

# Single-agent dynamic optimization models

---

C.Conlon - help from M. Shum and Paul Scott

Grad IO

## CCPs and Sufficient Statistics: Hotz-Miller (1993)

---

# Motivation

- In Rust, we started with a guess of parameters  $\theta$ , iterated on the Bellman operator to get  $EV_{\theta}(x, i)$  and then constructed CCP's  $P(i|x, \theta)$ .
- A disadvantage of Rust's approach is that it can be computationally intensive
  - With a richer state space, solving value function (inner fixed point) can take a very long time, which means estimation will take a very, very long time.
- Hotz and Miller's idea is to use observable data to form an estimate of (differences in) the value function from conditional choice probabilities (CCP's)
  - We observe  $P(i|x)$  directly in the data!
- The central challenge of dynamic estimation is computing continuation values. In Rust, they are computed by solving the dynamic problem. With Hotz-Miller (or the CCP approach more broadly), we “measure” continuation values using a function of CCP's.

## Rust's Theorem 1: Values to CCP's

- In Rust (1987), CCPs can be derived from the value function:

$$p_j(x) = \frac{\partial}{\partial \pi_j(x)} W(\pi(x) + \beta E[V(x') | x, j])$$

where  $W(u) = \int \max_j \{u_j + \varepsilon_j\} dG(\varepsilon)$  is the surplus function.

- For the logit case:

$$p_j(x) = \frac{\exp(v_j(x))}{\sum_{j' \in J} \exp(v_{j'}(x))}$$

where the choice specific value function for action  $j$  in state  $x$  is

$$v_j(x) \equiv \pi_j(x) + \beta E[V(x') | x, j]$$

## HM's Proposition 1: CCP's to Values

- Notice that CCP's are unchanged by subtracting some constant from every conditional (choice-specific) value function. Thus, consider

$$D_{j,0}v(x) \equiv v_j(x) - v_0(x)$$

where 0 denotes some reference action.

- Let  $Q : \mathbb{R}^{|\mathbf{I}|-1} \rightarrow \Delta^{|\mathbf{I}|}$  be the mapping from the differences in conditional (choice-specific) values to CCP's.
- Note: we're taking for granted that the distribution of  $\varepsilon$  is identical across states, otherwise  $Q$  would be different for different  $x$ .

### Hotz-Miller Inversion

$Q$  is invertible.

## HM inversion with logit errors

- Again, let's consider the case of where  $\varepsilon$  is i.i.d type I EV.
- Expression for CCP's:

$$p_j(x) = \frac{\exp(v_j(x))}{\sum_{j' \in \mathbf{J}} \exp(v_{j'}(x))}.$$

- The HM inversion follows by taking logs and differencing across actions:

$$\ln p_j(x) - \ln p_0(x) = v_j(x) - v_0(x)$$

- Thus, in the logit case (this looks a lot like Berry (1994)):

$$Q_j^{-1}(p) = \ln p_j - \ln p_0$$

- From now on, I will use  $\phi(p)$  to denote  $Q^{-1}(p)$ .

# Arcidiacono and Miller's Lemma

An equivalent result to the HM inversion was introduced by Arcidiacono and Miller (2011). It's worth introducing here because it makes everything from now on much simpler and more elegant.

## Arcidiacono Miller Lemma

For any action-state pair  $(a, x)$ , there exists a function  $\psi$  such that

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

Proof:

$$\begin{aligned} V(x) &= \int \max_j \{v_j(x) + \varepsilon_j\} dG(\varepsilon_j) \\ &= \int \max_j \{v_j(x) - v_a(x) + \varepsilon_j\} dG(\varepsilon_j) - v_a(x) \end{aligned}$$

## Important relationships

- The Hotz-Miller Inversion allows us to map from CCP's to differences in conditional (choice-specific) value functions:

$$\phi_{ja}(p(x)) = v_j(x) - v_a(x)$$

- The Arcidiacono and Miller Lemma allows us to relate ex ante and conditional (choice specific) value functions:

$$V(x) = v_j(x) + \psi_j(p(x))$$

- For the logit case:

$$\phi_{ja}(p(x)) = \ln(p_j(x)) - \ln(p_a(x))$$



## Estimation example: finite state space I

- Let's suppose that  $X$  is a finite state space. Furthermore, let's "normalize" the payoffs for a reference action  $\pi_0(x) = 0$  for all  $x$ .
  - We'll discuss soon whether this should really be called a "normalization"
- Using vector notation, recall the definition of the choice-specific value function for the reference action:

$$v_0 = \pi_0 + \beta F_0 V$$

$$v_0 = \beta F_0 V$$

- Using the Arcidiacono-Miller Lemma,

$$V - \psi_0(p) = \beta F_0 V$$

$$\Rightarrow$$

$$V = (I - \beta F_0)^{-1} \psi_0(p)$$

## Estimation example: finite state space II

- Now we have an expression for the ex ante value function that only depends on objects we can estimate in a first stage:

$$V = (I - \beta F_0)^{-1} \psi_0(p)$$

- To estimate the utility function for the other actions,

$$v_j = \pi_j + \beta F_j V$$

$$V - \psi_j(p) = \pi_j + \beta F_j V$$

$$\pi_j = -\psi_j(p) + (I - \beta F_j) V$$

$$\pi_j = -\psi_j(p) + (I - \beta F_j) (I - \beta F_0)^{-1} \psi_0(p)$$

# Identification of Models I

- If we run through the above argument with  $\pi_0$  fixed to an arbitrary vector  $\tilde{\pi}_0$  rather than 0, we will arrive at the following:

$$\pi_j = A_a \tilde{\pi}_0 + b_j$$

where  $A_a$  and  $b_a$  depend only on things we can estimate in a first stage:

$$A_j = (1 - \beta F_j) (1 - \beta F_0)^{-1}$$

$$b_j = A_j \psi_0(p) - \psi_j(p)$$

- We can plug in any value for  $\tilde{\pi}_0$ , and each value will lead to a different utility function (different values for  $\pi_j$ ). Each of those utility functions will be perfectly consistent with the CCP's we observe.

## Identification of Models II

- Another way to see that the utility function is under-identified: If there are  $|X|$  states and  $|J|$  actions, the utility function has  $|X| |J|$  parameters. However, there are only  $|X| (|J| - 1)$  linearly independent choice probabilities in the data, so we have to restrict the utility function for identification.
- Magnac and Thesmar (2002) make this point as part of their broader characterization of identification of DDC models. Their main result says that we can specify a vector of utilities for the reference action  $\tilde{\pi}$ , a distribution for the idiosyncratic shocks  $G$ , and a discount factor, and we will be able to find a model rationalizing the CCPs that features  $(\tilde{\pi}, \beta, G)$ .

# Identification of Counterfactuals

- Note that imposing a restriction like  $\forall x : \pi_0(x) = 0$  is NOT a normalization in the traditional sense. If we were talking about a static normalization, each  $x$  would represent a different utility function, and  $\pi_0(x) = 0$  would simply be a level normalization. However, in a dynamic model, the payoffs in one state affect the incentives in other states, so this is a substantive restriction.
- What is less clear *a priori* is whether these restrictions matter for counterfactuals. It turns out that some (but not all!) counterfactuals ARE identified, in spite of the under-identification of the utility function. What this means is that whatever value  $\tilde{\pi}_0$  we impose for the reference action, the model will not only rationalize the observed CCP's but also predict the same counterfactual CCP's. Kalouptsidei, Scott, and Souza-Rodrigues (2016) sort out when counterfactuals of DDC models are identified and when they are not.

## Extensions

- Just like we can write the BLP (1995) problem as an MPEC, we can do the same with Rust (1987). [See my MPEC notes].
- We cheated a bit because we assumed that not only were actions discrete but so was the state space. This trick is often attributed to Pesendorfer and Schmidt-Dengler (ReStud 2008).
- If the state space is not discrete we need to do some forward-simulation [next slide]. (Hotz, Miller, Sanders, Smith ReStud 1994).
- Others have extended these ideas to dynamic games. See Aguirragabiria and Mira (Ecma 2002/2008) and Bajari Benkard and Levin (Ecma 2007).
- Srisuma and Linton (2009) [very hard] show how to use Friedholm integral equations of 2nd kind to extend to continuous case.

## From NDP Notes: Policy Iteration (Howard 1960)

An alternative to value function iteration is policy function iteration.

- Make a guess for an initial policy, call it  $a(s) = \arg \max_a U(a, s)$  that maps each state into an action
- Assume the guess is stationary compute the implied  $V(a, s)$
- Improvement Step: improve on policy  $a_0$  by solving

$$a' = \arg \max_a U(a, s) + \beta \sum_{s'} V(a, s') p(s'|s, a)$$

- Helpful to define  $p(\tilde{s}'|s)$  as transition probability under optimal choice  $a(s)$ .
- Determine if  $\|a' - a\| < \epsilon$ . If yes then we have found the optimal policy  $a^*$  otherwise we need to go back to step 2.

## From NDP Notes: Policy Iteration (Howard 1960)

Policy Iteration is even easier if choices AND states are discrete.

- For Markov transition matrix  $\sum_j P_{ij} = 1$ , we want  $\pi P = \pi$
- $\lim_{t \rightarrow \infty} P^t = \pi$  where the  $j$ th element of  $\pi$  represents the long run probability of state  $j$ .
- We want the eigenvalue for which  $\lambda = 1$ .

Now updating the value function is easy for  $k$ th iterate of PI

$$\begin{aligned} V^k(s) &= Eu(a^k(s), s) + \beta \tilde{P}^k V^k(s) \\ \Rightarrow V^k(s) &= [1 - \beta \tilde{P}^k]^{-1} Eu(a^k(s), s) \end{aligned}$$

- Very fast when  $\beta > 0.95$  and  $s$  is relatively small. (Rust says 500 more like 5000).
- Inverting a large matrix is tricky



# Continuous State Space

When state space is continuous instead of discrete:

Approximation to the problem:

$$\hat{V}(s) = \max_{a \in \hat{A}(s)} \left[ (1 - \beta)u(s, a) + \beta \sum_{k=1}^N \hat{V}(s'_k) p_N(s'_k | s, a) \right]$$

Exact Problem

$$V(s) = \max_{a \in A(s)} \left[ (1 - \beta)u(s, a) + \beta \int V(s') p(ds' | s, a) \right]$$

- Now we need to do actual numerical integration instead of just summation.
- We cannot use the  $[I - \beta P]^{-1}$  to get the ergodic distribution.
- Usually requires interpolating between grid points to evaluate  $EV(\cdot)$ .

## Forward Simulation

In practice, "truncate" the infinite sum at some period  $T$ :

$$\begin{aligned}\tilde{V}(x, d = 1; \theta) = & \\ & u(x, d = 1; \theta) + \beta E_{x'|x, d=1} E_{d'|x'} E_{\epsilon''|d', x'} [u(x', d'; \theta) + \epsilon' \\ & + \beta E_{x''|x', d''} E_{d''|x''} E_{\epsilon'|d'', x''} [u(x'', d''; \theta) + \epsilon'' + \dots \\ & \beta E_{x^T|x^{T-1}, d^{T-1}} E_{d^T|x^T} E_{\epsilon^T|d^T, x^T} [u(x^T, d^T; \theta) + \epsilon^T]]]\end{aligned}$$

Also, the expectation  $E_{\epsilon|d, x}$  denotes the expectation of the  $\epsilon$  conditional choice  $d$  being taken, and current mileage  $x$ . For the logit case, there is a closed form:

$$E[\epsilon|d, x] = \gamma - \log(\text{Pr}(d|x))$$

where  $\gamma$  is Euler's constant ( $0.577 \dots$ ) and  $\text{Pr}(d|x)$  is the choice probability of action  $d$  at state  $x$ .

## Forward Simulation

Choice-specific value functions can be simulated by (for  $d = 1, 2$ ):

$$\begin{aligned}\tilde{V}(x, d; \theta) \approx & \frac{1}{S} \sum_s [u(x, d; \theta) + \beta[u(x'^s, d'^s; \theta) + \gamma - \log(\hat{P}(d'^s|x'^s)) \\ & + \beta[u(x''^s, d''^s; \theta) + \gamma - \log(\hat{P}(d''^s|x''^s)) + \beta \cdots ]]]\end{aligned}$$

where:

- $x'^s \sim \hat{G}(\cdot|x, d)$
- $d'^s \sim \hat{p}(\cdot|x'^s), x''^s \sim \hat{G}(\cdot|x'^s, d'^s)$
- & etc.

In short, you simulate  $\tilde{V}(x, d; \theta)$  by drawing  $S$  **sequences** of  $(d_t, x_t)$  with a initial value of  $(d, x)$ , and computing the present-discounted utility correspond to each sequence.

Then the simulation estimate of  $\tilde{V}(x, d; \theta)$  is obtained as the sample average.

- Given an estimate of  $\tilde{V}(\cdot, d; \theta)$ , you can get the *predicted choice probabilities*:

$$\tilde{p}(d = 1|x; \theta) \equiv \frac{\exp \left( \tilde{V}(x, d = 1; \theta) \right)}{\exp \left( \tilde{V}(x, d = 1; \theta) \right) + \exp \left( \tilde{V}(x, d = 0; \theta) \right)} \quad (1)$$

and analogously for  $\tilde{p}(d = 0|x; \theta)$ . Note that the predicted choice probabilities are different from  $\hat{p}(d|x)$ , which are the actual choice probabilities computed from the actual data. The predicted choice probabilities depend on the parameters  $\theta$ , whereas  $\hat{p}(d|x)$  depend solely on the data.

## Hotz-Miller (1993) to Aguirregabiria and Mira (2002)

- Choice probabilities conditional on any value of observed state variables are uniquely determined by the vector of normalized value functions
- HM show invertibility proposition (under some conditions).
- If mapping is one-to-one we can also express value function in terms of choice probabilities.

$$\begin{aligned}V_{\theta}(x) &= h(P(d|x, \theta), s, \theta) \\P(d|x, \theta) &= g(V_{\theta}(x), s, \theta) \\ \Rightarrow P(d|x, \theta) &= g(h(P(d|x, \theta), s, \theta), s, \theta)\end{aligned}$$

- The above fixed point relation is used in Aguirregabiria and Mira (2002) in their NPL Estimation algorithm.

## Hotz-Miller (1993) to Aguirregabiria and Mira (2002)

$$P^{k+1}(d|x, \theta) = g(h(\hat{P}^k(d|x, \theta), s, \theta), s, \theta)$$

- Key point here is that the functions  $h(\cdot)$  and  $g(\cdot)$  are quite easy to compute (compared to the inverse  $f^{-1}$ ).
- We can substantially improve estimation speed by replacing  $P$  with  $\hat{P}$  the Hotz-Miller simulated analogue.
- The idea is to reformulate the problem from **value space** to **probability space**.
- When initializing the algorithm with consistent nonparametric estimates of CCP, successive iterations return a sequence of estimators of the structural parameters
- Call this the  $K$  stage policy iteration (PI) estimator.

## Hotz-Miller (1993) to Aguirregabiria and Mira (2002)

- This algorithm nests Hotz Miller ( $K = 1$ ) and Rust's NFXP ( $K = \infty$ ).
- Asymptotically everything has the same distribution, but finite sample performance may be increasing in  $K$  (at least in Monte Carlo).
- The Nested Pseudo Likelihood (NPL) estimator of AM ( $K = 2$ ) seems to have much of the gains.