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


Definition: ML

- Given the log-likelihood function

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

the maximum likelihood estimator of \( \theta \) 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

\[
g_i(\theta) = \frac{\partial \ln P_i(\theta)}{\partial \theta} \quad \leftarrow \text{score of observation } i
\]
Definition: MSL

• Let

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

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

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

or equivalently is defined implicitly by

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

where

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

Properties of MSL

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

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

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

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

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

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

• The MOM estimator is implicitly defined by

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

MOM in a Linear Regression Model

• In a linear regression model

\[
y_i = x_i' \beta + \epsilon_i
\]

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

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

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 \( \theta_{MOM} \)
MOM and ML

- ML is a special case of MOM using scores as instruments, i.e.,
\[ \bar{x}_j = \frac{\partial \ln P_{y_j}(\theta)}{\partial \theta} \]
since
\[ 0 = \sum_{i=1}^{N} \sum_{j=1}^{J} \left[ y_{ij} - P_{y_j}(\theta_{\text{MOM}}) \right] \bar{x}_j \]
\[ = \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ij} \frac{\partial \ln P_{y_j}(\theta_{\text{MOM}})}{\partial \theta_{\text{MOM}}} - \sum_{i=1}^{N} \sum_{j=1}^{J} P_{y_j}(\theta_{\text{MOM}}) \frac{\partial \ln P_{y_j}(\theta_{\text{MOM}})}{\partial \theta_{\text{MOM}}} \]
\[ = \sum_{i=1}^{N} g_i(\theta_{\text{MOM}}) - \sum_{i=1}^{N} \sum_{j=1}^{J} \frac{\partial P_{y_j}(\theta_{\text{MOM}})}{\partial \theta_{\text{MOM}}} \]
\[ = \sum_{i=1}^{N} g_i(\theta_{\text{MOM}}) \Rightarrow \text{scores are ideal instruments} \]

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
\[ \sum_{i=1}^{N} \sum_{j=1}^{J} \left[ y_{ij} - \tilde{P}_{y_j}(\theta_{\text{MOM}}) \right] \bar{x}_j = 0 \]
- 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
  \[
  \sum_{i=1}^{N} g_i(\theta_{MS}) = 0
  \]
  so that
  \[
  \theta_{MS} = \theta_{ML}
  \]
- Methods of Simulated Scores (MSS) estimator is implicitly defined as
  \[
  \sum_{i=1}^{N} \tilde{g}_i(\theta_{MSS}) = 0
  \]
  where $\tilde{g}_i(\theta)$ is a simulator of the score function $g_i(\theta)$
Properties of MSS

• MSS is equivalent to MSL if

\[ \tilde{g}_i(\theta) = \tilde{g}_i(\theta) = \frac{\partial \ln P(\theta)}{\partial \theta} \]

• 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:

\[ g_i(\theta) = \frac{\partial \ln P_i(\theta)}{\partial \theta} = \left( \frac{1}{P_y(\theta)} \right) \frac{\partial P_i(\theta)}{\partial \theta} \]

\[ \frac{\partial P_y(\theta)}{\partial \theta} \] is unbiased

\[ \frac{1}{P_y(\theta)} = \text{Expected number of draws until } j \text{ is selected} \]

readily simulated, but not smooth
Deriving Properties of Estimators

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

\[ 0 = g(\theta_k) = \frac{1}{N} \sum_{i=1}^{N} g_i(\theta_k); k = ML, MOM \]

For ML:
\[ g_i(\theta_{ML}) = \frac{\partial \ln P_i(\theta_{ML})}{\partial \theta_{ML}} \]

For MOM:
\[ g_i(\theta_{MOM}) = \sum_{j=1}^{J} [y_{ij} - P_g(\theta_{MOM})] \bar{x}_{ij} \]

For both we assume that: \( E[g_i(\theta')] = 0 \)

Central Limit Theorem

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

\[ E(t_i) = \mu \]
\[ Var(t_i) = \sigma^2 \]
\[ t = \frac{1}{N} \sum_{i=1}^{N} t_i \]

then
\[ \sqrt{N}(t - \mu) \xrightarrow{d} N(0, \sigma^2) \]

so that
\[ t \sim N\left( \mu, \frac{\sigma^2}{N} \right) \]
Central Limit Theorem (cont’d)

- We can apply this result to obtain to the statistic
  \[ g(\theta^*) = \frac{1}{N} \sum_{i=1}^{N} g_i(\theta^*) \]
  so that
  \[ \sqrt{N} \left[ g(\theta^*) - 0 \right] \xrightarrow{d} N(0, V) \]
  where
  \[ V = E \left[ g_i(\theta^*) g_i(\theta^*)' \right] \]
  with
  \[ g(\theta^*) \sim N \left( 0, \frac{V}{N} \right) \]
  for large enough \( N \)

Central Limit Theorem (cont’d)

- We can use these results to derive distributions for alternative estimators:
  \[ 0 = g(\hat{\theta}) \approx g(\theta^*) + h(\theta^*) \left[ \hat{\theta} - \theta^* \right] \]
  \[ \hat{\theta} - \theta^* \approx -h^{-1}(\theta^*) g(\theta^*) \]
  where
  \[ h(\theta^*) = \frac{\partial g(\theta^*)}{\partial \theta} \xrightarrow{N \to \infty} H = E \left( \frac{\partial g_i(\theta^*)}{\partial \theta} \right) \]
Central Limit Theorem (cont’d)

This implies that

\[ \sqrt{N}(\hat{\theta} - \theta^*) \approx -\sqrt{N}h^{-1}(\theta^*)g(\theta^*) \xrightarrow{d} N\left(0, H^{-1}VH^{-1}\right) \]

So that

\[ \hat{\theta} \sim N\left(\theta^*, \frac{1}{N}H^{-1}VH^{-1}\right) \]

For ML, \( H=-V \)

\[ \hat{\theta}_{\text{ML}} \sim N\left(\theta^*, \frac{1}{N}V^{-1}\right) \]

Simulation Estimator Properties

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

\[ \tilde{g}(\theta^*) = g(\theta^*) + \left[ E \tilde{g}(\theta^*) - g(\theta^*) \right] + \left[ \tilde{g}(\theta^*) - E \tilde{g}(\theta^*) \right] = A + B + C \]

- \( A \) denotes the unsimulated mean
- \( B \) denotes the simulation bias
- \( C \) denotes the simulation noise
Distributional Results

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

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

where

\[ \sqrt{N} A \xrightarrow{d} N(0, V) \]

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

\[ \sqrt{N} C \xrightarrow{d} N(0, V_C(R)) \]

\[ V_C(R) = E[V_{nc}(R)] \quad V_{nc}(R) = \text{Var}\left( \tilde{g}_n(\theta^*) - E\left[ \tilde{g}_n(\theta^*) \right] \right) \]

\[ \frac{dV_{nc}(R)}{dR} < 0 \]

MSM

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

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

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

If \( R \) increases with \( N \), then

\[ \hat{\theta}_{\text{MSM}} \overset{a}{\sim} N \left( \theta^*, \frac{1}{N} H^{-1} VH^{-1} \right) \]

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

With an unbiased score simulator

\[ \sqrt{N} (\hat{\theta}_{MSS} - \theta^*) \approx -h^{-1}(\theta^*) \left[ \sqrt{NA} + \sqrt{NC} \right] \]

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

\[ \hat{\theta}_{MSS} \overset{a}{\sim} N \left( \theta^*, \frac{1}{N} V^{-1} \right) \]

If \( R \) increases with \( N \), then

\[ \hat{\theta}_{MSS} \overset{a}{\sim} N \left( \theta^*, \frac{1}{N} V^{-1} \right) \]

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