# Structural Metrics Reading Group
# Method of Simulated Moments

Olivia Bordeu and Lillian Rusk

April, 2020

# Recap of MSL

- Last week Francesco introduced us maximum simulated liklihood (MSL)
- MSL is MLE except that simulated probabilities replace true probabilities
- Recall,

$$SL(\theta|x) = \check{f}(x|\theta) = \prod_{i=1}^{N} \check{P}_i\left(x_i;\theta\right)$$

$$SLL(\theta|x) = \log \check{f}(x|\theta) = \sum_{i=1}^{N} \log \check{P}_i\left(x_i;\theta\right)$$

$$\widehat{\theta}_{MSL} \in \arg\max_{\widetilde{\theta}} SL(\widetilde{\theta}|x) \Longleftrightarrow \widehat{\theta}_{MSL} \in \arg\max_{\widetilde{\theta}} \mathrm{SLL}(\widetilde{\theta}|x)$$

where $\check{P}_i$ is a simulated approximation to $P_i$

# An undesirable property

- Suppose $\check{P}_i$ is an unbiased simulator of $P_i(\theta)$
- Since the log operator is non-linear, $log\check{P}_i$ is not unbiased for $logP_i(\theta)$
- Bias in $log\check{P}_i \implies$ bias in $\theta_{MSL}$
  - Bias diminishes with simulation draws (R), but for a fixed R, the estimator is biased
  - For fixed R, $\theta_{MSL}$ is inconsistent
- Is there an alternative simulated estimator that is consistent for fixed R?
  - Drum roll...

# Method of simulated moments (MSM)

- MSM is GMM but it replaces the model moments with simulated moments
- MSM is to GMM what MSL was to MLE
- MSM can be used to estimate likelihoods, and it will attain the goal of being consistent for fixed R

# Let's first refresh our memory of GMM

- GMM estimates parameters by minimizing distance btw model and data moments
- $\mu(x|\theta)$ Model moments under parameter $\theta$
- $\mu(x)$ Same moment calculated from data x

$$\hat{\theta}_{GMM} := \arg\min_{\theta} ||\mu(x|\theta) - \mu(x)||$$

- Estimator depends on choice of distance measure

# Motivation for method of simulated moments (MSM)

- Roadblocks to GMM
  - Calculating the moments is too cumbersome
    - Ex. Multiple integrals over nonlinear functions (McFadden (1989))
  - Model includes a latent variable
    - Laroque and Salanié (1993)
  - No analytic representation exists
- MSM to the rescue!

# MSM overview

1. Generate simulated datasets
   - For some choice of $\hat{\theta}$, simulate the model data R times $\tilde{x} = \{\tilde{x}_1, ..., \tilde{x}_r, ..., \tilde{x}_R\}$
   - One simulation is one dataset with as many obs as the observed sample
2. Calculate associated moments for the simulated data
$$\mu^R(\tilde{x}|\hat{\theta}) = \frac{1}{R}\sum_r m(\tilde{x}_r|\hat{\theta})$$
3. Compute moments for observed data $\mu(x)$
4. Calculate the distance between simulated and observed moments
$$||\mu^R(\tilde{x}|\hat{\theta}) - \mu(x)||$$
5. Search over $\hat{\theta}$ to minimize the distance calculcated in step 4
$$\hat{\theta}_{MSM} := \arg\min_{\theta} ||\mu^R(\tilde{x}|\theta) - \mu(x)||$$

# Practioner's notes

- **Random draws**: When simulating, you want random draws to be held constant so the only thing changing in the minimization problem is the value of the vector of parameters
- **Normalizing moments:** Most common distance measure is $L^2$ norm. To avoid unintended moment weighting due to units, convert differences in moments to percent deviations
  - $e(\tilde{x}, x|\theta) := \frac{\hat{m}(\tilde{x}|\theta) - m(x)}{m(x)}$ referred to as the "moment error function"
  - Then,

$$\hat{\theta}_{MSM} := \arg\min_{\theta} e(\tilde{x}, x|\theta)' W e(\tilde{x}, x|\theta)$$

# Comparison to MSL

- Consistency: MSM wins
  - $\check{P}_n(\theta)$ enters linearly so if $\check{P}_n(\theta)$ is unbiased $\Longrightarrow$ MSM is unbiased
  - Since there is no simulation bias, MSM is consistent even when R is fixed
- Efficiency: MSL wins
  - MSM is less efficient than MSL
    - Recall that GMM is less efficient unless ideal instruments (scores) are used
    - Scores are a function of true $lnP_n(\theta)$
- In summary, both approaches can be justified (See Adda and Cooper (2003))

# Limitations to MSM

- Researcher must make following choices:
  - Which moments to use
  - Number of simulations
  - Weighting matrix, W,
  - Optimization algorithm
- It is not well-understood how these choices affect performance of MSM (See Eisenhauer, Heckman, and Mosso (2015)

# A Simple Example: Mixed Logit (again!)

Let us first consider the simple logit case where $i \in \{1, .., N\}$ people choose from $j \in \{1, ..., J\}$ alternatives:

$$V_{ij} = X_{ij}'\beta + \epsilon_{ij} \quad \text{with } \epsilon_{ij} \sim T1EV, \quad \text{where } P_j(X_{ij}|\beta) = \frac{e^{X_{ij}'\beta}}{\sum_k e^{X_{ik}'\beta}}$$

We could estimate $\theta = \beta$ using GMM, where we use the implied choice probabilities as moments:

$$\hat{\theta}_{GMM} := \arg\min_{\theta} \Big[\sum_i^N \sum_j^J \Big(D_{ij} - P(X_{ij}|\theta)\Big) z_{ij}\Big]' W \Big[\sum_i^N \sum_j^J \Big(D_{ij} - P(X_{ij}|\theta)\Big) z_{ij}\Big]$$

GMM encompasses the Maximum Likelihood Estimator, where we can use the FOC of the log-likelihood as the moment:

$$\hat{\theta}_{ML} \in \arg\max_{\theta} \underbrace{\sum_{i=1}^N \sum_{j=0}^J D_{ij} \log P_j(X_{ij}|\theta)}_{=LL(\theta|x)} \iff \hat{\theta}_{GMM} := \arg\min_{\theta} || \frac{\partial LL(\theta|x)}{\partial \theta} ||$$

# A Simple Example: Mixed Logit (again!)

However, with random coefficients, the choice probabilities might not have a closed form solution:

$$V_{ij} = X'_{ij}\beta_i + \epsilon_{ij} \quad \text{with } \epsilon_{ij} \sim T1EV \quad \text{and } \beta_i \sim f(\beta;\theta)$$

Then, we have that:

$$P(D_{ij} = 1 | X_{ij};\theta) = \int \frac{e^{X'_{ij}\beta}}{\sum_k e^{X'_{ik}\beta}} f(\beta_i) d\beta_i$$

# A Simple Example: Mixed Logit (again!)

Following the same procedure as in Maximum Simulated Likelihood, we approximate $P(D_{ij}|X_{ij};\theta)$ through simulation:

1. Draw $R \gg 0$ independent draws $\beta^r$ from $f(\beta;\theta)$
2. Compute the choice probability as:

$$\check{P}_j(X_{ij}|\theta) = \frac{1}{R}\sum_{r=1}^{R}\frac{e^{X_{ij}'\beta^r}}{\sum_k e^{X_{ik}'\beta^r}}$$

Then, the Method of Simulated Moments estimator of $\theta$ is given by:

$$\hat{\theta}_{MSM} := \arg\min_{\theta}\Big[\sum_i^N\sum_j^J\Big(D_{ij} - \check{P}_j(X_{ij}|\theta)\Big)z_{ij}\Big]' W_R\Big[\sum_i^N\sum_j^J\Big(D_{ij} - \check{P}_j(X_{ij}|\theta)\Big)z_{ij}\Big]$$

# A Simple Example: Mixed Logit (again!)

Remember, in the Maximum Simulated Likelihood case, the estimator of $\theta$ was given by:

$$\hat{\theta}_{MSL} := \arg\max_{\theta} SLL(\theta|x)$$

where

$$SLL(\theta|x) = \sum_{i=1}^{N} \sum_{j=0}^{J} D_{ij} \log \check{P}(x|\theta)$$

- Question: Does MSM encompass MSL in this case? If the likelihood does not have a closed form solution, we can't use the first order condition as a moment in the MSM.

# The Weighting Matrix $W_R$

- Optimal weighting matrix: Smallest possible asymptotic variance of $\theta(W)$.
- The optimal weighting matrix is the inverse variance covariance matrix of the moments at the optimal moments.
- If $e(\tilde{x}, x|\theta)$ is the moment error function, then:

$$\hat{\theta}_{MSM} := \arg\min_{\theta} e(\tilde{x}, x|\theta)' W e(\tilde{x}, x|\theta)$$

$$W^{opt} = \left( \frac{1}{N} e(\tilde{x}, x|\theta_0) e(\tilde{x}, x|\theta_0)' \right)^{-1}$$

- When $R$ increases to infinity $\implies$ the variance of the MSM estimator is the same as the variance of the GMM estimator

# The Weighting Matrix $W_R$: Two-step estimator

1. Use $W = I$ to estimate a first-step SMM estimator for $\theta$:

$$\hat{\theta}_1 = \arg\min_{\theta} e(\tilde{x}, x|\theta)' I e(\tilde{x}, x|\theta)$$

2. Using $\theta_1$, estimate $\hat{W}$:

$$\hat{W}(\hat{\theta}_1) = \left(\frac{1}{N} e(\tilde{x}, x|\hat{\theta}_1) e(\tilde{x}, x|\hat{\theta}_1)'\right)^{-1}$$

3. Lastly, re-estimate the MSM estimator using the optimal two-step weighting matrix.

$$\hat{\theta}_2 = \arg\min_{\theta} e(\tilde{x}, x|\theta)' \hat{W}(\hat{\theta}_1) e(\tilde{x}, x|\theta)$$

- $\hat{\theta}_2$ is called the two-step MSM estimator
- If we would iterate over this procedure until $\hat{W}_{i+1}$ is very close to $\hat{W}_i$: Iterated variance covariance estimator of W

# Another Example: Competition between Walmart and Kmart

- **Paper**: What Happens When Walmart Comes to Town by Panle Jia (ECMA 2008)

Consider two chains competing in $M$ markets: Kmart and Walmart. Each chain faces the following problem:

$$\max_{D_1,\ldots,D_M} \Pi_i = \sum_{m=1}^{M} \left[ D_m \left( \beta_i X_m + \eta_{i,m} + \delta_{ij} D_{j,m} + \delta_{ii} \sum_{l \neq m} \frac{D_{i,l}}{Z_{ml}} \right) \right]$$

- $X_m$: market characteristics
- $\eta_{i,m}$: firm and market specific profit shock
- $\delta_{ij} D_{j,m}$: My profits are affected in market $m$ if my competitor decides to enter market $m$
- $\delta_{ii} \sum \frac{D_{i,l}}{Z_{ml}}$: The decision to open a store in market $m$ increases the profits on other markets through the chain effect.

# Another Example: Competition between Walmart and Kmart

We want to estimate $\theta = (\beta_i, \delta_{ij}, \delta_{ii})$ where $i \in \{\text{Walmart}, \text{Kmart}\}$

- This is a very complicated problem!
- There is no close solution for the equilibrium objects of this economy.
- Jia proposed a complicated algorithm to determine the equilibrium.
- MSM gives the flexibility necessary to estimate $\theta$, she uses the following moments:
  - Total number of Kmart stores and Walmart stores
  - Market structure: Number of markets where only one chain enters, where both enters
  - Chain equilibrium profits
  - Interaction between market characteristic's and equilibrium objects

# Use of MSM and further resources

- Applications
  - Models of job search (Flinn and Mabli, 2008)
  - Educational and occupational choices (Adda et al., 2011, 2013)
  - Household choices (Flinn and Del Boca, 2012)
  - Stochastic volatility models (Andersen et al., 2002; Raknerud and Skare, 2012)
  - Dynamic stochastic general equilibrium models (Ruge-Murcia, 2012)
- Textbook treatments
  - Train (2003), Adda and Cooper (2003) and Davidson and MacKinnon (2004)
- Practioner's guide
  - https://notes.quantecon.org/submission/5b3db2ceb9eab00015b89f93