Econ 673: Microeconometrics

Chapter 7: Estimation with Simulation

Simulation Estimation

- Maximum Simulated Likelihood is convenient for many limited dependent variable models
- In this chapter, we look at alternative estimation procedures based on simulations
  - Maximum Simulated Likelihood (MSL)
  - Method of Simulated Moments (MSM) – analogue to Method of Moments (MOM) estimation
  - Method of Simulated Scores (MSS)
- We will look at the
  - Statistical properties and
  - Trade-offs

  in using each method

Sources

- *Train, K., (2003), *Discrete Choice Methods with Simulation*, Cambridge, MA: Cambridge University Press, Ch. 10.
- Stern, S., (1997) “Simulation Based Estimation,” *Journal of Economic Literature*, 35: 2006-2039.

Definition: ML

- Given the log-likelihood function

$$LL(\theta) = \sum_{i=1}^{N} \ln\left[P_i(\theta)\right]$$

the maximum likelihood estimator of θ is given by

$$\theta_{ML} = \arg\max_{\theta} LL(\theta) = \arg\max_{\theta} \sum_{i=1}^{N} \ln\left[P_i(\theta)\right]$$

- Assuming that the log-likelihood function is smooth, the ML estimator is equivalently defined implicitly by

$$\sum_{i=1}^{N} g_i(\theta_{ML}) = 0$$

where

$$g_i(\theta) \equiv \frac{\partial \ln P_i(\theta)}{\partial \theta}$$

← *score* of observation *i*

Definition: MSL

- Let

$$SLL(\theta) = \sum_{i=1}^{N} \ln\left[\breve{P}_i(\theta)\right]$$

denote the simulated log likelihood function, where $\breve{P}_i(\theta)$ is a simulated estimate of $P_i(\theta)$, then

$$\theta_{SML} = \arg\max_{\theta} SLL(\theta) = \arg\max_{\theta} \sum_{i=1}^{N} \ln\left[\breve{P}_i(\theta)\right]$$

or equivalently is defined implicitly by

$$\sum_{i=1}^{N} \breve{g}_i(\theta_{ML}) = 0$$

where

$$\breve{g}_i(\theta) \equiv \frac{\partial \ln \breve{P}_i(\theta)}{\partial \theta}$$

Properties of MSL

- The key issue with MSL is that it is biased, i.e., while

$$E\left[\breve{P}_i(\theta)\right] = P_i(\theta)$$

$$E\left[\ln \breve{P}_i(\theta)\right] \neq \ln P_i(\theta)$$

though the bias diminishes as the number of draws used in simulation (R) increases.

- Properties of MSL
  - If R is fixed, MSL is inconsistent
  - If R rises with N, then MSL is consistent
  - If R rises faster than sqrt(N), then MSL is asymptotically equivalent to ML

Definition: MOM

- Methods of moments (MOM) estimation is driven by the notion that the residuals should be uncorrelated with those factors exogenous to the model.
- For discrete choice problems, this corresponds to

$$\sum_{i=1}^{N}\sum_{j=1}^{J}\left[y_{ij}-P_{ij}\left(\theta\right)\right]\tilde{x}_{ij}=0$$

where $\tilde{x}_{ij}$ denotes our exogenous factors (or *instruments*)

- The MOM estimator is implicitly defined by

$$\sum_{i=1}^{N}\sum_{j=1}^{J}\left[y_{ij}-P_{ij}\left(\theta_{MOM}\right)\right]\tilde{x}_{ij}=0$$

MOM in a Linear Regression Model

- In a linear regression model

$$y_i = x_i'\beta + \varepsilon_i$$

The MOM estimator be solves

$$\sum_{i=1}^{N}\left[y_i - x_i'\beta_{MOM}\right]\tilde{x}_i = 0$$

$$\Rightarrow \sum_{i=1}^{N}\tilde{x}_i y_i = \sum_{i=1}^{N}\tilde{x}_i x_i'\beta_{MOM}$$

$$\Rightarrow \beta_{MOM} = \left[\sum_{i=1}^{N}\tilde{x}_i x_i'\right]^{-1}\sum_{i=1}^{N}\tilde{x}_i y_i = \beta_{IV}$$

Instrumental variables estimator

MOM in a Linear Regression Model (cont'd)

- If the explanatory variables are exogenous, then the ideal instruments are the explanatory variables; i.e.,

$$\tilde{x}_i^* = x_i$$

and

$$\beta_{MOM}^* = \left[\sum_{i=1}^{N} x_i x_i' \right]^{-1} \sum_{i=1}^{N} x_i y_i = \beta_{OLS}$$

MOM in a Discrete Choice Model

- The MOM estimator is implicitly defined by

$$\sum_{i=1}^{N} \sum_{j=1}^{J} \left[y_{ij} - P_{ij} \left(\theta_{MOM} \right) \right] \tilde{x}_{ij} = 0$$

where the $\tilde{x}_{ij}$ are exogenous instruments, independent of the residuals is the population.

- MOM consistent as long as instruments independent of residuals
- Unlike linear regression case, we cannot solve explicitly for θ_{MOM}

MOM and ML

- ML is a special case of MOM using scores as instruments, i.e.,

$$\tilde{x}_{ij} = \frac{\partial \ln P_{ij}(\theta)}{\partial \theta}$$

since

$$\begin{aligned}
0 &= \sum_{i=1}^{N}\sum_{j=1}^{J}\left[y_{ij} - P_{ij}(\theta_{MOM})\right]\tilde{x}_{ij} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij}\frac{\partial \ln P_{ij}(\theta_{MOM})}{\partial \theta_{MOM}} - \sum_{i=1}^{N}\sum_{j=1}^{J} P_{ij}(\theta_{MOM})\frac{\partial \ln P_{ij}(\theta_{MOM})}{\partial \theta_{MOM}} \\
&= \sum_{i=1}^{N} g_i(\theta_{MOM}) - \sum_{i=1}^{N}\sum_{j=1}^{J}\frac{\partial P_{ij}(\theta_{MOM})}{\partial \theta_{MOM}} \\
&= \sum_{i=1}^{N} g_i(\theta_{MOM}) \qquad \Rightarrow \text{scores are ideal instruments}
\end{aligned}$$

Definition: MSM

- The Method of Simulated Moments is defined by replacing the true choice probabilities with their simulated counterpart; i.e.,
- The MSM estimator is implicitly defined by

$$\sum_{i=1}^{N}\sum_{j=1}^{J}\left[y_{ij} - \breve{P}_{ij}(\theta_{MOM})\right]\tilde{x}_{ij} = 0$$

- MSM consistent as long as instruments independent of simulated residuals

Properties of MSM

- A key feature of the MSM estimator is that the choice probabilities enter linearly, avoiding the bias of MSL.
- Tradeoff is that ideal instruments are a function of $\ln P_{ij}$
- Properties of MSM
  - Even with *R* is fixed, MSM is consistent
  - Inefficiency results from the use of less than ideal instruments

Definition: Methods of Scores (MS) and Simulated Scores (MSS)

- Methods of Scores (MS) estimator is implicitly defined as

$$\sum_{i=1}^{N} g_i\left(\theta_{MS}\right) = 0$$

so that

$$\theta_{MS} = \theta_{ML}$$

- Methods of Simulated Scores (MSS) estimator is implicitly defined as

$$\sum_{i=1}^{N} \hat{g}_i\left(\theta_{MSS}\right) = 0$$

where $\hat{g}_i\left(\theta\right)$ is a simulator of the score function $g_i\left(\theta\right)$

Properties of MSS

- MSS is equivalent to MSL if

$$\hat{g}_i(\theta) = \breve{g}_i(\theta) = \frac{\partial \ln \breve{P}(\theta)}{\partial \theta}$$

- However, given an unbiased score simulator, MSS is consistent for a fixed *R* and efficient as long as *R* rises at any rate with *N*

- The trouble is finding an unbiased score simulator

An Unbiased Score Simulator

- One possible score simulator is based on decomposition:

$$g_i(\theta) = \frac{\partial \ln P_{ij}(\theta)}{\partial \theta} = \left(\frac{1}{P_{ij}(\theta)}\right)\frac{\partial P_{ij}(\theta)}{\partial \theta}$$

$\frac{\partial \breve{P}_{ij}(\theta)}{\partial \theta}$ is unbiased

$\frac{1}{P_{ij}(\theta)} =$ Expected number of draws until *j* is selected

readily simulated, but not smooth

Deriving Properties of Estimators

- Both non-simulated estimators (ML=MS and MOM) take the general form

$$0 = g\left(\theta_k\right) = \tfrac{1}{N}\sum_{i=1}^{N} g_i\left(\theta_k\right); k = ML, MOM$$

For ML: $$g_i\left(\theta_{ML}\right) = \frac{\partial \ln P_i\left(\theta_{ML}\right)}{\partial \theta_{ML}}$$

For MOM: $$g_i\left(\theta_{MOM}\right) = \sum_{j=1}^{J}\left[y_{ij} - P_{ij}\left(\theta_{MOM}\right)\right]\tilde{x}_{ij}$$

For both we assume that: $E\left[g_i\left(\theta^*\right)\right] = 0$

Central Limit Theorem

- Standard central limit theorem results give us that for a random variable t_i such that:

$$E\left(t_i\right) = \mu$$

$$Var\left(t_i\right) = \sigma^2$$

$$t \equiv \frac{1}{N}\sum_{i=1}^{N} t_i$$

then

$$\sqrt{N}\left(t - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right)$$

so that

$$t \overset{a}{\sim} N\left(\mu, \frac{\sigma^2}{N}\right)$$

Central Limit Theorem (cont'd)

- We can apply this result to obtain to the statistic

$$g\left(\theta^*\right) = \frac{1}{N}\sum_{i=1}^{N} g_i\left(\theta^*\right)$$

so that

$$\sqrt{N}\left[g\left(\theta^*\right) - 0\right] \xrightarrow{d} N\left(0, V\right)$$

where

$$V = E\left[g_i\left(\theta^*\right) g_i\left(\theta^*\right)'\right]$$

$$= \mathbf{V} \equiv E\left(-\frac{\partial^2 \ln P_i\left(\theta^*\right)}{\partial\theta\partial\theta'}\right) \text{ for ML}$$

with

$$g\left(\theta^*\right) \overset{a}{\sim} N\left(0, \frac{V}{N}\right)$$

for large enough N

Central Limit Theorem (cont'd)

- We can us these results to derive distributions for alternative estimators:

$$0 = g\left(\hat{\theta}\right) \approx g\left(\theta^*\right) + h\left(\theta^*\right)\left[\hat{\theta} - \theta^*\right]$$

$$\hat{\theta} - \theta^* \approx -h^{-1}\left(\theta^*\right) g\left(\theta^*\right)$$

where

$$h\left(\theta^*\right) \equiv \frac{\partial g\left(\theta^*\right)}{\partial\theta} \xrightarrow{N\to\infty} H = E\left(\frac{\partial g_i\left(\theta^*\right)}{\partial\theta}\right)$$

Central Limit Theorem (cont'd)

This implies that

$$\sqrt{N}\left(\hat{\theta}-\theta^{*}\right)\approx-\sqrt{N}h^{-1}\left(\theta^{*}\right)g\left(\theta^{*}\right)$$

$$\xrightarrow{d} N\left(0,H^{-1}VH^{-1}\right)$$

So that

$$\hat{\theta}\overset{a}{\sim}N\left(\theta^{*},\tfrac{1}{N}H^{-1}VH^{-1}\right)$$

For ML, H=-**V**

$$\hat{\theta}_{ML}\overset{a}{\sim}N\left(\theta^{*},\tfrac{1}{N}\mathbf{V}^{-1}\right)$$

Simulation Estimator Properties

- The properties of our simulation based estimators stem from the decomposition

$$\begin{aligned}\breve{g}\left(\theta^{*}\right)&=g\left(\theta^{*}\right)+\left[E_{R}\breve{g}\left(\theta^{*}\right)-g\left(\theta^{*}\right)\right]+\left[\breve{g}\left(\theta^{*}\right)-E_{R}\breve{g}\left(\theta^{*}\right)\right]\\&=A+B+C\end{aligned}$$

- A denotes the unsimulated mean
- B denotes the simulation bias
- C denotes the simulation noise

Distributional Results

$$\sqrt{N}\left(\hat{\theta}-\theta^*\right) \approx -\sqrt{N}h^{-1}\left(\theta^*\right)\breve{g}\left(\theta^*\right)$$

$$= -h^{-1}\left(\theta^*\right)\left[\sqrt{N}A+\sqrt{N}B+\sqrt{N}C\right]$$

where

$$\sqrt{N}A \xrightarrow{d} N\left(0,V\right)$$

$$\sqrt{N}B = \sqrt{N}\left(\frac{V_B}{R}\right)$$

$$\sqrt{N}C \xrightarrow{d} N\left(0,V_C\left(R\right)\right)$$

$$V_C\left(R\right) = E\left[V_{nC}\left(R\right)\right] \qquad V_{nC}\left(R\right) = Var\left(\breve{g}_n\left(\theta^*\right) - E\left[\breve{g}_n\left(\theta^*\right)\right]\right)$$

$$\frac{dV_{nC}\left(R\right)}{dR} < 0$$

MSM

$$\sqrt{N}\left(\hat{\theta}_{MSM}-\theta^*\right) \approx -h^{-1}\left(\theta^*\right)\left[\sqrt{N}A+\sqrt{N}C\right]$$

$$\xrightarrow{d} N\left(0, H^{-1}\left[V+V_C\left(R\right)\right]H^{-1}\right)$$

$$\hat{\theta}_{MSM} \overset{a}{\sim} N\left(\theta^*, \tfrac{1}{N}H^{-1}\left[V+V_C\left(R\right)\right]H^{-1}\right)$$

If *R* increases with *N*, then

$$\hat{\theta}_{MSM} \overset{a}{\sim} N\left(\theta^*, \tfrac{1}{N}H^{-1}VH^{-1}\right)$$

but inefficiency results due to the use of less than ideal weights

MSS

With an unbiased score simulator

$$\sqrt{N}\left(\hat{\theta}_{MSM}-\theta^*\right)\approx -h^{-1}\left(\theta^*\right)\left[\sqrt{N}A+\sqrt{N}C\right]$$

$$\xrightarrow{d} N\left(0, H^{-1}\left[\mathbf{V}+V_C\left(R\right)\right]H^{-1}\right)$$

$$\hat{\theta}_{MSM}\overset{a}{\sim}N\left(\theta^*,\tfrac{1}{N}H^{-1}\left[\mathbf{V}+V_C\left(R\right)\right]H^{-1}\right)$$

If *R* increases with *N*, then

$$\hat{\theta}_{MSM}\overset{a}{\sim}N\left(\theta^*,\tfrac{1}{N}\mathbf{V}^{-1}\right)$$

but inefficiency results due to the use of less than ideal weights