Generalized Method of Moments

GMM in Applied Settings

David Hao Zhang[1]

Harvard University

September 16, 2018[2]

[1] dzhang@hbs.edu

[2] Based on previous notes by Ashvin Gandhi, Daniel Pollmann, Tom Wollmann, and Michael Sinkinson.

General Advice

Sections:

- Not weekly.
- Not required.
- Hopefully helpful.

Problem Sets:

- Read the papers.
- Work together, but *do not copy code or content.*
- For derivations, show your work.
- For programs, comment and package your code.
- Include your code in your writeup (LaTeX package *mcode*).

Introduction/What is GMM?

- Introduced in Chamberlain's Econ 2120, Lecture Note 14.
- GMM identifies parameters via the expectation:

$$\mathbb{E}\left[\psi\left(w_i;\theta_0\right)\right] = 0.$$

- Weighted penalty for deviation from the moments.

$$\theta_0 = \arg\min_{\theta} \mathbb{E}\left[\psi\left(w_i;\theta\right)\right]' C \mathbb{E}\left[\psi\left(w_i;\theta\right)\right],$$

  where C is positive definite.
- Heuristically, one can think of GMM as imposing less structure than MLE but more structure than non-parametric estimation.

Typical moments in IO applications

- Look at the models we estimate for zero-correlation conditions. One obvious example is "unobserved heterogeneity" in product characteristics, ξ. Look at what it is uncorrelated with and form a moment from that.
- Nash conditions: Equilibrium conditions which we assume to hold on the supply side, such as Differentiated Products Bertrand Equilibrium.
- Consumer optimality: Consumer may optimally stockpile, for example, based on sales frequencies and amounts.

Identification and Consistency

- Identification is achieved for GMM case if

  $$\mathbb{E}\left[\psi\left(w_i;\theta\right)\right] = 0$$

  only holds at the true value $\theta = \theta_0$, and that at all other values of the parameter vector, it does not hold.

- Consistency holds if:

  $$\hat{\theta} \xrightarrow{p} \theta_0$$

- Formally, this is the same as saying

  $$\lim_{N\to\infty} \Pr\left[\left\|\hat{\theta} - \theta_0\right\| > \varepsilon\right] = 0, \, \forall\varepsilon > 0.$$

- Under appropriate assumptions, GMM is consistent.

Efficiency

- We want to know whether our estimates are as precise as possible. The ML estimator achieves the Cramer-Rao lower bound on variance among all unbiased estimators in the parametric setting:

$$Var\left(\hat{\theta}(X)\right) \geq \Im\left(\theta_0\right)^{-1}$$

$$\Im\left(\theta\right) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta\partial\theta'}\ln f\left(X|\theta\right)\right]$$

  - We call $\Im\left(\theta\right)$ the Fisher Information matrix, and $\Im\left(\theta_0\right)^{-1}$ is the Cramer-Rao bound.
- The GMM estimator attains the semi-parametric efficiency bound (Chamberlain, 1987), which is the lower bound on variance for an estimator using only the information contained in the moment restrictions.
- In the over-identified case will require a two-step estimator (which we will discuss shortly) for efficiency.

Estimation

- We compute the empirical mean of the moment function, and select $\hat{\theta} = \arg\min_\theta Q_{C,n}(\theta)$, where
$$Q_{C,n}(\theta) = \left[\frac{1}{n}\sum_{i=1}^{n}\psi(w_i,\theta)\right]' C \left[\frac{1}{n}\sum_{i=1}^{n}\psi(w_i,\theta)\right]$$
for some positive definite $M \times M$-matrix C.
- The weighting matrix matrix C assigns "importance" to satisfying the different moment conditions.

Asymptotic variance

- Under appropriate assumptions,
  $$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, V\right),$$
  where
  $$V = \left(\Gamma' C \Gamma\right)^{-1} \Gamma' C \Delta C \Gamma \left(\Gamma' C \Gamma\right)^{-1}.$$
- $\Gamma = \mathbb{E}\left[\frac{\partial \psi}{\partial \theta}\left(x, \theta_0\right)\right]$ $(M \times K)$: gradient of the moment function with respect to the parameters
- $\Delta = \mathbb{E}\left[\psi\left(x, \theta_0\right) \psi\left(x, \theta_0\right)'\right]$ $(M \times M)$: outer product of the moments
- Note that this is only one component of error. There is also:
  - sampling error (if your data is a sample of the population).
  - simulation error (if you compute the moments via simulation).

Idea for Two-step GMM

Just-identified case, $C = I$:

$$\begin{aligned} V &= \left(\Gamma' C \Gamma\right)^{-1} \Gamma' C \Delta C \Gamma \left(\Gamma' C \Gamma\right)^{-1} \\ &= \Gamma^{-1} C^{-1} \Gamma'^{-1} \Gamma' C \Delta C \Gamma \Gamma^{-1} C^{-1} \Gamma'^{-1} \\ &= \Gamma^{-1} \Delta \Gamma'^{-1} \\ &= \left(\Gamma' \Delta^{-1} \Gamma\right)^{-1}, \end{aligned}$$

Over-identified case, $C = \Delta^{-1}$:

$$\begin{aligned} V &= \left(\Gamma' C \Gamma\right)^{-1} \Gamma' C \Delta C \Gamma \left(\Gamma' C \Gamma\right)^{-1} \\ &= \left(\Gamma' \Delta^{-1} \Gamma\right)^{-1} \Gamma' \Delta^{-1} \Delta \Delta^{-1} \Gamma \left(\Gamma' \Delta^{-1} \Gamma\right)^{-1} \\ &= \left(\Gamma' \Delta^{-1} \Gamma\right)^{-1} \end{aligned}$$

The proof that $\left(\Gamma' C \Gamma\right)^{-1} \Gamma' C \Delta C \Gamma \left(\Gamma' C \Gamma\right)^{-1} - \left(\Gamma' \Delta^{-1} \Gamma\right)^{-1} \geq 0$ (positive semi-definite) can be found in virtually every econometrics text or lecture notes. This proves that $C = \Delta^{-1}$ is indeed optimal.

Choice of weighting matrix via Two-step GMM

- We would like $C \propto \Delta^{-1}$. Recall that Δ is the expectation of the covariance matrix at θ_0.
- Problem: we don't know θ_0.
- Solution: Form a consistent estimate $\hat{\Delta}$ using a consistent though inefficient estimate of θ_0.

Two-step GMM:

- Step 1: Estimate $\hat{\theta}_{GMM1}$ by minimizing $Q_C(\theta)$ with any arbitrary choice of (positive semi-definite) C, such as the identity matrix.
- Step 2: Estimate the optimal weighting matrix as:

$$\hat{\Delta}^{-1} = \left\{ \mathbb{E}_n \left[\psi\left(w_i, \hat{\theta}_{GMM1}\right) \psi\left(w, \hat{\theta}_{GMM1}\right)' \right] \right\}^{-1}$$

  and use this to then solve for $\hat{\theta}_{GMM2} = \arg\min_\theta Q_{\hat{\Delta}^{-1}}(\theta)$.

A Simple Example

- Suppose we have the following model:

  $$y_i = x_i'\beta + \epsilon_i,$$

  where $\mathbb{E}\left(\epsilon_i | x_i\right) = 0$.
- Then, $\mathbb{E}\left(y_i - x_i'\beta | x_i\right) = 0 \Rightarrow \mathbb{E}\left[\left(y_i - x_i'\beta\right) h\left(x_i\right)\right] = 0$ for any function $h(\cdot)$, in particular $h(x) = x$.
- Hence,

  $$\mathbb{E}\left[\psi\left(w_i;\theta\right)\right] = 0,$$

  where $\psi\left(w_i;\theta\right) = \left(y_i - x_i'\beta\right) x_i$.
- In a more general problem, using "optimal instruments" means optimal choice of $h(\cdot)$, an approximation to which we will discuss later.

Linear IV example

- Now, suppose $\mathbb{E}\left(\epsilon_i|x_i\right) \neq 0$, but we have a (relevant) instrument z such that $\mathbb{E}\left(\epsilon_i|z_i\right) = 0$ (exclusion restriction).
  - Standard tool is TSLS
- In the GMM framework, we can use the moment function $\psi\left(w_i, \theta\right) = \left(y_i - x_i'\beta\right) z_i$.
- If only some elements of the K-vector x_i are endogenous, z_i will also include the remaining subset. If $\dim\left(z_i\right) = \dim\left(x_i\right)$, the model is just-identified; for $\dim\left(z_i\right) > \dim\left(x_i\right)$, it is over-identified.

Analytical solution to the linear GMM

- Chamberlain Lecture Notes 12 and 14.
- For x_t a set of explanatory variables in market t, and z_t a set of instruments, all in column form:

$$y_t = x_t'\beta + \epsilon_t, S_{zy} = \frac{1}{N}\sum_{t=1}^{N} z_t y_t, S_{zx} = \frac{1}{N}\sum_{t=1}^{N} z_t x_t'$$

$$\hat{\beta} = (S_{zx}' C S_{zx})^{-1} S_{zx}' C S_{zy}$$

$$\hat{S} = \frac{1}{N}\sum_{t=1}^{N} z_t z_t' \hat{\epsilon}_t^2$$

$$Cov(\hat{\beta}) = (S_{zx}' C S_{zx})^{-1} S_{zx}' C \hat{S} C S_{zx} (S_{zx}' C S_{zx})^{-1}$$

  Where C is the weight matrix. The weight matrix can be estimated after the first-step via: $\hat{C} = \hat{S}^{-1}$. $Cov(\hat{\beta})$ should be estimated in the second step with the second step $\hat{S}$.
- Note that this is equivalent to 2SLS if errors are homoscedastic (but they may not be!).

Logit example – linear GMM

Suppose instead, we have:

- Only market level data (market shares)
- Endogeneity of certain characteristics (need to instrument)

Harder to construct a likelihood function. Then what?

Let $\delta_j = \beta X_j + \xi_j$, so that market shares (aggregated across all consumers i) is:

$$s_j = \frac{\exp(\delta_j)}{1 + \sum_k \exp(\delta_k)} \tag{1}$$

Given share of outside good s_0, we can easily recover δ_j using an idea from Berry (1994):

$$\delta_j = \log(s_j) - \log(s_0) \tag{2}$$

So that given instruments Z_j which are assumed to be independent of ξ_j, the moment condition is:

$$E(\xi_j Z_j) = E((\delta_j - \beta X_j) Z_j) = 0 \tag{3}$$

Which is a linear GMM following the standard IV case!

Logit example – nested fixed point

Is there another way to get δ_j? Starting from any δ_k^0, define the following iterative procedure:

$$\delta_j^k = \delta_j^{k-1} + \log(s_j) - \log(\hat{s}_j(\delta^{k-1})) \tag{4}$$

Where s_j is the observed market share and $\hat{s}_j(\delta^{k-1})$ is the computed market share based on last iteration of δs. By Berry, Levinsohn, and Pakes (1995), this is a contraction mapping of modulus less than 1. To compute the fixed point, keep looping until for very small $\epsilon = 10^{-14}$:

$$|\max_j \left(\delta_j^k - \delta_j^{k-1}\right)| \leq \epsilon \tag{5}$$

And once we obtained δ_j^k, we can use it to form moments:

$$E(\xi_j Z_j) \approx E((\delta_j^k - \beta X_j) Z_j) = 0 \tag{6}$$

Where the approximation can be as accurate as you want by setting small ϵ.

An Approximation to Optimal Instruments

More generally, suppose we want to estimate α, β using the following moment condition:

$$E(\xi_j H_j(Z)) = E[(\delta_j - \beta X_j - \alpha p_j) H_j(Z)] = 0 \tag{7}$$

Chamberlain (1987) tells us that, with $T(z)' T(z) = \Delta^{-1}$ as a normalizing matrix, the optimal set of instruments is:

$$H_j(z) = E\left[\frac{\partial \xi_j(\theta_0)}{\partial \theta} | Z\right] T(z_j) \tag{8}$$

The approximation of Berry, Levinsohn, and Pakes (1999):

1. Obtain an initial estimate of $\hat{\alpha}, \hat{\beta}$.
2. Use the initial estimate to construct $\hat{\delta}_j = \hat{\beta} X_j + \hat{\alpha} p_j$, $(\xi = 0)$.
3. Solve the FOC of the model to find $\hat{p}, \hat{s}$ as a function of $\alpha, \beta, \hat{\delta}, X$.
4. Get $\hat{\xi}_j(\alpha, \beta) = \hat{\delta}_j(\alpha, \beta) - \beta X_j - \alpha \hat{p}_j(\alpha, \beta)$, and take the derivatives $\frac{\partial \hat{\xi}_j}{\partial \alpha}, \frac{\partial \hat{\xi}_j}{\partial \beta} | \hat{\alpha}, \hat{\beta}$ as an approximation to $E\left[\frac{\partial \xi_j(\theta_0)}{\partial \theta} | Z\right]$.

Implementing GMM in Matlab

- The primary Matlab functions you should be familiar with are "fminsearch" and "fminunc".
- Basic syntax example of how to use it (just-identified case):

```
beta = fminsearch(@(b) (X'*(Y-X*b))'*(X*(Y-X*b)),betastart,myopts)
```

- In our simple example:
  - Y is a column vector and X is a matrix where each row is an observation
  - The answer will be stored in a variable "beta"
  - "@(b)" means the routine will attempt to minimize the expression (X'*(Y-X*b'))'*(X'*(Y-X*b')) with respect to "b"
  - The starting guess for "b" will be the value held in the vector "betastart"
  - The routine will follow the specifications in the options set "myopts", which is set before this using a command like

```
myopts = optimset('TolFun',10^-12, 'MaxFunEvals',1000000,'MaxIter',1000)
```

Matlab: More complicated minimization

- The "fminsearch" command can also evaluate a named function. This is useful if your moments are hard to evaluate. In that case, you would create a separate .m file for the function. Here's an example of moment_function.m file:

```
function [val] = moment_function(beta, X, S, alpha, P)
    % Do manipulations with the input arguments beta, X, S, alpha, P.
    % Suppose you evaluate a moment condition for each observation
    % into a vector called "moment"
    ...
    val = mean(moment);
```

- I could then call this function from "fminsearch" using:

```
beta = fminsearch(@(b) moment_function(b, X, S, alpha, P), betastart, myopts)
```

- Question: How would you implement 2-step GMM?

Evaluating gradients

- Necessary for Γ in asymptotic variance.
- Exact differentiation (analytic derivatives) is always preferred to numerical differentiation due to approximation error. This is also runs *much* faster. Logit models (including BLP) does allow one to compute exact gradients – just differentiate the logit!
- If not practical, use finite differences with h:
  - Forward difference formula:
    $$f'(x) \approx \frac{f(x+h) - f(x)}{h}$$
  - Symmetric difference formula (more accurate):
    $$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}$$
  - See Judd (1998, Ch. 7) for details.

Sensitivity evaluation – is GMM a “black box”?

Main result of Andrews, Gentzkow, and Shapiro (2017): for any local perturbation to the true model leading to the moments converging asymptotically to $\tilde{\psi}$ instead of 0, the first-order asymptotic bias to the estimates $\tilde{\theta}$ is:

$$E(\tilde{\theta}) = \Lambda E(\tilde{\psi}) \tag{9}$$

Where $\Lambda = -(\Gamma' C \Gamma)^{-1} \Gamma' C$ is the sensitivity of estimated parameters to the model.

- For OLS, $\Lambda = -\Gamma^{-1} = -E(XX')$. Omitted variable intuition: the bias from not including an endogenous variable is related to its covariance with included variables.
- Andrews, Gentzkow, Shapiro (2014) generalizes this intuition to GMM.

References I

Andrews, Isaiah, Matthew Gentzkow, and Jesse Shapiro. 2017. "Measuring the Sensitivity of Parameter Estimates to Estimation Moments." *Quarterly Journal of Economics*.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1999. "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy." *American Economic Review*.

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *The RAND Journal of Economics*.

Chamberlain, Gary. 1987. "Asymptotic efficiency in estimation with conditional moment restrictions." *Journal of Econometrics*.