Lecture 5: Few more estimation methods: BBL and MSM

With application to Zurcher model

Fedor Iskhakov, Australian National University
Anders Munk-Nielsen, University of Copenhagen
John Rust, Georgetown University
Bertel Schjerning, University of Copenhagen

Empirical IO: Dynamic Structural Models
Department of Economics (NHH)
29 Nov - 3 Dec., 2021

BBL

Bajari, Benkard and Levin (2007)
Estimating Dynamic Models of Imperfect Competition
Econometrica, 75(5), 1331-1370

Main Contributions of Bajari, Benkard, Levin (2007)

Two-step algorithm for estimating dynamic games

- In the first step, the policy functions and the law of motion for the state variables are estimated.
- In the second step, the remaining structural parameters are estimated using the optimality conditions for equilibrium.
- The second step estimator is a simple simulated minimum distance estimator.
- The algorithm applies to a broad class of models, including industry competition models with both discrete and continuous controls such as the Ericson and Pakes (1995) model.

Monte Carlo experiments BBL test the algorithm on

- a dynamic discrete choice model with normally distributed errors
- a dynamic oligopoly model similar to that of Pakes and McGuire (1994).

Application of BBL approach to the bus engine problem

- Suppose we knew the choice-specific value functions $v(x, 1, \theta^*)$ (value of replacing a bus with x miles since last engine replacement) and $v(x, 0, \theta^*)$ (value of not replacing the bus engine)
- Optimal decision to keep the engine for given $\epsilon = (\epsilon(0), \epsilon(1))$ is

$$\delta(x, \epsilon, \theta^*) = \begin{cases} 1 & \text{if } v(x, 1, \theta^*) + \epsilon(1) \leq v(x, 0, \theta^*) + \epsilon(0) \\ 0 & \text{otherwise} \end{cases}$$

- Let $\delta^c(x, \epsilon, \theta^*) = 1 - \delta(x, \epsilon, \theta^*)$
  By definition this is *suboptimal decision rule. Hence we have for any* (x, ϵ)

$$v(x, \delta(x, \epsilon, \theta^*), \theta^*) + \epsilon(\delta(x, \epsilon, \theta^*)) \leq v(x, \delta^c(x, \epsilon, \theta^*), \theta^*) + \epsilon(\delta^c(x, \epsilon, \theta^*))$$

  i.e. the value of taking the optimal action is always less than the value of taking a suboptimal action prescribed by $\delta^c(x, \epsilon, \theta^*)$.

Application of BBL approach to the bus engine problem

- Define the function $g(x,\epsilon,\theta)$ by

$$\begin{aligned} g(x,\epsilon,\theta) &= v(x,\delta^c(x,\epsilon,\theta^*),\theta)+\epsilon(\delta^c(x,\epsilon,\theta^*)) \\ &- v(x,\delta(x,\epsilon,\theta^*),\theta)+\epsilon(\delta(x,\epsilon,\theta^*)) \end{aligned}$$

- Notice that at $\theta=\theta^*$ we have $g(x,\epsilon,\theta^*)\geq 0$ for any *perturbed policy* $\delta^c(x,\epsilon,\theta)$
- If $H(x,\epsilon,\delta^c)$ is some CDF over the space of policies, we have

$$\theta^* \in \underset{\theta}{argmin} \int_x \int_\epsilon (\min[g(x,\epsilon,\theta),0])^2 \, dH(x,\epsilon) \quad \text{(MIE)}$$

- Thus, if we can somehow be able to compute/simulate $v(x,1,\theta)$ and $v(x,0,\theta)$ for any θ and uncover/estimate and simulate the true optimal decision rule $\delta(x,\epsilon,\theta^*)$ for any value of (x,ϵ) then we can compute $g(x,\epsilon,\theta)$ and then search for the value θ^* that minimizes the criterion in the moment inequality estimator (MIE).

Application of BBL approach to the bus engine problem

- How do we uncover $\delta(x, \epsilon, \theta^*)$ if we do not know θ^*?
- Recall from the Hotz-Miller inversion of the CCPs: then we have

$$\log(P(1|x, \theta^*)/P(0|x, \theta^*)) = v(x, 1, \theta^*) - v(x, 0, \theta^*)$$

- Then we have

$$\begin{aligned}\delta(x, \epsilon, \theta^*) &= \begin{cases} 1 & \text{if } v(x, 1, \theta^*) - v(x, 0, \theta^*) \leq \epsilon(0) - \epsilon(1) \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} 1 & \text{if } \log(P(1|x, \theta^*)/P(0|x, \theta^*)) \leq \epsilon(0) - \epsilon(1) \\ 0 & \text{otherwise} \end{cases}\end{aligned}$$

- Thus if we can non-parametrically estimate $P(1|x, \theta^*)$ and simulate draws of $\epsilon = (\epsilon(0), \epsilon(1))$, then we can simulate values of $\delta(x, \epsilon, \theta^*)$ for any (x, ϵ) even though we do not know θ^*.

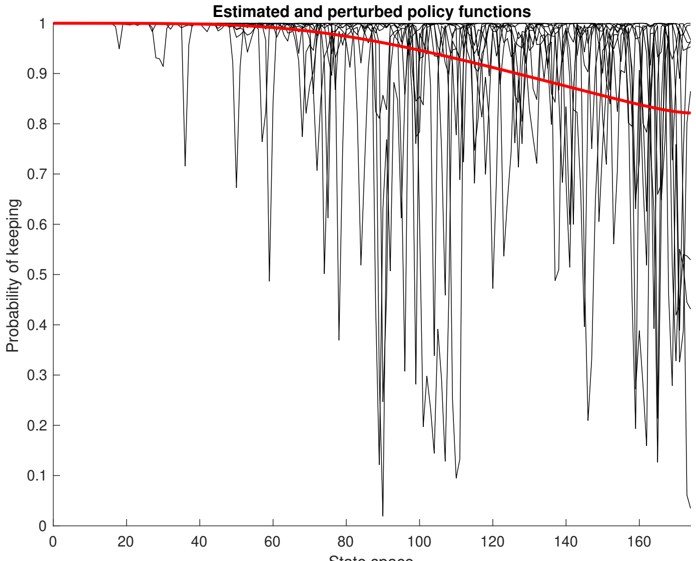
Application of BBL approach to the bus engine problem

- In the first stage we cannot exactly recover $P(1|x,\theta^*)$ from data, only a noisy estimate of it, $\hat{P}(1|x,\theta^*)$
- Need consistent estimate of the optimal decision rule, $\hat{\delta}(x,\epsilon)$ (that [hopefully] converges uniformly to the true decision rule $\delta(x,\epsilon,\theta^*)$ with number of observations $N \to \infty$
- Also, numerical and simulation noise $(\hat{v}(x,1,\theta), \hat{v}(x,0,\theta))$
- As a result the moment inequality function $\hat{g}(x,\epsilon,\theta)$ does not necessarily satisfy $\hat{g}(x,\epsilon,\theta) \geq 0$ at $\theta = \theta^*$
- However, as $N \to \infty$ we can expect that $\hat{g}(x,\epsilon,\theta^*) \to g(x,\epsilon,\theta^*) \geq 0$ with probability 1, so the moment inequalities will be satisfied at the true parameter values.

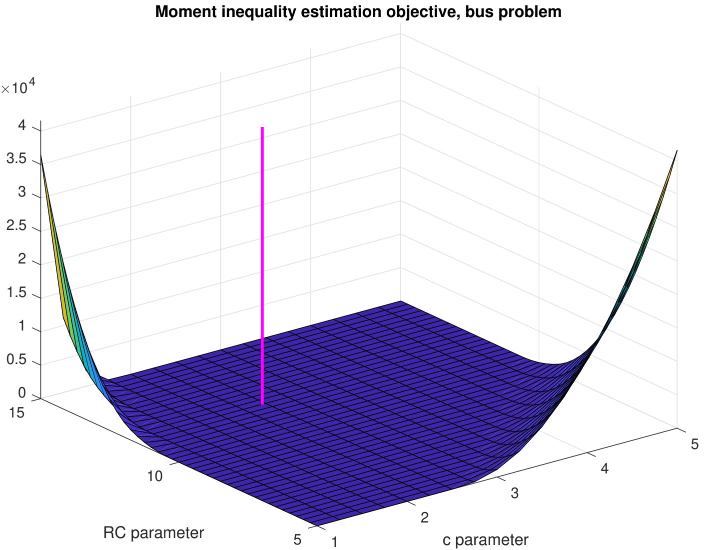
Application of BBL approach to the bus engine problem

- Now suppose we can instead *solve* the model, so that for any θ:
- We can calculate $(v(x, 1, \theta), v(x, 0, \theta))$ and the optimal policy $\delta(x, \epsilon, \theta)$
- Then it is easy to see we can evaluate the moment inequality function $g(x, \epsilon, \theta)$for any θ, (x, ϵ) and for any perturbed policy $\delta^c(x, \epsilon, \theta)$.
- Hence, we can evaluate the MIE objective function

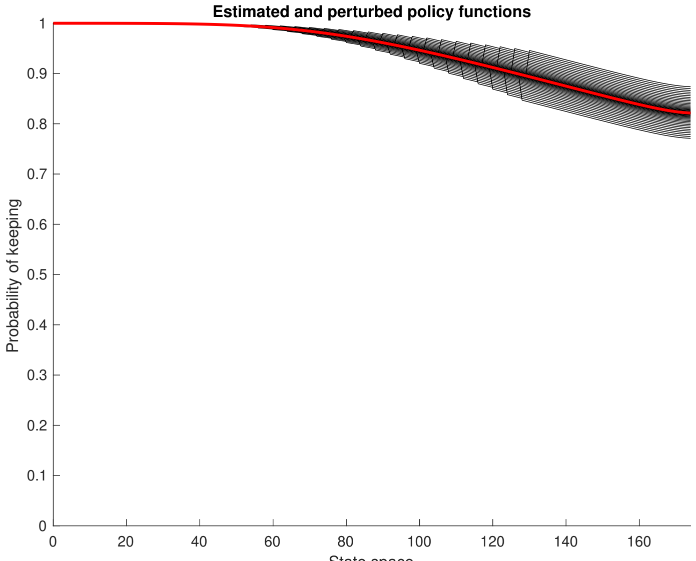
Policy perturbations, bus engine problem



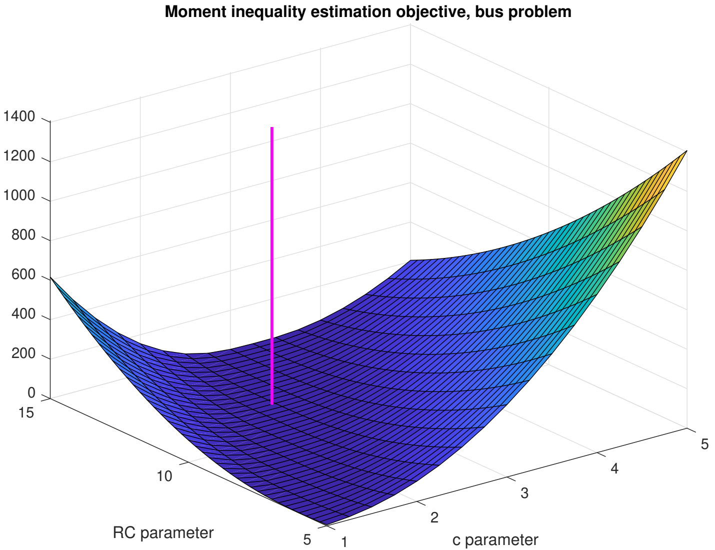
Moment inequality objective, bus engine problem



Policy perturbations, bus engine problem



Moment inequality objective, bus engine problem



Specific issues with BBL

From private communication with Sanjog Misra, Prof of Marketing at the University of Chicago Booth School of Business

1. "The notion of a 'perturbation' from the estimated optimal policy to create violations and consequently inequalities is not well articulated in BBL"
2. "What we found was that movements in the perturbation could swing results quite a bit."
3. "The perturbations have to be chosen carefully to obtain point estimates. This gives it a flavor of calibration rather than estimation."
4. "The objective function of the moment based inequality estimator is terrible (in the sense that it is flat). For some of the agents (Application to firm compensation, Misra and Nair, 2011, estimated the model agent by agent) I had to resort to a visual search (luckily we had a 2-dim search) to find parameter regions that were close enough to the optimal and then start the optimization there."

BBL: conclusion

- Two-step estimator with similar first stage to CCP and other
- Lots of *tuning in* required
- Can be applied to models which are hopeless to be solved explicitly

- Use with care when other methods are helpless

Matlab Implementation

MSM

McFadden (1989)
A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration
Econometrica, 57(5), 995

Pakes and Pollard (1989)
Simulation and the Asymptotics of Optimization Estimators
Econometrica, 57(5), 1027-1057

McFadden's MSM estimator for discrete models

- McFadden (1989) introduced the method of simulated moments (MSM) which can enable consistent estimation of discrete choice models with only 1 simulated utility draw per observation. The basic idea of MSM is that we can use the law of large numbers to average out simulation error in the same way we use the LLN to average out sampling error.
- MSM first introduced for estimation of static discrete choice models with non-extreme valued error terms with high dimensional choice sets
- However people soon realized that the idea of MSM is very general and extends to a much broader class of problem
- Jean-Francois Richard: "If you can simulate it, you can estimate it"
- MSM – "Swiss Army Knife" of structural estimation

McFadden's MSM estimator for discrete models

- Basic idea is very similar to *calibration* used in macroeconomics:
  1. you, as econometrician, determine a vector of *actual moments* characterizing/summarizing the data you observe and want to explain with your model
  2. For any parameter value θ for your structural model, solve the model and "simulate data" from the model to construct a correponding vector of *simulated moments*
  3. Search over θ for a value $\hat{\theta}_{\text{msm}}$ that enables the simulated moments to "best fit" the actual moments
- Actual data: $\{x_t\}$, $t = 1, \ldots, T$ (cross section, panel, or time series)
- Define moments as a $J \times 1$ mapping $h(x_t) \in R^J$. The moments are

$$\overline{h}_T = \frac{1}{T} \sum_{t=1}^{T} h(x_t)$$

Solving and simulating your structural model

- Now, suppose we have a structural model that depends on a vector θ of "structural parameters" we want to estimate
- Can be dynamic/static, singe/multiple agent, discrete/continuous choices and states
- Let $\{\tilde{x}_t^s(\theta)\}$, $t = 1, \ldots, T$, $s = 1, \ldots, S$ be S *IID* sets of *simulated data* from your structural model for the guess of the parameters θ.
- Similar to McFadden's case, we can get by with as few as $S = 1$ simulation of the model for the T observations/time periods.
- Form *simulated moments* $\overline{h}_T^S(\theta)$ as follows

$$\overline{h}_T^S(\theta) = \frac{1}{S} \sum_{s=1}^{S} \frac{1}{T} \sum_{t=1}^{T} h(\tilde{x}_t^s(\theta)).$$

Equations and unknowns

- Let there be K unknown parameters to be estimated, $\theta \in R^K$.
- Must have $J \geq K$ "equations"

$$\overline{h}_T = \overline{h}_T^S(\hat{\theta}_{\text{msm}}).$$

  This is the simulation analog of the classical "method of moments" introduced by Pafnuty Chebyshev in 1887.
- Usually it's the *overidentified case* where $J > K$
- So in the general case, MSM mimics that GMM strategy of Sargan/Hansen and we define the MSM estimator as

$$\hat{\theta}_{\text{msm}} = \underset{\theta}{argmin} \left([\overline{h}_T - \overline{h}_T^S(\theta)]' W_T [\overline{h}_T - \overline{h}_T^S(\theta)] \right)$$

  where W_T is a $J \times J$ positive semi-definitive weighting matrix.

Key Assumptions for MSM Estimatation

- **Correct specification** There is a value θ^* such that $\{\tilde{x}_t\} \sim \{\tilde{x}_t(\theta^*)\}$, i.e. that simulated data from the structural model at θ^* has the same probability distribution (stochastic process) as the actual data.
- **Identification** If $\theta \neq \theta^*$ then $E\{h(\tilde{x}(\theta)\} \neq E\{h(\tilde{x}(\theta^*)\}$.
- **Differentiability** The gradient $\nabla E\{h(\tilde{x}(\theta)\}$ $(J \times K)$ exists and is a continuous function of θ.
- **Full Rank** The generalized inverse of the asymptotic covariance matrix of the moments Ω, Ω^+, exists and the $K \times K$ matrix Λ given by
$$\Lambda = [\nabla E\{h(\tilde{x}(\theta^*)\}'[\Omega^+]\nabla E\{h(\tilde{x}(\theta^*)\}]$$
exists and is invertible.
- **Theorem** *Under the assumptions above (and other technical assumptions to guarantee LLN/CLT hold) we have*
$$\sqrt{T}[\hat{\theta}_{\mathrm{msm}} - \theta^*] \Longrightarrow N(0, (1 + 1/S)\Lambda^{-1}).$$

Low cost/penalty for simulating moments

- Notice that even with a single simulation "per observation" ($S = 1$), the MSM is consistent and asymptotically normally distributed and the penalty for using only a single simulation is just a doubling of the asymptotic covariance matrix of $\hat{\theta}_{\text{msm}}$ compared to the case of "exact integration" ($S = \infty$), since $(1 + 1/S) = 2$ when $S = 1$.
- Power of MSM approach: we can estimate models where it is very difficult/impossible to form a likelihood, so it is a flexible/convenient way to handle a host of econometric problems such as
  1. Measurement error
  2. Censoring/attrition/missing data
  3. Endogeneity
- Downside of MSM: its power can enable you to estimate *statistically degenerate models* i.e. models where certain observations have zero probability of occurring under the model due to lack of sufficient flexibility, no unobserved shocks, etc

Example: MSM estimation of the bus replacement problem

- The data: T buses, N_i observed pairs of mileage and replacement decisions

$$\{x_{it}, d_{it}\}_{i=1,..,T;t=1,..,N_i}$$

- **Moments**
  - Consider $J+1$ *mileage bins* defined by mileage cutoffs $(c_0, c_1, c_2, \dots, c_{J+1})$ where $c_0 = 0$ and c_{J+1} = maximum mileage
  - J moments defined as fractions of buses in each but one of the mileage bins in the long run (stationary disctribution)

$$\bar{h}_j = \frac{1}{T} \sum_{i=1}^{T} \left[\frac{1}{N_i} \sum_{t=1}^{N_i} 1\{c_{j-1} \leq x_{it} \leq c_j\} \right]$$

Example: MSM estimation of the bus replacement problem

- **Simulated moments**
  1. Solve the model to compute choice and transition probabilities for give parameter θ
  2. Simulate mileage $\tilde{x}_{it}(\theta)$ and replacements $\tilde{d}_{it}(\theta)$ for N buses and T periods using a collection of IID random numbers *with fixed seed* to avoid simulation noise during the estimation
  3. Compute simulated moments

$$\bar{h}_j^S(\theta) = \frac{1}{T}\sum_{i=1}^{T}\left[\frac{1}{N_i}\sum_{t=1}^{N_i} 1\{c_{j-1} \leq \tilde{x}_{it}(\theta) \leq c_j\}\right]$$

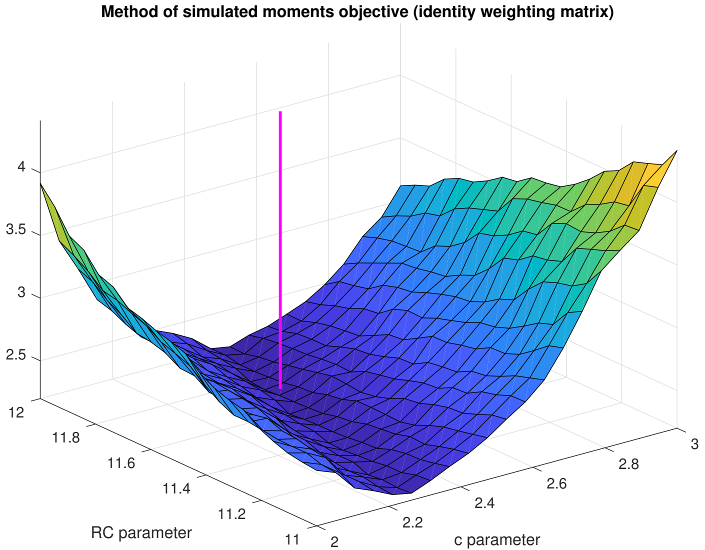
- The MSM criterion

$$Q(\theta) = g(\theta)Wg(\theta)', \quad g(\theta) = \bar{h} - \bar{h}^S(\theta)$$

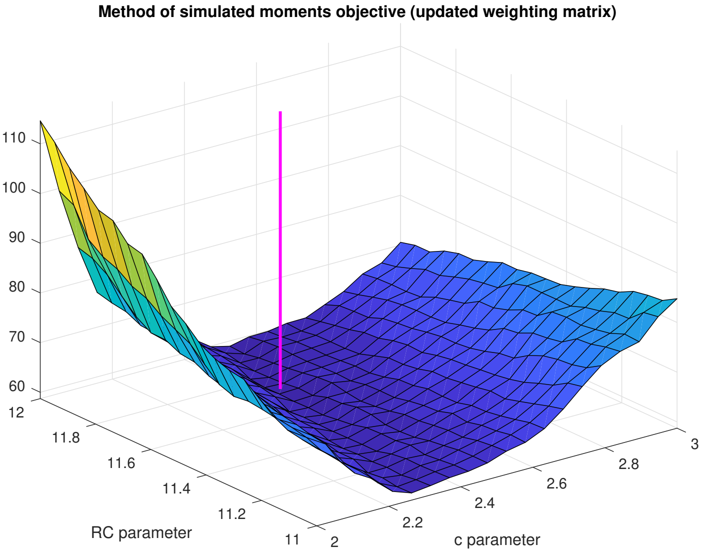
$g(\theta)$ is a 1 by J vector
W is J by J weighting matrix

Data and simulated moments, bus engine problem

MSM criterion, identity weighting matrix



MSM criterion, updated weighting matrix



Matlab Implementation