GMM FOR DUMMIES

AMY RHEE

1. Introduction

1.1. **Background.** Suppose that we have a parameter vector $\theta = (\theta_1, .., \theta_p)$ with p components. For $1 \leq j \leq p$, define the j^{th} *moment*

$$\mathbb{E}_\theta[X^j] = \int x^j dF_\theta(x) \equiv \alpha_j(\theta)$$

Similarly, define the j^{th} *sample moment*

$$\hat{\alpha}_j = \frac{1}{n}\sum_i^n X_i^j$$

1.2. **The Idea behind the Generalized Method of Moments.** GMM is based off of the idea that there exists a *true* parameter $\theta_0 \in \Theta$ that uniquely sets a K-dimensional vector of population moments to zero, i.e.

$$\mathbb{E}[g(X_i; \theta_0)] = 0$$

where our data are $\{X_i\}_{i=1}^N$, and we have p parameters and $K \geq p$ moment conditions.

We say that $\hat{\theta}_{mm}$ is a *Method of Moments estimator* if it is the value of θ such that

$$\alpha_1(\hat{\theta}_{mm}) = \hat{\alpha}_1, \ldots, \alpha_p(\hat{\theta}_{mm}) = \hat{\alpha}_p$$

Simply put, we find $\hat{\theta}_{mm}$ by setting population moments equal to the corresponding sample moments.

We say that $\hat{\theta}$ is a *GMM estimator* if it solves the following problem:

$$\begin{aligned}\hat{\theta} &= \arg\min_\theta Q_n(\theta) \\ &= \arg\min_\theta g_n(\theta)'\hat{W}g_n(\theta)\end{aligned}$$

where $g_n(\theta) = \frac{1}{n}\sum_i^n g(X_i; \theta)$ is our sample objective function and $\hat{W}$ is a $K \times K$ positive-semidefinite matrix that assigns a weight to each moment condition, telling the problem how much to penalize the violation of one moment condition relative to another.

Date: March 22, 2023.

The corresponding population objective function is

$$Q(\theta) = g(\theta)'Wg(\theta)$$

where $g(\theta) = \mathbb{E}[g(X_t;\theta)]$ and W is the plim of $\hat{W}$.

2. Familiar Examples

In some cases, we can think of $\mathbb{E}[g(X_i;\theta_0)] = 0$ as being a FOC for the minimization of some objective function $m(X_t;\theta)$. Informally, the GMM estimator makes sample moment conditions as "true" as possible.

2.1. **Normal Distribution.**

Suppose we have data $\{X_1, .., X_n\} \sim \mathcal{N}(\mu, \sigma^2)$. Then, $\mathbb{E}[X] = \mu$ and $\mathbb{V}(X) = \sigma^2$. We can rewrite the variance so that we have it as a function of the second moment and the first moment, i.e. $\mathbb{V}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \implies \mathbb{E}[X^2] = \mathbb{V}(X) + \mathbb{E}[X]^2$. Therefore, we have two *population* moment conditions:

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

Now let's write their *sample* analogues:

$$\frac{1}{n}\sum_i^n X_i = \hat{\mu}$$
$$\frac{1}{n}\sum_i^n X_i^2 = \hat{\sigma}^2 + \hat{\mu}^2$$

We have two equations in two unknowns $(\hat{\mu}, \hat{\sigma}^2)$, and as we probably already know, the solution to this system is

$$\hat{\mu} = \frac{1}{n}\sum_i^n X_i \equiv \bar{X}_n$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_i^n (X_i - \bar{X}_n)^2$$

We can also put this in our $g(\cdot)$ notation:

$$g_n^1(X_1, ..., X_n; \mu, \sigma) = (\frac{1}{n}\sum_i^n X_i) - \mu$$
$$g_n^2(X_1, ..., X_n; \mu, \sigma) = (\frac{1}{n}\sum_i^n X_i^2) - (\frac{1}{n}\sum_i^n X_i)^2 - \sigma^2$$

2.2. **OLS estimator: Bivariate case.** With our OLS estimator, the model is

$$y_i = x_i\beta + \varepsilon_i$$

For our conditional independence assumption to hold and our OLS estimator to be consistent, we need orthogonality between β and ε_i, i.e. the familiar exogeneity condition $\mathbb{E}[x_i'\varepsilon_i]$. To use GMM, we want to write this condition in terms of things we can observe:

$$\begin{aligned} &\mathbb{E}[x_i'(y_i - x_i\beta)] = 0 \\ \implies\ & g(y_i, x_i, \beta) = x_i'(y_i - x_i\beta) \end{aligned}$$

3. Unfamiliar Examples

3.1. Suppose we have two samples, A and B. Observations in sample A are assumed to have been drawn from a distribution with a mean equal to μ. Observations in sample B are assumed to have been drawn from a distribution with a mean equal to $\mu+5$. All observations are uncorrelated. The sample means are $\bar{y}_A = 7$ and $\bar{y}_B = 9$ and their variances are estimated to be $Var(\bar{y}_A) = 1$ and $Var(\bar{y}_B) = 0.2$.

3.1.1. *What is the quadratic form that is a function of μ that is minimized at the asymptotically efficient GMM estimate of μ?*

$$\begin{bmatrix} \bar{y}_A - \mathbb{E}[y_A] \\ \bar{y}_B - \mathbb{E}[y_B] \end{bmatrix}' \begin{bmatrix} 1 & 0 \\ 0 & 0.2 \end{bmatrix}^{-1} \begin{bmatrix} \bar{y}_A - \mathbb{E}[y_A] \\ \bar{y}_B - \mathbb{E}[y_B] \end{bmatrix} = \begin{bmatrix} \bar{y}_A - \mu \\ \bar{y}_B - (\mu+5) \end{bmatrix}' \begin{bmatrix} 1 & 0 \\ 0 & 0.2 \end{bmatrix}^{-1} \begin{bmatrix} \bar{y}_A - \mu \\ \bar{y}_B - (\mu+5) \end{bmatrix} =$$
$$\begin{bmatrix} 7-\mu \\ 4-\mu \end{bmatrix}' \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 7-\mu \\ 9-(\mu+5) \end{bmatrix}$$

Note that this matrix product yields a scalar!

3.1.2. *What is the numerical value of this GMM estimate?* Take the FOC and solve for $\hat{\mu}$:

$$\begin{aligned} \min_{\mu}\{(7-\mu)^2 + 5(9-(\mu+5))^2\} &= \min_{\mu}\{(7-\mu)^2 + 5(4-\mu)^2\} \\ &\implies -2(7-\hat{\mu}) - 10(4-\hat{\mu}) = 0 \\ &\implies \hat{\mu} = 54/12 \end{aligned}$$