Lesson 3: GMM Estimation

1

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.

$$f_X(x; heta_1,\ldots, heta_k)$$

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

$$egin{aligned} \mu_1 &= E[X] = g_1(heta_1,\ldots, heta_k) \ \mu_2 &= E[X^2] = g_2(heta_1,\ldots, heta_k) \ dots \ \mu_k &= E[X^k] = g_k(heta_1,\ldots, heta_k) \end{aligned}$$

The method of moments (cont.)

• Estimate the population moment with the sample moment

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

• Solve the system of equations

$$egin{aligned} \hat{\mu}_1 &= g_1(\hat{ heta}_1,\ldots,\hat{ heta}_k) \ \hat{\mu}_2 &= g_2(\hat{ heta}_1,\ldots,\hat{ heta}_k) \ dots \ \hat{\mu}_k &= g_k(\hat{ heta}_1,\ldots,\hat{ heta}_k) \end{aligned}$$

Example, normal distribution

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

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

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

• 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 imes 1) i.i.d. vector of random variables for observation i.
- heta is a (P imes 1) vector of unknown coefficients (parameters).
- $g(w_i, heta)$ is a (L imes 1) vector of functions $g:\mathbb{R}^M imes\mathbb{R}^P o\mathbb{R}^L$ $L\ge P$
- 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).

$$\mathbb{E}[g(w_i, heta_0)]=0$$

Which in the sample can be written as

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

We want to choose $\hat{ heta}$ such that $N^{-1}\sum_{i=1}^N g(w_i,\hat{ heta})$ 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.

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

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.

$$\hat{ heta_1} = rg\min_{ heta} \left[g_T(heta)
ight]' W \Big[g_T(heta) \Big]$$

First Order Condition

$$rac{\partial g_T(heta)}{\partial heta} W g_T(heta) = a g_T(heta) = 0$$

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

$$\hat{ heta_2} = rg\min_{ heta} \left[g_T(heta)
ight]' \hat{S}^{-1} \Big[g_T(heta) \Big]$$

where \hat{S} is the sample covariance of the moments given $\hat{\theta_1}$, is consistent and asymptotically normal. Define

$$d = rac{\partial g_T(heta)}{\partial heta}$$

Then the asymptotic variance of $\hat{ heta_2}$ is

$$\hat{V}(\hat{ heta_2}) = rac{1}{T} \Big[d' \hat{S}^{-1} d \Big]^{-1}$$

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 $T o\infty$:

$$\sqrt{T}\,ar{g}_T \stackrel{d}{
ightarrow} N(0,\Sigma), \quad ext{where } ar{g}_T = rac{1}{T}\sum_{t=1}^T g_t$$

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

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

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)
 - \circ Autocorrelation ($E[g_tg_{t-j}']
 eq 0$)
- HAC estimator: Newey-West (1987) kernel estimator:

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

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

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

Probability Limit (plim)

• Definition:
$$\hat{ heta}_T \stackrel{p}{
ightarrow} heta_0$$
 if:

$$orall \epsilon > 0, \quad \lim_{T o \infty} P(|| \hat{ heta}_T - heta_0 || > \epsilon) = 0$$

• Key properties:

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

$$\operatorname{plim} h(\hat{ heta}) = h(heta_0)$$

• Critical for GMM: Weighting matrix $W_T \stackrel{p}{\to} W$, and consistency of $\hat{\theta}$.

Formal Derivation of GMM

Based on Hansen (1982)

Moment Conditions

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

 $E[g_t(heta_0)]=0$

where $g_t(heta)$ is a m imes 1 vector of moment conditions.

• Sample analog (average over T observations):

$$g_T(heta) = rac{1}{T}\sum_{t=1}^T g_t(heta)$$

GMM Objective Function

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

$$Q_T(heta) = g_T(heta)' W_T g_T(heta)$$

First-Order Condition (FOC)

• Derivative of $Q_T(heta)$ w.r.t. heta (a p imes 1 vector):

$$rac{\partial Q_T}{\partial heta} = 2G_T(heta)' W_T g_T(heta) = 0$$

where $G_T(heta) = rac{1}{T} \sum_{t=1}^T rac{\partial g_t(heta)}{\partial heta}$ ($m imes p$ Jacobian).

• FOC defines the estimator $\hat{\theta}$:

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

Asymptotic Distribution

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

$$g_T(\hat{ heta}) pprox g_T(heta_0) + G_T(heta_0)(\hat{ heta} - heta_0)$$

• Substitute into FOC:

$$G_T(\hat{ heta})' W_T \left[g_T(heta_0) + G_T(heta_0) (\hat{ heta} - heta_0)
ight] = 0$$

Asymptotic Distribution (Cont.)

- Rearrange for
$$\hat{ heta}$$
 (as $T o \infty$):

$$\sqrt{T}(\hat{ heta} - heta_0) pprox - \left(G'_T W_T G_T
ight)^{-1} G'_T W_T \sqrt{T} g_T(heta_0)$$

where $G_T = E\left[rac{\partial g_t(heta_0)}{\partial heta}
ight].$

Central Limit Theorem (CLT)

• Under regularity conditions:

$$\sqrt{T}\,g_T(heta_0)\stackrel{d}{ o} N(0,S)$$
 where $S=\lim_{T o\infty}\mathrm{Var}\left(\sqrt{T}g_T(heta_0)
ight)$ (long-run variance).

Asymptotic Variance

• Combine CLT with expansion:

$$egin{aligned} &\sqrt{T}(\hat{ heta}- heta_0) \stackrel{d}{ o} N\left(0,\ (G'WG)^{-1}G'WSWG(G'WG)^{-1}
ight) \ &\circ \ G = \mathbb{E}\left[rac{\partial g_t(heta_0)}{\partial heta}
ight] \ &\circ \ W = \mathrm{plim}\ W_T \end{aligned}$$

• Recall the **plim** (probability limit) operator measures convergence in probability.

Efficient GMM

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

$$\operatorname{Avar}(\hat{ heta}) = \left(G'S^{-1}G
ight)^{-1}$$

Standard Errors (Detailed)

• Estimated asymptotic variance:

$$\widehat{\operatorname{Avar}}(\hat{ heta}) = rac{1}{T} \left(\hat{G}' \hat{W} \hat{G}
ight)^{-1} \hat{G}' \hat{W} \hat{S} \hat{W} \hat{G} \left(\hat{G}' \hat{W} \hat{G}
ight)^{-1}$$

 $= rac{1}{T} \sum_{t=1}^{T} rac{\partial g_t(\hat{ heta})}{\Delta t}$

$$\circ \hat{G} = rac{1}{T} \sum_{t=1}^{T} rac{\partial g_t(\theta)}{\partial heta}$$

$$\circ \hat{S}$$
: HAC estimator (e.g., Newey-West)

$$\circ \; \hat{W} = \hat{S}^{-1}$$
 for efficient GMM

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

Two-Step GMM Procedure

- 1. First step: Estimate $\hat{ heta}^{(1)}$ 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 $W_T = \hat{S}^{-1}$.

Overidentification Test (J-Test)

• Test statistic:

$$J = T \cdot Q_T(\hat{ heta}) \stackrel{d}{
ightarrow} \chi^2_{m-p}$$

- Tests whether all moments are jointly zero.
- \circ Degrees of freedom: #moments #parameters.

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

$$TQ_T(\hat{ heta}) \stackrel{d}{
ightarrow} \chi^2_{L-P}$$

Example, OLS using GMM

• Consider the simple linear regression model

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

The OLS conditions are

$$\mathbb{E}[X'\epsilon] = 0 \ \mathbb{E}[\epsilon] = 0$$

Replace

$$g(w_i, heta) = egin{bmatrix} X_i' \epsilon_i \ \epsilon_i \end{bmatrix}$$

Example, OLS using GMM (cont.)

Then the GMM estimator in the first step is

$$egin{aligned} \hat{eta}_1 &= rg\min_eta \left[N^{-1} \sum_{i=1}^N egin{bmatrix} X_i' \epsilon_i \ \epsilon_i \end{bmatrix}
ight]' I igg[N^{-1} \sum_{i=1}^N igg[X_i' (y_i - X_i eta) \ (y_i - X_i eta) \end{bmatrix} igg]' I igg[N^{-1} \sum_{i=1}^N igg[X_i' (y_i - X_i eta) \ (y_i - X_i eta) \end{bmatrix} igg]' I igg[N^{-1} \sum_{i=1}^N igg[X_i' (y_i - X_i eta) \ (y_i - X_i eta) \end{bmatrix} igg]' I igg[N^{-1} \sum_{i=1}^N igg[X_i' (y_i - X_i eta) \ (y_i - X_i eta) \end{bmatrix} igg]' I igg[N^{-1} \sum_{i=1}^N igg[X_i' (y_i - X_i eta) \ (y_i - X_i eta) \end{bmatrix} igg]' I igg[N^{-1} \sum_{i=1}^N igg[X_i' (y_i - X_i eta) \ (y_i - X_i eta) \end{bmatrix} igg]' I igg[N^{-1} \sum_{i=1}^N igg[X_i' (y_i - X_i eta) \ (y_i - X_i eta) \end{bmatrix} igg]' I igg[N^{-1} \sum_{i=1}^N igg[X_i' (y_i - X_i eta) \ (y_i - X_i eta) \end{bmatrix} igg] igg] H igg] H igg] H igg[X_i' (y_i - X_i eta) \ (y_i - X_i eta) \end{bmatrix} igg] H igg[X_i' (y_i - X_i eta) \ (y_i - X_i eta) \end{bmatrix} igg] H igg[X_i' (y_i - X_i eta) \ (y_i - X_i eta) \ (y_i - X_i eta) \end{bmatrix} igg]$$

Example, OLS using GMM (cont.)

Second step, given $\hat{eta_1}$ compute the covariance matrix of the moments

$$\hat{S} = rac{1}{N}\sum_{i=1}^N egin{bmatrix} X_i'(y_i - X_i \hat{eta_1}) \ (y_i - X_i \hat{eta_1}) \end{bmatrix} egin{bmatrix} X_i'(y_i - X_i \hat{eta_1}) \ (y_i - X_i \hat{eta_1}) \end{bmatrix}'$$

Then the GMM estimator is

$$\hat{eta_2} = rgmin_eta \left[N^{-1} \sum_{i=1}^N iggl[rac{X_i'(y_i - X_ieta)}{(y_i - X_ieta)}
ight]
ight]' \hat{S}^{-1} \Big[N^{-1} \sum_{i=1}^N iggl[rac{X_i'(y_i - X_ieta)}{(y_i - X_ieta)} \Big] \Big]$$

with covariance matrix

$$\hat{V}(\hat{eta_2}) = rac{1}{N} \Big[d' \hat{S}^{-1} d \Big]^{-1}$$

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.

References

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