Lecture 3: Generalized Method of Moments

Chris Conlon

February 14, 2021

NYU Stern

GMM: Intro

In the most basic setup we begin with some data w_i where $i = 1, \ldots, N$. Our economic model provides the following restriction on our data:

$$\mathbb{E}[g(w_i, \theta_0)] = 0$$

- At the true parameter value $\theta_0 \in \mathbb{R}^k$ our moment conditions $g(w_i, \theta)$ are on average equal to zero.
- What does "on average" mean? In theory, $g(w_i, \theta_0)$ is a random variable and we are making a statement about its first moment. This is what we mean when we write $\mathbb{E}[\cdot]$.

GMM: IID Normal

Let's estimate the parameters of an IID normal $(x_1, \ldots, x_n)$. Recall the moments of the normal:

$$\mathbb{E}[X_i] = \mu \quad \mathbb{E}[X_i^2] = \mu^2 + \sigma^2$$

We could form two moments by solving the expressions above for zero:

$$g_n^1(x_1, \ldots, x_n, \mu, \sigma) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) - \mu$$
$$g_n^2(x_1, \ldots, x_n, \mu, \sigma) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \mu^2 - \sigma^2$$

This gives us two equations and two unknowns which we can solve for (μ, σ^2).
Of course you probably knew how to estimate the parameters of a normal...

GMM: IID Normal

Let's estimate the parameters of an IID normal $(x_1, \ldots, x_n)$. Recall the moments of the normal:

$$\mathbb{E}[X_i] = \mu \quad \mathbb{E}[X_i^2] = \mu^2 + \sigma^2$$

We could form two moments by solving the expressions above for zero:

$$g_n^1(x_1, \ldots, x_n, \mu, \sigma) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) - \mu$$
$$g_n^{2'}(x_1, \ldots, x_n, \mu, \sigma) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 - \sigma^2$$

This gives us two equations and two unknowns which we can solve for (μ, σ^2).
Of course you probably knew how to estimate the parameters of a normal...

GMM: Sample Moments

In practice, it is helpful to consider the sample analogue, which we abbreviate with the shorthand $g_N(\theta) \in \mathbb{R}^q$, where $g_N(\theta)$ is a q-dimensional vector of moment conditions.

$$\mathbb{E}[g(w_i, \theta)] \approx \frac{1}{N} \sum_{i=1}^{N} g(w_i, \theta) \equiv g_N(\theta)$$

Other Definitions

- We define the Jacobian: $D(\theta) \equiv \mathbb{E}[\frac{\partial g(w_i,\theta)}{\partial \theta}]$, which is a $q \times k$ matrix.
- Evaluated at the optimum, $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} g(w_i, \theta_0) \xrightarrow{d} N(0, S)$ where $S = E[g(w_i, \theta_0) g(w_i, \theta_0)']$ is a $q \times q$ matrix.[1]
- In other words, the moment conditions which are 0 in expectation at θ_0 are normally distributed with some covariance S
- Later, we will refer to a weighting matrix W_N which is a $q \times q$ positive semi-definite matrix. It tells us how much to penalize the violations of one moment condition relative to another (in quadratic distance).

[1] Technical conditions to establish this are written down later.

Examples

It is easy to see some very simple examples:

OLS Here $y_i = x_i\beta + \epsilon_i$. Exogeneity implies that $E[x_i'\epsilon_i] = 0$. We can write this in terms of just observables and parameters as $E[x_i'(y_i - x_i\beta)] = 0$ so that $g(y_i, x_i, \beta) = x_i'(y_i - x_i\beta)$.

IV Again $y_i = x_i\beta + \epsilon_i$. Now, endogeneity implies that $E[x_i'\epsilon_i] \neq 0$. However there are some instruments z_i which may be partly contained in x_i and partly excluded from y_i, so that $E[z_i'\epsilon_i] = 0$. $E[z_i'(y_i - x_i\beta)] = 0$ so that $g(y_i, x_i, z_i, \beta) = z_i'(y_i - x_i\beta)$.

Examples (continued)

Maximum Likelihood

$g(w_i, \theta) = \frac{\partial \log f(w_i, \theta)}{\partial \theta}$ where $f(w_i, \theta)$ is the density function so that $\log f(w_i, \theta)$ is the contribution of observation i to the log-likelihood. Here we set the expected (average) derivative of the log-likelihood (score) function to zero.

Examples (continued)

Euler Equations

Assume we have a CRRA utility function $u(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$ and an agent who maximizes the expected discounted value of their stream of consumption. This leads to an Euler Equation:

$$E\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}R_{t+1} - 1 \middle| \Omega_t\right] = 0$$

where Ω_t is the "Information Set" (sigma algebra) of everything known to the agent up until time t (include full histories). We can write a moment restriction of the form for any measurable $z_t \in \Omega_t$.

$$E\left[z_t\left(\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}R_{t+1} - 1\right)\right] = 0$$

In the original work by Hansen (1982) on GMM, this $g(c_t, c_{t+1}, R_{t+1}, \beta, \gamma)$ was used to estimate (β, γ).

GMM Estimator

Here is the GMM estimator:

$$\hat{\theta} = \arg\min_{\theta} Q_N(\theta) \quad Q_N(\theta) = g_N(\theta)' W_N g_N(\theta)$$

Technical Conditions

These are a set of sufficient conditions to establish consistency and asymptotic normality of the GMM estimator. These conditions are stronger than necessary, but they establish the requisite LLN and CLT.

1. $\theta \in \Theta$ is compact.
2. $W_N \overset{p}{\to} W$.
3. $g_N(\theta) \overset{p}{\to} E[g(z_i, \theta)]$ (uniformly)
4. $E[g(z_i, \theta)]$ is continous.
5. We need that $E[g(z_i, \theta_0)] = 0$ and $W_N E[g(z_i, \theta)] \neq 0$ for $\theta \neq \theta_0$ (global identification condition).
6. $g_N(\theta)$ is twice continuously differentiable about θ_0.
7. θ_0 is not on the boundary of Θ.
8. $D(\theta_0) W D(\theta_0)'$ is invertible (non-singular).
9. $g(z_i, \theta)$ has at least two moments finite and finite derivatives at all $\theta \in \Theta$.

GMM: Asymptotics

The first five conditions give us consistency $\hat{\theta} \xrightarrow{p} \theta_0$ as $N \to \infty$. All nine conditions give us asymptotic normality.

$$\begin{aligned} \sqrt{N}(\hat{\theta} - \theta) &\xrightarrow{d} N(0, V_\theta) \\ V_\theta &= \underbrace{(DWD')^{-1}}_{\text{bread}} \underbrace{(DWSW'D')}_{\text{filling}} \underbrace{(DWD')^{-1}}_{\text{bread}} \end{aligned}$$

It is common to refer parts of the variance as the *bread* and the *filling* or *meat*, together this is referred to as the *sandwich* estimator of the variance.

GMM: Identification

- The global identification condition is difficult to understand, for the linear model we can replace it with a (local) condition on Jacobian of the moment conditions.
- Recall the Jacobian: $D \equiv \frac{\partial g(w_i,\theta)}{\partial \theta}$, which is a $q \times k$ matrix.
- We call the problem under-identified if $rank(D) < k$, just-identified if $rank(D) = k$ and over-identified if $rank(D) > k$.
  - In the under-identified case, there may be many such $\hat{\theta}$ where $g(w_i, \hat{\theta}) = 0$.
  - In the just-identified case, it should be possible to find a $\hat{\theta}$ where $g_N(\hat{\theta}) = 0$.
  - We are primarily interested in the over-identified case where we will generally not find $\hat{\theta}$ which satisfies the moment conditions $g_N(\hat{\theta}) \neq 0$.
- Instead, we search for $\hat{\theta}$ which minimizes the violations of the moment conditions. We write this as a quadratic form for some positive definite matrix W_N which is $q \times q$.

$$\hat{\theta} = \arg\min_{\theta} Q_N(\theta) \quad Q_N(\theta) = g_N(\theta)' W_N g_N(\theta)$$

Thanks!