Heteroscedasticity-Based Identification (Lewbel, 2012)

Notation. Let $(Y_1, Y_2)'$ be endogenous, X exogenous (contains a constant), and the $r \times 1$ vector $Z = g(X)$ any *mean-zero* function of X with $\mathbb{E}(ZZ')$ full rank.[1] Define the *linear-projection residuals*

$$W_j \;=\; Y_j - X'\mathbb{E}(XX')^{-1}\mathbb{E}(XY_j), \qquad \varepsilon_j \;=\; Y_j - \beta_j' X - \gamma_j Y_{3-j}, \quad j = 1, 2.$$

1. Structural Forms

$$\begin{aligned}
\textbf{Measurement error:} \quad & Y_1 = X'\beta_1 + Y_2^*\gamma_1 + V_1, & Y_2 = Y_2^* + U, \\
\textbf{Triangular:} \quad & Y_1 = X'\beta_1 + \gamma_1 Y_2 + \varepsilon_1, & Y_2 = X'\beta_2 + \varepsilon_2, \\
\textbf{Simultaneous:} \quad & Y_1 = X'\beta_1 + \gamma_1 Y_2 + \varepsilon_1, & Y_2 = X'\beta_2 + \gamma_2 Y_1 + \varepsilon_2.
\end{aligned}$$

Note: The measurement error model is rewritten as a triangular system with $\varepsilon_1 = -\gamma_1 U + V_1$ and $\varepsilon_2 = U + V_2$, where V_2 is the residual from projecting Y_2^* on X.

2. Core Assumptions

(A1) $\mathbb{E}|ZZ'| < \infty$ and $\mathbb{E}(ZZ')$ nonsingular.

(A2) $\mathbb{E}(Z\varepsilon_j) = 0$ and $\mathrm{Cov}(Z, \varepsilon_1\varepsilon_2) = 0$.

(A3) (Simultaneous case) the $r \times 3$ matrix $\left[\mathrm{Cov}(Z, W_1W_2) \;\; \mathrm{Cov}(Z, W_1^2) \;\; \mathrm{Cov}(Z, W_2^2)\right]$ has rank 2.

(A4) (γ_1, γ_2) and $(1/\gamma_2, 1/\gamma_1)$ cannot both lie in the parameter space unless $\gamma_1\gamma_2 = 1$.

Additional model-specific conditions

- Measurement error: U is independent of (X, Y_1, Y_2^*), $\mathbb{E}(XV_2) = 0$, $\mathrm{Cov}(Z, V_2^2) \neq 0$.
- Triangular: $\mathrm{Cov}(Z, \varepsilon_2^2) \neq 0$.
- Simultaneous: $\mathrm{Cov}(Z, \varepsilon_j^2) \neq 0$ for at least one j.

3. Internal Instruments (Two-Stage Least Squares)

For the triangular system (including measurement error as a special case):

$$\text{Instruments for } Y_2: \quad X \text{ and } (Z - \bar{Z})\hat{\varepsilon}_2$$

where $\hat{\varepsilon}_2$ are residuals from regressing Y_2 on X.

4. Point Identification

Theorem 1 (Triangular/Measurement Error)**.** *Under (A1)–(A2) and* $\mathrm{Cov}(Z, \varepsilon_2^2) \neq 0$, $(\beta_1, \beta_2, \gamma_1)$ *are point identified.*

Theorem 2 (Simultaneous)**.** *Under (A1)–(A4) and non-degenerate* $\mathrm{Cov}(Z, \varepsilon_j^2)$, $(\beta_1, \beta_2, \gamma_1, \gamma_2)$ *are point identified.*

5. GMM Moment Conditions

Parameter notation:

- $\mu = \mathbb{E}[Z]$ is the population mean of Z (unknown parameter)
- $\bar{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$ is the sample mean of Z (observable)
- In GMM: μ is estimated jointly with other parameters
- In 2SLS: $(Z - \bar{Z})\hat{\varepsilon}_2$ is used directly as an instrument

[1] Centring Z rules out a constant in Z and guarantees $\mathrm{Cov}(Z, \varepsilon_j) = 0$ while allowing $\mathrm{Cov}(Z, \varepsilon_j^2) \neq 0$.

(i) Triangular system (includes measurement error)

$$Q_{\mathrm{TRI}}(\theta) = \begin{pmatrix} X\varepsilon_1 \\ X\varepsilon_2 \\ Z-\mu \\ (Z-\mu)\varepsilon_1\varepsilon_2 \end{pmatrix}, \quad \mathbb{E}\big[Q_{\mathrm{TRI}}(\theta)\big] = 0, \quad \theta = (\beta_1, \beta_2, \gamma_1, \mu)'.$$

(ii) Simultaneous system

$$Q_{\mathrm{SIM}}(\theta) = \begin{pmatrix} X\varepsilon_1 \\ X\varepsilon_2 \\ Z-\mu \\ (Z-\mu)\varepsilon_1\varepsilon_2 \end{pmatrix}, \quad \mathbb{E}\big[Q_{\mathrm{SIM}}(\theta)\big] = 0, \quad \theta = (\beta_1, \beta_2, \gamma_1, \gamma_2, \mu)'.$$

Special case: If $\mathbb{E}[\varepsilon_1\varepsilon_2] = 0$, then μ can be dropped from θ, and the moment condition simplifies to $\mathbb{E}[Z\varepsilon_1\varepsilon_2] = 0$.

6. Set Identification (Relaxed A2)

Relaxed moment. Allow

$$|\operatorname{Corr}(Z, \varepsilon_1\varepsilon_2)| \;\le\; \tau\,|\operatorname{Corr}(Z, \varepsilon_2^2)|, \qquad 0 \le \tau < 1.$$

Theorem 3 (Bounds for the triangular model)**.** *Assume conditions for the triangular model hold except that the above relaxed moment replaces* $\operatorname{Cov}(Z, \varepsilon_1\varepsilon_2) = 0$ *and that* $\operatorname{Cov}(Z, W_2^2) \neq 0$, $\operatorname{Var}(W_2^2) > 0$*. Then* γ_1 *lies in the closed interval* Γ_1 *whose endpoints are the (real) roots of*

$$\frac{\big[\operatorname{Cov}(Z, W_1W_2)\big]^2}{\big[\operatorname{Cov}(Z, W_2^2)\big]^2} - \frac{\operatorname{Var}(W_1W_2)}{\operatorname{Var}(W_2^2)}\tau^2 + 2\Big(\frac{\operatorname{Cov}(W_1W_2, W_2^2)}{\operatorname{Var}(W_2^2)}\tau^2 - \frac{\operatorname{Cov}(Z, W_1W_2)}{\operatorname{Cov}(Z, W_2^2)}\Big)\gamma_1 + (1-\tau^2)\gamma_1^2 = 0. \quad (1)$$

Properties.

- $\tau = 0 \;\Rightarrow\; \Gamma_1$ collapses to the point-identified γ_1 of the triangular model.
- Γ_1 widens monotonically in τ.
- For any $\gamma_1 \in \Gamma_1$, $\beta_1 = \mathbb{E}(XX')^{-1}\mathbb{E}\big[X(Y_1 - Y_2\gamma_1)\big]$.

Estimation. Replace population moments in (**??**) by sample moments, solve the quadratic for the two roots, and obtain standard errors by the delta method.