# Linear Models Lecture 14: 2SLS, LATE, and the MTE Framework

Robert Gulotty

University of Chicago

February 26, 2026

# Control Function Regression

- Structural model with endogenous $X_2$:

$$Y = \boldsymbol{x}_1'\beta_1 + \boldsymbol{x}_2'\beta_2 + e$$

$$\boldsymbol{x}_2 = \Gamma_{12}'\boldsymbol{z}_1 + \Gamma_{22}'\boldsymbol{z}_2 + u_2$$

- The **control function** approach directly models the dependence between $e$ and $u_2$:

$$e = u_2'\alpha + v, \qquad \alpha = (E[u_2 u_2'])^{-1}E[u_2 e], \qquad E[u_2 v] = 0$$

- Substituting into the structural equation:

$$Y = X_1'\beta_1 + X_2'\beta_2 + u_2'\alpha + v$$

- After controlling for $u_2$: $E[X_1 v] = E[X_2 v] = E[u_2 v] = 0$.

# Control Function: Implementation

- **Step 1:** Estimate the first stage by OLS to get residuals:

$$\hat{u}_{2i} = X_{2i} - \hat{\Gamma}'_{12} Z_{1i} - \hat{\Gamma}'_{22} Z_{2i}$$

- **Step 2:** Include $\hat{u}_2$ as an additional regressor:

$$Y = X\hat{\beta} + \hat{U}_2\hat{\alpha} + \hat{v}$$

- This is "subtracting off the endogenous part" of $X_2$.
- Under homoskedasticity, the control function estimator of $\beta$ is numerically identical to 2SLS.
- The coefficient $\hat{\alpha}$ directly tests endogeneity — if $\alpha = 0$, then $X_2$ is exogenous.

# R: Control Function

```r
# Step 1: First stage
first_stage <- lm(X2 ~ Z1 + Z2, data = dat)
dat$u2_hat <- residuals(first_stage)

# Step 2: Structural equation with control
cf_reg <- lm(Y ~ X1 + X2 + u2_hat, data = dat)
summary(cf_reg)
# coef on u2_hat tests endogeneity (t-test on alpha)
# coef on X2 is the IV estimate of beta_2
```

The control function makes identification transparent: we "control for" the endogenous component of $X_2$. The $t$-test on $\hat{\alpha}$ is equivalent to the Hausman test.

# IV Has No Finite Moments

- A striking result (Kinal 1980, Kiviet): the IV estimator has at most $q = l - k$ moments, where $q$ is the number of overidentifying restrictions.
- When just identified ($q = 0$): IV has no expectation and no variance.
- Why? Consider $\hat{\beta}_{IV} = (Z'X)^{-1}Z'Y$. The denominator $Z'X$ can be arbitrarily close to zero, creating Cauchy-like tails.
- Consequences:
  - The **bootstrap fails** in the just-identified case — it requires finite variance.
  - "Bias" is not well-defined in the usual sense — we work with approximate bias instead.

# Imprecision of IV

- Suppose $Z$ and $X$ are scalar, mean zero. Compare asymptotic variances:

$$Avar(\hat{\beta}_{OLS}) = \frac{\sigma_e^2}{n} \cdot \frac{1}{\text{var}(x)}$$

$$Avar(\hat{\beta}_{IV}) = \frac{\sigma_e^2}{n} \cdot \frac{1}{\text{var}(x)} \cdot \frac{1}{\rho_{xz}^2}$$

- Therefore:

$$\boxed{Avar(\hat{\beta}_{IV}) = Avar(\hat{\beta}_{OLS}) \cdot \frac{1}{\rho_{xz}^2}}$$

- As $\rho_{xz}^2 \to 0$: $Avar(\hat{\beta}_{IV}) \to \infty$.
- IV is **always** less precise than OLS — the cost of correcting for endogeneity.

# Decomposing the IV Estimation Error

- Write: $y = X\beta + e$ and $X = Z\pi + v$. Then:

$$\begin{aligned}
\hat{\beta}_{IV} &= (X'P_Z X)^{-1} X'P_Z y \\
&= \beta + (X'P_Z X)^{-1} X'P_Z e \\
&= \beta + (X'P_Z X)^{-1} (\pi' Z' + v')P_Z e \\
&= \beta + \underbrace{(X'P_Z X)^{-1} \pi' Z' e}_{\text{Term A}} + \underbrace{(X'P_Z X)^{-1} v' P_Z e}_{\text{Term B}}
\end{aligned}$$

- **Term A:** Involves $Z'e$. Since $E[Z'e] = 0$, this vanishes in expectation.
- **Term B:** Involves $v'P_Z e$. Since $v$ and $e$ are correlated (endogeneity!), this does **not** vanish.
- Term B is the source of finite-sample bias.

# Approximate Bias of IV

Taking approximate expectations (since exact ones may not exist):

$$\begin{aligned} E(\hat{\beta}_{IV}) - \beta &\approx (E(X'P_Z X))^{-1} E(v'P_Z e) \\ &= (E[(\pi'Z' + v')P_Z(Z\pi + v)])^{-1} E[v'P_Z e] \\ &= (E[\pi'Z'Z\pi] + E[v'P_Z v])^{-1} E[v'P_Z e] \end{aligned}$$

where the cross terms vanish because $E[Z'v] = 0$.
Using $E[v'P_Z v] = \sigma_v^2 \cdot p$ and $E[v'P_Z e] = \sigma_{ev} \cdot p$ (where $p = l$ is the number of instruments):

$$E(\hat{\beta}_{IV}) - \beta \approx (E[\pi'Z'Z\pi] + \sigma_v^2 p)^{-1} \sigma_{ev} \cdot p$$

# Bias in Terms of the F-Statistic

Dividing numerator and denominator by $\sigma_v^2 p$:

$$E(\hat{\beta}_{IV}) - \beta \approx \frac{1}{\left(\frac{E[\pi' Z' Z \pi]/p}{\sigma_v^2} + 1\right)} \cdot \frac{\sigma_{ev}}{\sigma_v^2} \approx \frac{1}{1 + F_{p,n-p}} \cdot \frac{\sigma_{ev}}{\sigma_v^2}$$

- $F$ is the first-stage F-statistic testing $H_0 : \pi = 0$.
- The term $\frac{\sigma_{ev}}{\sigma_v^2}$ is exactly the **OLS bias** (endogeneity).
- Key implications:
  - As $F \to \infty$: bias $\to 0$ (strong instruments eliminate bias)
  - As $F \to 0$: bias $\to \frac{\sigma_{ev}}{\sigma_v^2}$ = OLS bias (IV provides no correction)
  - Adding irrelevant instruments increases $p$, decreases $F$, increases bias

# Rule of Thumb and Modern Diagnostics

- **Staiger & Stock (1997):** Rule of thumb $F > 10$ for reliable IV.
- **Stock & Yogo (2005):** Critical values for the F-statistic that ensure IV bias is at most $x\%$ of OLS bias.
- **Lee et al. (2022):** The $tF$ procedure — adjust critical values for the $t$-test based on the first-stage $F$.
- The bias formula explains why:
  - "Fishing" for instruments is dangerous — more instruments $\Rightarrow$ more bias
  - A single strong instrument is often better than many weak ones
  - The just-identified case ($p = k_2$) minimizes the finite-sample bias problem

# Weak Instruments: OLS Asymptotic Bias

Scalar case: single $x$, single instrument $z$. An instrument is **weak** if $\rho_{zx}$ is small.

$$\begin{aligned}
\text{plim } \hat{\beta}_{OLS} &= \text{plim } \frac{\text{cov}(x, y)}{\text{var}(x)} = \text{plim } \frac{\text{cov}(x, x\beta + e)}{\text{var}(x)} \\
&= \beta + \text{plim } \frac{\text{cov}(x, e)}{\text{var}(x)} \\
&= \beta + \frac{\text{cov}(x, e)}{\sqrt{\text{var}(x)}\sqrt{\text{var}(e)}} \cdot \frac{\sqrt{\text{var}(e)}}{\sqrt{\text{var}(x)}} \\
&= \beta + \rho_{xe}\frac{\sigma_e}{\sigma_x}
\end{aligned}$$

The OLS asymptotic bias is $\rho_{xe}\frac{\sigma_e}{\sigma_x}$.

# Weak Instruments: IV Asymptotic Bias

$$\begin{aligned}
\text{plim}\ \hat{\beta}_{IV} &= \text{plim}\ \frac{\text{cov}(z,y)}{\text{cov}(z,x)} = \beta + \text{plim}\ \frac{\text{cov}(z,e)}{\text{cov}(z,x)} \\
&= \beta + \frac{\rho_{ze}}{\rho_{zx}} \cdot \frac{\sigma_e}{\sigma_x} \\
&= \beta + \frac{\rho_{ze}}{\rho_{zx}\rho_{xe}} \cdot \underbrace{\rho_{xe}\frac{\sigma_e}{\sigma_x}}_{ABias(\hat{\beta}_{OLS})}
\end{aligned}$$

- If the instrument is perfectly valid ($\rho_{ze} = 0$): IV is consistent regardless of $\rho_{zx}$.
- If the instrument is even slightly invalid ($\rho_{ze} \neq 0$): weak relevance ($\rho_{zx} \approx 0$) amplifies the bias.

# When IV Is Worse Than OLS

$$\frac{ABias(\hat{\beta}_{IV})}{ABias(\hat{\beta}_{OLS})} = \frac{\rho_{ze}}{\rho_{zx}\rho_{xe}}$$

If $\frac{\rho_{ze}}{\rho_{zx}\rho_{xe}} \geq 1$, then IV is more biased than OLS.

- **Example:** $\rho_{xe} = 0.5$ (very endogenous $X$), $\rho_{ze} = 0.01$ (barely invalid $Z$), $\rho_{zx} = 0.019$ (weak instrument):

$$\frac{0.01}{0.019 \times 0.5} = 1.052 \quad \Rightarrow \quad \text{IV bias exceeds OLS bias}$$

- Approximate relationship to the F-statistic:

$$\frac{ABias(\hat{\beta}_{IV})}{ABias(\hat{\beta}_{OLS})} \approx \frac{1}{F}$$

An $F$ of 100 means IV is $\sim$1% as biased as OLS. An $F$ of 5 means $\sim$20%.

# Testing Endogeneity: Durbin-Wu-Hausman Test

- If $X$ is exogenous, both OLS and IV are consistent, but OLS is efficient (BLUE).
- If $X$ is endogenous, only IV is consistent.
- The **Hausman test** compares the two:

$$H = (\hat{\beta}_{IV} - \hat{\beta}_{OLS})'[Avar(\hat{\beta}_{IV}) - Avar(\hat{\beta}_{OLS})]^{-1}(\hat{\beta}_{IV} - \hat{\beta}_{OLS}) \sim \chi^2_{k_2}$$

- Rejecting $H_0$: either $X$ is endogenous **or** $Z$ is an invalid instrument.
- Equivalent to testing $\hat{\alpha} = 0$ in the control function regression.

# Testing Overidentifying Restrictions: Sargan Test

- When $l > k$, we have $q = l - k$ "extra" moment conditions.
- If the model is correct, all moment conditions should be approximately satisfied.
- The **Sargan test** (also called Hansen's $J$-test):

$$J = n \cdot \hat{e}' P_Z \hat{e} / \hat{\sigma}^2 \sim \chi^2_q \quad \text{under } H_0$$

where $\hat{e} = Y - X\hat{\beta}_{2SLS}$.

- Rejecting $H_0$: at least one instrument is invalid (correlated with $e$).
- Limitation: If **all** instruments are invalid in the same way, the test has no power.

**GMM connection:** The Sargan/$J$-test is the overidentification test of GMM.

# R: IV Estimation and Diagnostics

```r
library(estimatr); library(lmtest)

# 2SLS with robust SEs
iv_fit <- iv_robust(Y ~ X2 + X1 | Z2 + X1, data = dat)
summary(iv_fit)  # coefs, robust SEs, first-stage F

# First-stage F-test (check instrument strength)
first <- lm(X2 ~ Z2 + X1, data = dat)
linearHypothesis(first, "Z2 = 0")

# OLS for comparison
ols_fit <- lm(Y ~ X2 + X1, data = dat)
```

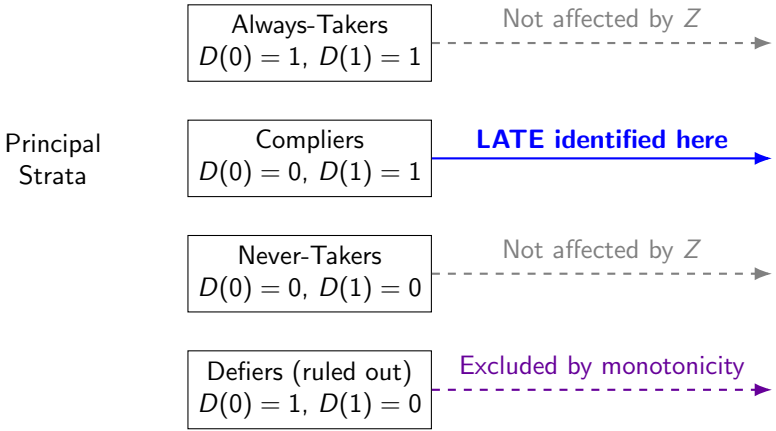
# What Does IV Estimate Under Heterogeneity?

- So far: IV estimates the "structural parameter" $\beta$.
- But what if treatment effects are **heterogeneous**? $\beta_i$ varies across individuals.
- Angrist & Imbens (1994): with a binary instrument, IV estimates the **Local Average Treatment Effect**:

$$\hat{\beta}_{IV} \xrightarrow{p} \text{LATE} = E[Y_i(1) - Y_i(0) \mid \text{Compliers}]$$

- "Compliers" = individuals whose treatment status is changed by the instrument.

# Compliance Types



# LATE Assumptions (Angrist-Imbens)

**Four assumptions** for LATE identification:

1. **Independence:** $(Y_i(0), Y_i(1), D_i(0), D_i(1)) \perp Z_i$
2. **Exclusion:** $Z_i$ affects $Y_i$ only through $D_i$
3. **Monotonicity:** $D_i(1) \geq D_i(0)$ for all $i$ (no defiers)
4. **First stage:** $E[D_i \mid Z_i = 1] \neq E[D_i \mid Z_i = 0]$

Under these assumptions, the Wald estimator identifies:

$$\frac{E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]}{E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0]} = E[Y_i(1) - Y_i(0) \mid \text{Compliers}]$$

LATE is **instrument-dependent**: different instruments $\Rightarrow$ different compliers $\Rightarrow$ different LATEs.

# Example: Returns to Education

- Model: $\text{lwage}_i = \alpha + \gamma_1 \text{educ}_i + v_i \cdot \text{educ}_i + u_i$
- Heterogeneous returns: $\beta_i = \gamma_1 + v_i$
- If we instrument education with quarter of birth (Angrist & Krueger 1991):
  - Compliers = those who stay in school because of compulsory schooling laws
  - LATE = return to education for **marginal students** (those induced to stay)
  - This may differ from ATE or ATT
- **Policy relevance:** LATE answers "what is the return for people affected by the policy?" — exactly the right parameter for evaluating that policy.

# The Generalized Roy Model

- Potential outcomes: $Y_i(1)$, $Y_i(0)$ with heterogeneous gains $\Delta_i = Y_i(1) - Y_i(0)$.
- The treatment decision is based on **net utility**:

$$D_i^* = \mu_D(Z_i) - U_{Di}$$

$$D_i = \mathbb{1}[D_i^* > 0] = \mathbb{1}[\mu_D(Z_i) > U_{Di}]$$

- $\mu_D(Z_i)$: observable component of the treatment decision (driven by instruments).
- $U_{Di}$: unobservable **resistance to treatment** — the utility cost of participating.
  - Low $U_{Di}$: individual has low cost $\Rightarrow$ eager to participate.
  - High $U_{Di}$: individual has high cost $\Rightarrow$ reluctant to participate.
- Normalize $U_D \sim U[0, 1]$ (via probability integral transform).

# Utility and Selection

- In the Roy framework, agents choose treatment when net utility is positive:

$$D_i = 1 \quad \Leftrightarrow \quad \underbrace{\text{Expected benefit of treatment}}_{\text{depends on } \Delta_i} > \underbrace{\text{Cost of treatment}}_{U_{Di}}$$

- The propensity score $P(Z_i) = \Pr(D_i = 1 \mid Z_i) = \mu_D(Z_i)$ summarizes the instrument.
- An individual participates iff $U_{Di} \leq P(Z_i)$.
- **Key insight:** $U_D$ orders individuals from most to least eager.
  - $U_D \approx 0$: would participate under almost any instrument value ("always-takers")
  - $U_D \approx 1$: would almost never participate ("never-takers")
  - $U_D \in [P(z_0), P(z_1)]$: participate when $Z = z_1$ but not $Z = z_0$ ("compliers")

# The Marginal Treatment Effect (Heckman & Vytlacil 2005)

## Definition (Marginal Treatment Effect)

$$\Delta^{MTE}(x, u_D) = E[Y(1) - Y(0) \mid X = x, U_D = u_D]$$

- MTE is the treatment effect for individuals **at the margin** — those who are just indifferent between treatment and control when their unobserved resistance equals $u_D$.
- As $u_D$ increases from 0 to 1, we trace out effects from the most eager to the most reluctant:
  - Low $u_D$: low-cost individuals (high utility from treatment) — often highest returns.
  - High $u_D$: high-cost individuals — often lowest returns (if gains and costs are correlated).
- MTE is typically **downward-sloping**: those who self-select into treatment tend to benefit most (essential heterogeneity).

# All Treatment Parameters as Weighted MTE

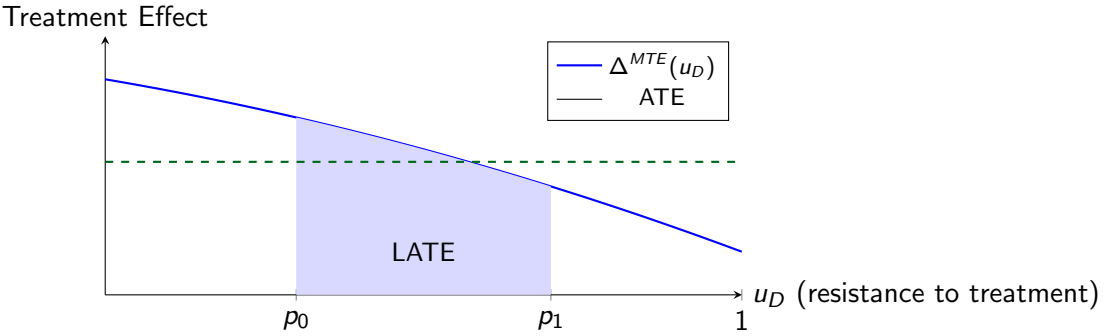
- Heckman & Vytlacil's key insight: **every** treatment parameter is a weighted average of MTE:

$$\Delta^j(x) = \int_0^1 \Delta^{MTE}(x, u_D)\, \omega_j(x, u_D)\, du_D$$

| **Parameter** | **Weight** $\omega_j$ | **Who?** |
|---|---|---|
| ATE | 1 | Everyone |
| ATT | $\frac{u_D}{E[D\|X]}$ (overweights low $u_D$) | Treated |
| ATU | $\frac{1-u_D}{E[1-D\|X]}$ (overweights high $u_D$) | Untreated |
| LATE | $\frac{1}{u'_D - u_D}\mathbb{1}[u_D \leq t \leq u'_D]$ | Compliers |

- LATE is MTE averaged over a **specific interval** of $u_D$ values.

# Visualizing the MTE Curve




LATE averages MTE over complier interval $[p_0, p_1]$. ATE averages the entire curve.

# What Does Conventional IV Recover?

- With a continuous instrument, 2SLS recovers a **variance-weighted average** of LATEs:

$$\beta_{IV} = \int_0^1 \Delta^{MTE}(u_D)\, \omega_{IV}(u_D)\, du_D$$

  where $\omega_{IV}$ weights by the "first-stage intensity" at each margin.
- **Key points** (Hull 2024):
  - Under monotonicity + exclusion: IV estimates a positively-weighted average of individual effects
  - Without monotonicity: weights can be negative (hard to interpret)
  - The "model" matters for extrapolation beyond the complier population

## Policy Relevance

- Different policies target different margins of the MTE curve.
- The **Policy-Relevant Treatment Effect** (PRTE):

$$\Delta^{PRTE} = \int_0^1 \Delta^{MTE}(u_D)\, \omega_{PRTE}(u_D)\, du_D$$

  where $\omega_{PRTE}$ depends on how the policy shifts participation.
- LATE from one instrument is generally **not** the right parameter for a different policy.
- Estimating the full MTE curve (under stronger assumptions) allows us to evaluate **any** proposed policy.

The MTE framework unifies the structural and treatment-effects traditions: it shows exactly what each estimand identifies and what assumptions are needed.

# R: Estimating MTE with ivmte

Angrist & Evans (1998): effect of fertility on labor supply. $Y$ = worked, $D$ = morekids, $Z$ = samesex.

```
library(ivmte)  # Mogstad, Santos, Torgovitsky (2018)

# ATT via MTE extrapolation
att <- ivmte(data = AE, target = "att",
  m0 = ~ u + yob, m1 = ~ u + yob,
  ivlike = worked ~ morekids + samesex,
  propensity = morekids ~ samesex + yob)

# ATE -- just change target
ate <- ivmte(data = AE, target = "ate",
  m0 = ~ u + yob, m1 = ~ u + yob,
  ivlike = worked ~ morekids + samesex,
  propensity = morekids ~ samesex + yob)
```

# R: ivmte — LATE and Richer Specifications

```
# LATE for specific complier group
late <- ivmte(data = AE, target = "late",
  late.from = c(samesex = 0),
  late.to   = c(samesex = 1),
  m0 = ~ u + I(u^2) + yob,
  m1 = ~ u + I(u^2) + yob,
  ivlike = worked ~ morekids + samesex,
  propensity = morekids ~ samesex + yob)
```

- `m0`, `m1`: marginal treatment response functions; `u` is the unobserved resistance $U_D$.
- `target`: `"ate"`, `"att"`, `"atu"`, `"late"`, or `"genlate"`.
- Polynomial terms in `u` allow flexible MTE curves.
- The package implements Mogstad, Santos & Torgovitsky (2018, *Econometrica*) — bounds when point identification fails.

# Why We Need GMM: The MTE Estimation Problem

- The MTE framework poses an estimation challenge that 2SLS alone cannot solve.
- In `ivmte`, the `ivlike` argument specifies **multiple IV-like moment conditions**:

$$E[Z_i(Y_i - m_0(\boldsymbol{x}_i, U_{Di}; \theta)(1 - D_i) - m_1(\boldsymbol{x}_i, U_{Di}; \theta)D_i)] = 0$$

- We want to recover $\theta$ (the MTR parameters) from these moments, but:
  1. We may have **more moments than parameters** (overidentification)
  2. The model is **nonlinear** in $\theta$ (MTR functions can be flexible polynomials in $u_D$)
  3. Under heteroskedasticity, different moments have different precision
- 2SLS handles (1) but only for linear models. We need a **general framework** for combining moment conditions optimally.

# From IV Moments to GMM

- Recall: 2SLS solves the overidentified moment conditions

$$E[Z_i(Y_i - X_i'\beta)] = 0$$

  using the weighting matrix $(Z'Z/n)^{-1}$.
- The **Generalized Method of Moments** generalizes this:
  1. **Efficiency:** Under heteroskedasticity, 2SLS is not efficient. GMM uses the optimal weighting $\hat{W} = \hat{\Omega}^{-1}$.
  2. **Generality:** GMM applies to **any** moment conditions $E[g(W_i, \theta)] = 0$ — including the nonlinear moments from MTE estimation.
  3. **Testing:** The GMM $J$-statistic generalizes the Sargan test.
- MTE estimation via `ivmte` is a GMM problem: find MTR parameters that best fit the IV-like moments, subject to shape constraints.

# Preview: The GMM Estimator

- Given moment conditions $E[g(W_i, \theta_0)] = 0$ with $\dim(g) > \dim(\theta)$:

$$\hat{\theta}_{GMM} = \arg\min_{\theta} \left[ \frac{1}{n} \sum_{i=1}^{n} g(W_i, \theta) \right]' \hat{W} \left[ \frac{1}{n} \sum_{i=1}^{n} g(W_i, \theta) \right]$$

- Special cases:
  - Linear IV moments + $\hat{W} = (Z'Z/n)^{-1}$: **2SLS**
  - Linear IV moments + $\hat{W} = \hat{\Omega}^{-1}$: **efficient IV-GMM**
  - MTR moments + shape constraints: **MTE estimation** (Mogstad, Santos & Torgovitsky 2018)
- The $J$-test: $J = n \cdot \bar{g}(\hat{\theta})' \hat{W} \bar{g}(\hat{\theta}) \sim \chi^2_q$ tests overidentifying restrictions.

**Next week:** Full development of GMM — estimation, inference, and applications.

# Summary

1. The **control function** makes endogeneity correction explicit by adding $\hat{u}_2$ as a regressor.
2. IV has no finite moments when just identified — bootstrap fails.
3. IV is always less precise than OLS: $Avar(\hat{\beta}_{IV}) = Avar(\hat{\beta}_{OLS})/\rho^2_{xz}$.
4. IV bias $\approx$ OLS bias $\times \frac{1}{1+F}$; weak instruments ($F < 10$) are dangerous.
5. The **Hausman test** detects endogeneity; **Sargan test** detects invalid instruments.
6. Under heterogeneous effects, IV estimates **LATE** — the effect for compliers.
7. The **MTE framework**: all treatment parameters are weighted averages of $\Delta^{MTE}(u_D)$, indexed by unobserved resistance (utility cost).
8. 2SLS is a special case of **GMM** with a specific weighting matrix.