CCP Estimators

Robert A. Miller

Structural Econometrics

October 2017

Criteria for Evaluating Estimators

Three criteria for assessing an estimator

- Three criteria for evaluating different estimators are:
  1. Large sample properties:
     - Does the estimator converge to the identified set?
     - If so, what is the rate of convergence?
     - What is the asymptotic distribution of the estimator?
  2. Finite sample properties:
     - At what sample size do the finite sample properties accurately reflect the asymptotic distribution?
     - For a given sample size, what is the standard deviation and mean squared error of the estimator ?
  3. Implementation:
     - Is the estimator defined by an algorithm or only a set of conditions to be satisfied?
     - Are numerical approximations involved?
     - Does the estimator require tuning parameters or instruments?

Large Sample or Asymptotic Properties

In what sense does an estimator converge, and what does it converge to?

- There are several types of convergence, such as: almost sure, in mean square, and in probability.
- Given a type of convergence, we ask:
  1. Does the estimator converge to a set that includes the identified set? In other words is the estimator is tight?
  2. Is the set of parameters to which the estimator converges included in the identified set? In other words is the estimator is sharp?
- If both conditions are satisfied, then we say the estimator is consistent.
- For example if the identified set is a singleton, that is the model is pointwise identified, then an estimator is consistent if it converges to that singleton.

Large Sample or Asymptotic Properties

The rate of convergence

- The other two criteria are extensively analyzed in econometric theory, and can typically be applied in a straightforward way to dynamic discrete choice models in a straightforward way.
- For example, suppose the parameter space is Θ, the data is generated by $\theta_0 \in \Theta$, the model in point identified, and the estimator, denoted by θ_N is consistent with:

$$\theta_N \underset{p}{\longrightarrow} \theta_0$$

- The rate of convergence is defined by N^α where:

$$\alpha = \arg\sup_a \left[N^a \left(\theta_N - \theta_0 \right) \right] \underset{p}{\longrightarrow} 0$$

- Structural estimates of dynamic discrete choice models are typically $\sqrt{N}$ consistent.

Large Sample or Asymptotic Properties

The asymptotic distribution

- Suppose θ_N converges in probability to θ_0 at rate α.
- Let ξ be drawn from the limiting distribution of $N^\alpha(\theta_N - \theta_0)$:

$$N^\alpha(\theta_N - \theta_0) \underset{d}{\longrightarrow} \xi$$

- Structural estimates of dynamic discrete choice models are typically asymptotically normal.
- An estimator is asymptotically efficient if ξ is $\mathcal{N}\left(0, \mathcal{I}(\theta_0)^{-1}\right)$ where:

$$\mathcal{I}(\theta) \equiv E\left[\frac{\partial l(d, x | x_1 ; \theta)}{\partial \theta}\frac{\partial l(d, x | x_1 ; \theta)}{\partial \theta}'\right] = -E\left[\frac{\partial^2 l(d, x | x_1 ; \theta)}{\partial \theta \partial \theta'}\right]$$

  and the likelihood is based on the sequence (d, x) conditional on the state at date one, x_1.
- The ML estimator for dynamic discrete choice models typically attain $\mathcal{I}(\theta_0)^{-1}$ the Cramer-Rao lower bound.

Implementation

Does an algorithm define the estimator?

- Ideally an estimator is defined by an algorithm that depends on the data for each sample size N. In that case the estimator:
  1. can be implemented mechanically, so is easy to explain;
  2. is easy to replicate on the same and on different data sets, a virtue in scientific enquiry.
- Cell estimators and hence unrestricted ML estimators satisfy this definition.
- An OLS estimator also satisfies the first definition because algorithms exist to invert matrices exactly, within a finite number of steps.
- Similarly Gaussian methods, successively substituting out parameters, solve linear systems quickly within a finite number of steps.

Implementation

Is the estimator defined by a set of conditions it must satisfy?

- A weaker, more inclusive definition is that an estimator solves a set of conditions jointly satisfied by the parameter values and the data.
- Since the algorithm used to implement the estimator is not defined, such estimators are almost invariably, less transparent, and therefore harder to replicate with data.
- Extremum estimators for nonlinear models defined this way include:
  - nonlinear least squares;
  - full solution estimators to dynamic discrete choice models;
  - CCP estimators in which G or β is estimated.
- It is useful to know whether a unique solution exists. For example:
  - Is the minimization (maximization) problem strictly convex (concave)?
- If not, can all the parameters, bar one or two, be solved in terms of the one or two remaining parameters?
  - In the first case, the concentrated objective function can be plotted.
  - In the second equi-value contours can be plotted.

Implementation

Are numerical approximations involved?

- Because ML estimation of dynamic discrete choice models is relatively imposing in terms of programming demands and computational time, researchers economize on both by using numerical approximations:
  1. approximating $E\left[\max\left\{x, y\right\}\right]$ with $\max\left\{E\left[x\right], E\left[y\right]\right\}$;
  2. approximating distant horizons with zero;
  3. linearizing the value function;
  4. interpolating the state space to obtain estimates of continuation values;
  5. approximating smoothed integrals with rectangles and quadrilaterals;
  6. reducing the impact of the state space by treating the continuation value as a sufficient statistic for the state space;
  7. more generally only allowing the individuals to condition on a smaller set of values than there are state variables.
- These approximation errors open a gap between the defined estimator and its numerical counterpart.

Maximum Likelihood Estimation

Data and outside knowledge

- Suppose the data comes from a long panel (either stationary or complete panel histories for finite lived agents).
- Also assume we know:
  1. the discount factor β
  2. the distribution of disturbances $G_t(\epsilon | x)$
  3. $u_{1t}(x)$ (or more generally one of the payoffs for each state and time).
  4. $u_{1t}(x) = 0$ (for notational convenience)
- Since the panel is long, $p_t(x)$ and hence $\psi_{jt}(x)$ are identified.
- There are, of course, alternative assumptions that deliver identification, and the methods described below are generic.

Maximum Likelihood Estimation

The likelihood

- To simplify the notation, consider a sample of N independently drawn observations on the whole history $t \in \{1, \ldots, T\}$ of individuals $n \in \{1, \ldots, N\}$, with data on their state variables decisions denoted by x_{nt}, and decisions denoted by d_{njt}.
- The joint probability distribution of the decisions and outcomes is:

$$\prod_{n=1}^{N} \prod_{t=1}^{T} \left(\sum_{j=1}^{J} \sum_{x'=1}^{X} d_{njt} I \left\{ x_{n,t+1} = x' \right\} p_{jt}(x) f_{jt}(x'|x) \right)$$

- Taking logs yields:

$$\sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J} d_{njt} \left\{ \log \left[p_{jt}(x_{nt}) \right] + \sum_{x=1}^{X} I \left\{ x_{n,t+1} = x \right\} \log \left[f_{jt}(x|x_{nt}) \right] \right\}$$

Maximum Likelihood Estimation

The reduced form

- Note the choice probabilities are additively separable from the transition probabilities in the formula for the joint distribution of decisions and outcomes.
- Hence the estimation of the joint likelihood splits into one piece dealing with the choice probabilities conditional on the state, and another dealing with the transition conditional on the choice and state.
- Maximizing each additive piece separately with respect to $f_j(x'|x)$ and $p_t(x_{nt})$ we obtain the unrestricted ML estimators:

$$\widehat{f}_{jt}\left(x' \left| x\right.\right) = \frac{\sum_{n=1}^{N} I\left\{x_{nt} = x, d_{njt} = 1, x_{n,t+1} = x'\right\}}{\sum_{n=1}^{N} I\left\{x_{nt} = x, d_{njt} = 1\right\}}$$

and:

$$\widehat{p}_{jt}\left(x\right) = \frac{\sum_{n=1}^{N} I\left\{x_{nt} = x, d_{njt} = 1\right\}}{\sum_{n=1}^{N} I\left\{x_{nt} = x\right\}}$$

Maximum Likelihood Estimation

Estimating an intermediate probability distribution

- Following the notation of Lecture 9, let $\kappa_{jt\tau}(x_{t+\tau+1}|x_t)$ denote the probability of reaching $x_{t+\tau+1}$ at $t+\tau+1$ from x_t by following action j at t and then always choosing the first action:

$$\kappa_{jt\tau}(x_{t+\tau+1}|x_t) \equiv \begin{cases} f_{jt}(x_{t+1}|x_t) & \tau = 0 \\ \sum_{x=1}^{X} f_{1,t+\tau}(x_{t+\tau+1}|x)\kappa_{jt,\tau-1}(x|x_t) & \tau = 1, \ldots \end{cases}$$

- Thus we can recursively estimate $\kappa_{jt\tau}(x_{t+\tau+1}|x_t)$ with:

$$\widehat{\kappa}_{jt\tau}(x_{t+\tau+1}|x_t) \equiv \begin{cases} \widehat{f}_{jt}(x_{t+1}|x_t) & \tau = 0 \\ \sum_{x=1}^{X} \widehat{f}_{1,t+\tau}(x_{t+\tau+1}|x)\widehat{\kappa}_{jt,\tau-1}(x|x_t) & \tau = t+1, \ldots \end{cases}$$

- Similarly we estimate $\psi_{jt}(x_t)$ with $\widehat{\psi}_{jt}(x_t)$ using the $\widehat{p}_{jt}(x)$ estimates of the CCPs.

- From previous lectures:

$$\begin{aligned} u_{jt}(x_t) &= \psi_{1t}(x_t) - \psi_{jt}(x_t) \\ &+ \sum_{\tau=1}^{T-t} \sum_{x=1}^{X} \beta^{\tau-t} \psi_{1,t+\tau}(x) \left[\kappa_{t1,\tau-1}(x|x_t) - \kappa_{tj,\tau-1}(x|x_t) \right] \end{aligned}$$

- Substituting $\widehat{\kappa}_{\tau-1}(x|x_t, j)$ for $\kappa_{\tau-1}(x|x_t, j)$ and $\psi_{jt}(x_t)$ with $\widehat{\psi}_{jt}(x_t)$ then yields:

$$\begin{aligned} \widehat{u}_{jt}(x_t) &\equiv \widehat{\psi}_{1t}(x_t) - \widehat{\psi}_{jt}(x_t) \\ &+ \sum_{\tau=1}^{T-t} \sum_{x=1}^{X} \beta^{\tau-t} \widehat{\psi}_{1,t+\tau}(x) \left[\widehat{\kappa}_{t1,\tau-1}(x|x_t) - \widehat{\kappa}_{tj,\tau-1}(x|x_t) \right] \end{aligned}$$

- The stationary case is similar (and has the matrix representation we discussed in previous lectures).

Large Sample or Asymptotic Properties

Properties of the unrestricted ML estimator

- By the Law of Large Numbers $\widehat{f}_{jt}(x' | x)$ converges to $f_{jt}(x' | x)$ and $\widehat{p}_{jt}(x)$ converges to $p_{jt}(x)$, both almost surely.
- By the Central Limit Theorem both estimators converge at $\sqrt{N}$ and and have asymptotic normal distributions.
- Both $\widehat{f}_{jt}(x' | x)$ and $\widehat{p}_{jt}(x)$ are ML estimators for $f_{jt}(x' | x)$ and $p_{jt}(x)$ and obtain the Cramer-Rao lower bound asymptotically.
- Since and $u_{jt}(x)$ is exactly identified, it follows by the invariance principle that $\widehat{u}_{jt}(x)$ is consistent and asymptotically efficient for $u_{jt}(x_t)$, also attaining its Cramer Rao lower bound.
- The same properties apply to the stationary model.

- In practice applications further restrict the parameter space.
- For example assume $\theta \equiv \left(\theta^{(1)}, \theta^{(2)}\right) \in \Theta$ is a closed convex subspace of Euclidean space, and:
  - $u_{jt}(x) \equiv u_j(x, \theta^{(1)})$
  - $f_{jt}(x|x_{nt}) \equiv f_{jt}(x|x_{nt}, \theta^{(2)})$
- We now define the model by (T, β, θ, g).
- Assume the DGP comes from (T, β, θ_0, g) where $\theta_0 \in \Theta^{(interior)}$.
- The ML estimator, denoted by θ_{ML}, maximizes:

$$\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J} d_{njt}\left\{\ln\left[p_{jt}(x_{nt}, \theta)\right] + \sum_{x=1}^{X} I\left\{x_{n,t+1} = x\right\}\ln\left[f_{jt}\left(x|x_{nt}, \theta^{(2)}\right)\right]\right\}$$

over $\theta \in \Theta$ where $p_t(x, \theta)$ are the CCPs for (T, β, θ, g).

- A common variation on the ML estimator is:
  1. estimate $f_{jt}(x|x_{nt}, \theta^{(2)})$ from the state transitions.
  2. obtain a limited information ML estimator $\theta^{(2)}_{LIML}$.
  3. estimate $\theta^{(1)}$ by searching over $p_t(x, \theta^{(1)}, \theta^{(2)}_{LIML})$.
- More precisely we define:

$$\theta^{(2)}_{LIML} \equiv \arg\max_{\theta_2} \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J} \sum_{x=1}^{X} I\left\{x_{n,t+1} = x\right\} d_{njt} \log\left[f_{jt}(x|x_{nt}, \theta^{(2)})\right]$$

$$\widehat{\theta}^{(1)}_{ML} \equiv \arg\max_{\theta_1} \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J} d_{njt} \left\{\log\left[p_{jt}(x_{nt}, \theta^{(1)}, \theta^{(2)}_{LIML})\right]\right\}$$

- Note that:
  - when $\theta^{(2)}_0$, that is $f_{jt}(x|x_{nt})$, is known, $\widehat{\theta}^{(1)}_{ML} = \theta^{(1)}_{ML}$;
  - otherwise $\widehat{\theta}^{(1)}_{ML}$ is less efficient but computationally simpler than $\theta^{(1)}_{ML}$;
  - nevertheless both estimators solve for the optimal rule many times.

Quasi Maximum Likelihood Estimation

The steps (Hotz and Miller, 1993)

- The essential difference between this estimator and ML is this estimator substitutes an estimator of the continuation value into the likelihood rather than computing it from the optimal policy function:
  1. Estimate the reduced form $\widehat{p}$ and $\widehat{f}$ (or $\theta^{(2)}_{LIML}$) as above;
  2. Apply the Representation Theorem to obtain expressions for $v_{jt}(x_t) - v_{kt}(x_t)$;
  3. Substitute the reduced form estimates into these differences to obtain $\widehat{v}_{jt}\left(x, \theta^{(1)}\right) - \widehat{v}_{kt}\left(x, \theta^{(1)}\right)$ for any given $\theta^{(1)}$;
  4. Replace $v_{jt}(x_t)$ with $\widehat{v}_{jt}\left(x, \theta^{(1)}\right)$ in the random utility model (RUM) to obtain an estimate $\widehat{p}_{jt}\left(x, \theta^{(1)}\right)$ for any given $\theta^{(1)}$;
  5. Maximize the quasi-likelihood with respect to $\theta^{(1)}$.
- In effect we estimate a static RUM where differences in current utilities $u_j\left(x, \theta^{(1)}\right) - u_k\left(x, \theta^{(1)}\right)$ are augmented by a *dynamic correction factor* estimated in the first stage off the reduced form.

- In the second step, appealing to the Representation theorem, and the slides above $\widehat{v}_{jt}\left(x,\theta^{(1)}\right)-\widehat{v}_{kt}\left(x,\theta^{(1)}\right)=$

$$u_j\left(x,\theta^{(1)}\right)-u_k\left(x,\theta^{(1)}\right)-\sum_{\tau=1}^{T-t}\sum_{x=1}^{X}\beta^{\tau}\widehat{\psi}_{1,t+\tau}(x)\left[\begin{array}{c}\widehat{\kappa}_{kt,\tau-1}(x|x_t)\\-\widehat{\kappa}_{jt,\tau-1}(x|x_t)\end{array}\right]$$

- In the last two steps we define:

$$\widehat{p}_{jt}\left(x,\theta^{(1)}\right)\equiv\int_{\epsilon_t}\prod_{k=1}^{J}I\left\{\begin{array}{l}\epsilon_{kt}-\epsilon_{jt}\\\leq\widehat{v}_{jt}\left(x,\theta^{(1)}\right)-\widehat{v}_{kt}\left(x,\theta^{(1)}\right)\end{array}\right\}dG_t\left(\epsilon_t\,|x_t\right)$$

and:

$$\theta_{QML}^{(1)}\equiv\arg\max_{\theta_1}\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J}d_{njt}\left\{\log\left[\widehat{p}_{jt}(x_{nt},\theta^{(1)},\theta_{LIML}^{(2)})\right]\right\}$$

Quasi Maximum Likelihood Estimation

Exploiting finite dependence in QML Estimation

- Supposing there is ρ-period dependence, we could form:

$$
\begin{aligned}
& \widetilde{v}_{jt}\left(x, \theta^{(1)}\right) - \widetilde{v}_{it}\left(x, \theta^{(1)}\right) \\
= \; & u_j\left(x, \theta^{(1)}\right) - u_i\left(x, \theta^{(1)}\right) \\
& + \sum_{\tau=1}^{\rho} \sum_{(k, x_{t+\tau})}^{(J,X)} \beta^{\tau} \left\{ \begin{array}{c} \left[u_i\left(x_{t+\tau}, \theta^{(1)}\right) + \widehat{\psi}_{k\tau}(x_{t+\tau}) \right] \times \\ \left[\begin{array}{l} \omega_{ikt\tau}\left(x_t, x_{t+\tau}\right) \widehat{\kappa}_{it,\tau-1}(x_{t+\tau}|x_t) \\ -\omega_{jkt\tau}\left(x_t, x_{t+\tau}\right) \widehat{\kappa}_{jt,\tau-1}(x_{t+\tau}|x_t) \end{array} \right] \end{array} \right\}
\end{aligned}
$$

- Then $\widetilde{p}_{jt}\left(x, \theta_1\right)$ is formed in an analogous manner to $\widehat{p}_{jt}\left(x, \theta_1\right)$
- The last maximization step is defined in a similar way.
- Note this estimator does not have exactly the same interpretation as $\theta_{QML}^{(1)}$, because the dynamic selection correction involves the current utility parameters.

Quasi Maximum Likelihood Estimation

Adjusting the asymptotic covariance for pre-estimation (Hotz and Miller, 1993)

- Form $P\left(\theta^{(1)}, p, f\right)$, a mapping from $\Theta^{(1)} \times P \times F$ to P with:

$$\kappa_{jt\tau}(x_{t+\tau+1}|x_t) \equiv \begin{cases} f_{jt}(x_{t+1}|x_t) & \tau = 0 \\ \sum_{x=1}^{X} f_{1,t+\tau}(x_{t+\tau+1}|x)\kappa_{jt,\tau-1}(x|x_t) & \tau = t+1, \ldots \end{cases}$$

$$\begin{aligned} v_{jt}\left(x, \theta^{(1)}\right) - v_{kt}\left(x, \theta^{(1)}\right) &= u_j\left(x, \theta^{(1)}\right) - u_k\left(x, \theta^{(1)}\right) \\ &\quad - \sum_{\tau=1}^{T-t} \sum_{x=1}^{X} \beta^{\tau} \psi_{1,t+\tau}(x) \left[\begin{array}{l} \kappa_{kt,\tau-1}(x|x_t) \\ -\kappa_{jt,\tau-1}(x|x_t) \end{array} \right] \end{aligned}$$

$$p_{jt}\left(x, \theta^{(1)}\right) \equiv \int_{\epsilon_t} \prod_{k=1}^{J} I\left\{ \begin{array}{l} \epsilon_{kt} - \epsilon_{jt} \\ \leq v_{jt}\left(x, \theta^{(1)}\right) - v_{kt}\left(x, \theta^{(1)}\right) \end{array} \right\} g_t\left(\epsilon_t \left| x_t \right.\right) d\epsilon_t$$

Quasi Maximum Likelihood Estimation

Adjusting the asymptotic covariance for pre-estimation

- Let:

$$\pi_{1n}\left(\theta^{(1)}, p, f\right) = W_N\left\{z_n \otimes \left[d_n - P\left(\theta^{(1)}, p, f\right)\right]\right\}$$

where W_N is a weighting matrix and z_n are instruments.

- Define the CCP estimator for $\theta^{(1)}_{CCP}$ by solving:

$$\sum_{n=1}^{N} \pi_{1n}\left(\theta^{(1)}, \widehat{p}, \widehat{f}\right) = 0$$

Large Sample or Asymptotic Properties

The asymptotic covariance matrix

- Write the cell estimators as the solution to:

$$0 = \sum_{n=1}^{N} \pi_{2n}\left(\widehat{p}, \widehat{f}\right) = \sum_{n=1}^{N} \begin{bmatrix} I_n^d \left(d_n - \widehat{p}\right) \\ I_n^f \left(f_n - \widehat{f}\right) \end{bmatrix}$$

  where:

  - I_n^d is a $(J-1)XT$ dimensional row vector indicator function matching the state variables of n to the relevant CCP component(s) in p;
  - I_n^f is a $(J-1)X^2T$ dimensional row vector indicator function matching state variables and decision(s) of n to f components;
  - f_n is the outcome from the n making a choice given her state variables.

- For $k \in \{1,2\}$ and $k' \in \{1,2\}$ define:

$$\Omega_{kk'} \equiv E\left[\pi_{kn}\pi'_{k'n}\right] \qquad \Gamma_{11} \equiv E\left[\frac{\partial \pi_{1n}}{\partial \theta^{(1)}}\right] \qquad \Gamma_{12} \equiv E\left[\frac{\partial \pi_{1n}}{\partial p}, \frac{\partial \pi_{1n}}{\partial f}\right]$$

- Then the asymptotic covariance matrix for $\theta_{CCP}^{(1)}$, denoted by Σ_1, is:

$$\Sigma_1 = \Gamma_{11}^{-1}\left[\Omega_{11} + \Gamma_{12}\left(\Omega_{22} - \Omega_{21} - \Omega_{12}\right)\Gamma'_{12}\right]\Gamma_{11}^{-1\prime}$$

Minimum Distance Estimators

Minimizing the difference between unrestricted and restricted current payoffs

- Another approach is to match up the parametrization of $u_{jt}(x_t)$, denoted by $u_{jt}(x_t, \theta^{(1)})$, to its representation as closely as possible:
  1. Form the vector function where $\Psi(p, f)$ by stacking:

$$\begin{aligned}\Psi_{jt}(x_t, p, f) \equiv\ & \psi_{1t}(x_t) - \psi_{jt}(x_t) \\ & + \sum_{\tau=1}^{T-t} \sum_{x=1}^{X} \beta^{\tau} \psi_{1,t+\tau}(x) \begin{bmatrix} \kappa_{kt,\tau-1}(x|x_t) \\ -\kappa_{jt,\tau-1}(x|x_t) \end{bmatrix}\end{aligned}$$

  2. Estimate the reduced form $\widehat{p}$ and $\widehat{f}$.
  3. Minimize the quadratic form to obtain:

$$\begin{aligned}\theta_{MD}^{(1)} &= \underset{\theta^{(1)} \in \Theta^{(1)}}{\arg\min} \left[u(x, \theta^{(1)}) - \Psi\left(\widehat{p}, \widehat{f}\right)\right]' \widetilde{W} \left[u(x, \theta^{(1)}) - \Psi\left(\widehat{p}, \widehat{f}\right)\right] \\ &= \underset{\theta^{(1)} \in \Theta^{(1)}}{\arg\min} \left[u(x, \theta^{(1)})' \widetilde{W} u(x, \theta^{(1)}) - 2\Psi\left(\widehat{p}, \widehat{f}\right)' \widetilde{W} u(x, \theta^{(1)})\right]\end{aligned}$$

  where $\widetilde{W}$, is a square $(J-1)\,TX$ weighting matrix.

Minimum Distance Estimators

Notes on minimizing the difference between unrestricted and restricted payoffs

- From the Representation theorem $u_{jt}(x_t, \theta_0^{(1)}) = \Psi_{jt}(x_t, p, f_0)$ if p are the CCPs for (T, β, θ_0, g).
- Furthermore $u_{jt}(x)$ is exactly identified from $\Psi_{jt}(x, p, f_0)$ without imposing any additional restrictions.
- Therefore parameterizing u with $\theta_0^{(1)}$ imposes overidentifying restrictions so $\theta_{MD}^{(1)}$ is consistent if the restrictions are true.
- Note $\theta_{MD}^{(1)}$ has a closed form if $u(x; \theta_0^{(1)})$ is linear in $\theta_0^{(1)}$.

Minimum Distance Estimators

A minimum distance estimator that exploits finite dependence (Altug and Miller, 1998)

- We can adapt this estimator by exploiting finite dependence.
- Suppose utility is stable over time with $u_{jt}\left(x,\theta^{(1)}\right)=u_j\left(x,\theta^{(1)}\right)$, and that $\omega_{ikt\tau}\left(x_t,x_{t+\tau}\right)$ achieves ρ dependence for all K choices at each x at t:
  1. Form the $K\left(\mathcal{T}-\rho\right)X$ utility vector $\Lambda\left(\theta^{(1)},p,f\right)$ from

$$\Lambda_{ijt}\left(x_t,\theta^{(1)},p,f\right)=$$

$$u_j\left(x_t,\theta^{(1)}\right)-u_i\left(x_t,\theta^{(1)}\right)-\psi_{jt}(x_t)+\psi_{it}(x_t)$$

$$\sum_{\tau=1}^{\rho}\sum_{(k,x_\tau)}^{(J,X)}\beta^{\tau}\left\{\begin{array}{c}\left[u_k\left(x_{t+\tau},\theta^{(1)}\right)+\psi_{k,t+\tau}(x_{t+\tau})\right]\times\\ \left[\begin{array}{c}\omega_{ikt\tau}\left(x_t,x_{t+\tau}\right)\kappa_{it,\tau-1}(x_{t+\tau}|x_t)\\ -\omega_{jkt\tau}\left(x_t,x_{t+\tau}\right)\kappa_{jt,\tau-1}(x_{t+\tau}|x_t)\end{array}\right]\end{array}\right\}$$

  2. Given $K\left(\mathcal{T}-\rho\right)X$ weighting matrix $\overline{W}$, minimize:

$$\Lambda\left(\theta^{(1)},\widehat{f},\widehat{p}\right)'\overline{W}\Lambda\left(\theta^{(1)},\widehat{f},\widehat{p}\right)$$

Simulated Moments Estimators

- We could form a Methods of Simulated Moments (MSM) estimator from:
  1. Simulate a lifetime path from x_{nt_n} onwards for each j, using $\widehat{f}$ and $\widehat{p}$.
  2. Obtain estimates of $\widehat{E}\left[\epsilon_{jt} \middle| d_{jt}^{o} = 1, x_t\right]$.
  3. Stitch together a simulated lifetime utility outcome from the j^{th} choice at t_n onwards for n, denoted $\widehat{v}_{nj} \equiv \widehat{v}_{jt_n}\left(x_{nt_n}; \theta^{(1)}, \widehat{f}, \widehat{p}\right)$.
  4. Form the $J-1$ dimensional vector $h_n\left(x_{nt_n}; \theta^{(1)}, \widehat{f}, \widehat{p}\right)$ from:

$$
\begin{aligned}
h_{nj}\left(x_{nt_n}; \theta^{(1)}, \widehat{f}, \widehat{p}\right) \equiv\ & \widehat{v}_{jt_n}\left(x_{nt_n}, \theta^{(1)}, \widehat{f}, \widehat{p}\right) - \widehat{v}_{Jt_n}\left(x_{nt_n}, \theta^{(1)}, \widehat{f}, \widehat{p}\right) \\
& + \widehat{\psi}_{jt}(x_{nt_n}) - \widehat{\psi}_{Jt}(x_{nt_n})
\end{aligned}
$$

  5. Given a weighting matrix W_S and an instrument vector z_n minimize:

$$
N^{-1}\left[\sum_{n=1}^{N} z_n h_n\left(x_{nt_n}; \theta^{(1)}, \widehat{f}, \widehat{p}\right)\right]' W_S \left[\sum_{n=1}^{N} z_n h_n\left(x_{nt_n}; \theta^{(1)}, \widehat{f}, \widehat{p}\right)\right]
$$

- In the first step, given the state simulate a choice using $\widehat{p}$, and simulate the next state using $\widehat{f}$. In this way generate $\widehat{x}_{ns}$ and $\widehat{d}_{ns} \equiv \left(\widehat{d}_{n1s}, \ldots, \widehat{d}_{nJs}\right)$ for all $s \in \{t_n + 1, \ldots, T\}$.
- Generating this path does not exploit knowledge of G, only the CCPs.
- In the second step $\widehat{E}\left[\epsilon_{jt} \middle| d^o_{jt} = 1, x_t\right] \equiv$

$$p_{jt}^{-1}(x_t) \int_{\epsilon_t} \prod_{k=1}^{J} I\left\{\widehat{\psi}_{jt}(x_t) - \widehat{\psi}_{kt}(x_t) \leq \epsilon_{jt} - \epsilon_{kt}\right\} \epsilon_{jt} g(\epsilon_t)\, d\epsilon_t$$

- In Step 4 $\widehat{v}_{jt}\left(x_{nt_n}, \theta^{(1)}, \widehat{f}, \widehat{p}\right)$ is stitched together as:

$$u_{jt}(x_{nt_n}, \theta^{(1)}) + \sum_{s=t+1}^{T} \sum_{j=1}^{J} \beta^{t-1} 1\left\{\widehat{d}_{njs} = 1\right\} \left\{ \begin{array}{l} u_{js}(\widehat{x}_{ns}, \theta^{(1)}) \\ + \widehat{E}\left[\epsilon_{js} \middle| \widehat{x}_{ns}, \widehat{d}_{njs} = 1\right] \end{array} \right\}$$

- The solution has a closed form if $u_{jt}(x, \theta^{(1)})$ is linear in $\theta^{(1)}$.

Simulated Moments Estimators

Another MSM estimator

- Indeed ϵ_t could be simulated as well:
  1. Draw a realization $\widehat{\epsilon}$ from $G(\epsilon)$ for each $s \in \{t_n, \ldots, T\}$ and n.
  2. Set:
  $$\widehat{d}_{njs} = \prod_{k=1}^{J} I\left\{\widehat{\psi}_{js}(\widehat{x}_{ns}) - \widehat{\psi}_{ks}(\widehat{x}_{ns}) \leq \widehat{\epsilon}_{njs} - \widehat{\epsilon}_{nks}\right\}$$
  and stitch together:
  $$u_{jt}(x_{nt_n}, \theta^{(1)}) + \sum_{s=t+1}^{T}\sum_{j=1}^{J}\beta^{t-1} 1\left\{\widehat{d}_{njs} = 1\right\}\left\{u_{js}(\widehat{x}_{ns}, \theta^{(1)}) + \widehat{\epsilon}_{js}\right\}$$
  3. Minimize an analogous quadratic form to obtain $\theta^{(1)}$.
- Bajari, Benkard and Levin (2007) estimate an approximate reduced form of the policy function without exploiting the CCPs (pages 1341-1342, 2007), but acknowledge: "Our method requires that one be able to consistently estimate each firm's policy function, so this may limit our ability to estimate certain models (page 1345, 2007)."

Large Sample or Asymptotic Properties

Adjusting the asymptotic covariance for simulation as well

- Simulation adds an additional, independent source of variation to the sample moments and hence the estimated asymptotic standard errors.
- Following the definition given in Lecture 7 suppose $\widehat{\theta}^{(1)}$ minimizes:

$$N^{-1}\left[\sum_{n=1}^{N} z_n h_n\left(x_{nt_n},\theta^{(1)},\widehat{f},\widehat{p}\right)\right]' W_S\left[\sum_{n=1}^{N} z_n h_n\left(x_{nt_n},\theta^{(1)},\widehat{f},\widehat{p}\right)\right]$$

- Then the additional component to the covariance matrix for $\widehat{\theta}^{(1)}$ is:

$$\Sigma_1^S \equiv S^{-1}\left(\Upsilon' W_S \Upsilon\right)^{-1}\Upsilon' W_S E\left[z_n h_n h_n' z_n'\right] W_S \Upsilon\left(\Upsilon' W_S \Upsilon\right)^{-1}$$

where S is the number of simulations (per observation):

$$\Upsilon = E\left[\frac{z_n \partial h_n\left(x_{nt_n},\theta_0^{(1)}, f_0, p_0\right)}{\partial\theta^{(1)}}\right]$$

- Note that $\Sigma_1^S \rightarrow 0$ as $S \rightarrow \infty$.

Summary

Efficiency versus computational ease

- There is a trade off between efficiency and computational ease:
  1. ML estimator;
  2. 2 step asymptotically efficient CCP estimators based on a Newton step;
  3. QML estimators (still nonlinear);
  4. MD estimators (closed form if utility is linear in the parameters and minimal use of functional form of G);
  5. Simulation CCP estimators (similar to MD estimators)
- Within each of these categories finite dependence can be exploited if that property holds.