Econometrics II

Lecture 3 - Generalized Method of Moments

Lasha Chochua

2025

Estimation Methods We've Covered

- **Moment Estimators**
  - Based on the method of moments.
  - Use sample moments to estimate population parameters.
  - Match sample moments to theoretical moments:

$$\frac{1}{n}\sum_{i=1}^{n} m(X_i, \theta) = 0$$

- **Maximum Likelihood Estimators (MLE)**
  - Choose parameter θ to maximize the likelihood function.
  - Log-likelihood for iid data:

$$\ell(\theta) = \sum_{i=1}^{n} \log f(X_i; \theta)$$

  - MLE: $\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \ell(\theta)$

Estimation Methods We've Covered (Cont.)

- **Ordinary Least Squares (OLS)**
  - Linear regression: minimize the sum of squared residuals.

$$\hat{\beta}_{\text{OLS}} = \arg\min_{\beta} \sum_{i=1}^{n} (Y_i - X_i^\top \beta)^2$$

  - Closed-form solution:

$$\hat{\beta}_{\text{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

- **Instrumental Variables (IV)**
  - Used when regressors are endogenous.
  - Requires instruments Z uncorrelated with the error term.
  - Two-stage procedure:
    1. Regress X on Z to get predicted values $\hat{X}$
    2. Regress Y on $\hat{X}$:

$$\hat{\beta}_{\text{IV}} = (Z'X)^{-1}Z'Y$$

Generalized Method of Moments (GMM)

- GMM is a flexible and powerful estimation method in econometrics.
- It generalizes the classical method of moments:
  - Allows for more moment conditions than unknowns (overidentified).
  - Permits nonlinear functions of data and parameters.
- GMM includes as special cases:
  - Ordinary Least Squares (OLS)
  - Instrumental Variables (IV)
  - Multivariate regression
  - Two-Stage Least Squares (2SLS)
- GMM handles both linear and nonlinear models.
- In this lecture, we focus on linear GMM.

Historical Background

- GMM introduced in econometrics by Lars Hansen (1982).
- Builds on work by:
  - Amemiya (1974, 1977)
  - Gallant (1977)
  - Gallant & Jorgenson (1979)
- Related to estimating equations in statistics (see Godambe, 1991).
- Also connected to work by:
  - White (1980, 1982)
  - White & Domowitz (1984)

Moment Equation Models

- As said, all models we've studied so far can be written as **moment equation models**.
- These models solve a system of moment equations:

$$\mathbb{E}[g_i(\beta)] = 0 \tag{1}$$

- Here:
    - $g_i(\beta)$ is a known $\ell \times 1$ function of the ith observation.
    - β is a $k \times 1$ parameter vector.
    - $\beta \in \mathcal{B}$, a parameter space.
- Example: In IV models,

$$g_i(\beta) = Z_i(Y_i - X_i'\beta)$$

Identification in Moment Models

- We say β is **identified** if the moment condition has a unique solution.
- Let $\ell =$ number of moment conditions, $k =$ number of parameters:
  - If $\ell = k$: **just identified** — just enough information.
  - If $\ell > k$: **overidentified** — more information than needed.
  - If $\ell < k$: **underidentified** — not enough information.
- Typically assume: $\ell \geq k$

Method of Moments Estimators

- Focus on the just-identified case: $\ell = k$.
- Define the **sample moment condition**:

$$\bar{g}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta) \tag{2}$$

- The **method of moments estimator (MME)** solves:

$$\bar{g}_n(\widehat{\beta}_{\text{mm}}) = 0 \tag{3}$$

- These equations are called **estimating equations**.
- In some cases, $\widehat{\beta}_{\text{mm}}$ has a closed-form; in others, numerical methods are needed.

Examples of MME

- **Mean only**:

$$g_i(\mu) = Y_i - \mu \quad \Rightarrow \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{4}$$

- **Mean and variance**:

$$g_i(\mu, \sigma^2) = \begin{pmatrix} Y_i - \mu \\ (Y_i - \mu)^2 - \sigma^2 \end{pmatrix} \quad \Rightarrow \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\mu})^2 \tag{5}$$

MME for OLS

- Set:

$$g_i(\beta) = X_i(Y_i - X_i'\beta) \tag{6}$$

- Then the MME is:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{7}$$

OLS with Variance

- Set:

$$g_i(\beta, \sigma^2) = \begin{pmatrix} X_i(Y_i - X_i'\beta) \\ (Y_i - X_i'\beta)^2 - \sigma^2 \end{pmatrix} \tag{8}$$

- Then:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i'\hat{\beta})^2 \tag{9}$$

Method of Moments: Multivariate Examples

- **Multivariate Least Squares (vector form)**:

$$g_i(\beta) = \bar{X}_i(Y_i - \bar{X}_i\beta) \tag{10}$$

$$\hat{\beta} = \left(\sum_{i=1}^{n} \bar{X}_i'\bar{X}_i\right)^{-1} \left(\sum_{i=1}^{n} \bar{X}_i'Y_i\right) \tag{11}$$

- **Multivariate Least Squares (matrix form)**:

$$g_i(B) = \text{vec}\left(X_i(Y_i' - X_i'B)\right) \tag{12}$$

$$\hat{B} = \left(\sum_{i=1}^{n} X_iX_i'\right)^{-1} \left(\sum_{i=1}^{n} X_iY_i'\right) \tag{13}$$

Definition: The vec Operator

- For a matrix $A \in \mathbb{R}^{m \times n}$,
$$\text{vec}(A) \in \mathbb{R}^{mn \times 1} \tag{14}$$
is formed by stacking the columns of A into a single column vector.

Seemingly Unrelated Regressions (SUR)

- Moment condition:

$$g_i(\beta, \Sigma) = \begin{pmatrix} \bar{X}_i \Sigma^{-1} (Y_i - \bar{X}_i \beta) \\ \text{vec}\left(\Sigma - (Y_i - \bar{X}_i \beta)(Y_i - \bar{X}_i \beta)'\right) \end{pmatrix} \tag{15}$$

- MME:

$$\hat{\beta} = \left(\sum_{i=1}^{n} \bar{X}_i' \hat{\Sigma}^{-1} \bar{X}_i \right)^{-1} \left(\sum_{i=1}^{n} \bar{X}_i' \hat{\Sigma}^{-1} Y_i \right) \tag{16}$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{X}_i \hat{\beta})(Y_i - \bar{X}_i \hat{\beta})' \tag{17}$$

IV as Method of Moments

- Set:

$$g_i(\beta) = Z_i(Y_i - X_i'\beta) \tag{18}$$

$$\hat{\beta} = (\mathbf{Z'X})^{-1}\mathbf{Z'Y} \tag{19}$$

Overidentified Moment Equations

- When $\ell > k$, the model is **overidentified**.
- The sample moment condition:

$$\bar{g}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} Z_i(Y_i - X_i'\beta) = \frac{1}{n}(\mathbf{Z}'\mathbf{Y} - \mathbf{Z}'\mathbf{X}\beta) \tag{20}$$

- In this case, we cannot find $\hat{\beta}$ that sets $\bar{g}_n(\beta) = 0$.
- Idea: Choose $\hat{\beta}$ that makes $\bar{g}_n(\beta)$ **as small as possible**.

Interpreting GMM as Least Squares

- Define:
  - $\mu = \mathbf{Z}'\mathbf{Y} \in \mathbb{R}^{\ell \times 1}$
  - $G = \mathbf{Z}'\mathbf{X} \in \mathbb{R}^{\ell \times k}$
  - $\eta = \mu - \mathbf{G}\beta \in \mathbb{R}^{\ell \times 1}$
- Then moment condition becomes:

$$\mu = \mathbf{G}\beta + \eta \tag{21}$$

- This is a **linear regression**:
  - Regress $\mu \in \mathbb{R}^{\ell \times 1}$ on $\mathbf{G} \in \mathbb{R}^{\ell \times k}$.
  - Estimate β by minimizing the sum of squared residuals:

$$\hat{\beta} = (\mathbf{G}'\mathbf{G})^{-\mathbf{1}}\mathbf{G}'\mu \quad \text{(dimensions: } k \times 1) \tag{22}$$

- This minimizes:

$$\eta'\eta = (\mu - \mathbf{G}\beta)'(\mu - \mathbf{G}\beta) \tag{23}$$

Generalized Method of Moments (GMM)

- When errors are heteroskedastic, we minimize a **weighted** sum of squares $\eta'\mathbf{W}\eta$ (recall GLS):
$$\hat{\beta}_{\text{gmm}} = (\mathbf{G}'\mathbf{W}\mathbf{G})^{-\mathbf{1}}\mathbf{G}'\mathbf{W}\mu = (\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{X})^{-\mathbf{1}}(\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{Y}) \tag{24}$$
- GMM minimizes:
$$J(\beta) = n\,\bar{g}_n(\beta)'\mathbf{W}\bar{g}_n(\beta) \tag{25}$$
- The factor "n" is not important for the definition of the estimator but is convenient for the distribution theory.

GMM Estimator

Definition 1 *(Generalized Method of Moments Estimator)*

The GMM estimator is defined as:

$$\hat{\beta}_{\mathsf{gmm}} = \arg\min_{\beta} J(\beta) \tag{26}$$

GMM and MME Connection

- When $\ell = k$, the model is **just-identified**.
  - Then $J(\hat{\beta}_{\text{mm}}) = 0$
  - GMM reduces to MME: $\hat{\beta}_{\text{gmm}} = \hat{\beta}_{\text{mm}}$
- So: **MME is a special case of GMM**
- In the overidentified case, GMM depends on the choice of weight matrix W.
- GMM was introduced by **Lars Hansen (1982)**.
  - Developed its asymptotic theory, efficient weighting, and overidentification tests.

Linear Moment Models

- The moment equation framework supports both **linear** and **nonlinear** models.
- When moment equations are **linear in parameters**, we gain:
  - Explicit solutions for estimators.
  - Simpler asymptotic theory.
- Examples of models with **linear moment conditions**:
  - Sample mean and variance
  - OLS and IV
  - Multivariate least squares
  - 2SLS
- Nonlinear examples include:
  - Sample variance as a parameter
  - Seemingly unrelated regressions (SUR)
  - Models with generated regressors

Example: Overidentified IV Model

- Focus on the moment condition:

$$g_i(\beta) = Z_i(Y_i - X_i'\beta) \tag{27}$$

- Where:
  - Z_i is $\ell \times 1$
  - X_i is $k \times 1$
- This is a **linear** function of β.
- Overidentification: $\ell > k$

GMM Estimator

- Given the IV moment condition:

$$g_i(\beta) = Z_i(Y_i - X_i'\beta) \tag{28}$$

- The GMM criterion is:

$$J(\beta) = n(\mathbf{Z}'\mathbf{Y} - \mathbf{Z}'\mathbf{X}\beta)'\mathbf{W}(\mathbf{Z}'\mathbf{Y} - \mathbf{Z}'\mathbf{X}\beta) \tag{29}$$

- The GMM estimator minimizes $J(\beta)$.

First Order Conditions

- Gradient with respect to β:

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \beta} J(\hat{\beta}) && (30)\\
&= 2 \frac{\partial}{\partial \beta} \bar{g}_n(\hat{\beta}) \mathbf{W} \bar{g}_n(\hat{\beta}) && (31)\\
&= -2 \left(\frac{1}{n} \mathbf{X}' \mathbf{Z} \right) \mathbf{W} \left(\frac{1}{n} \mathbf{Z}' (\mathbf{Y} - \mathbf{X} \hat{\beta}) \right) && (32)
\end{aligned}
$$

GMM Estimator

Theorem 1 *(For the overidentified IV model)*

The GMM estimator is:

$$\hat{\beta}_{\text{gmm}} = (\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{X})^{-\mathbf{1}}(\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{Y}) \tag{33}$$

Properties and Special Cases

- GMM depends on the weight matrix $\mathbf{W}$ only **up to scale**.
  - If $\mathbf{W}$ is replaced by $c\mathbf{W}$, the estimator remains unchanged.
- When W is fixed: **one-step GMM**.
- For just-identified case ($k = \ell$):

$$\hat{\beta}_{\text{gmm}} = (\mathbf{Z'X})^{-\mathbf{1}}\mathbf{W}^{-\mathbf{1}}(\mathbf{X'Z})^{-\mathbf{1}}(\mathbf{X'ZWZ'Y}) = (\mathbf{Z'X})^{-\mathbf{1}}\mathbf{Z'Y} = \hat{\beta}_{\text{iv}} \tag{34}$$

- The GMM estimator resembles the 2SLS estimator. In fact they are equal when $\mathbf{W} = (\mathbf{Z'Z})^{-\mathbf{1}}$. This means that the 2SLS estimator is a one-step GMM estimator for the linear model.

Equivalence of GMM and 2SLS

Theorem 2

If $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}$, then:

$$\hat{\beta}_{\text{gmm}} = \hat{\beta}_{\text{2sls}} \tag{35}$$

Furthermore, if $k = \ell$, then:

$$\hat{\beta}_{\text{gmm}} = \hat{\beta}_{\text{iv}} \tag{36}$$

Asymptotic Distribution of GMM Estimator

- Let:

$$\mathbf{Q} = \mathbb{E}[ZX'] \in \mathbb{R}^{\ell \times k}$$
$$\Omega = \mathbb{E}[ZZ'e^2] \in \mathbb{R}^{\ell \times \ell}$$

- Then we have the following convergence results:

$$\left(\frac{1}{n}\mathbf{X}'\mathbf{Z}\right)\mathbf{W}\left(\frac{1}{n}\mathbf{Z}'\mathbf{X}\right) \xrightarrow{p} \mathbf{Q}'\mathbf{W}\mathbf{Q} \tag{37}$$

$$\left(\frac{1}{n}\mathbf{X}'\mathbf{Z}\right)\mathbf{W}\left(\frac{1}{\sqrt{n}}\mathbf{Z}'\mathbf{e}\right) \xrightarrow{d} \mathcal{N}(0, \mathbf{Q}'\mathbf{W}\Omega\mathbf{W}\mathbf{Q}) \tag{38}$$

Asymptotic Distribution of GMM

Theorem 3 *(Asymptotic Distribution of GMM Estimator)*

Under Assumption 12.2 in the book, as $n \to \infty$:

$$\sqrt{n}(\widehat{\beta}_{\mathsf{gmm}} - \beta) \xrightarrow{d} \mathcal{N}(0, \mathbf{V}_{\beta}) \tag{39}$$

where

$$\mathbf{V}_{\beta} = (\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-\mathbf{1}}(\mathbf{Q}'\mathbf{W}\Omega\mathbf{W}\mathbf{Q})(\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1} \tag{40}$$

Notes on the Variance Formula

- The asymptotic variance has **"sandwich form"**.
- Derivation assumes W is non-random, but the result holds when:

$$\widehat{W} \xrightarrow{p} W \quad \text{(positive definite)} \tag{41}$$

- Rescaling W does **not affect** the estimator:
  - e.g., replacing $\widehat{W} = (Z'Z)^{-1}$ with $\widehat{W} = (n^{-1}Z'Z)^{-1}$
- This rescaling is typically ignored in implementation.

Efficient GMM

- The asymptotic variance of $\hat{\beta}_{\text{gmm}}$ depends on the weight matrix W:

$$\mathbf{V}_{\beta} = (\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-\mathbf{1}}(\mathbf{Q}'\mathbf{W}\Omega\mathbf{W}\mathbf{Q})(\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-\mathbf{1}} \tag{42}$$

- The **optimal weight matrix** is:

$$\mathbf{W}_0 = \Omega^{-1} \tag{43}$$

  because it **minimizes** $\mathbf{V}_{\beta}$.

- When $\mathbf{W} = \Omega^{-1}$ (or a consistent estimator), we call the estimator **Efficient GMM**:

$$\hat{\beta}_{\text{gmm}} = \left(\mathbf{X}'\mathbf{Z}\Omega^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}(\mathbf{X}'\mathbf{Z}\Omega^{-1}\mathbf{Z}'\mathbf{Y}) \tag{44}$$

Efficient GMM: Asymptotic Variance

- When $\mathbf{W} = \Omega^{-1}$, the asymptotic variance simplifies to:

$$\mathbf{V}_\beta = (\mathbf{Q}'\Omega^{-1}\mathbf{Q})^{-1}(\mathbf{Q}'\Omega^{-1}\Omega\Omega^{-1}\mathbf{Q})(\mathbf{Q}'\Omega^{-1}\mathbf{Q})^{-1} = (\mathbf{Q}'\Omega^{-1}\mathbf{Q})^{-1}. \tag{45}$$

- This is the variance of the **efficient GMM estimator**.
- It achieves the lowest possible asymptotic variance among GMM estimators.

Efficient GMM: Asymptotic Variance

Theorem 4 *(Asymptotic Distribution of Efficient GMM)*

Under Assumption 12.2 and $\mathbf{W} = \Omega^{-1}$, as $n \to \infty$:

$$\sqrt{n}(\hat{\beta}_{\mathsf{gmm}} - \beta) \xrightarrow{d} \mathcal{N}(0, \mathbf{V}_{\beta}) \tag{46}$$

where:

$$\mathbf{V}_{\beta} = (\mathbf{Q}'\Omega^{-1}\mathbf{Q})^{-1} \tag{47}$$

Efficient GMM: Asymptotic Variance

Theorem 5 *(Efficient GMM)*

Under Assumption 12.2 in the book, for any $\mathbf{W} > 0$:

$$(\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-\mathbf{1}}(\mathbf{Q}'\mathbf{W}\,\mathbf{W}\mathbf{Q})(\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1} - (\mathbf{Q}'\Omega^{-1}\mathbf{Q})^{-1} \geq 0 \tag{48}$$

The inequality is strict if $\mathbf{W} \neq \Omega^{-1}$. Hence:

$$\text{avar}(\hat{\beta}_{\text{gmm}}) \leq \text{avar}(\tilde{\beta}_{\text{gmm}}) \tag{49}$$

where $\tilde{\beta}_{\text{gmm}}$ is any other GMM estimator.

Efficiency and Semiparametric Optimality

- The smallest possible GMM covariance matrix (in the positive definite sense) is achieved when:

$$\mathbf{W}_0 = \Omega^{-1} \tag{50}$$

- In practice, Ω is unknown — but it can be **consistently estimated**.
- For any $\widehat{\mathbf{W}} \xrightarrow{p} \mathbf{W}_0$, the asymptotic distribution from Theorem 4 still holds.
- Any $\hat{\beta}_{\text{gmm}}$ constructed with a consistent estimator of $\mathbf{W}_0$ is called an **efficient GMM estimator**.

Semiparametric Efficiency

- "Efficient" means the smallest asymptotic variance among GMM estimators **using the same moment conditions**.
- This is a **weak optimality** concept (only compares over $\widehat{W}$).
- But: GMM is **semiparametrically efficient** (Chamberlain, 1987):
  - If $\mathbb{E}[g_i(\beta)] = 0$ is all that's known, GMM attains the **lowest possible variance**:

$$(\mathbf{G}'\Omega^{-1}\mathbf{G})^{-1}, \quad \text{where } \mathbf{G} = \mathbb{E}\left[\frac{\partial}{\partial\beta'} g_i(\beta)\right] \tag{51}$$

- **Conclusion**:
  - No semiparametric estimator has a smaller variance than efficient GMM — unless additional assumptions are imposed.

Efficient GMM vs 2SLS

- In linear IV models, 2SLS is a standard estimator for β.
- 2SLS is a **GMM estimator** with weight matrix:

$$\widehat{\mathbf{W}} = (\mathbf{Z}'\mathbf{Z})^{-1} \quad \text{or equivalently} \quad \widehat{\mathbf{W}} = (n^{-1}\mathbf{Z}'\mathbf{Z})^{-1} \tag{52}$$

  since **scaling doesn't matter** in GMM.

- If $\widehat{\mathbf{W}} \xrightarrow{p} (\mathbb{E}[\mathbf{Z}\mathbf{Z}'])^{-1}$, then 2SLS is asymptotically GMM with:

$$\mathbf{W} = (\mathbb{E}[\mathbf{Z}\mathbf{Z}'])^{-1} \tag{53}$$

Comparing to Efficient GMM

- Efficient GMM uses the weight matrix:

$$\mathbf{W} = \left(\mathbb{E}[ZZ'e^2]\right)^{-1} \tag{54}$$

- If we assume **conditional homoskedasticity**:

$$\mathbb{E}[e^2 \mid Z] = \sigma^2 \tag{55}$$

  then:

$$\mathbb{E}[ZZ'e^2] = \sigma^2\mathbb{E}[ZZ'] \quad \Rightarrow \quad \mathbf{W} = (\mathbb{E}[ZZ'])^{-1} \tag{56}$$

- Therefore, under conditional homoskedasticity, 2SLS **uses the efficient weight matrix** since scaling does not matter.

Theorem 6 *(2SLS is Efficient under Homoskedasticity)*

Under Assumption 12.2 in the book and $\mathbb{E}[e^2 \mid Z] = \sigma^2$,

$$\hat{\beta}_{\text{2sls}} \text{ is an efficient GMM estimator.} \tag{57}$$

Interpretation and Implications

- This result shows:
  - 2SLS is **efficient** under homoskedasticity.
  - In that case, there is **no reason to prefer efficient GMM** over 2SLS.
- But: If the errors are **conditionally heteroskedastic**, then:
  - $\mathbb{E}[e^2 \mid Z] \neq \sigma^2$
  - $\mathbf{W} \neq (\mathbb{E}[ZZ'])^{-1}$
  - So 2SLS is **no longer efficient**, and GMM with a consistent $\widehat{W}$ performs better.
- Bottom line:
  - 2SLS $=$ efficient GMM $\iff$ conditional homoskedasticity

Estimating the Efficient Weight Matrix

- To construct the **efficient GMM estimator**, we need a consistent estimator:

$$\widehat{\mathbf{W}} \quad \text{of} \quad \mathbf{W}_0 = \Omega^{-1}$$

- The typical approach is to:
  - Estimate Ω via residuals
  - Set $\widehat{\mathbf{W}} = \widehat{\Omega}^{-1}$
- This leads to the **two-step GMM estimator**:
  - Use a consistent one-step estimator $\hat{\beta}$ (e.g. 2SLS)
  - Construct $\widehat{\mathbf{W}}$ using that estimate

Two-Step GMM Estimator: Linear Case

- Use 2SLS to compute residuals:

$$\tilde{e}_i = Y_i - X_i' \widehat{\beta}_{\text{2sls}}$$

- Define:

$$\tilde{g}_i = g_i(\widehat{\beta}) = Z_i \tilde{e}_i, \quad \bar{g}_n = \frac{1}{n} \sum_{i=1}^{n} \tilde{g}_i$$

- Two common estimators of Ω:

$$\widehat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} \tilde{g}_i \tilde{g}_i' \tag{58}$$

$$\widehat{\Omega}^* = \frac{1}{n} \sum_{i=1}^{n} (\tilde{g}_i - \bar{g}_n)(\tilde{g}_i - \bar{g}_n)' \tag{59}$$

Interpretation of Estimators

- $\widehat{\Omega}$ is an **uncentered covariance estimator**.
- $\widehat{\Omega}^*$ is a **centered (robust)** covariance estimator.
- Both are consistent when $\mathbb{E}[Ze] = 0$ (correct specification).
- Under misspecification ($\mathbb{E}[Ze] \neq 0$), only $\widehat{\Omega}^*$ consistently estimates:

$$\text{var}[Ze]$$

- Hence:
  - $\widehat{\Omega}$ estimates $\mathbb{E}[ZZ'e^2]$
  - $\widehat{\Omega}^*$ is **robust** and often preferred in practice
  - Most software packages default to $\widehat{\Omega}$ (58)
  - If the model is just-identified, $\bar{g}_n = 0$ so (58) = (59)

Constructing the Two-Step Estimator

- Choose the weight matrix:

$$\widehat{\mathbf{W}} = \widehat{\Omega}^{-1} \quad \text{or} \quad \widehat{\mathbf{W}} = (\widehat{\Omega}^{*})^{-1}$$

- Then plug into:

$$\hat{\beta}_{\text{gmm}} = (\mathbf{X}'\mathbf{Z}\widehat{\mathbf{W}}\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}\widehat{\mathbf{W}}\mathbf{Z}'\mathbf{Y})$$

- Since $\hat{\beta}_{\text{2sls}}$ is consistent, and the covariance matrix estimators are consistent, it follows that $\widehat{\mathbf{W}} \xrightarrow{p} \Omega^{-1}$.
- Therefore, the **two-step GMM estimator satisfies the conditions of Theorem 4**.

Efficient Weight Matrix

Theorem 7 *(Asymptotic Normality of Two-Step GMM)*

Under Assumption 12.2 in the book and $\Omega > 0$, if $\widehat{\mathbf{W}} = \widehat{\Omega}^{-1}$ or $\widehat{\mathbf{W}} = (\widehat{\Omega}^{*})^{-1}$, then as $n \to \infty$:

$$\sqrt{n}(\widehat{\beta}_{\mathsf{gmm}} - \beta) \xrightarrow{d} \mathcal{N}(0, \mathbf{V}_{\beta})$$

where:

$$\mathbf{V}_{\beta} = (\mathbf{Q}'\Omega^{-1}\mathbf{Q})^{-1}$$

Key Points on Two-Step GMM

- Asymptotically efficient and robust to heteroskedasticity.
- Common in practice due to:
  - Simplicity
  - Generality
  - Available software support
- Often preferred when testing (e.g. overidentification) since $\widehat{\Omega}^*$ is robust under misspecification.

Iterated GMM

- The **asymptotic distribution** of the two-step GMM estimator does **not** depend on the initial one-step estimator.
- However, the **actual value** of the two-step estimator **does** depend on that initial choice, especially in finite samples.
  - This dependence can lead to **inefficiency**.

Iteration Procedure

- Given an initial GMM estimate $\hat{\beta}_{\text{gmm}}$, we:
  1. Recompute the weight matrix $\widehat{W}$ using updated residuals
  2. Re-estimate $\hat{\beta}_{\text{gmm}}$
  3. Repeat until convergence
- This produces the **iterated GMM estimator**:
  - A common approach to efficient GMM
  - Especially used to **stabilize estimates in finite samples**

Practical Note

- According to Hansen & Lee (2021):
  - The iterated GMM estimate is **invariant** to whether $\widehat{\mathbf{W}}$ is based on **centered** or **uncentered** residuals
  - However, **standard errors and test statistics do depend** on this choice

Covariance Matrix Estimation

- To estimate the asymptotic variance of $\widehat{\beta}_{\text{gmm}}$, we plug in **consistent estimators** into the asymptotic variance formula.

Two-Step GMM Estimator

- The asymptotic variance estimator is:

$$\widehat{\mathbf{V}}_{\beta} = \left(\widehat{\mathbf{Q}}'\widehat{\mathbf{W}}\widehat{\mathbf{Q}}\right)^{-1}\left(\widehat{\mathbf{Q}}'\widehat{\mathbf{W}}\widehat{\Omega}\widehat{\mathbf{W}}\widehat{\mathbf{Q}}\right)\left(\widehat{\mathbf{Q}}'\widehat{\mathbf{W}}\widehat{\mathbf{Q}}\right)^{-1} \tag{60}$$

- Where:

$$\widehat{\mathbf{Q}} = \frac{1}{n}\sum_{i=1}^{n} Z_i X_i' \tag{61}$$

- $\widehat{W}$ is constructed using either:
  - The **uncentered** estimator (58)
  - The **centered** estimator (59)
- Residuals:

$$\hat{e}_i = Y_i - X_i'\widehat{\beta}_{\mathsf{gmm}} \tag{62}$$

Iterated GMM Estimator

- The asymptotic variance for efficient GMM is simplified:

$$\widehat{\mathbf{V}}_\beta = \left(\widehat{\mathbf{Q}}'\widehat{\Omega}^{-1}\widehat{\mathbf{Q}}\right)^{-1} = \left(\left(\frac{1}{n}\mathbf{X}'\mathbf{Z}\right)\widehat{\Omega}^{-1}\left(\frac{1}{n}\mathbf{Z}'\mathbf{X}\right)\right)^{-1} \tag{63}$$

- $\widehat{\Omega}$ may be:
  - Uncentered:

$$\widehat{\Omega} = \frac{1}{n}\sum_{i=1}^{n}\tilde{g}_i\tilde{g}_i' \tag{64}$$

  - Centered:

$$\widehat{\Omega}^* = \frac{1}{n}\sum_{i=1}^{n}(\tilde{g}_i - \bar{g}_n)(\tilde{g}_i - \bar{g}_n)' \tag{65}$$

- Use **final residuals**:

$$\hat{e}_i = Y_i - X_i'\widehat{\beta}_{\text{gmm}} \tag{66}$$

Standard Errors and Centering

- Asymptotic standard errors:

$$\text{se}(\hat{\beta}_j) = \sqrt{\left[\frac{1}{n}\widehat{\mathbf{V}}_{\beta}\right]_{jj}} \tag{67}$$

- Centered $\widehat{\Omega}^*$:
  - Yields smaller standard errors
  - Leads to more "significant" inference
- Uncentered $\widehat{\Omega}$:
  - Gives larger standard errors
  - More conservative in inference

Wald Test

- Define a transformation of the parameter:

$$r(\beta) : \mathbb{R}^k \to \Theta \subset \mathbb{R}^q, \quad \theta = r(\beta) \tag{68}$$

- The GMM estimator of θ is:

$$\hat{\theta}_{\text{gmm}} = r(\hat{\beta}_{\text{gmm}}) \tag{69}$$

- By the delta method:

$$\sqrt{n}(\hat{\theta}_{\text{gmm}} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathbf{V}_\theta), \quad \mathbf{V}_\theta = \mathbf{R}'\mathbf{V}_\beta\mathbf{R} \tag{70}$$

where:

$$\mathbf{R} = \frac{\partial}{\partial \beta'} r(\beta), \quad \hat{\mathbf{R}} = \frac{\partial}{\partial \beta'} r(\hat{\beta}_{\text{gmm}}) \tag{71}$$

Wald Test

- The estimated asymptotic covariance matrix is:

$$\widehat{\mathbf{V}}_{\theta} = \hat{\mathbf{R}}'\widehat{\mathbf{V}}_{\beta}\hat{\mathbf{R}} \tag{72}$$

- If θ is scalar, then the standard error is:

$$\text{se}(\hat{\theta}_{\text{gmm}}) = \sqrt{n^{-1}\widehat{\mathbf{V}}_{\theta}} \tag{73}$$

Wald Test Statistic

- We test the hypothesis:

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0 \tag{74}$$

- The Wald statistic is:

$$W = n(\hat{\theta} - \theta_0)' \widehat{\mathbf{V}}_{\theta}^{-1} (\hat{\theta} - \theta_0) \tag{75}$$

- Let $G_q(u)$ denote the cumulative distribution function of the χ^2_q distribution.

Wald Test

Theorem 8 *(Asymptotic Distribution of Wald Test Statistic)*

Under Assumption 12.2, Assumption 7.3 in the book, and under H_0, as $n \to \infty$:

$$W \xrightarrow{d} \chi^2_q \tag{76}$$

For c such that $\alpha = 1 - G_q(c)$,

$$\mathbb{P}(W > c \mid H_0) \to \alpha \tag{77}$$

so the test "Reject H_0 if $W > c$" has asymptotic size α.

Overidentification Test

- Overidentified models have $\ell > k$, meaning there are more moment conditions than parameters.
- This makes it possible to **test the validity** of the model:

$$H_0 : \mathbb{E}[Ze] = 0 \tag{78}$$

- This is the **overidentifying restriction** — under the null, **all moment conditions are valid**.

Example: Linear Model with Invalid Moment

- Consider the model:

$$Y = \beta_1' X_1 + \beta_2' X_2 + e \tag{79}$$

- Assume both X_1 and X_2 are exogenous:

$$\mathbb{E}[X_1 e] = 0 \quad \text{and} \quad \mathbb{E}[X_2 e] = 0 \tag{80}$$

- Suppose we impose the restriction $\beta_2 = 0$. Then the model simplifies to:

$$Y = \beta_1' X_1 + e \tag{81}$$

Consequences of Incorrect Exclusion

- The estimator of β_1 in the restricted model uses:

$$\mathbb{E}[X_1(Y - \beta_1' X_1)] = 0 \tag{82}$$

- However, if $\beta_2 \neq 0$, the moment condition:

$$\mathbb{E}[X_2(Y - \beta_1' X_1)] = 0 \tag{83}$$

will generally **not hold**.

Why the Moment Fails

- Substituting Y from the full model:

$$Y - \beta_1' X_1 = \beta_2' X_2 + e \tag{84}$$

- Then:

$$\mathbb{E}[X_2(Y - \beta_1' X_1)] = \mathbb{E}[X_2(\beta_2' X_2 + e)] \neq 0 \tag{85}$$

unless $\beta_2 = 0$.

Interpretation

- The restriction $\beta_2 = 0$ **removes** one regressor.
- The moment condition based on X_2 becomes **testable**.
- This is an **overidentifying restriction**: an assumption not needed for identification but testable.

Overidentification in GMM

- Overidentification arises when:

$$\# \text{ of instruments } (\ell) > \# \text{ of parameters } (k)$$

- The extra moment conditions allow us to **test** model validity using:

$$H_0 : \mathbb{E}[Ze] = 0 \tag{86}$$

- The **J-statistic** is constructed from the sample analog of $\mathbb{E}[Ze]$ and tests whether all moment conditions are jointly valid.

GMM Criterion as Test Statistic

- Sample moment condition:

$$\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\beta}_{\mathsf{gmm}}) \xrightarrow{p} \mathbb{E}[Ze] \tag{87}$$

- Use the **GMM criterion function**:

$$J(\hat{\beta}_{\mathsf{gmm}}) = n\bar{g}_n' \widehat{\Omega}^{-1}\bar{g}_n \tag{88}$$

- This is a **quadratic form** in $\bar{g}_n$, natural for testing $H_0 : \mathbb{E}[Ze] = 0$

Theorem 9 *(Asymptotic Distribution of the Overidentification Test)*

Under Assumption 12.2 in the book, as $n \to \infty$:

$$J(\hat{\beta}_{\text{gmm}}) \xrightarrow{d} \chi^2_{\ell-k} \tag{89}$$

For c such that $\alpha = 1 - G_{\ell-k}(c)$,

$$\mathbb{P}(J > c \mid H_0) \to \alpha \tag{90}$$

So the test "Reject H_0 if $J > c$" has asymptotic size α.

Interpretation

- The degrees of freedom equal the number of **overidentifying restrictions**: $\ell - k$
- A large J statistic indicates that **not all moment conditions are valid**
- The test doesn't tell you **which** moment fails — only that at least one does

Endogeneity Test

- Consider the structural model:

$$Y = Z_1'\beta_1 + Y_2'\beta_2 + e \tag{91}$$

- We assume Z_1 (included instruments) and Z_2 (excluded instruments) are exogenous:

$$\mathbb{E}[Z_1 e] = 0, \quad \mathbb{E}[Z_2 e] = 0 \tag{92}$$

- The question is whether Y_2 is endogenous.

Null and Alternative

- Null hypothesis: $H_0 : \mathbb{E}[Y_2 e] = 0$ (i.e., Y_2 is exogenous)
- Alternative: $H_1 : \mathbb{E}[Y_2 e] \neq 0$ (i.e., Y_2 is endogenous)
- This is a **subset overidentification test**: we test a subset of moment conditions for validity.

Test Construction

1. Estimate the model using **efficient GMM** with instruments (Z_1, Z_2) for (Z_1, Y_2).
   Let $\hat{J}$ be the resulting GMM criterion value.
2. Estimate again using instruments (Z_1, Z_2, Y_2) for (Z_1, Y_2).
   Let $\tilde{J}$ be the new GMM criterion.
3. Compute the test statistic:

$$C = \tilde{J} - \hat{J} \tag{93}$$

Theorem 10 *(Endogeneity Test as Difference in J Statistics)*

Under Assumption 12.2 in the book and if $\mathbb{E}[Z_2 Y_2']$ has full rank k_2, then as $n \to \infty$:

$$C \xrightarrow{d} \chi^2_{k_2} \tag{94}$$

For c such that $\alpha = 1 - G_{k_2}(c)$,

$$\mathbb{P}(C > c \mid H_0) \to \alpha \tag{95}$$

So the test "Reject H_0 if $C > c$" has asymptotic size α.

Practical Notes

- This is a **special case of the GMM overidentification test**.
- It isolates the moment conditions involving Y_2 and tests whether they are valid.

Reading

- **Hansen (2022), Econometrics.**
  - Sections: 13.1 - 13.12, 13.14, 13.21, and 13.23