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:

$$\mu_1 = E[X] = g_1(\theta_1, \ldots, \theta_k)$$
$$\mu_2 = E[X^2] = g_2(\theta_1, \ldots, \theta_k)$$
$$\vdots$$
$$\mu_k = E[X^k] = g_k(\theta_1, \ldots, \theta_k)$$

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

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

Probability Concepts for GMM

CLT, HAC, and Probability Limits

Central Limit Theorem (CLT)

- **Key result**: For i.i.d. data with $E[g_t] = 0$ and $\mathrm{Var}(g_t) = \Sigma$, as $T \to \infty$:

$$\sqrt{T}\,\bar{g}_T \xrightarrow{d} N(0, \Sigma), \quad \text{where } \bar{g}_T = \frac{1}{T}\sum_{t=1}^{T} g_t$$

- **General CLT**: For dependent data (e.g., time series), if g_t is stationary and weakly dependent:

$$\sqrt{T}\,\bar{g}_T \xrightarrow{d} N(0, S), \quad S = \sum_{j=-\infty}^{\infty} E[g_t g_{t-j}']$$

  where S is the **long-run variance**.

- Critical for deriving asymptotic distributions in GMM.

Heteroskedasticity and Autocorrelation (HAC)

- **Problem**: In time series/finance data, moments often exhibit:
  - Heteroskedasticity (varying variance)
  - Autocorrelation ($E[g_t g'_{t-j}] \neq 0$)
- **HAC estimator**: Newey-West (1987) kernel estimator:

$$\hat{S} = \hat{\Gamma}_0 + \sum_{j=1}^{m} \left(1 - \frac{j}{m+1}\right)(\hat{\Gamma}_j + \hat{\Gamma}'_j)$$

  where $\hat{\Gamma}_j = \frac{1}{T} \sum_{t=j+1}^{T} g_t g'_{t-j}$.

- **Truncation parameter**: m (e.g., $m = \lfloor 4(T/100)^{2/9} \rfloor$).
- Ensures $\hat{S}$ consistently estimates S for GMM standard errors.

Probability Limit (plim)

- **Definition**: $\hat{\theta}_T \xrightarrow{p} \theta_0$ if:

$$\forall \epsilon > 0, \quad \lim_{T \to \infty} P(||\hat{\theta}_T - \theta_0|| > \epsilon) = 0$$

- **Key properties**:
  i. plim of sample mean: $\text{plim} \frac{1}{T} \sum_{t=1}^{T} g_t = E[g_t]$
  ii. Slutsky's theorem: If $\text{plim}\, \hat{\theta} = \theta_0$ and h is continuous,

$$\text{plim}\, h(\hat{\theta}) = h(\theta_0)$$

- **Critical for GMM**: Weighting matrix $W_T \xrightarrow{p} W$, and consistency of $\hat{\theta}$.

Formal Derivation of GMM

Based on Hansen (1982)

Moment Conditions

- **Population moments**: True parameter θ_0 satisfies:

$$E[g_t(\theta_0)] = 0$$

  where $g_t(\theta)$ is a $m \times 1$ vector of moment conditions.

- **Sample analog** (average over T observations):

$$g_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} g_t(\theta)$$

GMM Objective Function

- **Weighting matrix**: Choose W_T (positive definite, $m \times m$).
- **Quadratic form** to minimize:

$$Q_T(\theta) = g_T(\theta)' W_T g_T(\theta)$$

First-Order Condition (FOC)

- **Derivative** of $Q_T(\theta)$ w.r.t. θ (a $p \times 1$ vector):

$$\frac{\partial Q_T}{\partial \theta} = 2G_T(\theta)'W_T g_T(\theta) = 0$$

where $G_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial g_t(\theta)}{\partial \theta}$ ($m \times p$ Jacobian).

- **FOC** defines the estimator $\hat{\theta}$:

$$G_T(\hat{\theta})'W_T g_T(\hat{\theta}) = 0$$

Asymptotic Distribution

- **Taylor expansion** of $g_T(\hat{\theta})$ around θ_0:

$$g_T(\hat{\theta}) \approx g_T(\theta_0) + G_T(\theta_0)(\hat{\theta} - \theta_0)$$

- Substitute into FOC:

$$G_T(\hat{\theta})' W_T \left[g_T(\theta_0) + G_T(\theta_0)(\hat{\theta} - \theta_0) \right] = 0$$

Asymptotic Distribution (Cont.)

- Rearrange for $\hat{\theta}$ (as $T \to \infty$):

$$\sqrt{T}(\hat{\theta} - \theta_0) \approx -\left(G_T' W_T G_T\right)^{-1} G_T' W_T \sqrt{T} g_T(\theta_0)$$

where $G_T = E\left[\frac{\partial g_t(\theta_0)}{\partial \theta}\right]$.

Central Limit Theorem (CLT)

- Under regularity conditions:

$$\sqrt{T}\, g_T(\theta_0) \xrightarrow{d} N(0, S)$$

where $S = \lim_{T\to\infty} \mathrm{Var}\left(\sqrt{T} g_T(\theta_0)\right)$ (long-run variance).

Asymptotic Variance

- Combine CLT with expansion:

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0,\ (G'WG)^{-1}G'WSWG(G'WG)^{-1}\right)$$

  - $G = \mathbb{E}\left[\frac{\partial g_t(\theta_0)}{\partial \theta}\right]$
  - $W = \text{plim}\ W_T$
  - Recall the **plim** (probability limit) operator measures convergence in probability.

Efficient GMM

- **Optimal weighting matrix**: $W = S^{-1}$ minimizes asymptotic variance.
- **Asymptotic variance** becomes:

$$\mathrm{Avar}(\hat{\theta}) = \left(G' S^{-1} G\right)^{-1}$$

Standard Errors (Detailed)

- **Estimated asymptotic variance**:

$$\widehat{\text{Avar}}(\hat{\theta}) = \frac{1}{T}\left(\hat{G}'\hat{W}\hat{G}\right)^{-1}\hat{G}'\hat{W}\hat{S}\hat{W}\hat{G}\left(\hat{G}'\hat{W}\hat{G}\right)^{-1}$$

  - $\hat{G} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial g_t(\hat{\theta})}{\partial \theta}$
  - $\hat{S}$: HAC estimator (e.g., Newey-West)
  - $\hat{W} = \hat{S}^{-1}$ for efficient GMM

- **Standard errors**: Square roots of diagonal elements divided by T.

Two-Step GMM Procedure

1. **First step**: Estimate $\hat{\theta}^{(1)}$ using $W_T = I$ (identity matrix).
2. **Compute residuals**: $g_t(\hat{\theta}^{(1)})$ to estimate $\hat{S}$.
3. **Second step**: Re-estimate $\hat{\theta}$ using $W_T = \hat{S}^{-1}$.

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?
  1. Use only 1 step.
  2. Add small noise to the variance matrix.
  3. Use a "generalized" inverse.

References

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4), 1029-1054.