Generalized Method of Moments (GMM)

Dr. Yang Liang

1 1. What is GMM?

Idea

The **Generalized Method of Moments (GMM)** is a flexible estimation framework that uses the **moment conditions** implied by economic theory or statistical models.

If a model implies:

$$E[m(W_i, \theta_0)] = 0,$$

where $m(\cdot)$ is a $q \times 1$ vector of functions of the data W_i and parameters θ, and $q \geq k = \dim(\theta)$, then GMM estimates θ_0 by making the sample analog as close to zero as possible.

$$\hat{\theta}_{GMM} = \arg\min_{\theta} \; g_n(\theta)' W_n \, g_n(\theta),$$

where

$$g_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} m(W_i, \theta),$$

and W_n is a positive definite **weighting matrix**.

—

Intuition

- The model says: "At the truth, certain sample moments should equal zero." - GMM chooses $\hat{\theta}$ to make the sample moments as close to zero as possible. - Flexible: does not require specifying a full likelihood (unlike MLE).

—

2 2. OLS as GMM

Model

$$y_i = x_i'\beta + u_i, \quad E[u_i | x_i] = 0.$$

Moment Condition

Conditional mean assumption implies:

$$E[x_i u_i] = E[x_i(y_i - x_i'\beta)] = 0.$$

So define $m_i(\beta) = x_i(y_i - x_i'\beta)$.

Sample moment:

$$g_n(\beta) = \frac{1}{n}X'(Y - X\beta).$$

GMM Criterion

$$Q_n(\beta) = g_n(\beta)'Wg_n(\beta).$$

Taking $W = (E[x_i x_i'])^{-1}$ (or sample analog $(X'X/n)^{-1}$):

$$Q_n(\beta) = (Y - X\beta)'X(X'X)^{-1}X'(Y - X\beta).$$

Minimization

Differentiate and set to zero:

$$\frac{\partial Q_n}{\partial \beta} = -2X'(Y - X\beta) = 0 \quad \Rightarrow \quad \hat{\beta}_{OLS} = (X'X)^{-1}X'Y.$$

Therefore: OLS is GMM with moment $E[x_i(y_i - x_i'\beta)] = 0$ and weight $W = (X'X/n)^{-1}$.

—

3 3. GLS as GMM

Model

$$y = X\beta + u, \quad E[u|X] = 0, \quad Var(u|X) = \Omega.$$

Here Ω may be non-scalar (heteroskedastic or correlated).

Moment Condition

Same as OLS:

$$E[X'u] = 0.$$

But now moments have covariance:

$$E[X'uu'X] = X'\Omega X.$$

Efficient GMM Weight

Optimal weight $W = (X'\Omega X)^{-1}$ corresponds to weighting each moment by its variance.
Criterion:

$$Q(\beta) = (Y - X\beta)'\Omega^{-1}(Y - X\beta).$$

Minimization gives:

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y.$$

Therefore: GLS is GMM with moment $E[X'u] = 0$ and optimal weighting matrix $W = \Omega^{-1}$.
—

4 4. IV / 2SLS as GMM

Model

$$Y = X\beta + u, \quad E[Zu] = 0, \quad E[ZX] \text{ full rank.}$$

Here Z are instruments ($n \times q$, $q \geq k$).

Moment Condition

$$E[Z'(Y - X\beta)] = 0.$$

Sample analog:

$$g_n(\beta) = \frac{1}{n}Z'(Y - X\beta).$$

GMM Criterion

$$Q_n(\beta) = g_n(\beta)'Wg_n(\beta) = (Y - X\beta)'ZWZ'(Y - X\beta).$$

First-order condition:

$$X'ZWZ'(Y - X\beta) = 0.$$

Solution

$$\hat{\beta}_{GMM} = (X'ZWZ'X)^{-1}X'ZWZ'Y.$$

Special Cases:

- If $W = (Z'Z)^{-1}$, then
$$\hat{\beta}_{GMM} = (X'P_ZX)^{-1}X'P_ZY = \hat{\beta}_{2SLS},$$
where $P_Z = Z(Z'Z)^{-1}Z'$.
- If $W = \Omega^{-1}$, where $\Omega = E[Z'uu'Z]$, then we get the **efficient GMM / optimal IV** estimator.

—

Relationship Among OLS, GLS, IV

Estimator	Moment Condition	Weighting Matrix W	Estimator
OLS	$E[X'(Y - X\beta)] = 0$	$(X'X/n)^{-1}$	$(X'X)^{-1}X'Y$
GLS	$E[X'(Y - X\beta)] = 0$	Ω^{-1}	$(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$
IV (2SLS)	$E[Z'(Y - X\beta)] = 0$	$(Z'Z/n)^{-1}$	$(X'P_ZX)^{-1}X'P_ZY$

All are special cases of:

$$\hat{\beta}_{GMM} = (X'ZWZ'X)^{-1}X'ZWZ'Y,$$

where (Z, W) define the model's moment restrictions and weighting scheme.

—

5 5. Intuition and Efficiency

- Each estimator arises from choosing **different moments and weights**.
- OLS uses its own regressors as instruments ($Z = X$), with homoskedastic errors.
- GLS uses known error covariance Ω as a weight.
- IV/2SLS uses external instruments Z when regressors are endogenous.

Efficient GMM: When the weighting matrix equals the inverse of the true moment covariance, $W = \Omega^{-1}$, GMM achieves the Cramér–Rao lower bound among all estimators that use the same moments.

—

6 6. Visual Summary

$$\boxed{\text{GMM unifies all linear estimators:} \quad \hat{\beta}_{GMM} = (X'ZWZ'X)^{-1}X'ZWZ'Y}$$

$$\begin{aligned}
Z = X,\ W = (X'X)^{-1} &\quad \Rightarrow \text{OLS} \\
Z = X,\ W = \Omega^{-1} &\quad \Rightarrow \text{GLS} \\
Z \neq X,\ W = (Z'Z)^{-1} &\quad \Rightarrow \text{IV / 2SLS} \\
Z \neq X,\ W = \Omega^{-1} &\quad \Rightarrow \text{Efficient GMM (Optimal IV)}
\end{aligned}$$

7 3. GMM and Extremum Estimators

Definition

An **extremum estimator** solves:

$$\hat{\theta} = \arg\max_{\theta} Q_n(\theta)$$

$(\text{or } \arg\min)$. Examples: MLE, NLS, GMM.

GMM is a particular extremum estimator where:

$$Q_n(\theta) = -g_n(\theta)' W_n g_n(\theta),$$

so it fits perfectly within the extremum framework.

—

Consistency and Asymptotic Normality

Under regularity conditions:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, (G'WG)^{-1} G'W\Omega WG(G'WG)^{-1}\right),$$

where:

$$G = E\left[\frac{\partial m(W_i, \theta_0)}{\partial \theta'}\right], \quad \Omega = E[m(W_i, \theta_0) m(W_i, \theta_0)'].$$

—

Optimal GMM

Efficiency in GMM depends on choice of W. The asymptotic variance is minimized when:

$$W = \Omega^{-1}.$$

Then:

$$\text{Avar}(\hat{\theta}_{opt}) = (G'\Omega^{-1}G)^{-1}.$$

—

8 4. Detailed Econometrics: Steps, Testing, and Implementation

Step 1: Model moments

Specify q moment conditions:

$$E[m(W_i, \theta_0)] = 0.$$

—

Step 2: Sample moments

Compute sample analog:

$$g_n(\theta) = \frac{1}{n} \sum_i m(W_i, \theta).$$

—

Step 3: Choose initial weighting matrix

- Often start with $W_n = I_q$ (identity) or $(Z'Z/n)^{-1}$. - Compute preliminary GMM estimate $\tilde{\theta}$.

—

Step 4: Estimate the optimal weighting matrix

Estimate

$$\hat{\Omega} = \frac{1}{n}\sum_i m(W_i, \tilde{\theta})m(W_i, \tilde{\theta})'.$$

Set

$$\hat{W} = \hat{\Omega}^{-1}.$$

—

Step 5: Two-step efficient GMM

Re-estimate:

$$\hat{\theta}_{GMM} = \arg\min_{\theta} g_n(\theta)'\hat{W}\, g_n(\theta).$$

Iterating again yields the same asymptotic efficiency.

—

Step 6: Over-identification Test (Sargan–Hansen J test)

If $q > k$, the model is over-identified.

$$J = n\, g_n(\hat{\theta}_{GMM})'\hat{W}\, g_n(\hat{\theta}_{GMM}) \sim \chi^2_{q-k}.$$

Interpretation: - Small J: moments consistent $\Rightarrow$ instruments valid. - Large J: reject over-identifying restrictions $\Rightarrow$ possible invalid instruments.

—

Step 7: Robust Standard Errors

Variance–covariance (sandwich) estimator:

$$\widehat{\mathrm{Var}}(\hat{\theta}_{GMM}) = (G'\hat{W}G)^{-1}\, G'\hat{W}\hat{\Omega}\hat{W}G\, (G'\hat{W}G)^{-1}.$$

In practice, software packages compute this automatically.

—

9 5. GMM in Practice

- **When to use:** models defined by moment restrictions (e.g., IV, dynamic panels, consumption Euler equations).
- **Strengths:** general, robust, efficient under weak distributional assumptions.
- **Weaknesses:** finite-sample bias, sensitivity to weighting matrix, weak identification.

—

10 6. Summary Table

Estimator	Moment condition	Weight W	Efficiency criterion
OLS	$E[X'u] = 0$	$(X'X)^{-1}$	Homoskedasticity
GLS	$E[X'u] = 0$	Ω^{-1}	Heteroskedasticity / known Ω
IV / 2SLS	$E[Z'u] = 0$	$(Z'Z)^{-1}$	Linear IV
Efficient GMM	$E[m(W_i, \theta)] = 0$	Ω^{-1}	Optimal weighting

—

11 7. Intuitive Summary

GMM unifies classical estimators: OLS, GLS, IV are all special cases. It replaces "derive a formula" with "impose a moment condition."

GMM is flexible: Works for nonlinear, dynamic, and structural models. In large samples, efficient GMM = MLE when both are correctly specified.

Philosophy: *"Estimate parameters so that the theoretical moments match the data moments as closely as possible."*