Heteroscedasticity-Based Identification

June 12, 2025

1 Preliminaries and Notation

This document summarizes the heteroscedasticity-based identification strategy for models with endogenous regressors, as developed by Lewbel (2012). The method provides a way to construct valid instruments from the model's data when traditional external instruments are unavailable.

Throughout, $Z = g(X)$ denotes a transformation of the exogenous variables, constructed to be mean-zero.[1] All instruments are therefore understood to be mean-zero.

2 Structural Forms and Reduced-Form Residuals

Let the endogenous vector be $(Y_1, Y_2)'$, X a vector of exogenous variables (including a constant), and $Z = g(X)$ as above.

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

Projecting each Y_j on X yields *reduced-form residuals* $W_j \triangleq Y_j - X'(\mathbb{E}[XX'])^{-1}\mathbb{E}[XY_j]$. A short calculation gives

$$W_1 = \frac{\varepsilon_1 + \gamma_1\varepsilon_2}{1 - \gamma_1\gamma_2}, \qquad W_2 = \frac{\varepsilon_2 + \gamma_2\varepsilon_1}{1 - \gamma_1\gamma_2},$$

with the triangular case obtained by setting $\gamma_2 = 0$.

Remark 1. *The algebra follows directly from solving the two-equation system for the reduced form and collecting the error terms.*

3 Core Assumptions for Point Identification

The method relies on the following key assumptions:

(A1) **Strict exogeneity.** $\mathbb{E}[\varepsilon_j \mid X] = 0$ for $j = 1, 2$. *Time-series note:* in Section 8 we weaken this to a martingale-difference assumption and add HAC inference.

(A2) **Covariance restriction.** $\mathrm{Cov}(Z, \varepsilon_1\varepsilon_2) = 0$. This holds automatically if the errors have a common factor structure $\varepsilon_j = a_j u + \eta_j$ where the common factor u is uncorrelated with and mean independent of Z.

(A3) **Instrument relevance via heteroscedasticity.**

- *Triangular:* $\mathrm{Cov}(Z, \varepsilon_2^2) \neq 0$.
- *Simultaneous:* the $r \times 2$ matrix $\Phi_W = [\,\mathrm{Cov}(Z, W_1^2) \;\; \mathrm{Cov}(Z, W_2^2)\,]$ has rank 2.

(A4) **Normalization (simultaneous case).** The parameter space for (γ_1, γ_2) precludes the observationally equivalent pair $(1/\gamma_2, 1/\gamma_1)$ unless $\gamma_1\gamma_2 = 1$.

[1] For simplicity, we write Z to mean the centered variable $Z - \mathbb{E}[Z]$. In practice, this means we use the sample-demeaned version, $Z_i - \bar{Z}$, in all calculations.

4 Triangular System: Closed-Form Identification and 2SLS

4.1 Closed-form

$$\gamma_1 = \frac{\mathrm{Cov}(Z, W_1 W_2)}{\mathrm{Cov}(Z, W_2^2)}. \tag{1}$$

Remark 2 (Why (1) identifies γ_1). *Under (A2) the numerator simplifies to* $\mathrm{Cov}(Z, \varepsilon_1\varepsilon_2) + \gamma_1 \mathrm{Cov}(Z, \varepsilon_2^2) = \gamma_1 \mathrm{Cov}(Z, \varepsilon_2^2)$. *Because the denominator is non-zero by (A3),* γ_1 *is point identified.*

4.2 Feasible two-step 2SLS

1. **Generate residuals.** Regress Y_2 on X via OLS and store the residuals $\hat{\varepsilon}_2 = Y_2 - X'\hat{\beta}_2^{\mathrm{OLS}}$.
2. **Construct the heteroscedasticity-based instrument.** The generated instrument is $IV = (Z - \bar{Z})\hat{\varepsilon}_2$.
3. **First stage.** Regress the endogenous variable Y_2 on the exogenous variables and the generated instrument, $[X, IV]$, to obtain fitted values $\hat{Y}_2$.
4. **Second stage.** Regress Y_1 on $[X, \hat{Y}_2]$ to estimate (β_1, γ_1).

Practical guidance. Weak-instrument concerns apply because the instrument is generated. Report the first-stage F-statistic on IV and, where necessary, use weak-IV-robust inference. Standard errors should account for the first-stage estimation step, typically via GMM or bootstrap.

5 GMM Moment Conditions

The identification strategy can be expressed more generally using GMM. The core idea is to form moment conditions that are zero at the true parameter values θ.

5.1 Triangular system

The parameter vector is $\theta = (\beta_1', \gamma_1, \beta_2')'$. The moment vector is formed by multiplying instruments by the model's structural errors, which are defined as $\varepsilon_1(\theta) = Y_1 - X'\beta_1 - Y_2\gamma_1$ and $\varepsilon_2(\theta) = Y_2 - X'\beta_2$.

$$Q_{\mathrm{TRI}}(\theta) = \begin{pmatrix} X \cdot \varepsilon_1(\theta) \\ X \cdot \varepsilon_2(\theta) \\ Z \cdot \varepsilon_1(\theta) \cdot \varepsilon_2(\theta) \end{pmatrix}. \tag{2}$$

The instruments are constructed directly from the exogenous data:

- For the two mean equations, the instruments are the exogenous variables X.
- For the covariance restriction, the instrument is the heteroscedasticity driver Z.

Since the errors are functions of data and parameters, the entire moment vector is observable for any candidate θ.

5.2 Simultaneous system

The parameter vector is $\theta = (\beta_1', \gamma_1, \beta_2', \gamma_2)'$. The structural errors are now $\varepsilon_1(\theta) = Y_1 - X'\beta_1 - Y_2\gamma_1$ and $\varepsilon_2(\theta) = Y_2 - X'\beta_2 - Y_1\gamma_2$.

$$Q_{\mathrm{SIM}}(\theta) = \begin{pmatrix} X \cdot \varepsilon_1(\theta) \\ X \cdot \varepsilon_2(\theta) \\ Z \cdot \varepsilon_1(\theta) \cdot \varepsilon_2(\theta) \end{pmatrix}. \tag{3}$$

The instruments are again constructed from exogenous data: X for the mean equations and Z for the covariance restriction. Estimation proceeds by finding the parameter vector $\hat{\theta}$ that minimizes the sample analog of the moment conditions, $\bar{Q}(\theta)'W\bar{Q}(\theta)$, where W is a weighting matrix.

6 Set Identification Under a Relaxed Covariance Restriction

When assumption (A2) is weakened to allow for a small correlation, $|\operatorname{Corr}(Z, \varepsilon_1\varepsilon_2)| \leq \tau|\operatorname{Corr}(Z, \varepsilon_2^2)|,\ \tau \in [0, 1)$, the parameter γ_1 in the triangular model is set-identified rather than point-identified.

Theorem 1 (Bounds for γ_1 with $\tau > 0$)**.** *γ_1 is contained in the closed interval whose endpoints are the (real) roots of the quadratic equation in γ_1:*

$$\frac{\operatorname{Cov}(Z, W_1W_2)^2}{\operatorname{Cov}(Z, W_2^2)^2} - \frac{\operatorname{Var}(W_1W_2)}{\operatorname{Var}(W_2^2)}\tau^2 + 2\left(\frac{\operatorname{Cov}(W_1W_2, W_2^2)}{\operatorname{Var}(W_2^2)}\tau^2 - \frac{\operatorname{Cov}(Z, W_1W_2)}{\operatorname{Cov}(Z, W_2^2)}\right)\gamma_1 + (1 - \tau^2)\gamma_1^2 = 0.$$

The interval collapses to the point estimate from (1) when $\tau = 0$ and widens as $\tau \to 1$.

Identification of other parameters. The remaining parameters are identified conditional on a value of γ_1 from its identified set.

- The parameter β_2 is always point-identified by OLS: $\hat{\beta}_2 = (\mathbb{E}[X'X])^{-1}\mathbb{E}[X'Y_2]$.
- For each value $\gamma_{1,k}$ in the identified interval for γ_1, there is a corresponding identified value for β_1, given by $\beta_{1,k} = (\mathbb{E}[X'X])^{-1}\mathbb{E}[X'(Y_1 - Y_2\gamma_{1,k})]$.

The result is an identified set of parameter pairs (β_1, γ_1) corresponding to the interval for γ_1.

7 Two Illustrative Extensions

The core idea can be adapted to other contexts by creatively defining the heteroscedasticity-generating variable Z.

7.1 Conditional heteroscedasticity (Prono)

Context and Model. In a time-series setting, the structural model is a triangular system:

$$\begin{aligned} Y_{1t} &= X_t'\beta_1 + \gamma_1 Y_{2t} + \varepsilon_{1t} \\ Y_{2t} &= X_t'\beta_2 + \varepsilon_{2t} \end{aligned}$$

The key insight is that the error variance may be time-varying and predictable. Prono's extension assumes ε_{2t} follows a GARCH process, where its conditional variance is a function of past errors and variances:

$$\operatorname{Var}(\varepsilon_{2t} \mid \mathcal{F}_{t-1}) = \sigma_{2t}^2 = \omega + \alpha\varepsilon_{2,t-1}^2 + \beta\sigma_{2,t-1}^2.$$

Procedure. The fitted conditional variance from the GARCH model serves as the heteroscedasticity driver.

1. Estimate the second equation by OLS to get residuals $\hat{\varepsilon}_{2t} = Y_{2t} - X_t'\hat{\beta}_2$.
2. Fit a GARCH(1,1) model to the residuals $\hat{\varepsilon}_{2t}$ to obtain the series of fitted conditional variances, $\hat{\sigma}_{2t}^2$.
3. This fitted variance is the heteroscedasticity driver: set $Z_t = \hat{\sigma}_{2t}^2$.
4. Construct the generated instrument: $IV_t = (Z_t - \bar{Z})\hat{\varepsilon}_{2t}$.
5. Proceed with 2SLS as in Section 4.2, using $[X_t, IV_t]$ as instruments for Y_{2t} in the first structural equation. Use HAC-robust standard errors.

7.2 Regime heteroscedasticity (Rigobon)

Context and Model. The model is a simultaneous system where the error variances differ across observable, discrete regimes (e.g., pre- and post-policy change, or high- vs. low-volatility periods).

$$Y_1 = X'\beta_1 + \gamma_1 Y_2 + \varepsilon_1$$
$$Y_2 = X'\beta_2 + \gamma_2 Y_1 + \varepsilon_2$$

The key assumption is that for at least one error term ε_j, its variance changes across regimes s, while the covariance between the errors remains constant:

$$\mathrm{Var}(\varepsilon_j \mid s) \neq \mathrm{Var}(\varepsilon_j \mid s') \quad \text{for } s \neq s'$$
$$\mathrm{Cov}(\varepsilon_1, \varepsilon_2 \mid s) = \text{constant for all } s.$$

Procedure. The regime indicators are used to generate the instrument.

1. Estimate the second equation by OLS to get residuals $\hat{\varepsilon}_2$.
2. Create a set of dummy variables $\{D_1, \ldots, D_S\}$ for the regimes.
3. The heteroscedasticity drivers are these dummies: set Z to be the vector of centered dummies, $Z_s = D_s - p_s$, where p_s is the sample proportion of observations in regime s.
4. Construct the generated instrument(s): $IV_s = Z_s\hat{\varepsilon}_2$.
5. Proceed with 2SLS, using $[X, IV_1, \ldots, IV_S]$ as instruments for Y_2 in the main equation.

8 Time-Series Variant with Log-Linear Conditional Variances

Context and Model. This extension formalizes identification within a GMM framework for a stationary and mixing time series. The structural model is a simultaneous system:

$$Y_{1t} = X_t'\beta_1 + \gamma_1 Y_{2t} + \varepsilon_{1t}$$
$$Y_{2t} = X_t'\beta_2 + \gamma_2 Y_{1t} + \varepsilon_{2t}$$

The key assumption is that the conditional variances are an explicit log-linear function of the exogenous variables X_t:

$$\log \sigma_{jt}^2 = X_t'\delta_j, \quad \text{where} \quad \varepsilon_{jt} \mid \mathcal{F}_{t-1} \sim (0, \sigma_{jt}^2) \quad \text{for } j = 1, 2.$$

Since (δ_1, δ_2) are assumed non-zero, X_t is correlated with ε_{jt}^2, satisfying the instrument relevance condition (A3). The heteroscedasticity driver is defined as the centered exogenous variables, $Z_t = X_t - \mathbb{E}[X_t]$.

8.1 Moment vector

The full set of parameters is $\theta = (\beta_1', \beta_2', \gamma_1, \gamma_2, \delta_1', \delta_2')'$. The moment conditions implied by the model are $\mathbb{E}[Q_t(\theta)] = 0$, where:

$$Q_t(\theta) = \begin{pmatrix} X_t\varepsilon_{1t}(\theta) \\ X_t\varepsilon_{2t}(\theta) \\ Z_t\varepsilon_{1t}(\theta)\varepsilon_{2t}(\theta) \\ Z_t\Big(\varepsilon_{1t}(\theta)^2 - e^{X_t'\delta_1}\Big) \\ Z_t\Big(\varepsilon_{2t}(\theta)^2 - e^{X_t'\delta_2}\Big) \end{pmatrix}.$$

Instruments as Functions of Observables. The instruments are derived from the exogenous variables X_t. The terms inside the expectation are functions of these instruments and the structural errors (e.g., $\varepsilon_{1t}(\theta) = Y_{1t} - X_t'\beta_1 - Y_{2t}\gamma_1$), which are themselves functions of the observable data and the parameter vector θ.

- For the first two mean equations, the instruments are X_t.
- For the error product and variance specification moments, the instrument is $Z_t = X_t - \mathbb{E}[X_t]$, estimated in-sample as $X_t - \bar{X}$.

Therefore, the sample average of $Q_t(\theta)$ is a computable criterion function for any candidate parameter values.

8.2 Estimation

Estimation proceeds via a multi-step GMM procedure:

1. Obtain initial estimates of mean parameters via OLS/IV, and get residuals $\hat{\varepsilon}_{jt}$.
2. Estimate the variance parameters δ_j by regressing $\log \hat{\varepsilon}_{jt}^2$ on X_t.
3. Use the full set of moment conditions defined in $Q_t(\theta)$ to form a GMM criterion function.
4. Minimize the GMM objective function using a HAC-robust weighting matrix (e.g., Newey-West) to obtain efficient, consistent, and asymptotically normal estimates of $\hat{\theta}$. A conventional bandwidth choice is $\ell_T = \lfloor 4(T/100)^{2/9} \rfloor$.

9 Diagnostic Tests and Practical Checks

- **Instrument relevance.** Report the first-stage F-statistic on the generated instrument(s); if $F < 10$, standard inference is unreliable, and weak-IV-robust tests should be used.
- **Instrument validity.** If there are more heteroscedasticity drivers Z than needed for identification (i.e., the model is overidentified), a Hansen J-test of overidentifying restrictions can be used to test the validity of the moment conditions.
- **Endogeneity of Y_2.** The endogeneity of Y_2 can be tested using a difference-in-Hansen test (C-statistic) or a Hausman-style test comparing the OLS and 2SLS estimates of γ_1.
- **Heteroscedasticity of ε_2.** The crucial assumption (A3) can be checked with a Breusch–Pagan or White test for heteroscedasticity, by regressing the squared residuals $\hat{\varepsilon}_2^2$ on the proposed driver(s) Z. A significant relationship provides evidence for instrument relevance.