Dynamic Games

Ron Yang

April 6, 2019

This section draws on Robin's slides, as well as POB/BBL for the first half, and Ericson-Pakes for the second half. I focus on material you need for the problem set.

I try to make the notation as internally consistent as possible, and as consistent with the previous section notes in the class, so this deviates slightly from the notation in the slides or in the original papers.

1 Multiple Agent Dynamic Discrete Choice

1.1 Recap: Single Agent DDC

Given the previous section on the single-agent dynamic discrete choice framework, the techniques we need for (1) multiple agents and (2) continuous choice can be seen as a straightforward extension. Recall the structure of single-agent DDC estimation:

1. First, we estimate conditional value functions using choice probabilities.
2. Second, given conditional value functions and parameters, our model implies choice probabilities. We minimize an objective function which finds parameters which fit the data.

In the Rust Nested Fixed Point method, we use a fixed-point iteration for the first step, and we use maximum likelihood in the second step. In the Hotz-Miller method, we use an inversion for the first step, while in the Hotz-Miller-Sanders-Smith method, we use simulation for the first step[1]. For the CCP methods, we used minimum distance for the second step, fitting observed and predicted choice probabilities. It is useful to think about these two steps in the structure separately; we can "mix and match" first and second stage steps according to the problem.

1.2 Inversion

The Pakes-Ostrovsky-Berry (POB) method uses inversion, in the spirit of Hotz-Miller. The notation in POB is a bit specific to their model (VE, VC, etc.), so I try to write more generally. Let there be states $x \in X$, actions $a \in A$, discount factor δ, and flow payoffs π.

First, define the conditional *continuation* value function. This value function captures the expected continuation value from doing action a in state x.

$$V_a(x) = E[\pi(x') + \delta V(x') \mid a, x]$$

[1] Aside: What is the intuitive relationship between inversion and simulation? We can think about inversion as representing an infinite discounted sum. In the single-variable case for scalar x, let $X = \lim_{n\to\infty} \sum_{t=0}^{n} \beta^t x$. If $\beta < 1$, the limit X exists and

$$X = x + \beta X \implies X = \frac{1}{1-\beta} x$$

In the multiple-variable case for $N \times 1$ vector x, and state-transition matrix T, $X = \lim_{n\to\infty} \sum_{t=0}^{n} \beta^t T^t x$. If $I - \beta T$ is invertible, the limit X exists and

$$X = x + \beta T X \implies X = (I - \beta T)^{-1} x$$

Since value functions are infinite discounted sums, inversion and simulation methods naturally arise in CCP estimation.

Suppose that the value function given a state $x \in X$ can be decomposed into a conditional value function $V_a(x)$ and a function of choice probabilities $\psi_a(p(x))$. For example, POB use an exponential exit value to perform this decomposition.

$$V(x) = V_a(x) + \psi_a(p(x))$$

Stack the definition of the conditional continuation value function across states, and plug in the decomposition. Let M_a be a $|X| \times |X|$ transition matrix where $M_a(x, x')$ is the probability of going from state s to state x', conditional on the agent choosing action a.

$$V_a = M_a(\pi + \delta V) = M_a(\pi + \delta V_a + \delta \psi_a(p))$$

Rearrange and invert to find the familiar POB formula.

$$V_a = (I - \delta M_a)^{-1} M_a(\pi + \delta \psi_a(p))$$

Given this formula and guesses for the parameters in π and ψ_a, we can estimate M_a and p nonparametrically and compute an estimate of V_a.[2]

The difference between this formula and the Hotz-Miller inversion we recovered in the previous problem set reflects the difference between the conditional value function and the conditional continuation value function. Since POB are using the "next period's value function," they need the additional M_a term to reflect state transitions between this period and next.

1.3 Simulation

The Bajari-Benkard-Levin (BBL) method uses simulation, in the spirit of Hotz-Miller-Sanders-Smith. Again, I try to write this more generally than the original paper.[3]

Let there be J agents. Let $\epsilon_{j,t}$ be an $|A| \times 1$ private shock to π (e.g. entry costs or scrap values), and $\sigma(x, \epsilon)$ be the optimal policy. As in HMSS, begin by estimating conditional choice probabilities $\hat{p}(x_t \mid \theta_0)$ and conditional state transitions $\hat{M}$. Use the CCPs to back out the optimal policy function $\sigma(x, \epsilon)$.

Then, to estimate the conditional *continuation* value function $V^*_{a,i}(x \mid \theta)$ for agent $i \in J$ by simulation,

1. For each simulation draw $s \in S$, draw a sequence of states, actions for each agent j, and shocks: $\{x_t, \{a_t\}_j, \{\epsilon_t\}_j\}_s$.
   (a) Initialize the initial state at $x_0 = x$.
   (b) For each simulation period $t \in [0, \ldots, \infty]$,
      i. For each agent $j \in J$,
         A. If $j = i$ and $t = 0$, let $a_{jt} = a$. Else, choose $a_{jt} = \sigma(x_t, \epsilon_{jt})$.
         B. Draw $\epsilon_{j,t+1}$.
      ii. Draw $x_{t+1} \sim \hat{M}(x_{t+1} \mid x_t, \{a_t\}_j, \theta_0)$.
      iii. If x_{t+1} is a terminal state for agent i, stop.
2. For each simulated sequence $\{x_t, \{a_t\}_j, \{\epsilon_t\}_j\}_s$, compute the conditional continuation value function.

$$V^s_{a,i}(x \mid \{x_t, \{a_t\}_j, \{\epsilon_t\}_j\}_s, \theta) = \sum_{t=1}^{\infty} \beta^t [\pi_{a_t}(x_t, \epsilon^{a_t}_{i,t} \mid \theta)]$$

[2] Remember to condition on the agent's action a when estimating M. See POB, especially footnotes 6 and 7.

[3] When A is discrete, this estimation proceeds as in HMSS. When A is continuous, BBL needs a monotonicity condition to "invert" the CCPs to recover the policies.

Suppose that $\pi(a \mid x, \epsilon)$ has increasing differences in (a, ϵ). This implies that $\sigma(x, \epsilon)$ is monotonically increasing in ϵ. The conditional choice distribution can be written as $F(a \mid x) = P(\sigma(x \mid \epsilon) \leq a \mid x) = G(\sigma^{-1}(a \mid x) \mid x, \theta)$. Invert to recover the policy function.

$$\sigma(x, \epsilon) = F_i^{-1}(P(\sigma(x, \epsilon) \leq a_i \mid x)$$

3. Average over simulation draws to get an estimate of the conditional continuation value function.

$$V^*_{a,i}(x \mid \theta) = \frac{1}{S} \sum_s V^s_{a,i}(x \mid \{x_t, \{a_t\}_j, \{\epsilon_t\}_j\}_s, \theta)$$

1.4 Objective Functions

There are a variety of possible objective functions to use. Let $Q(\theta \mid (a_{it}, x_t))$ be the objective function to minimize given a set of observed actions a_{it} and states x_t. POB propose the first two, BBL propose the latter two.

- **Pseudo Maximum Likelihood**: Maximize the log-likelihood of observed choices.

$$Q(\theta) = -\sum_t \sum_{i \in J} \log(P(a_{it} \mid x_t, \theta)) \iff \theta^* = \arg\min_\theta Q(\theta)$$

  POB recommends against this objective function because estimation error in the first stage will (1) violate usual regularity conditions which ensure efficiency of pseudo-MLE estimators, (2) cause likelihood-based estimators to have poor finite-sample performance, and (3) cause the model to ascribe 0 probability to some observed events.

- **Method-of-Moments / Pseudo-χ^2 in Choice Probabilities**: Minimize a norm in the differences between model-predicted choice probabilities $P(a \mid x, \theta)$ and actual choice probabilities $\hat{P}(a \mid x)$.

$$Q(\theta) = ||\hat{P}(a \mid x) - P(a \mid x, \theta)|| \iff \theta^* = \arg\min_\theta Q(\theta)$$

  POB suggests that first-stage estimation error will also cause finite-sample issues, because the estimation error introduces a bias (see page 388 of POB for details). To resolve this, POB suggest using an instrument which is orthogonal to this estimation error. For example, a constant instrument implies pooling evenly probabilities across states.

- **Method-of-Moments in Deviations**: Minimize a norm in the utility gained by deviating to non-equilibrium strategies. Let $g(\{\sigma', x, i\} \mid \theta)$ define the utility gain to an agent i from deviating from the observed policy σ to deviant policy σ' in state x.

$$g(\{\sigma', x, i\} \mid \theta) = V_i(x, \sigma' \mid \theta) - V_i(x, \sigma \mid \theta)$$

  Then given a distribution H over these deviation-state-agent tuples $\{\sigma', x, i\}$, we can form an objective.

$$Q(\theta) = \int ||g(\{\sigma', x, i\})|| dH(\{\sigma', x, i\}) \iff \theta^* = \arg\min_\theta Q(\theta)$$

- **Bounds Estimation in Deviations**: If there exist θ such that the Method-of-Moments in Deviations objective is zero (i.e. no deviations considered yield a utility gain), then the coefficients may only be set-identified. In this case, we estimate an identified set.

$$Q(\theta) = \int ||g(\{\sigma', x, i\})|| dH(\{\sigma', x, i\}) \iff \Theta^* = \{\theta \mid Q(\theta) \leq 0\}$$

2 Markov-Perfect Industry Dynamics (Ericson-Pakes)

2.1 Motivation: The role of T in simulation methods

In the second problem set, the last question asked: at what level of T (the number of future periods to simulate) did the CCP estimates converge? Depending on the implementation, typical answers fell in the 200 to 500 range.

One initial guess might be that after a few hundred periods, the future states would be discounted so much that they would be irrelevant for decisions made today. However, with the $\beta = 0.9999$ specification, even $0.9999^{500} \approx 95\%$ is far from 0. If not discounting, then, what might explain the declining effect of T?

2.2 Markov Chains

To answer this question, I begin by laying some foundations of discrete time Markov processes. This is a bit notation and definition heavy, but bear with me.

Consider a finite set of states X and a random sequence of states $\{x_0, x_1, ...\}$.

Definition The n**-step transition probability** between state i and state j is

$$p_{ij}^{(n)} = P(x_n = j \mid x_0 = i)$$

Definition A state j is **accessible** from state i if $p_{ij}^{(n)} > 0$ for some $n \geq 0$.

Definition State i and j **communicate** if they are accessible from each other.

Claim Communication is an equivalence relation. *(Convince yourself that this is true).*

Definition A **communicating class** is an equivalence class with the communicate relation, such that all members of a communicating class communicate with each other.

Definition Let $f_{ii} = P(\exists n \geq 1 \quad s.t. \quad x_n = i \mid x_0 = i)$. A state i is **recurrent** if $f_{ii} = 1$ and **transient** if $f_{ii} < 1$.

Claim All communicating classes are either recurrent or transient. *(Convince yourself that this is true).*

Claim All finite Markov Chains have at least one recurrent class. *(Convince yourself that this is true).*

Question What does the Markov Chain representation of the Rust bus replacement model look like? What are the recurrent classes?

2.3 Equilibrium Dynamics

Given the groundwork we laid in 2.2, we can now discuss the concept of dynamic equilibrium in Ericson-Pakes. I assume we already have access to a set of equilibrium policy functions for all players, so I can focus on the state transitions.

Suppose we receive a sequence of states $\{x_0, x_1, \dots\}$ from a Markov Chain. When we consider the behavior of this sequence in the limit, what object are we learning about? Ericson-Pakes contrasts two potential objects of inquiry.

Limit Structure One limiting object is the "limit structure": the state that the sequence visits in the limit.

$$x \subseteq X \quad s.t. \quad \lim_{n \to \infty} P(x_n = x) = 1$$

This object allows us to ask questions of the form, "in the limit, will there be a non-zero number of firms in this industry?" or "in the limit, what will be the prevailing price in the industry?" When there is a single limiting state, the limit structure provides a precise prediction about the long-run characteristics of this market. Such limit structures are studied in Jovanovic (1982) and Hopenhayn (1992).

However, studying the limit structure tells us little about industries & markets that are inherently dynamic. Suppose that in the long run, a market cycles between having one firm charging monopoly prices and having two firms charging competitive prices. The relative share of time spent in each of these two states may have great importance for consumer welfare, but cannot be expressed using a limit structure.

Limit Sequence of Structures Another limiting object is the "limit sequence of structures": the distribution of state sequences in the limit.

$$\mu^* : X \to \mathbb{R} \quad s.t. \quad \forall x \in X, \mu_n(x) \to \mu^*(x)$$

I will unpack this definition step by step. First, I add some notation. Let the $|X| \times |X|$ matrix M be the state transition matrix implied by the equilibrium policies, and let the initial probability distribution over states at time 0 be μ_0. Then, the distribution of states after n periods will be

$$\mu_n(\nu) = M^n \mu_0$$

Under some assumptions, Ericson-Pakes characterize the resulting equilibrium in the following theorem.

Ericson-Pakes Theorem 2

1. *There exists a unique recurrent communicating class $R \subseteq X$.*
2. *There exists a distribution over states μ^* such that $\mu^*(x) = [mM(x,x)]^{-1}$ for $x \in R$ and $\mu^*(x) = 0$ for $x \notin R$, where $mM(x,x')$ is the expected time it takes to reach state x' from state x.*
3. *$\forall x \in X$, $\mu_n(x) \to \mu^*(x)$. That is, $\mu^*(x)$ is the limiting distribution of states.*
4. *μ^* is the distribution of a stationary, ergodic Markov process such that $M\mu^* = \mu^*$.*

This object allows us to ask questions about quantities, e.g. "in the limit, what is the average number of firms in this industry", and also about dynamics, e.g. "in the limit, do we observe a constant state, a cycle through several states, or punctuated shifts between disparate states?"

The limiting object is therefore a regular evolution through the states in the recurrent class, rather than a single deterministic state. In addition to allowing us to capture a richer set of real-world phenomena, this object encourages us to think about structure, behavior, and welfare jointly dynamically. For example, this suggests that a static regression of outcomes on market structure (a la SCP paradigm) does not identify a coherent object when the true model is dynamic.

Ergodicity The ergodicity of the limit process yields two features. First, as $n \to \infty$, time averages of states will approximate the ergodic distribution. If $I_n(x) = \mathbb{1}_{x_n = x}$ is an indicator for the state x, then

$$\lim_{n \to \infty} \frac{1}{n} \sum_n I_n(x) = \mu^*(x)$$

Second, the process will have two distinct stages. The process will initially depend on the distribution μ_0, but this dependence will die out as $n \to \infty$ and μ_n approaches the ergodic distribution μ^*. In the initial periods, the process may transit through various transitory states and classes, but eventually the process will only take on states in the recurrent class R. The rate of convergence, and the nature of the recurrent class, will depend on the particular model and application.

Estimation The previous analysis in this section has fixed an equilibrium and studied its characteristics. We now consider how this analysis can help us build consistent estimators under multiplicity of equilibria[4].

When motivating CCP estimation, we relied on the idea that the data will "identify" the equilibrium policies. Pakes-Ostrovsky-Berry formalize this intuition in the following proposition.

Pakes-Ostrovsky-Berry Proposition 1: *Assume that agents' beliefs about other agents' policies only depends on the state. Take an equilibrium E with a recurrent class $R \subseteq X$ and a data-generating process D. If alternate equilibrium E' yields a data-generating process D' which is consistent with D on R, then for any state $x \in R$, the policy functions of all players are the same in E and E'.*

The intuition of the proof is that within a recurrent class, all states will be visited infinitely many times. With sufficient data, we can recover the data generating process. If we have the data generating process,

[4] BBL simply assumes that the data are generated by a unique equilibrium (Assumption ES). "This assumption is relatively unrestrictive if the model has a unique equilibrium."

we can recover the agents' beliefs, and the beliefs imply a unique policy function. If two equilibria generate the same data generating process in the recurrent class, then they must have the same policy functions in recurrent class.

The additional assumption rules out state variables that agents observe but the econometrician does not, since that would prevent beliefs from being identified from observable states. This becomes especially problematic in settings where serially correlated unobserables are important.

2.4 Application

Let's return to our motivating example: the Rust model in the second problem set.

We used simulation of future states to estimate the value function given model parameters, by computing simulated paths of future states from each initial state, and applying the parametric flow payoffs in each state.

$$\hat{v}(x_0 \mid \theta) = \sum_s \sum_{t=0}^{T} \pi(x_{st} \mid \theta)$$

This implies that the relevant feature of each simulation is the set of future states that are visited from any initial state, $\{x_{st}\}$. The relative magnitude of these value functions then affects choice probabilities, because the agent is more likely to choose policies which bring it to high-value states. Specifically, the probability of replacing an engine weighs the value of the 0-mileage state against the value of higher mileage states.

Think of the initial distribution of states as μ_0 such that $\mu_0(x_0) = 1$ and $\mu_0(x') = 0$ for all $x' \neq x_0$. Our discussion of Markov Perfect industry dynamics suggests that for low t, μ_t still depends on μ_0, and therefore adding the t-th period to our simulation will differentially shift the value functions for different initial states x_0. However, for sufficiently high t, μ_t will approach the ergodic distribution μ^* and be independent of μ_0. Adding the t-th period will then shift all value functions symmetrically, with no effect on choice probabilities.

In Figure 2, I plot the average simulated mileage bin after t periods for a bus that starts in the 1st, 25th, and 50th mileage bins. Although the average states depend on the initial state for low t, the distributions quickly converge to the ergodic distribution and $\mu_t^1 \approx \mu_t^{25} \approx \mu_t^{50}$ for $t \geq 150$.

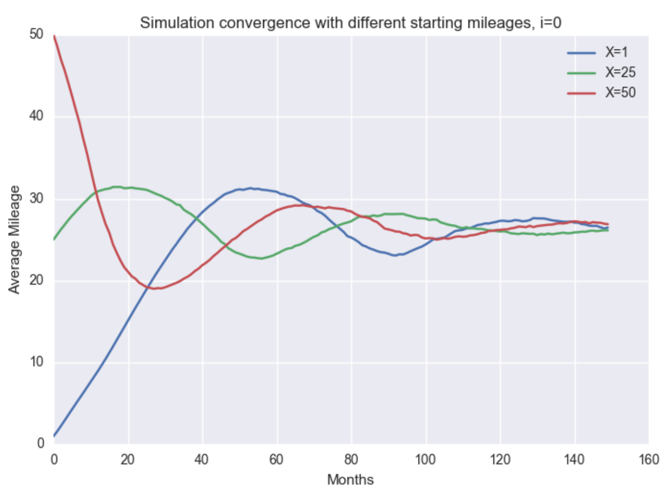

Figure 1: Convergence of Average States

In Figure 3, I plot the average simulated cumulative number of replacements after t periods for a bus that starts in the 1st, 25th, and 50th mileage bins. A similar pattern emerges: after a certain number of periods, the conditional probability of replacement becomes the same regardless of the initial state.

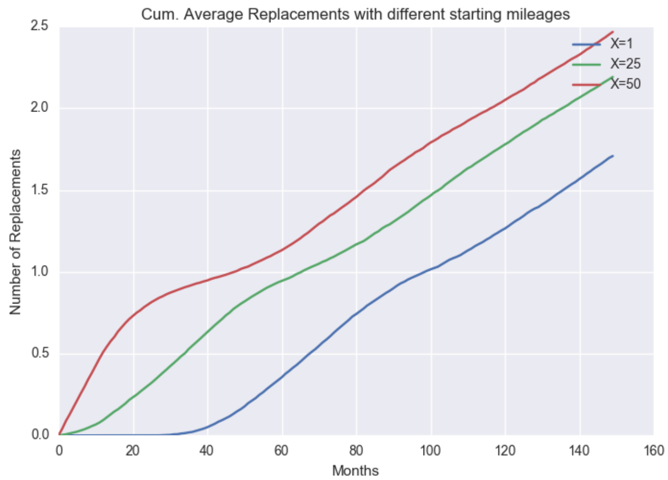

Figure 2: Convergence of Average Policies

Importantly, the convergence to the ergodic distribution happens much faster than β^t converges to 0. This explains why T stops affecting our estimates even when $\beta^T >> 0$.