Structural Estimation 3: Generalized Method of Moments

Christine Braun

So Far

- So far we have talked about Maximum Likelihood Estimation
- **Today:**
  - Generalized Method of Moments (GMM)
- **Next Time:**
  - Simulated Method of Moments (SMM)
  - Touch on indirect inference (SMM is indirect inference)

Generalized Method of Moments

- Y_t: n-dimensional vector of observations
  - t does not have to mean time, could be people
  - unemployment, wages, duration, observables characteristics, ect..
- θ_0: vector of true parameters
- $g(Y_t, \theta)$: a vector valued function of data and parameters
  - such that $E[g(Y_t, \theta_0)] = 0$
  - where does g come from?

Generalized Method of Moments

- Basic idea is we replace $E[\cdot]$ with empirical analog

$$E[g(Y_t, \theta)] \to \frac{1}{T} \sum_{t=1}^{T} g(Y_t, \theta)$$

- The GMM estimate of θ_0 is

$$\hat{\theta} = \underset{\theta \in \Theta}{\text{argmin}} \left(\frac{1}{T} \sum_{t=1}^{T} g(Y_t, \theta) \right)' W \left(\frac{1}{T} \sum_{t=1}^{T} g(Y_t, \theta) \right)$$

  where W is the weighting matrix.

- In practice we replace W with $\hat{W}$ computed using the data

Asymptotic Distribution of GMM Estimator

- The asymptotic distribution of GMM Estimator is

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N\big(0, (J'WJ)^{-1}J'W\Omega WJ(J'WJ)^{-1}\big)$$

- $J = E[\nabla_\theta g(Y_t, \theta)]$: jacobian of g

- $\Omega = E[g(Y_t, \theta_0)g(Y_t, \theta_0)']$

- If we have $W = \Omega^{-1}$

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N\big(0, (J'\Omega J)^{-1}\big)$$

GMM In Practice

- We can not set $W = \Omega^{-1}$, we don't know Ω
- **Iterated GMM:**
  1: Take $\hat{W}_{(1)} = I$ (identity matrix) estimate $\hat{\theta}_{(1)}$
  2: Calculate

$$\hat{W}_{(2)} = \left(\frac{1}{T} \sum_{t=1}^{T} g(Y_t, \hat{\theta}_{(1)}) g(Y_t, \hat{\theta}_{(1)})' \right)^{-1}$$

  3: Repeat 1 & 2, each time with $\hat{W}_{(i+1)}(\hat{\theta}_{(i)})$ until convergence
- **Continuously updating GMM:** Estimate as

$$\hat{\theta} = \operatorname*{argmin}_{\theta \in \Theta} \left(\frac{1}{T} \sum_{t=1}^{T} g(Y_t, \theta) \right)' \hat{W}(\theta) \left(\frac{1}{T} \sum_{t=1}^{T} g(Y_t, \theta) \right)$$

Same Simple Example: Model

- Model
  - unemployed workers receive job offers at rate λ
  - job offers are drawn from an exogenous wage distribution $F(w)$
  - jobs get destroyed at rate δ
  - workers discount at rate r

Same Simple Example: Model

- Value functions and reservation wage

$$rU = b + \lambda \int_{w_R}^{\infty} E(w) - U \; dF(w)$$

$$rE(w) = w + \delta[U - E(w)]$$

$$w_R = b + \frac{\lambda}{r + \delta} \int_{w_R} w - w_R \; dF(w)$$

Same Simple Example: Model

- What are the parameters of the model that we want to estimate?
  - λ: arrival rate of job offers
  - b: unemployment flow utility
  - r: discount rate
  - δ: separation rate
  - $F(w)$: wage offer distribution
    - let's make the same assumption about the dist.
    - $F(w) \sim \ln N(\mu, \sigma)$

Same Simple Example, Same Identification Issues

- We will use data4.csv to estimate parameters
  - column 1: dummy =1 if unemployed
  - column 2: unemployment duration
  - column 3: wages of employed
  - column 4: employment duration
- We have the same identification issues as MLE
  - w_R is a function of all the parameters
  - use $\hat{w}_R = \min\{w_1, w_2, ..., w_N\}$
  - set $r = 0.05$

Parameters and Moments

- We have 4 parameters to estimate
    - λ, δ, μ, σ
- What moments can we use?

Parameters and Moments

- We have 4 parameters to estimate
  - λ, δ, μ, σ
- What moments can we use?
  1. unemployment rate
  2. expected unemployment duration
  3. expected employment duration
  4. first moment of wage
  5. second moment of wage or variance

GMM estimator notation

- $\{Y_t\}$: the observables
  - for us: $Y_i = \{u_i, tu_i, w_i, te_i\}$ for $i = 1, ..., N = 10,000$
- θ: $(\lambda,\ \delta,\ \mu,\ \sigma)$
- $g(Y_i, \theta)$: function of data and parameters such that $E[g(Y_i, \theta)] = 0$
  - for us: difference between empirical moment and theoretical (calculated from the model) moment

GMM estimator notation: Theoretical Moments

- unemployment rate

$$\frac{\delta}{\delta + \lambda[1 - F(w_R; \mu, \sigma)]}$$

- expected unemployment duration

$$\frac{1}{\lambda[1 - F(w_R; \mu, \sigma)]}$$

- expected employment duration

$$\frac{1}{\delta}$$

- first moment truncated log-normal

$$E[w; \mu, \sigma] = \exp\left(\mu + \frac{\sigma^2}{2}\right) \frac{\Phi\left(\frac{\mu + \sigma^2 - \ln(w_R)}{\sigma}\right)}{1 - \Phi\left(\frac{\ln(w_R) - \mu}{\sigma}\right)}$$

- second moment of truncated log-normal

$$E[w^2; \mu, \sigma] = \exp(2\mu + 2\sigma^2) \frac{\Phi\left(\frac{\mu + 2\sigma^2 - \ln(w_R)}{\sigma}\right)}{1 - \Phi\left(\frac{\ln(w_R) - \mu}{\sigma}\right)}$$

GMM estimator notation: g

- $g(\{Y_i\}, \theta)$ returns a $(M \times N)$ vector, for us $(5 \times 10,000)$
- $g(Y_i, \theta)$ returns a $(M \times 1)$

$$g(Y_i, \theta) = \begin{bmatrix} u_i - \frac{\delta}{\delta + \lambda[1 - F(w_R; \mu, \sigma)]} \\ tu_i - \frac{1}{\lambda[1 - F(w_R; \mu, \sigma)]} \\ te_i - \frac{1}{\delta} \\ w_i - E[w; \mu, \sigma] \\ w_i^2 - E[w^2; \mu, \sigma] \end{bmatrix}$$

- Let $\tilde{N} = [N_u, N_u, N_e, N_e, N_e]$ and N_u, N_e is the number of unemployed and employed then

$$\tilde{N}^{-1} \sum_{i=1}^{N} g(Y_i, \theta) \to E[g(Y_i, \theta)]$$

Estimation in Matlab

- Use data4.csv
- File 1: SE3_main.m
- File 2: g_function.m
  - inputs: parameters, data, w_R estimate
  - output: $(M \times N)$ matrix of moments
- File 3: GMM.m
  - inputs: parameters, data, w_R estimate, $\tilde{N}$
  - outputs: weighted squared distance
- First estimate with $W = I$ then calculate efficient W and re-estimate

Estimation in Matlab: Standard errors

- First we will need the Jacobian Matrix
  - Add on: Adaptive Robust Numerical Differentiation
  - jacobianest(fun,x0)
- The function we are differentiating

$$\tilde{N}^{-1}\sum_{i=1}^{N} g(Y_i,\theta)$$

  should return a $(M \times dim(\theta))$ matrix. For us: (5×4)
- Evaluating at $x0 =$ GMM_ests1

Estimation in Matlab: Standard errors

- Estimate of Ω matrix

$$\hat{\Omega} = \tilde{N}^{-1} \sum_{i=1}^{N} g(Y_i, \theta) g(Y_i, \theta)'$$

- Variance-Covariance Matrix (with $W = I$)

$$\hat{V} = (J'WJ)^{-1} J'W\hat{\Omega}WJ(J'WJ)^{-1}$$

- Standard errors

$$std = \sqrt{\frac{diag(\hat{V})}{N}}$$

Estimation in Matlab: Answers

	GMM		MLE	
Parameter	Estimate	Std. Err.	Estimate	Std. Err.
λ	0.2994	0.0117	0.2820	0.0127
δ	0.0222	0.0002	0.0225	0.0009
μ	2.2043	0.0195	2.2339	0.0119
σ	0.4023	0.0087	0.3794	0.0043

- Data is generated using same underlying parameters
- Asymptotically MLE std. err. smaller than GMM std. err.
- MLE is the minimum variance unbiased estimator
- Note: we are using more information in the GMM (te_i)

Estimation in Matlab: Updated weighting matrix

- Calculate new weighting matrix

$$\hat{W} = \hat{\Omega}^{-1}$$

$$\hat{W} = \begin{bmatrix} 9.1490 & 0.0000 & -0.0000 & 0.0011 & -0.0000 \\ 0.0000 & 0.0257 & -0.0000 & 0.0000 & -0.0000 \\ -0.0000 & -0.0000 & 0.0005 & -0.0002 & 0.0000 \\ 0.0012 & 0.0000 & -0.0002 & 1.5591 & -0.0457 \\ -0.0000 & -0.0000 & 0.0000 & -0.0457 & 0.0014 \end{bmatrix}$$

- Estimate with new weighting matrix

Estimation in Matlab: Answers

	GMM $W = \hat{\Omega}^{-1}$		GMM $W = I$	
Parameter	Estimate	Std. Err.	Estimate	Std. Err.
λ	0.2981	0.0115	0.2994	0.0117
δ	0.0222	0.0002	0.0222	0.0002
μ	2.2060	0.0194	2.2043	0.0195
σ	0.4016	0.0087	0.4023	0.0087

- standard errors get slightly smaller
- we can repeat again, but when do we stop?

$$||W_{(i+1)} - W_{(i)}|| < \varepsilon$$

GMM vs MLE

- **MLE Strengths**
  - more statistical significance
  - less sensitive to parameter or model normalizations
  - less bias and more efficiency with small samples
- **MLE Weaknesses**
  - require strong distributional assumptions
  - likelihood function can become highly non-linear

GMM vs MLE

- **GMM Strengths**
  - minimal distributional assumptions
  - more flexible identification
  - strongly consistent with large samples
- **GMM Weaknesses**
  - less statistical significance
  - more sensitive to normalizations
  - often large bias and inefficiency with small samples

Choosing between GMM and MLE

1. How much data do you have?
2. How complex/non-linear is the model?
3. How comfortable are you making distributional assumptions?
    - wages are log-normal is not so controversial
    - what about an ability or human capital distribution?

A Note on $g(Y_t, \theta)$

- Sometimes we choose to minimize the moment error function

$$e(Y_t, \theta) = \frac{m(Y_t, \theta) - m(Y_t)}{m(Y_t)}$$

  - $m(Y_t, \theta)$: moments calculated using model
  - $m(Y_t)$: moments calculated using data

- Then the GMM estimate is

$$\hat{\theta} = \underset{\theta \in \Theta}{argmin} \left(\frac{1}{T} \sum_{t=1}^{T} e(Y_t, \theta) \right)' W \left(\frac{1}{T} \sum_{t=1}^{T} e(Y_t, \theta) \right)$$

- The error function is percent deviation from moment

- Puts all the moments in the same units
  - no moment gets unintended weighting due to units

- Can also start with a $W_1 = I/(\text{empirical moments})$

Next Time

- With this model we were able to find closed form solutions to the theoretical moments
- This will not always (rarely!) be the case
- Simulated Method of Moments (SMM)
  - given a set of parameters
  - simulate data from the model
  - calculate moments in simulated data
  - compare to moments from observed data