Numerical Methods-Lecture XIII: A Practical Discussion of GMM/MSM

(or; How to Estimate a Structural Model)

Trevor Gallen

Fall 2015

Introduction

- We will play fast and loose with the term "GMM"
- Much of the time I just mean a "minimum distance" estimator
- That said, GMM is the best
- A way to "calibrate" or estimate your model
- Take your model seriously and have it interact with the data
- Don't have to target entire distribution: require your model to say something without saying everything

GMM

- First, forget everything you know.
- In most uses, GMM is making your fit look like the data, just like least squares
- Only difference is the moment conditions
- Let's do an extremely transparent example (!?)

Example: Setup

- Agents like Medicaid, private health insurance, and consumption

$$U_i(m_i, p_i, c_i) = \log(c_i) + \psi_{1,i} m_i + \psi_{2,i} p_i$$

Where they have the budget constraint:

$$Inc_i = c_i + P_p p_i + P_m m_i$$

And where:

$$\begin{bmatrix} \psi_{1,i} \\ \psi_{2,i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\psi_1} & \sigma^2_{\psi_1 \psi_2} \\ \sigma^2_{\psi_1 \psi_2} & \sigma^2_{\psi_2} \end{bmatrix} \right)$$

$$Inc_i \sim \log \mathcal{N}(\mu_{Inc}, \sigma^2_{Inc})$$

Example: Estimation Assumptions

- Given P_p, P_m, μ_{ψ_1}, μ_{ψ_2}, $\sigma^2_{\psi_1}$, $\sigma^2_{\psi_2}$ $\sigma^2_{\psi_1\psi_2}$, μ_{Inc}, σ^2_{Inc}

$$U_i(m_i, p_i, c_i) = \log(c_i) + \psi_{1,i} m_i + \psi_{2,i} p_i$$

Where they have the budget constraint:

$$Inc_i = c_i + P_p p_i + P_m m_i$$

And where:

$$\begin{bmatrix} \psi_{1,i} \\ \psi_{2,i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\psi_1} & \sigma^2_{\psi_1\psi_2} \\ \sigma^2_{\psi_1\psi_2} & \sigma^2_{\psi_2} \end{bmatrix} \right)$$

$$Inc_i \sim \log\mathcal{N}(\mu_{Inc}, \sigma^2_{Inc})$$

Example: Estimation Procedure

1. Extract moments from “true dataset.”
2. Assume a set of parameters.
3. Simulate dataset.
4. Compare simulated parameters to true parameters.
5. Take weighted sum of squared differences.
6. With a new set of parameters, start back at 3
7. Continue until your simulated parameters are as close as can be

Example: Identification

- It's worth thinking about how we identify preferences
- We'll identify based off behavior: what people actually do
- Moreover, we assume they maximize, so take FOC's/maximize
- In this case, maximize.
- Want to estimate ψ_1, ψ_2, ψ_3, σ

Example: Estimation Moments

- How do you choose moments?
- I choose (haphazardly):
  1. Proportion of households that have no insurance
  2. Proportion of households with only one medicaid contract
  3. Proportion of households with 1 medicaid contract that have at least one private insurance contract
  4. Average level of consumption (3)
  5. Standard deviation of consumption ()
  6. Conditional on having any insurance, mean consumption
  7. Conditional on having any insurance, standard deviation of consumption
  8. Proportion of households that have more than 5 contracts
  9. Correlation of m and p
  10. Correlation of m and c
  11. Correlation of c and p

Example: Estimation Moments

- The core of GMM here is, after creating the simulated population:

$$f(\Theta) = \begin{bmatrix} [length(find(best\_m == 0 \& best\_p == 0))./num] \\ [length(find(best\_m == 1))./num] \\ \vdots \\ [mean(best\_c(find(best\_m + best\_p > 0)))] \end{bmatrix}$$

- And, with moment(:,1) holding the simulated moments and moment(:,2) holding the targets,

$$error = sum((moment(:,1) - moment(:,2)).^2)$$

Example: Estimation Moments

See Main.m and Estimation.m (and Data.m for the "true" data generating process!)

Okay...I think I get it(?)

Okay...I think I get it(?)

- Do you?

Okay...I think I get it(?)

- Do you?
- Also see a general equilibrium version of firm taxation in separate files Main.m and Estimator.m
- I still don't get the notation and the optimal part...

A little on notation

- You normally see GMM written like:

$$\beta_N = \arg \underset{\min}{\beta \in} \mathbb{P} g_N(\beta)' W g_N(\beta)$$

- Where $g_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} f(x_t, \beta)$, where your moment condition is: $E_t(f(x_t, \beta)) = 0$

- Let ME_1 stand for "Moment error 1", then $g_N(\beta)' W g_N(\beta)$ is really just:

$$\begin{bmatrix} ME_1 \\ ME_2 \\ ME_3 \\ \vdots \\ ME_r \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} ME_1 & ME_2 & ME_3 & \cdots & ME_r \end{bmatrix}$$

- Note that because of how I wrote the weighting matrix, this is just the sum of squared errors:

$$ME_1^2 + ME_2^2 + ME_3^2 + ... + ME_r^2$$

A little on the best choice of W

- You'll note that some of my moment choices are pretty correlated
- With this weighting, you give one data concept multiple weights
- But let's now take the variance-covariance matrix of our moments:

$$V = \left(\begin{bmatrix} ME_1 \\ ME_2 \\ ME_3 \\ \cdots \\ ME_r \end{bmatrix} \begin{bmatrix} ME_1 & ME_2 & ME_3 & \cdots & ME_r \end{bmatrix} \right)$$

  (Or, with many data points the sample mean of the same thing)
- Then use V as your new W
- What is this doing?

A little on the best choice of W

- This is just weighted least squares
- But more numerically, what V is getting at is the information in a given moment condition
- Every moment is saying something about your θ's
- What the weighting matrix does is listens more to:
  - Moments that are more consistent (not a lot of noise)
  - Moments that move a lot for small deviations of θ
- That second is particularly interesting, and should a bit like the information matrix?
- When your model blows up with small changes in the parameter, it means the parameter is very precisely pinned down
- Good and bad, bad and good.
- From a practical standpoint, $W = I$ with slight tweaks for moments of different absolute magnitude gets you pretty far

Calibration vs. Estimation

- Fundamentally, model uncertainty about parameters comes from sensitivity
- If changing a parameter a little doesn't harm your fit, then you don't really know what it is
- This may help the information matrix's relation to standard errors make sense to you!

Calibration vs. Estimation

- Fundamentally, model uncertainty about parameters comes from sensitivity
- If changing a parameter a little doesn't harm your fit, then you don't really know what it is
- This may help the information matrix's relation to standard errors make sense to you!
- It also makes clear that robustness checks are doing the same thing!

Calibration vs. Estimation

- Fundamentally, model uncertainty about parameters comes from sensitivity
- If changing a parameter a little doesn't harm your fit, then you don't really know what it is
- This may help the information matrix's relation to standard errors make sense to you!
- It also makes clear that robustness checks are doing the same thing!
- It's the *Change* in your score.

Calibration vs. Estimation

- Fundamentally, model uncertainty about parameters comes from sensitivity
- If changing a parameter a little doesn't harm your fit, then you don't really know what it is
- This may help the information matrix's relation to standard errors make sense to you!
- It also makes clear that robustness checks are doing the same thing!
- It's the *Change* in your score.
- For whatever reason, it's relatively rare to report standard errors when you don't touch microdata
- More typically, you see "calibration" and "robustness" rather than "estimation" and "standard errors."
- "Estimation is when you care about your standard errors."

Second, Firm-Size Model

- Many firms, with production function:

$$Y_i = A_i L_i^{\alpha}$$

- And profit function:

$$\pi_i = Y_i - wL_i$$

- Many households with labor supply FOC:

$$L = \frac{2w}{\psi + 0.002w^2}$$

- Supply and demand must hold:

$$\sum_{h \in \{HH\}} L_h = \sum_{f \in \{F\}} L_F$$