Lecture 10: The GMM Estimator - Part II

Raul Riva

FGV EPGE

November, 2025

Intro

Recap from Last Class

Last time we covered:

- The GMM framework: moment conditions $\mathbb{E}[\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)] = 0$
- The GMM estimator: $\hat{\boldsymbol{\theta}}(\mathbf{W}_T) = \arg\min_{\boldsymbol{\theta}} Q_T(\boldsymbol{\theta}, \mathbf{W}_T)$ for a general convergence sequence $\{\mathbf{W}_T\}$;
- Asymptotic distribution: $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_{\mathbf{0}}) \xrightarrow{d} N(0, \mathbf{V})$

where the asymptotic variance is:

$$\mathbf{V} \equiv (\mathbf{D}'\mathbf{W}\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}\mathbf{S}\mathbf{W}\mathbf{D}(\mathbf{D}'\mathbf{W}\mathbf{D})^{-1}$$

with $\mathbf{D} = \frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\mathbf{0}}} = \frac{\partial \mathbb{E}[\mathbf{h}(\boldsymbol{\theta}, \mathbf{w}_t)]}{\partial \boldsymbol{\theta}'}$ and $\mathbf{S} = \sum_{j=-\infty}^{\infty} \mathbb{E}[\mathbf{h}(\boldsymbol{\theta}_{\mathbf{0}}, \mathbf{w}_t)\mathbf{h}(\boldsymbol{\theta}_{\mathbf{0}}, \mathbf{w}_{t-j})']$

Flight Plan

- Optimal weighting matrix: $\mathbf{W}^* = \mathbf{S}^{-1}$
- This gives the efficient variance: $\mathbf{V}^* = (\mathbf{D}'\mathbf{S}^{-1}\mathbf{D})^{-1}$
- We showed that $\mathbf{V} - \mathbf{V}^* \geq 0$ (positive semidefinite)

Flight Plan

- Optimal weighting matrix: $\mathbf{W}^* = \mathbf{S}^{-1}$
- This gives the efficient variance: $\mathbf{V}^* = (\mathbf{D}'\mathbf{S}^{-1}\mathbf{D})^{-1}$
- We showed that $\mathbf{V} - \mathbf{V}^* \geq 0$ (positive semidefinite)

The big problem: we do not observe $\mathbf{S}$ + it depends on parameters we want to estimate!

Flight Plan

- Optimal weighting matrix: $\mathbf{W}^* = \mathbf{S}^{-1}$
- This gives the efficient variance: $\mathbf{V}^* = (\mathbf{D}'\mathbf{S}^{-1}\mathbf{D})^{-1}$
- We showed that $\mathbf{V} - \mathbf{V}^* \geq 0$ (positive semidefinite)

The big problem: we do not observe $\mathbf{S}$ + it depends on parameters we want to estimate!

Today:

1. Estimating $\mathbf{S}$ and implementing efficient GMM
2. Testing overidentifying restrictions (J-test)
3. Using multiple instruments in conditional moment models
4. Classical estimators as special cases of GMM

Questions?

Estimating S and Efficient GMM

The Feasible GMM Estimator

- We need a consistent estimator $\hat{\mathbf{S}}_T$ such that $\hat{\mathbf{S}}_T \xrightarrow{p} \mathbf{S}$;
- Natural idea: use the sample analog with estimated residuals;
- Very important: consistency does *not* depend on using the optimal matrix!

The Feasible GMM Estimator

- We need a consistent estimator $\hat{\mathbf{S}}_T$ such that $\hat{\mathbf{S}}_T \xrightarrow{p} \mathbf{S}$;
- Natural idea: use the sample analog with estimated residuals;
- Very important: consistency does *not* depend on using the optimal matrix!
- Define $\hat{\mathbf{h}}_t \equiv \mathbf{h}(\hat{\boldsymbol{\theta}}_T, \mathbf{w}_t)$ as the residuals from some initial estimator, e.g., using the identity matrix;
- For the **i.i.d. or MDS case**, a natural estimator is:

$$\hat{\mathbf{S}}_T^{iid} = \frac{1}{T}\sum_{t=1}^{T} \hat{\mathbf{h}}_t \hat{\mathbf{h}}_t'$$

The Feasible GMM Estimator

- We need a consistent estimator $\hat{\mathbf{S}}_T$ such that $\hat{\mathbf{S}}_T \xrightarrow{p} \mathbf{S}$;
- Natural idea: use the sample analog with estimated residuals;
- Very important: consistency does *not* depend on using the optimal matrix!
- Define $\hat{\mathbf{h}}_t \equiv \mathbf{h}(\hat{\boldsymbol{\theta}}_T, \mathbf{w}_t)$ as the residuals from some initial estimator, e.g., using the identity matrix;
- For the **i.i.d. or MDS case**, a natural estimator is:

$$\hat{\mathbf{S}}_T^{iid} = \frac{1}{T}\sum_{t=1}^{T} \hat{\mathbf{h}}_t \hat{\mathbf{h}}_t'$$

- Under standard regularity conditions, $\hat{\mathbf{S}}_T^{iid} \xrightarrow{p} \mathbf{S}$
- But what if the data is **not** i.i.d.?

HAC Estimation for Dependent Data

- For dependent data, deploy a **Heteroskedasticity and Autocorrelation Consistent (HAC)** estimator
- The Newey-West estimator is:

$$\hat{\mathbf{S}}_T^{NW} = \hat{\mathbf{\Gamma}}_0 + \sum_{j=1}^{q} \omega_j \left(\hat{\mathbf{\Gamma}}_j + \hat{\mathbf{\Gamma}}_j' \right)$$

where $\hat{\mathbf{\Gamma}}_j = \frac{1}{T} \sum_{t=j+1}^{T} \hat{\mathbf{h}}_t \hat{\mathbf{h}}_{t-j}'$ is the sample autocovariance at lag j

- The Bartlett kernel uses weights: $\omega_j = 1 - \frac{j}{q+1}$

HAC Estimation for Dependent Data

- For dependent data, deploy a **Heteroskedasticity and Autocorrelation Consistent (HAC)** estimator
- The Newey-West estimator is:

$$\hat{\mathbf{S}}_T^{NW} = \hat{\mathbf{\Gamma}}_0 + \sum_{j=1}^{q} \omega_j \left(\hat{\mathbf{\Gamma}}_j + \hat{\mathbf{\Gamma}}_j'\right)$$

where $\hat{\mathbf{\Gamma}}_j = \frac{1}{T}\sum_{t=j+1}^{T} \hat{\mathbf{h}}_t \hat{\mathbf{h}}_{t-j}'$ is the sample autocovariance at lag j

- The Bartlett kernel uses weights: $\omega_j = 1 - \frac{j}{q+1}$
- Choice of bandwidth q: typically $q \approx T^{1/4}$ (e.g., Andrews, 1991)
- **Intuition**: place less weight distant lags;

The Two-Step GMM Procedure

A practical algorithm to obtain the efficient GMM estimator:

The Two-Step GMM Procedure

A practical algorithm to obtain the efficient GMM estimator:

Step 1: Choose an arbitrary positive definite weighting matrix $\mathbf{W}_T^{(1)}$ (e.g., identity matrix $\mathbf{I}_r$)

- Minimize $Q_T(\boldsymbol{\theta}, \mathbf{W}_T^{(1)})$ to obtain $\hat{\boldsymbol{\theta}}^{(1)}$

The Two-Step GMM Procedure

A practical algorithm to obtain the efficient GMM estimator:

Step 1: Choose an arbitrary positive definite weighting matrix $\mathbf{W}_T^{(1)}$ (e.g., identity matrix $\mathbf{I}_r$)

- Minimize $Q_T(\boldsymbol{\theta}, \mathbf{W}_T^{(1)})$ to obtain $\hat{\boldsymbol{\theta}}^{(1)}$

Step 2: Compute $\hat{\mathbf{S}}_T$ using residuals $\hat{\mathbf{h}}_t = \mathbf{h}(\hat{\boldsymbol{\theta}}^{(1)}, \mathbf{w}_t)$

- Use $\hat{\mathbf{S}}_T^{iid}$ or $\hat{\mathbf{S}}_T^{NW}$ depending on context

The Two-Step GMM Procedure

A practical algorithm to obtain the efficient GMM estimator:

Step 1: Choose an arbitrary positive definite weighting matrix $\mathbf{W}_T^{(1)}$ (e.g., identity matrix $\mathbf{I}_r$)

- Minimize $Q_T(\boldsymbol{\theta}, \mathbf{W}_T^{(1)})$ to obtain $\hat{\boldsymbol{\theta}}^{(1)}$

Step 2: Compute $\hat{\mathbf{S}}_T$ using residuals $\hat{\mathbf{h}}_t = \mathbf{h}(\hat{\boldsymbol{\theta}}^{(1)}, \mathbf{w}_t)$

- Use $\hat{\mathbf{S}}_T^{iid}$ or $\hat{\mathbf{S}}_T^{NW}$ depending on context

Step 3: Set $\mathbf{W}_T^{(2)} = \hat{\mathbf{S}}_T^{-1}$ and minimize $Q_T(\boldsymbol{\theta}, \mathbf{W}_T^{(2)})$ to obtain $\hat{\boldsymbol{\theta}}^{(2)}$

The Two-Step GMM Procedure

A practical algorithm to obtain the efficient GMM estimator:

Step 1: Choose an arbitrary positive definite weighting matrix $\mathbf{W}_T^{(1)}$ (e.g., identity matrix $\mathbf{I}_r$)

- Minimize $Q_T(\boldsymbol{\theta}, \mathbf{W}_T^{(1)})$ to obtain $\hat{\boldsymbol{\theta}}^{(1)}$

Step 2: Compute $\hat{\mathbf{S}}_T$ using residuals $\hat{\mathbf{h}}_t = \mathbf{h}(\hat{\boldsymbol{\theta}}^{(1)}, \mathbf{w}_t)$

- Use $\hat{\mathbf{S}}_T^{iid}$ or $\hat{\mathbf{S}}_T^{NW}$ depending on context

Step 3: Set $\mathbf{W}_T^{(2)} = \hat{\mathbf{S}}_T^{-1}$ and minimize $Q_T(\boldsymbol{\theta}, \mathbf{W}_T^{(2)})$ to obtain $\hat{\boldsymbol{\theta}}^{(2)}$

Result: $\hat{\boldsymbol{\theta}}^{(2)}$ is asymptotically efficient!

Why Does Two-Step GMM Work?

- Key insight: consistency of $\widehat{\boldsymbol{\theta}}^{(1)}$ does **not** require optimal weighting
- Any positive definite $\mathbf{W}_T^{(1)}$ gives a consistent estimator (we showed this last class!)

Why Does Two-Step GMM Work?

- Key insight: consistency of $\hat{\boldsymbol{\theta}}^{(1)}$ does **not** require optimal weighting
- Any positive definite $\mathbf{W}_T^{(1)}$ gives a consistent estimator (we showed this last class!)
- Since $\hat{\boldsymbol{\theta}}^{(1)} \xrightarrow{p} \boldsymbol{\theta}_0$, we have:

$$\hat{\mathbf{h}}_t = \mathbf{h}(\hat{\boldsymbol{\theta}}^{(1)}, \mathbf{w}_t) \approx \mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)$$

Why Does Two-Step GMM Work?

- Key insight: consistency of $\hat{\boldsymbol{\theta}}^{(1)}$ does **not** require optimal weighting
- Any positive definite $\mathbf{W}_T^{(1)}$ gives a consistent estimator (we showed this last class!)
- Since $\hat{\boldsymbol{\theta}}^{(1)} \xrightarrow{p} \boldsymbol{\theta}_0$, we have:

$$\hat{\mathbf{h}}_t = \mathbf{h}(\hat{\boldsymbol{\theta}}^{(1)}, \mathbf{w}_t) \approx \mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)$$

- Therefore, $\hat{\mathbf{S}}_T \xrightarrow{p} \mathbf{S}$ even though we used “suboptimal” weights in Step 1
- Slutsky’s theorem ensures that using $\hat{\mathbf{S}}_T^{-1}$ in Step 2 gives the efficient estimator
- Both $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ are consistent, but $\hat{\boldsymbol{\theta}}^{(2)}$ has smaller asymptotic variance!

Iterated GMM

An alternative procedure is to keep iterating until convergence:

Iterated GMM

An alternative procedure is to keep iterating until convergence:

Algorithm:

- Start with $\hat{\boldsymbol{\theta}}^{(0)}$ (or use identity weighting in Step 0)
- For $k = 1, 2, 3, ...$:
  1. Compute $\hat{\mathbf{S}}_T^{(k)}$ using residuals from $\hat{\boldsymbol{\theta}}^{(k-1)}$
  2. Update: $\hat{\boldsymbol{\theta}}^{(k)} = \arg\min_{\boldsymbol{\theta}} Q_T(\boldsymbol{\theta}, (\hat{\mathbf{S}}_T^{(k)})^{-1})$
  3. Stop when $\|\hat{\boldsymbol{\theta}}^{(k)} - \hat{\boldsymbol{\theta}}^{(k-1)}\| < \epsilon$, where ϵ is specified by the user;

Iterated GMM

An alternative procedure is to keep iterating until convergence:

Algorithm:

- Start with $\hat{\boldsymbol{\theta}}^{(0)}$ (or use identity weighting in Step 0)
- For $k = 1, 2, 3, ...$:
    1. Compute $\hat{\mathbf{S}}_T^{(k)}$ using residuals from $\hat{\boldsymbol{\theta}}^{(k-1)}$
    2. Update: $\hat{\boldsymbol{\theta}}^{(k)} = \arg\min_{\boldsymbol{\theta}} Q_T(\boldsymbol{\theta}, (\hat{\mathbf{S}}_T^{(k)})^{-1})$
    3. Stop when $\|\hat{\boldsymbol{\theta}}^{(k)} - \hat{\boldsymbol{\theta}}^{(k-1)}\| < \epsilon$, where ϵ is specified by the user;
- The iterated GMM has the **same asymptotic distribution** as two-step GMM;
- Potential finite-sample improvements, but results are mixed in the literature;
- Computationally more expensive;

Continuously-Updated GMM (CUE)

- Alternative approach: update $\mathbf{W}_T$ at **each function evaluation** during optimization
- The CUE estimator is:

$$\hat{\boldsymbol{\theta}}^{CUE} = \arg\min_{\boldsymbol{\theta}} \hat{\mathbf{m}}_T(\boldsymbol{\theta})' \hat{\mathbf{S}}_T(\boldsymbol{\theta})^{-1} \hat{\mathbf{m}}_T(\boldsymbol{\theta})$$

where $\hat{\mathbf{S}}_T(\boldsymbol{\theta})$ is computed using the current value of $\boldsymbol{\theta}$

Continuously-Updated GMM (CUE)

- Alternative approach: update $\mathbf{W}_T$ at **each function evaluation** during optimization
- The CUE estimator is:

$$\hat{\boldsymbol{\theta}}^{CUE} = \arg\min_{\boldsymbol{\theta}} \hat{\mathbf{m}}_T(\boldsymbol{\theta})' \hat{\mathbf{S}}_T(\boldsymbol{\theta})^{-1} \hat{\mathbf{m}}_T(\boldsymbol{\theta})$$

where $\hat{\mathbf{S}}_T(\boldsymbol{\theta})$ is computed using the current value of $\boldsymbol{\theta}$

- Computationally intensive: need to compute and invert $\hat{\mathbf{S}}_T$ at each function **evaluation**;
- Better finite-sample properties and less finite-sample bias (Hansen, Heaton, Yaron, 1996)
- Same asymptotic distribution as two-step and iterated GMM;
- Major recent breakthrough by Moreira, Newey, and Sharifvaghefi (202x);

Chamberlain's Semiparametric Efficiency Bound

A fundamental theoretical result (Chamberlain, 1987):

- The efficient GMM estimator with $\mathbf{W}^* = \mathbf{S}^{-1}$ achieves the **semiparametric efficiency bound**;
- This means: among **all** estimators that only use the moment conditions $\mathbb{E}[\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)] = 0$, efficient GMM has the smallest asymptotic variance;

Chamberlain's Semiparametric Efficiency Bound

A fundamental theoretical result (Chamberlain, 1987):

- The efficient GMM estimator with $\mathbf{W}^* = \mathbf{S}^{-1}$ achieves the **semiparametric efficiency bound**;
- This means: among **all** estimators that only use the moment conditions $\mathbb{E}[\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)] = 0$, efficient GMM has the smallest asymptotic variance;

Intuition:

- GMM does not require knowing the full distribution of the data, only moments;
- You can't do better without making **stronger assumptions** (e.g., specifying the entire likelihood);
- If you're willing to assume more (parametric model), MLE might beat GMM;
- But if you only trust your moment conditions, efficient GMM is optimal!

Questions?

The J-Test for Overidentification

Testing the Model

Recall that GMM allows $r > a$ (more moments than parameters):

- This is called **overidentification**
- Natural question: are the extra moment conditions consistent with the data?

Testing the Model

Recall that GMM allows $r > a$ (more moments than parameters):

- This is called **overidentification**
- Natural question: are the extra moment conditions consistent with the data?
- If the model is correctly specified, all r moment conditions should be "close to zero"
- But we can only set a of them exactly to zero (through choice of $\hat{\boldsymbol{\theta}}_T$)
- The remaining $r - a$ restrictions are "overidentifying"

Testing the Model

Recall that GMM allows $r > a$ (more moments than parameters):

- This is called **overidentification**
- Natural question: are the extra moment conditions consistent with the data?
- If the model is correctly specified, all r moment conditions should be "close to zero"
- But we can only set a of them exactly to zero (through choice of $\hat{\boldsymbol{\theta}}_T$)
- The remaining $r - a$ restrictions are "overidentifying"

Idea: Test whether $Q_T(\hat{\boldsymbol{\theta}}_T, \mathbf{W}_T)$ is "too large"

- If the model is correct, this criterion should be small
- If the model is wrong, moment conditions will be incompatible and Q_T will be large

The J-Statistic

Hansen (1982) proposed the following test statistic:

$$J_T \equiv T \cdot Q_T(\hat{\boldsymbol{\theta}}_T^*, \hat{\mathbf{S}}_T^{-1}) = T \cdot \hat{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T^*)' \hat{\mathbf{S}}_T^{-1} \hat{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T^*)$$

where $\hat{\boldsymbol{\theta}}_T^*$ is the **efficient GMM estimator** (using $\mathbf{W}_T = \hat{\mathbf{S}}_T^{-1}$)

The J-Statistic

Hansen (1982) proposed the following test statistic:

$$J_T \equiv T \cdot Q_T(\hat{\boldsymbol{\theta}}_T^*, \hat{\mathbf{S}}_T^{-1}) = T \cdot \hat{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T^*)' \hat{\mathbf{S}}_T^{-1} \hat{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T^*)$$

where $\hat{\boldsymbol{\theta}}_T^*$ is the **efficient GMM estimator** (using $\mathbf{W}_T = \hat{\mathbf{S}}_T^{-1}$)

Under H_0: model is correctly specified, we have:

$$J_T \xrightarrow{d} \chi^2_{r-a}$$

The J-Statistic

Hansen (1982) proposed the following test statistic:

$$J_T \equiv T \cdot Q_T(\hat{\boldsymbol{\theta}}_T^*, \hat{\mathbf{S}}_T^{-1}) = T \cdot \hat{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T^*)' \hat{\mathbf{S}}_T^{-1} \hat{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T^*)$$

where $\hat{\boldsymbol{\theta}}_T^*$ is the **efficient GMM estimator** (using $\mathbf{W}_T = \hat{\mathbf{S}}_T^{-1}$)

Under H_0: model is correctly specified, we have:

$$J_T \xrightarrow{d} \chi^2_{r-a}$$

- Degrees of freedom = number of overidentifying restrictions
- **Reject** H_0 if $J_T > \chi^2_{r-a,1-\alpha}$ (right-tail test)
- **Important**: Must use efficient weighting matrix for the χ^2 result to hold!

Intuition Behind the J-Test

- We typically cannot make all $r \geq a$ moments exactly zero with only a parameters;
- J_T measures the weighted "distance from zero" of the sample moments;
- Large J_T suggests the moments are incompatible with each other;
- Small J_T suggests the model fits the data reasonably well;

Intuition Behind the J-Test

- We typically cannot make all $r \geq a$ moments exactly zero with only a parameters;
- J_T measures the weighted "distance from zero" of the sample moments;
- Large J_T suggests the moments are incompatible with each other;
- Small J_T suggests the model fits the data reasonably well;

Step-by-step procedure:

1. Estimate $\hat{\boldsymbol{\theta}}_T^*$ using efficient GMM (two-step or iterated GMM, for example);
2. Compute $\hat{\mathbf{S}}_T$ using residuals $\hat{\mathbf{h}}_t = \mathbf{h}(\hat{\boldsymbol{\theta}}_T^*, \mathbf{w}_t)$;
3. Compute the J-statistic:
$$J_T = T \cdot \hat{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T^*)' \hat{\mathbf{S}}_T^{-1} \hat{\mathbf{m}}_T(\hat{\boldsymbol{\theta}}_T^*)$$
4. Reject H_0 if $J_T > c_\alpha$ where $c_\alpha = \chi^2_{r-a,1-\alpha}$;

What If We Reject H_0?

Rejection of the J-test means: the moment conditions are incompatible with the data

What If We Reject H_0?

Rejection of the J-test means: the moment conditions are incompatible with the data

Possible interpretations:

1. The model is misspecified (wrong functional form, missing variables)
2. Some moment conditions are invalid
3. Finite-sample issues (the test can over-reject in small samples)

What If We Reject H_0?

Rejection of the J-test means: the moment conditions are incompatible with the data

Possible interpretations:

1. The model is misspecified (wrong functional form, missing variables)
2. Some moment conditions are invalid
3. Finite-sample issues (the test can over-reject in small samples)

Important caveats:

- The J-test has power against **many different alternatives**
- It is **not diagnostic**: doesn't tell you *which* moment condition is wrong
- You need economic theory or additional tests to diagnose the problem
- A failure to reject does **not** prove the model is correct!

Questions?

Conditional Moment Models and Instruments

Revisiting the Consumption Model

Recall from last class the consumption-based asset pricing model:

- Euler equation: $C_t^{-\gamma} = \beta E_t[(1 + r_{t+1})C_{t+1}^{-\gamma}]$
- We can rewrite this as a **conditional moment restriction**:

$$\mathbb{E}_t \left[\beta(1 + r_{t+1}) \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} - 1 \right] = 0$$

Revisiting the Consumption Model

Recall from last class the consumption-based asset pricing model:

- Euler equation: $C_t^{-\gamma} = \beta E_t[(1 + r_{t+1})C_{t+1}^{-\gamma}]$
- We can rewrite this as a **conditional moment restriction**:

$$\mathbb{E}_t \left[\beta(1 + r_{t+1}) \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} - 1 \right] = 0$$

Define the moment function:

$$\mathbf{h}_{t+1}(\boldsymbol{\theta}, \mathbf{w}_{t+1}) \equiv \beta(1 + r_{t+1}) \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} - 1, \qquad \mathbf{w}_{t+1} \equiv (r_{t+1}, C_{t+1}, C_t)$$

where $\boldsymbol{\theta} = (\beta, \gamma)'$

Revisiting the Consumption Model

Recall from last class the consumption-based asset pricing model:

- Euler equation: $C_t^{-\gamma} = \beta E_t[(1 + r_{t+1})C_{t+1}^{-\gamma}]$
- We can rewrite this as a **conditional moment restriction**:

$$\mathbb{E}_t \left[\beta(1 + r_{t+1}) \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} - 1 \right] = 0$$

Define the moment function:

$$\mathbf{h}_{t+1}(\boldsymbol{\theta}, \mathbf{w}_{t+1}) \equiv \beta(1 + r_{t+1}) \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} - 1, \qquad \mathbf{w}_{t+1} \equiv (r_{t+1}, C_{t+1}, C_t)$$

where $\boldsymbol{\theta} = (\beta, \gamma)'$

Key insight: The expectation is **conditional** on information at time t, $\mathcal{F}_t$

- This says: "given what we know at t, the expected value of $\mathbf{h}_{t+1}(\boldsymbol{\theta}_0, \mathbf{w}_{t+1})$ is zero"

From Conditional to Unconditional Moments

The law of iterated expectations tells us something powerful:

$$E_t[h_{t+1}(\boldsymbol{\theta}_0)] = 0 \implies \mathbb{E}[g(\mathcal{F}_t) \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$$

for **any** function g of variables in the information set $\mathcal{F}_t$!

From Conditional to Unconditional Moments

The law of iterated expectations tells us something powerful:

$$E_t[h_{t+1}(\boldsymbol{\theta}_0)] = 0 \implies \mathbb{E}[g(\mathcal{F}_t) \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$$

for **any** function g of variables in the information set $\mathcal{F}_t$!

Why? By the law of iterated expectations:

$$\mathbb{E}[g(\mathcal{F}_t) \cdot h_{t+1}(\boldsymbol{\theta}_0)] = \mathbb{E}[\mathbb{E}[g(\mathcal{F}_t) \cdot h_{t+1}(\boldsymbol{\theta}_0) | \mathcal{F}_t]] = \mathbb{E}[g(\mathcal{F}_t) \cdot \underbrace{E_t[h_{t+1}(\boldsymbol{\theta}_0)]}_{=0}] = 0$$

From Conditional to Unconditional Moments

The law of iterated expectations tells us something powerful:

$$E_t[h_{t+1}(\boldsymbol{\theta}_0)] = 0 \implies \mathbb{E}[g(\mathcal{F}_t) \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$$

for **any** function g of variables in the information set $\mathcal{F}_t$!

Why? By the law of iterated expectations:

$$\mathbb{E}[g(\mathcal{F}_t) \cdot h_{t+1}(\boldsymbol{\theta}_0)] = \mathbb{E}[\mathbb{E}[g(\mathcal{F}_t) \cdot h_{t+1}(\boldsymbol{\theta}_0)|\mathcal{F}_t]] = \mathbb{E}[g(\mathcal{F}_t) \cdot \underbrace{E_t[h_{t+1}(\boldsymbol{\theta}_0)]}_{=0}] = 0$$

Implication: We can create **many** unconditional moment conditions from one conditional restriction!

Examples of Valid Instruments

In the consumption model, $\mathcal{F}_t$ includes past consumption, returns, and other observables at t:

- $\mathbb{E}[C_t \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$
- $\mathbb{E}[C_{t-1} \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$
- $\mathbb{E}[r_t \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$
- $\mathbb{E}[(C_t/C_{t-1}) \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$
- $\mathbb{E}[r_t^2 \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$

Examples of Valid Instruments

In the consumption model, $\mathcal{F}_t$ includes past consumption, returns, and other observables at t:

- $\mathbb{E}[C_t \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$
- $\mathbb{E}[C_{t-1} \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$
- $\mathbb{E}[r_t \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$
- $\mathbb{E}[(C_t/C_{t-1}) \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$
- $\mathbb{E}[r_t^2 \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$

All of these are valid moment conditions! We can use any subset to estimate $\boldsymbol{\theta}$

Examples of Valid Instruments

In the consumption model, $\mathcal{F}_t$ includes past consumption, returns, and other observables at t:

- $\mathbb{E}[C_t \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$
- $\mathbb{E}[C_{t-1} \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$
- $\mathbb{E}[r_t \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$
- $\mathbb{E}[(C_t/C_{t-1}) \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$
- $\mathbb{E}[r_t^2 \cdot h_{t+1}(\boldsymbol{\theta}_0)] = 0$

All of these are valid moment conditions! We can use any subset to estimate $\boldsymbol{\theta}$

If we use q different functions of $\mathcal{F}_t$, we have $r = q$ moment conditions for $a = 2$ parameters

- **Exactly identified** if $q = 2$
- **Overidentified** if $q > 2$ (can use J-test!)

General Framework: Conditional Moments

More generally, many models specify:

$$\mathbb{E}[\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_{t+1})|\mathcal{F}_t] = 0$$

where $\mathbf{h}$ is a $k \times 1$ vector of moment functions

General Framework: Conditional Moments

More generally, many models specify:

$$\mathbb{E}[\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_{t+1}) | \mathcal{F}_t] = 0$$

where $\mathbf{h}$ is a $k \times 1$ vector of moment functions

Creating unconditional moments: Let $\mathbf{z}_t$ be an $\ell \times 1$ vector of variables in $\mathcal{F}_t$

- By law of iterated expectations: $\mathbb{E}[\mathbf{z}_t \otimes \mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_{t+1})] = 0$
- The Kronecker product $\otimes$ creates all interactions:

$$\mathbf{z}_t \otimes \mathbf{h}_{t+1} = \begin{bmatrix} z_{1,t} \cdot \mathbf{h}_{t+1} \\ z_{2,t} \cdot \mathbf{h}_{t+1} \\ \vdots \\ z_{\ell,t} \cdot \mathbf{h}_{t+1} \end{bmatrix}_{(\ell \cdot k) \times 1}$$

General Framework: Conditional Moments

More generally, many models specify:

$$\mathbb{E}[\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_{t+1})|\mathcal{F}_t] = 0$$

where $\mathbf{h}$ is a $k \times 1$ vector of moment functions

Creating unconditional moments: Let $\mathbf{z}_t$ be an $\ell \times 1$ vector of variables in $\mathcal{F}_t$

- By law of iterated expectations: $\mathbb{E}[\mathbf{z}_t \otimes \mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_{t+1})] = 0$
- The Kronecker product $\otimes$ creates all interactions:

$$\mathbf{z}_t \otimes \mathbf{h}_{t+1} = \begin{bmatrix} z_{1,t} \cdot \mathbf{h}_{t+1} \\ z_{2,t} \cdot \mathbf{h}_{t+1} \\ \vdots \\ z_{\ell,t} \cdot \mathbf{h}_{t+1} \end{bmatrix}_{(\ell \cdot k) \times 1}$$

This gives $r = \ell \cdot k$ unconditional moment conditions

The Instruments Trade-off

More instruments (larger ℓ):

- More information, potentially more efficient estimates;
- Can test overidentifying restrictions (J-test);
- Increases dimensionality of the optimization problem;
- May include weak or irrelevant instruments $\implies$ this scan screw-up everything!

The Instruments Trade-off

More instruments (larger ℓ):

- More information, potentially more efficient estimates;
- Can test overidentifying restrictions (J-test);
- Increases dimensionality of the optimization problem;
- May include weak or irrelevant instruments $\implies$ this scan screw-up everything!

Fewer instruments (smaller ℓ):

- Simpler estimation, less finite-sample bias;
- Focus on strongest/most relevant moment conditions;
- Less efficient asymptotically;
- Cannot test overidentification if $r = a$;

Practical Guidance

How to choose which functions of $\mathcal{F}_t$ to use?

1. **Theory first**: Use economic theory to guide which variables should be relevant
2. **Parsimony**: Start with a small set of the most relevant instruments
3. **Balance**: Use enough moments to test overidentifying restrictions, but not so many that finite-sample bias becomes severe
4. **Robustness**: Check sensitivity to different instrument sets

Practical Guidance

How to choose which functions of $\mathcal{F}_t$ to use?

1. **Theory first**: Use economic theory to guide which variables should be relevant
2. **Parsimony**: Start with a small set of the most relevant instruments
3. **Balance**: Use enough moments to test overidentifying restrictions, but not so many that finite-sample bias becomes severe
4. **Robustness**: Check sensitivity to different instrument sets

Example from consumption model:

- Conservative: Use just C_t and r_t (exactly identified if $k = 2$)
- Moderate: Add C_{t-1} and r_{t-1} (overidentified, can test!)
- Aggressive: Add many lags and transformations (potential finite-sample issues)

Questions?

Classical Estimators as GMM

Introduction

GMM provides a **unifying framework** for econometric estimation:

- Many familiar estimators are special cases of GMM
- Understanding this connection:
  - Provides intuition for GMM
  - Shows how to derive asymptotic variance easily
  - Suggests robustness checks (e.g., change weighting matrix)

Introduction

GMM provides a **unifying framework** for econometric estimation:

- Many familiar estimators are special cases of GMM
- Understanding this connection:
  - Provides intuition for GMM
  - Shows how to derive asymptotic variance easily
  - Suggests robustness checks (e.g., change weighting matrix)

We will show that OLS, IV/2SLS, GLS, and NLS all fit into the GMM framework!

OLS as GMM

Consider the linear regression model:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_0 + u_t \quad \text{with} \quad \mathbb{E}[\mathbf{x}_t u_t] = 0$$

OLS as GMM

Consider the linear regression model:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_0 + u_t \quad \text{with} \quad \mathbb{E}[\mathbf{x}_t u_t] = 0$$

Moment condition:

$$\mathbf{h}(\boldsymbol{\beta}, \mathbf{w}_t) = \mathbf{x}_t(y_t - \mathbf{x}_t'\boldsymbol{\beta})$$

OLS as GMM

Consider the linear regression model:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_0 + u_t \quad \text{with} \quad \mathbb{E}[\mathbf{x}_t u_t] = 0$$

Moment condition:

$$\mathbf{h}(\boldsymbol{\beta}, \mathbf{w}_t) = \mathbf{x}_t(y_t - \mathbf{x}_t'\boldsymbol{\beta})$$

- Number of moments: $r = p$ (dimension of $\mathbf{x}_t$)
- Number of parameters: $a = p$
- **Exactly identified**: weighting matrix doesn't matter!

OLS as GMM

Consider the linear regression model:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_0 + u_t \quad \text{with} \quad \mathbb{E}[\mathbf{x}_t u_t] = 0$$

Moment condition:

$$\mathbf{h}(\boldsymbol{\beta}, \mathbf{w}_t) = \mathbf{x}_t(y_t - \mathbf{x}_t'\boldsymbol{\beta})$$

- Number of moments: $r = p$ (dimension of $\mathbf{x}_t$)
- Number of parameters: $a = p$
- **Exactly identified**: weighting matrix doesn't matter!

First-order condition:

$$\sum_{t=1}^{T} \mathbf{x}_t(y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}) = 0 \implies \hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

This is the standard OLS estimator!

IV/2SLS as GMM

Consider the IV model:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_0 + u_t \quad \text{with} \quad \mathbb{E}[\mathbf{z}_t u_t] = 0$$

where $\mathbf{z}_t$ is an $\ell \times 1$ vector of instruments

IV/2SLS as GMM

Consider the IV model:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}_0 + u_t \quad \text{with} \quad \mathbb{E}[\mathbf{z}_t u_t] = 0$$

where $\mathbf{z}_t$ is an $\ell \times 1$ vector of instruments

Moment condition:

$$\mathbf{h}(\boldsymbol{\beta}, \mathbf{w}_t) = \mathbf{z}_t (y_t - \mathbf{x}_t' \boldsymbol{\beta})$$

IV/2SLS as GMM

Consider the IV model:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_0 + u_t \quad \text{with} \quad \mathbb{E}[\mathbf{z}_t u_t] = 0$$

where $\mathbf{z}_t$ is an $\ell \times 1$ vector of instruments

Moment condition:

$$\mathbf{h}(\boldsymbol{\beta}, \mathbf{w}_t) = \mathbf{z}_t(y_t - \mathbf{x}_t'\boldsymbol{\beta})$$

Exactly identified case ($\ell = p$):

- Weighting matrix doesn't matter
- Any $\mathbf{W}_T$ gives the same estimator

IV/2SLS as GMM

Consider the IV model:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_0 + u_t \quad \text{with} \quad \mathbb{E}[\mathbf{z}_t u_t] = 0$$

where $\mathbf{z}_t$ is an $\ell \times 1$ vector of instruments

Moment condition:

$$\mathbf{h}(\boldsymbol{\beta}, \mathbf{w}_t) = \mathbf{z}_t(y_t - \mathbf{x}_t'\boldsymbol{\beta})$$

Exactly identified case ($\ell = p$):

- Weighting matrix doesn't matter
- Any $\mathbf{W}_T$ gives the same estimator

Overidentified case ($\ell > p$):

- Optimal weighting: $\mathbf{W}^* = [\mathbb{E}(\mathbf{z}_t\mathbf{z}_t'\sigma^2)]^{-1} \propto \mathbb{E}(\mathbf{z}_t\mathbf{z}_t')^{-1}$
- Two-step GMM with $\mathbf{W}^{(1)} = (\frac{1}{T}\sum \mathbf{z}_t\mathbf{z}_t')^{-1}$ gives **2SLS**

GLS/FGLS as GMM

Consider the linear model with **known heteroskedasticity**:

$$\mathbf{y} = \mathbf{X}\beta_0 + \mathbf{u} \quad \text{with} \quad \mathbb{E}[\mathbf{u}|\mathbf{X}] = 0, \quad \mathbb{E}[\mathbf{u}\mathbf{u}'|\mathbf{X}] = \boldsymbol{\Omega}$$

Consider the linear model with **known heteroskedasticity**:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{u} \quad \text{with} \quad \mathbb{E}[\mathbf{u}|\mathbf{X}] = 0, \quad \mathbb{E}[\mathbf{u}\mathbf{u}'|\mathbf{X}] = \boldsymbol{\Omega}$$

Moment condition (stacking all observations):

$$\mathbf{h}(\boldsymbol{\beta}, \mathbf{w}) = \mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

This gives p moment conditions from T observations

GLS/FGLS as GMM

Consider the linear model with **known heteroskedasticity**:

$$\mathbf{y} = \mathbf{X}\beta_0 + \mathbf{u} \quad \text{with} \quad \mathbb{E}[\mathbf{u}|\mathbf{X}] = 0, \quad \mathbb{E}[\mathbf{u}\mathbf{u}'|\mathbf{X}] = \boldsymbol{\Omega}$$

Moment condition (stacking all observations):

$$\mathbf{h}(\beta, \mathbf{w}) = \mathbf{X}'(\mathbf{y} - \mathbf{X}\beta)$$

This gives p moment conditions from T observations

Optimal weighting:

- The long-run variance $\mathbf{S}$ accounts for $\mathbb{E}[\mathbf{u}\mathbf{u}'] = \boldsymbol{\Omega}$
- With known $\boldsymbol{\Omega}$: efficient GMM = **GLS** = $(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$
- With estimated $\hat{\boldsymbol{\Omega}}$: efficient GMM = **FGLS**

NLS as GMM

Consider the **nonlinear regression** model:

$$y_t = g(\mathbf{x}_t, \boldsymbol{\beta}_0) + u_t \quad \text{with} \quad \mathbb{E}[\mathbf{x}_t u_t] = 0$$

where $g(\cdot, \cdot)$ is a known nonlinear function

NLS as GMM

Consider the **nonlinear regression** model:

$$y_t = g(\mathbf{x}_t, \boldsymbol{\beta}_0) + u_t \quad \text{with} \quad \mathbb{E}[\mathbf{x}_t u_t] = 0$$

where $g(\cdot, \cdot)$ is a known nonlinear function

Moment condition:

$$\mathbf{h}(\boldsymbol{\beta}, \mathbf{w}_t) = \mathbf{x}_t(y_t - g(\mathbf{x}_t, \boldsymbol{\beta}))$$

NLS as GMM

Consider the **nonlinear regression** model:

$$y_t = g(\mathbf{x}_t, \beta_0) + u_t \quad \text{with} \quad \mathbb{E}[\mathbf{x}_t u_t] = 0$$

where $g(\cdot, \cdot)$ is a known nonlinear function

Moment condition:

$$\mathbf{h}(\beta, \mathbf{w}_t) = \mathbf{x}_t(y_t - g(\mathbf{x}_t, \beta))$$

- **Exactly identified** if $\dim(\mathbf{x}_t) = \dim(\beta)$
- First-order conditions give the **NLS estimator**
- Asymptotic variance formula comes directly from GMM theory

Questions?

The End

References

- Chapter 14 from Hamilton's book;
- Chapter 13 from Hansen's book;
- Hansen, L. P. (1982). "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*, 50(4), 1029-1054.
- Newey, W. K., & West, K. D. (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55(3), 703-708.
- Hansen, L. P., Heaton, J., & Yaron, A. (1996). "Finite-Sample Properties of Some Alternative GMM Estimators." *Journal of Business & Economic Statistics*, 14(3), 262-280.