Discussion of Efficient and Convergent Sequential Pseudo-Likelihood Estimation of Dynamic Discrete Games

Chris Conlon

NYU Stern

April 6, 2019

A solid IIOC tradition

As per usual:

- I am asked to discuss *A More Complicated than ever before Dynamic Game* paper (in 5 minutes).
- My usual bag of tricks:

A solid IIOC tradition

As per usual:

- I am asked to discuss *A More Complicated than ever before Dynamic Game* paper (in 5 minutes).
- My usual bag of tricks:
  - Write the game as a decision problem
  - Write the dynamic problem as a static one.
  - ie: Avoid dynamic games at all costs...

A solid IIOC tradition

As per usual:

- I am asked to discuss *A More Complicated than ever before Dynamic Game* paper (in 5 minutes).
- My usual bag of tricks:
  - Write the game as a decision problem
  - Write the dynamic problem as a static one.
  - ie: Avoid dynamic games at all costs...
- So I googled around for lecture notes on Aguirregabiria Mira (2002/2007) and I found...

A solid IIOC tradition

As per usual:

- I am asked to discuss *A More Complicated than ever before Dynamic Game* paper (in 5 minutes).
- My usual bag of tricks:
  - Write the game as a decision problem
  - Write the dynamic problem as a static one.
  - ie: Avoid dynamic games at all costs...
- So I googled around for lecture notes on Aguirregabiria Mira (2002/2007) and I found...
  - My own damn slides!

Often faced with extremum estimator problems in econometrics (ML, GMM, MD, etc.) that look like:

$$\hat{\theta} = \arg\max_{\theta} Q_n(\theta), \quad \theta \in \Theta \tag{1}$$

Many economic problems contain constraints, such as: market clearing (supply equals demand), consumer's consume their entire budget set, or firm's first order conditions are satisfied. A natural way to represent these problems is as constrained optimization.

Constrained Problems

MPEC

$$\hat{\theta} = \arg\max_{\theta, Y} Q_n(\theta, Y), \quad \text{s.t.} \quad G(Y, \theta) = 0, \quad \theta \in \Theta$$

Fixed Point / Implicit Solution

In much of the literature the tradition has been to express the solutions $G(Y, \theta) = 0$ implicitly as $G(\theta)$:

$$\hat{\theta} = \arg\max_{\theta} Q_n(\theta, G(\theta)), \quad \theta \in \Theta$$

Rust Problem

- Bus repairman sees mileage x_t at time t since last overhaul
- Repairman chooses between overhaul and normal maintenance

$$u(x_t, d_t, \theta^c, RC) = \begin{cases} -c(x_t, \theta^c) & \text{if} \quad d_t = 0 \\ -(RC + c(0, \theta^c) \quad) & \text{if} \quad d_t = 1 \end{cases}$$

- Repairman solves DP:

$$V_\theta(x_t) = \sum_{f_t, f_{t+1}, \ldots} E\left\{ \sum_{j=t}^{\infty} \beta^{j-t} [u(x_j, f_j, \theta) + \varepsilon_j(f_j)] | x_t \right\}$$

- Econometrician
  - Observes mileage x_t and decision d_t but not cost.
  - Assumes extreme value distribution for $\varepsilon_t(d_t)$
- Structural parameters to be estimated $\theta = (\theta^c, RC, \theta^p)$.
  - Coefficients of cost function $c(x, \theta^c) = \theta_1^c x + \theta_2^c x^2$
  - Overhaul cost RC; Transition probabilities in mileages $p(x_{t+1} | x_t, d_t, \theta^p)$

Rust Problem

- Data: time series $(x_t, d_t)_{t=1}^{T}$
- Likelihood function

$$
\begin{aligned}
\max_{\theta \geq 0} l(\theta) &= \sum_{t=2}^{T} \log Pr(d_t | x_t, \theta^c, RC) + \log p(x_t | x_{t-1}, d_{t-1}, \theta^p) \\
\text{with } Pr(d|x, \theta^c, RC) &= \frac{\exp[u(x, d, \theta^c, RC) + \beta EV_\theta(x, d)}{\sum_{d' \in \{0,1\}} \exp[u(x, d', \theta^c, RC) + \beta EV_\theta(x', d)} \\
EV_\theta(x, d) &= T_\theta(EV_\theta)(x, d)
\end{aligned}
$$

$$
\equiv \int_{x'=0}^{\infty} \log \left[\sum_{d' \in \{0,1\}} \exp[u(x, d', \theta^c, RC) + \beta EV_\theta(x', d)] \right] p(dx' | x, d, \theta^p)
$$

Rust Problem

- Outer Loop: Solve Log Likelihood

$$\max_{\theta \geq 0} l(\theta) = \sum_{t=2}^{T} \log Pr(d_t | x_t, \theta^c, RC) + \log p(x_t | x_{t-1}, d_{t-1}, \theta^p)$$

- Convergence test: $\|\nabla_\theta \mathcal{L}(\theta)\| \leq \epsilon_{out}$
- Inner Loop: Compute expected value function EV_θ for a given θ
- Main idea: Use values to get CCP's.
- EV_θ is the implicit expected value function defined by the Bellman equation or the fixed point function

$$EV_\theta = T_\theta(EV_\theta)$$

- Convergence test: $\left\| EV_\theta^{(k+1)} - EV_\theta^{(k)} \right\| \leq \epsilon_{in}$
- Start with contraction iterations and polish with Newton Steps

Hotz and Miller (1992) Prop 2, Arcidiacono-Miller Lemma

- Relate choice-specific value's to CCP's (just like Rust)

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

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

- In the logit-case HM inverse is easy to get choice specific value functions

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

- And we go from choice-specific to ex-ante value function

$$EV(x) = v_j(x) - \ln\left(p_j(x)\right) + \gamma$$

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
- If mapping is one-to-one we can also express value function in terms of choice probabilities.

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

- 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)

$$
\begin{aligned}
\hat{\theta} &= \arg\max_{\theta} Q_n(\theta, \hat{P}^k) \\
P^{k+1}(d|x,\theta) &= g(h(\hat{P}^k(d|x,\theta), s, \theta), s, \theta)
\end{aligned}
$$

- Key point here is that the functions $h(\cdot)$ and $g(\cdot)$ are quite easy to compute.
- How do we get initialize P? try $\hat{P}$ the Hotz-Miller simulated analogue.
- The idea is to reformulate the problem from value space to probability space.
- This algorithm nests Hotz Miller ($K = 1$) and Rust's NFXP ($K = \infty$).

- Form the augmented likelihood function for data $X = (x_t, d_t)_{t=1}^{T}$

$$\mathcal{L}(EV, \theta; X) = \prod_{t=2}^{T} P(d_t|x_t, \theta^c, RC)p(x_t|x_{t-1}, d_{t-1}, \theta^p)$$

$$\text{with } P(d|x, \theta^c, RC) = \frac{\exp[u(x, d, \theta^c, RC) + \beta EV(x, d)}{\sum_{d' \in \{0,1\}} \exp[u(x, d', \theta^c, RC) + \beta EV(x', d)}$$

- Rationality and Bellman equation imposes a relationship between θ and EV

$$EV = T(EV, \theta)$$

- Solve constrained optimization problem

$$\max_{(\theta, EV)} \mathcal{L}(EV, \theta; X)$$

$$\text{subject to } EV = T(EV, \theta)$$

Says no, don't do AM: go back to value space!

- Newton-Kantorovich update:

$$\Upsilon_k(\theta, \hat{\gamma}_{k-1}) = \hat{Y}_k - \left(\nabla_Y G\left(\hat{\theta}_{k-1}, \hat{Y}_{k-1}\right)\right)^{-1} G\left(\theta, \hat{Y}_{k-1}\right)$$

- Solve constrained problem iteratively:

$$\begin{aligned} \text{Step 1:} \quad & \hat{\theta}_k = \arg\max_{\theta \in \Theta} \quad \hat{Q}\left(\Upsilon_k(\theta, \hat{\gamma}_{k-1})\right) \\ \text{Step 2:} \quad & \hat{Y}_k = \Upsilon_k\left(\hat{\theta}_k, \hat{\gamma}_{k-1}\right) \end{aligned}$$

- Where $Y \equiv v_j$ (choice specific value function) and $G(\theta, Y) = v - \Phi(\theta, v)$ (eq in value functions).

Comments

1. If Q depends on θ only through $\Upsilon_k(\theta, \hat{\gamma}_{k-1})$. Is this iterative updating or is this successive approximation?
2. We usually think that $Q(\theta) \equiv l(\theta) = \sum_{i=1}^{N} \log Pr(d|x, \theta)$ (this is naturally in the space of CCP's).
    - Did the objective function change?
    - Did we not write down the Value $\rightarrow$ CCP step? My $g(\cdot)$ function.
3. Is the key the N-K update step? Or is it writing the problem in value space?
4. Are we secretly doing Su and Judd (2012)?
5. Are we re-inventing constrained Quasi-Newton barrier methods for Su and Judd (2012) problem?

Thanks!