Lesson 3: GMM Estimation

Some sources used in the slides

- Whited T. and Taylor L. Summer School in Structural Estimation.
- Wooldridge, J. M. (2001). Econometric analysis of cross section and panel data.
- Asset Pricing, Cochrane J. 2006.

Introduction

- GMM stands for Generalized Method of Moments. It is a generalization of the method of moments estimator.
- It was formalized by Hansen (1982), and since has become one of the most widely used methods of estimation for models in economics and finance.
- It is the basis for methods like the Simulated Method of Moments (SMM) and the Indirect Inference (II) estimator.
- The power of GMM is that it allows us to estimate models without having to specify the distribution of the data.

The method of moments estimator (Chebyshev)

- It was introduced by Pafnuty Chebyshev in 1887 in the proof of the central limit theorem.
- Suppose you need to estimate k unknown parameters $\theta_1, \ldots, \theta_k$ that characterize the distribution of a random variable X.

$$f_X(x; \theta_1, \ldots, \theta_k)$$

Now, assume that the first k moments can be expressed as a function of the parameters:

$$\begin{aligned} \mu_1 &= E[X] = g_1(\theta_1, \ldots, \theta_k) \\ \mu_2 &= E[X^2] = g_2(\theta_1, \ldots, \theta_k) \\ &\vdots \\ \mu_k &= E[X^k] = g_k(\theta_1, \ldots, \theta_k) \end{aligned}$$

The method of moments (cont.)

- Estimate the population moment with the sample moment

$$\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} x_i^j$$

- Solve the system of equations

$$\begin{aligned}
\hat{\mu}_1 &= g_1(\hat{\theta}_1, \ldots, \hat{\theta}_k) \\
\hat{\mu}_2 &= g_2(\hat{\theta}_1, \ldots, \hat{\theta}_k) \\
&\vdots \\
\hat{\mu}_k &= g_k(\hat{\theta}_1, \ldots, \hat{\theta}_k)
\end{aligned}$$

Example, normal distribution

$$\mu_1 = E[X] = \int_{-\infty}^{\infty} x f_X(x; \mu, \sigma) dx =$$
$$\mu_2 = E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x; \mu, \sigma) dx$$

- After observing a sample of n observations $\{x_1, \ldots, x_n\}$, we can estimate the population moments with the sample moments

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i$$
$$\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

- And solve numerically the system of equations.

GMM

- When the number of moments is equal to the number of parameters there is a unique solution to the system of equations.
- However, we cannot compute the standard errors of the estimates. For this task we need to use the GMM estimator, and include more moments.

GMM (cont.)

- Notation in Wooldride
- w_i is a $(M \times 1)$ i.i.d. vector of random variables for observation i.
- θ is a $(P \times 1)$ vector of unknown coefficients (parameters).
- $g(w_i, \theta)$ is a $(L \times 1)$ vector of functions $g : \mathbb{R}^M \times \mathbb{R}^P \to \mathbb{R}^L$ $L \geq P$
- Function g can be potentially non linear.
- Let θ_0 be the true value of θ.
- Let $\hat{\theta}$ be an estimator of θ.
- The hat and naught notation is used to denote estimators and true values, respectively.

Moment Restrictions

- GMM is based on the idea that the moment restrictions should be zero in expectation (e.g. the difference between the sample and population moments).

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

Which in the sample can be written as

$$\frac{1}{N}\sum_{i=1}^{N} g(w_i, \theta) = 0$$

We want to choose $\hat{\theta}$ such that $N^{-1}\sum_{i=1}^{N} g(w_i, \hat{\theta})$ is as close to zero as possible.

Criterion Function

- If we have more moments than parameters there might not be a solution to the system of equations, but we can make those moments as close to zero as possible.
- Hint, minimize a weighted sum of squared moments.
- How much importance you give to each moment will be discussed later.
- The estimator $\hat{\theta}$ uses the following function (criterion) as a function to minimize.

$$Q_N(\theta) = \left[N^{-1}\sum_{i=1}^{N} g(w_i,\theta)\right]'\hat{W}\left[N^{-1}\sum_{i=1}^{N} g(w_i,\theta)\right]$$

where $\hat{W}$ is a positive definite weighting matrix that converges in probbaility to W_0.

Asymptotic Properties

Hansen (1982) Large Sample Properties of Generalized Method of Moments, **Econometrica**. Two-stage procedure, for any positive semidefined matrix W e.g. I.

$$\hat{\theta}_1 = \arg\min_{\theta} \left[g_T(\theta)\right]' W \left[g_T(\theta)\right]$$

First Order Condition

$$\frac{\partial g_T(\theta)}{\partial \theta} W g_T(\theta) = a g_T(\theta) = 0$$

This estimator is consistent and asymptotically normal but not always efficient, the efficient estimator is obtained by estimating W as the inverse of covariance of moments $g_T(\hat{\theta}_1)$ and re-estimate.

Standard Errors

Hansen proved that the estimator

$$\hat{\theta}_2 = \arg\min_{\theta} \left[g_T(\theta)\right]' \hat{S}^{-1} \left[g_T(\theta)\right]$$

where $\hat{S}$ is the sample covariance of the moments given $\hat{\theta}_1$, is consistent and asymptotically normal. Define

$$d = \frac{\partial g_T(\theta)}{\partial \theta}$$

Then the asymptotic variance of $\hat{\theta}_2$ is

$$\hat{V}(\hat{\theta}_2) = \frac{1}{T}\left[d' \hat{S}^{-1} d\right]^{-1}$$

Goodness of Fit

- The GMM criterion function can be used to test the null hypothesis that the model is correctly specified.
- The test statistic is

$$TQ_T(\hat{\theta}) \xrightarrow{d} \chi^2_{L-P}$$

Example, OLS using GMM

- Consider the simple linear regression model

$$y = X\beta + \epsilon$$

The OLS conditions are

$$\begin{aligned} \mathbb{E}[X'\epsilon] &= 0 \\ \mathbb{E}[\epsilon] &= 0 \end{aligned}$$

Replace

$$g(w_i, \theta) = \begin{bmatrix} X_i'\epsilon_i \\ \epsilon_i \end{bmatrix}$$

Example, OLS using GMM (cont.)

Then the GMM estimator in the first step is

$$
\begin{aligned}
\hat{\beta}_1 &= \arg\min_{\beta} \left[N^{-1}\sum_{i=1}^{N}\begin{bmatrix} X_i'\epsilon_i \\ \epsilon_i \end{bmatrix}\right]' I \left[N^{-1}\sum_{i=1}^{N}\begin{bmatrix} X_i'\epsilon_i \\ \epsilon_i \end{bmatrix}\right] \\
&= \arg\min_{\beta} \left[N^{-1}\sum_{i=1}^{N}\begin{bmatrix} X_i'(y_i - X_i\beta) \\ (y_i - X_i\beta) \end{bmatrix}\right]' I \left[N^{-1}\sum_{i=1}^{N}\begin{bmatrix} X_i'(y_i - X_i\beta) \\ (y_i - X_i\beta) \end{bmatrix}\right]
\end{aligned}
$$

Example, OLS using GMM (cont.)

Second step, given $\hat{\beta}_1$ compute the covariance matrix of the moments

$$\hat{S} = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} X_i'(y_i - X_i\hat{\beta}_1) \\ (y_i - X_i\hat{\beta}_1) \end{bmatrix} \begin{bmatrix} X_i'(y_i - X_i\hat{\beta}_1) \\ (y_i - X_i\hat{\beta}_1) \end{bmatrix}'$$

Then the GMM estimator is

$$\hat{\beta}_2 = \arg\min_{\beta} \left[N^{-1} \sum_{i=1}^{N} \begin{bmatrix} X_i'(y_i - X_i\beta) \\ (y_i - X_i\beta) \end{bmatrix} \right]' \hat{S}^{-1} \left[N^{-1} \sum_{i=1}^{N} \begin{bmatrix} X_i'(y_i - X_i\beta) \\ (y_i - X_i\beta) \end{bmatrix} \right]$$

with covariance matrix

$$\hat{V}(\hat{\beta}_2) = \frac{1}{N} \left[d'\hat{S}^{-1} d \right]^{-1}$$

GMM in practice

- In many applications, the covariance matrix of the moments is numerically singular.
- How to solve it?
    i. Use only 1 step.
    ii. Add small noise to the variance matrix.
    iii. Use a "generalized" inverse.