$ is finite, and that it is a simple capital accumulation game (not necessary & not always appropriate), i.e. $\forall (m_i, m_{-i}) \in \mathcal{M}^n$, & $\forall\, \omega \in \Omega$

$$P_\omega(\cdot|m_i, m_{-i}, \omega) = P_\omega(\cdot|m_i, \omega),$$

• The public information, ξ, is used to predict competitor behavior and common demand and cost conditions (these evolve as an exogenous Markov process).

• A "state" of the system, is

$$s_t = \{J_{1,t}, \ldots, J_{n_t,t}\} \in \mathcal{S},$$

$\#\mathcal{S}$ is finite. $\Rightarrow$ any set of policies will insure that s_t will wander into a recurrent subset of $\mathcal{S}$, say $\mathcal{R} \subset \mathcal{S}$, in finite time, and after that $s_{t+\tau} \in \mathcal{R}$ w.p.1 forever.

• Note that the agents need not keep track of all of s_t, only $J_{i,t}$; i.e. $J_{i,t}$ is whatever management conditions on when forming dynamic policies.

• Let the agent's perception of the expected discounted value of current and future net cash flow were it to chose m at state J_i, be

$$W(m|J_i), \ \ \forall m \in \mathcal{M} \ \ \& \ \ \forall J_i \in \mathcal{J},$$

• and of expected profits be $\pi^E(m|J_i)$.

Our assumptions imply:

- Each agent choses an action which maximizes its perception of its expected discounted value, and
- For those states that are visited repeatedly (i.e. are in the recurrent class or $\mathcal{R}$), these perceptions are consistent with observed outcomes.

Formally

$$\mathbf{A}. \quad W(m^*|J_i) \geq W(m|J_i), \;\; \forall m \in \mathcal{M} \;\; \& \;\; \forall J_i \in \mathcal{J},$$

$$\mathbf{B}. \;\; \&, \; \forall J_i \;\; \text{which is a component of an} \;\; s \in \mathcal{R}$$

$$W(m(J_i)|J_i) = \pi^E(m|J_i) + \beta \sum_{J_i'} W(m^*(J_i')|J_i')p^e(J_i'|J_i),$$

where, if $p^e(\cdot)$ provides the empirical probability (the fraction of periods the event occurs)

$$\pi^E(m|J_i) \equiv \sum_{J_{-i}} E[\pi(\cdot)|J_i, J_{-i}]p^e(J_{-i}|J_i),$$

and

$$\left\{ p^e(J_{-i}|J_i) \equiv \frac{p^e(J_{-i}, J_i)}{p^e(J_i)} \right\}_{J_{-i}, J_i},$$

while

$$\left\{ p^e(J_i'|J_i) \equiv \frac{p^e(J_i', J_i)}{p^e(J_i)} \right\}_{J_i', J_i}. \; \spadesuit$$

"Experience Based Equilibrium"

• These are the conditions of a (restricted) EBE (Fershtman and Pakes, 2012). Conceptually similar to Fudenberg and Levine's (1993) self-confirming equilibria and (Osborne and Rubinstein, 1998) procedural rational equilibrium, but we allow state variables to evolve over time.

• Bayesian Perfect satisfy them, but so do weaker notions. In particular REBE does not require consistent perceptions for points that are not visited repeatedly. This opens up a different avenue for multiplicity as perceptions for points not visited can influence decisions at the points visited; a point we come back to below.

• Consider two computation algorithms.
  - The first is an algorithm that firms could also use to learn equilibrium play. Likely more useful for small perturbations, as to learn after a dramatic change would take a long time.
  - The second does not have the advantage of a learning rule, but goes directly to compute an equilibrium

Computational Algorithm.

• Asynchronous reinforcement learning algorithm (first used for dynamic games in Pakes and McGuire, 2001), earlier machine learning literature for single agent problems calls this "Q-learning".

• The fact that it is based on learning from realized data makes it a candidate to analyze perturbations to the environment (provided they are not large enough to induce experimentation), as well as to compute equilibrium to dynamic games.

• If there is more than one equilibria, the learning algorithm will pick one out. If algorithm is run many times from the same initial conditions, it will pick out a distribution of equilibria.

• Formally it circumvents the two sources of the curse of dimensionality in computing equilibrium.

Iterations

• The computation problem is now different for management then for the researcher as management just conditions on $J_{i,t}$ but the researcher must compute equilibria for $s_t = (J_{1,t} \dots J_{n_t})$. Here we deal with the computational problem for researchers. We come back to the firm's problem below.

• **Iterations are defined by**

- A location, say $L^k = (J_1^k, \dots J_{n(k)}^k) \in \mathcal{S}$: is the information sets of active agents .
- Objects in memory (i.e. M^k):
  (i) perceived evaluations, W^k,
  (ii) No. of visits to each point, h^k.

Must update (L^k, W^k, h^k)**.** Computational burden determined by; memory constraint, and compute time. I use a simple (not neccesarily optimal) structure to memory.

Update Location.

- Calculate "greedy" policies (policies are now "m") for each agent

$$m^*_{i,k} = arg \max_{m \in \mathcal{M}} W^k(m|J_{i,k})$$

- Take random draws on outcomes conditional on $m^*_{i,k}$:
- E.g.; if we invest in "payoff relevant" $\omega_{i,k} \in J_{i,k}$, draw $\omega_{i,k+1}$ conditional on $(\omega_{i,k}, m^*_{i,k})$.
- Use outcomes to update $L^k \to L^{k+1}$.

Update W^k.

• "Learning" interpretation: Simple case: assume agent observes the competitors' static controls m_{-i} (usually a price or bid) and initially we will assume the agent knows (perhaps through estimation) the primitives; $\pi_i(\cdot), p(\omega_{i,t+1}|\omega_{i,t}, m_{i,t})$. Can generalize and allow them not to be known.

• Its ex poste perception of what its value would have been had it chosen m is

$$V^{k+1}(J_{i,k}, m) =$$
$$\pi(\omega_{i,k}, m, m_{-i,k}, d_k) + \max_{\tilde{m}\in M} \beta W^k(\tilde{m}|J_{i,k+1}(m)),$$

where $J_i^{k+1}(m)$ is what the $k+1$ information would have been given m and *competitors actual play*.

Treat $V^{k+1}(J_{i,k})$ as a random draw from the possible realizations of $W(m|J_{i,k})$, and update W^k as in stochastic integration (Robbins and Monroe,1956)

$$W^{k+1}(m|J_{i,k}) =$$

$$\frac{1}{h^k(J_{i,k})} V^{k+1}(J_{i,k}, m) + \frac{(h^k(J_{i,k}) - 1)}{h^k(J_{i,k})} W^k(m|J_{i,k}),$$

or

$$W^{k+1}(m|J_{i,k}) - W^k(m|J_{i,k}) =$$

$$\frac{1}{h^k(J_{i,k})}[V^{k+1}(J_{i,k}, m) - W^k(m|J_{i,k})].$$

(other weights are more efficient as the early estimates of $V^{k+1}(J_{i,k}, m)$ are noisier than the later estimates, and it would be good to know how to use information on close states to update a given state ....)

Notes.

• If we have equilibrium valuations we tend to stay their, i.e. if * designates equilibrium

$$E[V^*(J_i, m^*)|W^*] = W^*(m^*|J_i).$$

So if it reaches equilibrium it tends to stay at equilibrium.

• As in all computational algorithms for dynamic games there is no guarantee that it will converge, but the learning interpretation provides a rational for using it regardless.

• In fact, the smoothing implicit in the draws tends to mitigate the convergence problem in pointwise algorithms, but it may take many iterations.

• Agents (not only the analyst) could use the algorithm to find equilibrium policies or adjust to perturbations in the environment.

- Formally algorithm has no curse of dimensionality.
  1. Computing continuation values: integration is replaced by averaging two numbers.
  2. algorithm eventually wanders into $\mathcal{R}$ and stays their, and $\#\mathcal{R} \leq \#\mathcal{J}$.

- Fershtman and Pakes (2012) also provide a test for equilibrium that has no curse of dimensionality.

- Still the number of states can grow large (typically grows linearly in the number of state variables).

- One way around this is to nest functional form approximations into this. There is a related Operations Research literature on computation of single agent problems often refered to as "TD learning" (see Sutton and Barto,1998).

Multiplicity of REBE.

• Recall that $\mathcal{R}$ is the recurrent class (the points that are visited repeatedly).

• $\mathcal{R}$ contains both "interior" and "boundary" points. Points at which there are feasible strategies which can lead outside of $\mathcal{R}$ are boundary points. Interior points are points that can only transit to other points in $\mathcal{R}$ no matter which (feasible) policy is chosen.

• Our conditions only insure that perceptions of outcomes are consistent with the results from actual play at interior points. Perceptions of outcomes for some feasible (but in-optimal) policy at boundary points are not tied down by actual outcomes.

• "MPBE" are a special case of (restricted) EBE and they have multiplicity. Here differing perceptions at boundary points can support a (possibly much) wider range of equilibria.

Narrowing the Set of Equilibria.

• Often if the firms are in an equilibrium, the data will reveal which of the possible equilibria they are in. So the problem really is only a problem when calculating counterfactuals. There are a number of ways of limiting the number of equilibria, but none that I am aware of that give us uniqueness.

- Constrain the initial condition to whatever is in the data (history limits the equilibria that can be generated).
- Asker, Fershtman, Jihye, and Pakes, (2020, *RAND*), compute equilibria in which firms experiment with $m_i \neq m_i^*$ at boundary points. Leads to a stronger notion of, and test for, equilibrium (Boundary Consistency).

• Boundary consistency insures that perceptions are consistent with the results from **actual play** for each **feasible** action at boundary points (and hence on $\mathcal{R}$) if one played the strategies in memory.

Boundary Consistency

- Let $B(J_i|\mathcal{W})$ be the set of actions at $J_i \in s \in \mathcal{R}$ which could generate outcomes which are not in the recurrent class (if it is not empty, J_i is a boundary point) and $B(\mathcal{W}) = \cup_{J_i \in \mathcal{R}} B(J_i|\mathcal{W})$. Then the extra condition needed to insure "Boundary Consistency" is:
- **Extra Condition.** Let τ index future periods, then $\forall (m, J_i) \in B(\mathcal{W})$

$$W(m^*|J_i) \geq$$

$$E\left[\sum_{\tau=0}^{\infty} \delta^{\tau} \pi\Big(m(J_{i,\tau}), m(J_{-i,\tau})\Big)\Big| J_i = J_{i,0}, \mathcal{W}\right],$$

where $E[\cdot|J_i, \mathcal{W}]$ takes expectations over future states starting at J_i using the policies generated by $\mathcal{W}$. ♠

A Note on Proceeding to Empirical Work (These are notes from Dubois Pakes, 2025?).

- Theory is agnostic with respect to the agent's information set, but empirical work must specify it.

- Suggestion:
  - start with an empirical model, where the firm maximizes given its perceptions, but the perceptions need not be "correct" or "equilibrium" perceptions.
  - This should use the data to determine which variables the firm's policies seem to respond to that are in your data. Typically includes a function of profits, institutional features and controls or states of competing firms.
  - Then add a serially correlated unobserved state to account for variables that the firm knows but are not in your data (this is almost always needed).

• In estimating the empirical model be careful to account for issues generated by the unobserved serially correlated states. Especially

- selection issues when there is either exit, or firms stopping to invest for a period.
- endogeneity of right hand side variables problems that arise from serial correlation of the unobservables.

• Use the estimated model and the initial condition in the data to compute the evolution the industry from the initial period. This should give you an idea of the fit of the empirical model.

• Compute an EBE (computational algorithm below), and do a similar excersize to compare its fit both

- to the data
- to the empirical model.

Counterfactuals.

• Counterfactuals re-parameterize the primitives of the problem. They are necessarily "out-of-sample" so use of the empirical model alone is inappropriate.

• There are two options for analyzing counterfactuals.

- Use the learning algorithm we went over.
- Use the direct algorithm I go to now.

• Nobody believes that the market switches to the new equilibrium instantaneously, so there is likely some learning process. The question is if a learning model available to the researcher is a good approximation.

• If the firm learns from information not in our data, the learning models we have access to may be inappropriate.

• Going directly to an equilibrium seems more appropriate for longer run predictions. Of course if there are continual changes in the environment that may not be the right way to do the analysis.

Algorithm: Overview (from Dubois-Pakes).

- Problems in computing fixed point for dynamic market equilibria.

1. Dimension of state space (how many points do we need policies for).
   - Do not compute policies for all possible states. Only compute separate policies for each product in each time period (small no. of points). This is "pointwise": i.e. no discretization or approximations.
2. Computing continuation values (dimension of integral).
   - Substitute simulation for integration. Average of simulated sequences of continuation values. Random variables that must be held constant across iterations. Can use last period's investment measures post sample incentives (so post sample is firm's perceptions).

- Algorithm starts at initial point in the data. It assumes each point visited is in the recurrent class of a boundary consistent equilibrium thereafter.

Algorithm Details.

• Algorithm is iterative. Iterations are indexed by l & associated with policy functions. Initial estimate is the empirical model's policy function.

• Assume that the policy function is

$$a_t = \theta_0 + \theta_1 \log\left(\frac{\partial \pi^l(\xi', J')}{\partial \xi'}\right) + w_t \beta_w$$

where

$$\omega_t = \rho \omega_{t-1} + \nu_t$$

and ν_t is independent of history.

• Start with the estimate of the policy function obtained from the empirical model, and holding the process generating ω_t equal to the empirical model's estimate.

• At each iteration.use the policy functions in memory to simulate K sample paths for profits, $\{\pi^{k,l}(\cdot)_{j,t}\}$ and the serially correlated unobservable, $\{\omega^{k,l}_{h,j,t}\}$ for each firm, keeping the underlying random draws the same across iterations (else it will not converge).

• Compute the averages over the sample paths of $\{\pi^{k,l}(\cdot)_{j,t}\}$ and $\{\omega^{k,l}_{h,j,t}\}$ say $\{\pi^{l}(\cdot)\}_{j,t}$, & $\{\omega^{l}_{h,j,t}\}_{h,j,t}$.

• Compute the average discounted value $W^{l}_{j,t} \equiv \sum_{\tau=1}^{\infty} \beta^{\tau} \pi^{l}(\cdot)_{j,t+\tau}$.

• The marginal return for $a_,t$ (which we set equal to one) is

$$\mathcal{E}\Big[\sum_{\tau=0}^{\infty} \beta^{\tau} \frac{\partial \pi(\cdot)_{t+\tau}}{\partial a_t} | J_{j,t}\Big] \approx \sum \beta^{\tau} \frac{\partial \pi^{k}(\cdot)_{t+\tau}}{\partial a_t}$$

- Terminal period's contribution: Use the last observed advertising, say T_j, to approximate for $\sum_{\tau=T_j}^{\infty}(\beta\rho)^{\tau}\frac{\partial\pi(\cdot)_{t+\tau}}{\partial\xi_{t+1}}\beta_{a,h}$ (eliminates need to assume firm's perceptions about post sample process is the same as in sample).

- Update the policy function. Regress the dependent variable

$$\sum \beta^{\tau}\frac{\partial\pi^{k}(\cdot)_{t+\tau}}{\partial a_t}-\omega_t$$

on

$$\approx \theta_0^{l+1}+\theta_1^{l+1}\log\left(\frac{\partial\pi^{l}(\xi',J')}{\partial\xi'}\right)+w_t\beta_w^{l+1}$$

- Finally compute

$$Y^{l}\equiv\frac{1}{\sum_{j,t}1}\sum_{j,t}\left(\frac{W_{j,t}^{l}-W_{j,t}^{l-1}}{W_{j,t}^{l-1}}\right)^{2}.$$

- If $Y^{l}<10^{-5}$ stop. If $Y^{l}>10^{-5}$ proceed to iteration $l+1$.