Lecture 15: Asymptotic Normality of the Instrumental Variables Estimator

Econometrics 2 – *Completing the IV Score Argument from Lecture 14*

Instructor: Prof. S. Arvanitis | **Notes transcribed by:** T. Kourtalis & Ch. Dourgoutis | **Digitisation & added notes:** T. Kourtalis

Semester: Spring 2026

▷ **Amber = Handwritten Notes (professor's words)** ◇ **Teal = Student's Notes**

Recall from Earlier Lectures

Where we are in the course:

- **Lecture 12:** IV consistency proved; $b_n \xrightarrow{p} \beta_0$ for any $\omega \succ 0$.
- **Lecture 14:** OLSE asymptotic normality derived in full; IV score decomposed and tagged as $(*)$, deferred here.

This lecture picks up at $(*)$, completes the IV asymptotic normality argument arriving at the full sandwich distribution for b_n, and then analyses how the asymptotic variance depends on the choice of weighting matrix ω.

1 Setup: Score and Hessian of the IV Objective

▷ **Handwritten Notes** (what the professor said)

Recall from Lecture 14 that for the IV objective function with weighting matrix ω:

$$\frac{\partial M_n(\beta)}{\partial \beta} = -2\,\frac{X_n' Z_n}{n}\,\omega\,\frac{Z_n'(Y_n - X_n\beta)}{n}$$

$$\frac{\partial^2 M_n(\beta)}{\partial \beta\,\partial \beta'} = 2\,\frac{X_n' Z_n}{n}\,\omega\,\frac{Z_n' X_n}{n}$$

Evaluating the scaled score at β_0:

$$\sqrt{n}\,\frac{\partial M_n(\beta_0)}{\partial \beta} = -2\,\frac{X_n' Z_n}{n}\,\omega\,\frac{Z_n'\varepsilon_n}{\sqrt{n}} = -2\left(\frac{1}{n}\sum_{i=1}^{n} X_{(i)}' Z_{(i)}\right)\omega\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{(i)}\varepsilon_{(i)}\right) \qquad (*)$$

◇ Student's Notes

Lecture 14 ended exactly here. Everything in this lecture is the completion of that argument.

The structure of $(*)$ is a product of three things: a $p \times q$ matrix $(\frac{X'Z}{n})$, a fixed $q \times q$ matrix (ω), and a $q \times 1$ vector $(\frac{Z'\varepsilon}{\sqrt{n}})$. The first factor is handled by LLN, the third by CLT, and ω sits fixed in the middle untouched. This split into LLN times CLT is the same idea used in Lecture 14 for OLS, except there the LLN piece was trivial since there was no pre-multiplier. The endogeneity of X forces routing through Z, introducing $\frac{X'Z}{n}$ as an extra piece needing its own LLN.

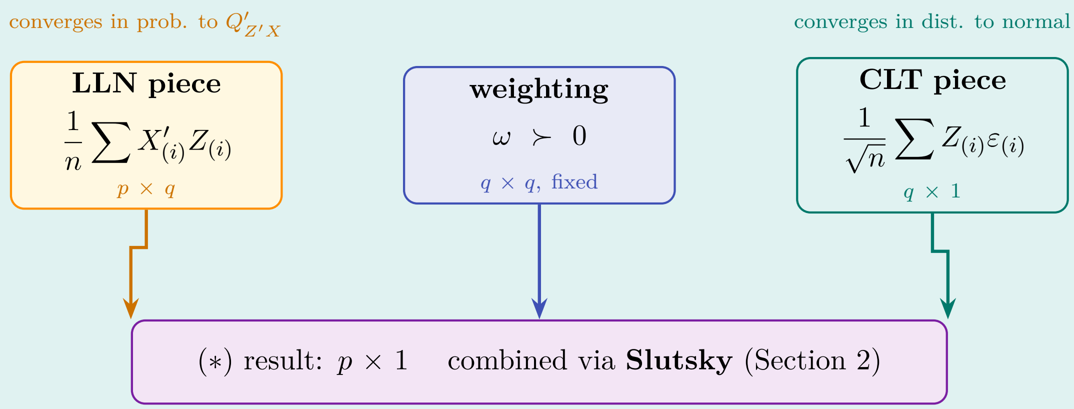

Figure 1: The three components of the IV scaled score $(*)$. The LLN piece (amber) converges in probability to $Q'_{Z'X}$. The CLT piece (teal) converges in distribution to a normal centered at zero by instrument exogeneity. Slutsky combines them.

2 Asymptotic Behavior of $(*)$

▷ Handwritten Notes (what the professor said)

Regarding the asymptotic behavior of $(*)$, we assume for reasons of simplicity that $(X_{(i)}, Z_{(i)}, \varepsilon_{(i)})$ are i.i.d. with respect to i.

2.1 The LLN Term

▷ Handwritten Notes (what the professor said)

Since $\mathbb{E}(Z'_{(i)}X_{(i)})$ exists, then because of Kolmogorov's Law of Large Numbers:

$$\frac{1}{n}\sum_{i=1}^{n} X'_{(i)}Z_{(i)} \xrightarrow{p} \mathbb{E}(X'_{(i)}Z_{(i)}) = Q'_{Z'X}$$

where from the derivation of consistency we had that $Q'_{Z'X}$ has rank p.

◇ Student's Notes

This is the same LLN that appeared in Lecture 12 when proving IV consistency. There, $\frac{Z'X}{n} \xrightarrow{p} Q_{Z'X}$ was used to establish that the IV objective limit function had curvature $Q'_{Z'X}\omega Q_{Z'X}$ and a unique minimum at β_0. Here the same convergence is recycled as the pre-multiplier in the score.
The rank condition on $Q'_{Z'X}$ is also from Lecture 12. Without it, $Q'_{Z'X} = 0$ and the score $(*)$ collapses to zero regardless of the CLT term, meaning b_n carries no information about β_0.

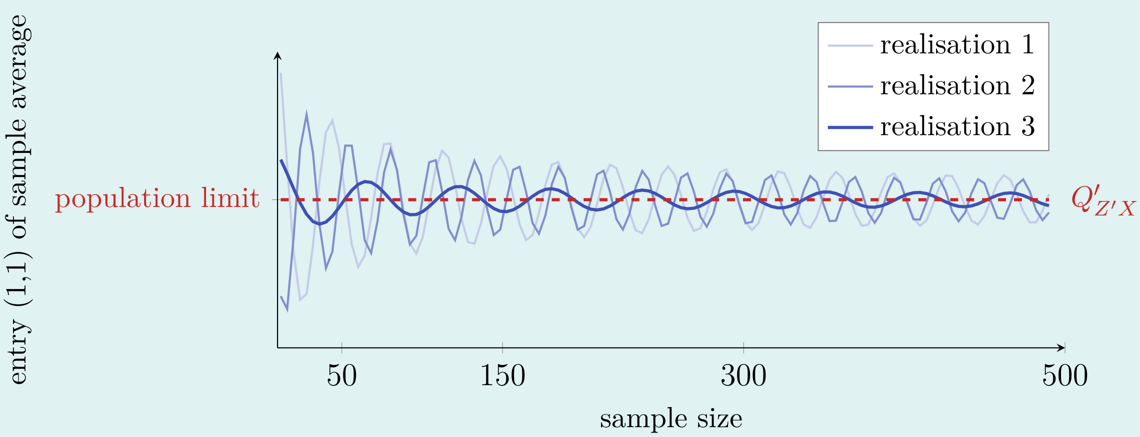

Figure 2: Three simulated paths of the $(1,1)$ entry of $\frac{1}{n}\sum X'_{(i)}Z_{(i)}$ as the sample size grows. All paths collapse onto the population limit $[Q'_{Z'X}]_{11}$ (rose dashed). This is the same LLN used in Lecture 12 for IV consistency, recycled here as the LLN piece of the score.

2.2 The CLT Term

▷ Handwritten Notes (what the professor said)

If furthermore it holds that:

$$\mathbb{E}(Z'_{(i)}\varepsilon^2_{(i)}Z_{(i)}) := V_{Z'\varepsilon}$$

exists and is non-singular, then because of the Central Limit Theorem we have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{(i)}\varepsilon_{(i)} \xrightarrow{d} Z^* \sim \mathcal{N}(0_{q\times 1},\, V_{Z'\varepsilon})$$

◇ Student's Notes

The CLT applies to $Z_{(i)}\varepsilon_{(i)}$ because it is i.i.d. with mean zero and finite variance $V_{Z'\varepsilon}$. The mean-zero property is instrument exogeneity $\mathbb{E}(Z'\varepsilon) = 0$, the foundation of the IV approach from Lecture 6. This is directly analogous to the OLS score variance $4\mathbb{E}(\varepsilon^2 X'X)$ from Lecture 14, with Z replacing X.

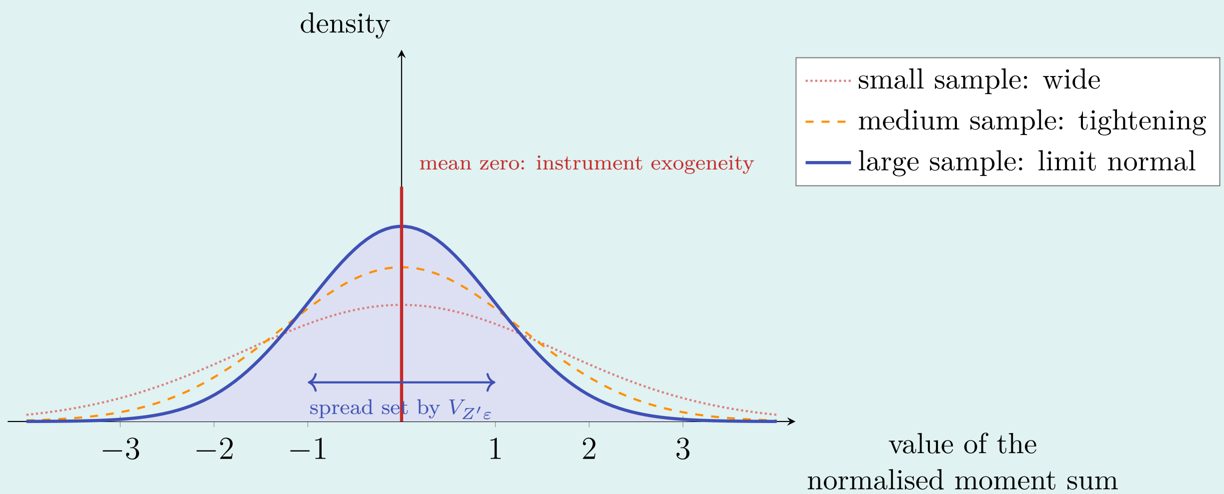

Figure 2: One component of the CLT term $\frac{1}{\sqrt{n}}\sum Z_{(i)}\varepsilon_{(i)}$ concentrating around the limit normal (indigo, filled) as the sample grows. All distributions are centered at zero by instrument exogeneity from Lecture 6.

2.3 Combining via Slutsky

▷ **Handwritten Notes** (what the professor said)

Therefore, since:

$$\frac{1}{n}\sum_{i=1}^{n} X'_{(i)}Z_{(i)} \xrightarrow{p} Q'_{Z'X} \qquad \text{and} \qquad \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{(i)}\varepsilon_{(i)} \xrightarrow{d} Z^*$$

we consequently have:

$$(*) \xrightarrow{d} -2\,Q'_{Z'X}\,\omega\,Z^* \sim \mathcal{N}(0_{p\times 1},\; 4\,Q'_{Z'X}\,\omega\,V_{Z'\varepsilon}\,\omega\,Q_{Z'X})$$

◇ **Student's Notes**

Slutsky's theorem: if $A_n \xrightarrow{p} A$ and $B_n \xrightarrow{d} B$, then $A_nB_n \xrightarrow{d} AB$. Here $A_n\omega B_n \xrightarrow{d} Q'_{Z'X}\omega Z^*$ with ω passing through unchanged. The variance follows from $\mathrm{Var}(CZ^*) = CVC'$ with $C = -2Q'_{Z'X}\omega$:

$$\mathrm{Var}(-2Q'_{Z'X}\omega Z^*) = 4\,Q'_{Z'X}\,\omega\,V_{Z'\varepsilon}\,\omega\,Q_{Z'X}$$

This is Assumption 3 of Lecture 13 now verified for IV. The score variance already depends on ω, the first sign that the asymptotic variance of b_n will not be ω-free like consistency was.

3 Hessian Convergence

▷ **Handwritten Notes** (what the professor said)

Finally we deal with the asymptotic behavior of the Hessian at the intermediate point β_n^*:

$$\frac{\partial^2 M_n(\beta_n^*)}{\partial\beta\,\partial\beta'} = 2\,\frac{X'_nZ_n}{n}\,\omega\,\frac{Z'_nX_n}{n} \xrightarrow{p} 2\,Q'_{Z'X}\,\omega\,Q_{Z'X} := H(\beta_0)$$

This limit is positive definite when $\omega \succ 0$ and $\mathrm{rank}(Q_{Z'X}) = p$.

◇ **Student's Notes**

Like the OLS Hessian in Lecture 14, the IV Hessian has no dependence on β, so β_n^* is irrelevant. Two LLN results plus CMT give the limit directly.

Positive definiteness follows from the Cholesky argument in Lecture 12. Write $\omega = LL'$. For any $v \neq 0_p$:

$$v'Q'_{Z'X}\omega Q_{Z'X}v = \|L'Q_{Z'X}v\|^2 > 0$$

since $\mathrm{rank}(Q_{Z'X}) = p$ means $Q_{Z'X}v \neq 0$. So $H(\beta_0) \succ 0$, which is Assumption 5 of

Lecture 13.
Unlike OLS where $H = 2\mathbb{E}(X'X)$ had no ω, the IV Hessian $H(\beta_0) = 2Q'_{Z'X}\omega Q_{Z'X}$ depends on ω. This propagates into the sandwich and makes the asymptotic variance of b_n sensitive to the choice of weighting matrix.

4 Final Result

▷ Handwritten Notes (what the professor said)

Consequently, and since:

0. $\beta_0 \in \text{Int}(\Theta)$
1. $(X_{(i)}, Z_{(i)}, \varepsilon_{(i)})$ are i.i.d. with respect to i
2. $Q_{Z'X} := \mathbb{E}(Z'_{(i)}X_{(i)})$ exists and has rank p
3. $V_{Z'\varepsilon} := \mathbb{E}(Z'_{(i)}\varepsilon^2_{(i)}Z_{(i)})$ exists and is invertible

then:

$$\sqrt{n}(b_n - \beta_0) \xrightarrow{d} \mathcal{N}\big(0_{p\times 1},\ (Q'_{Z'X}\,\omega\,Q_{Z'X})^{-1}\,Q'_{Z'X}\,\omega\,V_{Z'\varepsilon}\,\omega\,Q_{Z'X}\,(Q'_{Z'X}\,\omega\,Q_{Z'X})^{-1}\big)$$

Key Result

Asymptotic Distribution of the IV Estimator
Under Assumptions 0–3 above, with $\omega \succ 0$:

$$\sqrt{n}(b_n - \beta_0) \ \xrightarrow{d}\ \mathcal{N}\left(0_{p\times 1},\ \underbrace{(Q'_{Z'X}\,\omega\,Q_{Z'X})^{-1}}_{H^{-1}(\beta_0)}\underbrace{Q'_{Z'X}\,\omega\,V_{Z'\varepsilon}\,\omega\,Q_{Z'X}}_{V}\underbrace{(Q'_{Z'X}\,\omega\,Q_{Z'X})^{-1}}_{H^{-1}(\beta_0)}\right)$$

◇ Student's Notes

Component	OLS (Lecture 14)	IV (this lecture)
$H(\beta_0)$	$2\mathbb{E}(X'_{(i)}X_{(i)})$	$2Q'_{Z'X}\,\omega\,Q_{Z'X}$
V	$4\mathbb{E}(\varepsilon^2 X'_{(i)}X_{(i)})$	$4Q'_{Z'X}\,\omega\,V_{Z'\varepsilon}\,\omega\,Q_{Z'X}$
Depends on ω?	No	Yes (through H and V)
OLS special case	$Z = X$, $\omega = I$	Reduces to OLS result

Assumption 0 ($\beta_0 \in \text{Int}(\Theta)$) is numbered from zero because it is the interior condition from Lecture 13, not specific to IV. Assumptions 1–3 are the IV-specific conditions.

5 Observations on the Asymptotic Variance

▷ Handwritten Notes (what the professor said)

- The rate of convergence of b_n is $\sqrt{n}$ and is independent of ω.
- The asymptotic distribution is Gaussian, which is also independent of ω.
- The asymptotic variance depends on ω.

Question: Can the asymptotic variance be independent of ω?

When $p = q$ (exact identification):
The matrices $Q_{Z'X}$, ω, and $V_{Z'\varepsilon}$ all become square $p \times p$ and are all invertible. Therefore:

$$\begin{aligned}\text{AsyVar} &= (Q'_{Z'X}\,\omega\,Q_{Z'X})^{-1}\,Q'_{Z'X}\,\omega\,V_{Z'\varepsilon}\,\omega\,Q_{Z'X}\,(Q'_{Z'X}\,\omega\,Q_{Z'X})^{-1} \\ &= Q^{-1}_{Z'X}\,\omega^{-1}\,(Q'_{Z'X})^{-1}\,Q'_{Z'X}\,\omega\,V_{Z'\varepsilon}\,\omega\,Q_{Z'X}\,Q^{-1}_{Z'X}\,\omega^{-1}\,(Q'_{Z'X})^{-1} \\ &= Q^{-1}_{Z'X}\,V_{Z'\varepsilon}\,(Q'_{Z'X})^{-1}\end{aligned}$$

When $p = q$ and $\beta_0 \in \text{Int}(\Theta)$, the estimator becomes asymptotically independent of ω. This does not hold when β_0 is a boundary point.

When $q > p$ (over-identification):
The asymptotic variance depends on ω. The question of the optimal choice of ω therefore arises in order to achieve the smallest possible asymptotic variance.

◇ Student's Notes

The three observations structure the whole message of this section. The rate $\sqrt{n}$ and Gaussianity are free: you get them for any $\omega \succ 0$, regardless of how suboptimal your weighting matrix is. Only the spread of the limiting distribution varies with ω. This is a clean separation between what ω can and cannot affect.

The cancellation in the just-identified case.
The algebra for $p = q$ is worth following step by step. When $Q_{Z'X}$ is square and invertible:

$$(Q'_{Z'X}\,\omega\,Q_{Z'X})^{-1} = Q^{-1}_{Z'X}\,\omega^{-1}\,(Q'_{Z'X})^{-1}$$

Substituting into the sandwich and tracking the cancellations:

$$Q^{-1}_{Z'X}\,\underbrace{\omega^{-1}\,\underbrace{(Q'_{Z'X})^{-1}Q'_{Z'X}}_{I_p}\,\omega}_{I_p}\,V_{Z'\varepsilon}\,\underbrace{\omega\,\underbrace{Q_{Z'X}Q^{-1}_{Z'X}}_{I_p}\,\omega^{-1}}_{I_p}(Q'_{Z'X})^{-1} = Q^{-1}_{Z'X}\,V_{Z'\varepsilon}\,(Q'_{Z'X})^{-1}$$

The ω matrices cancel in pairs via identity matrices. This is not a numerical coincidence. When $q = p$, the sample moment conditions $\frac{1}{n}Z'(Y - Xb_n) = 0$ constitute a square system with exactly p equations and p unknowns, so it can be solved exactly. Any positive definite ω is merely a reweighting of equations that all hold exactly at the same solution, so the weighting cannot change the answer.

When $q > p$ the system is overdetermined: $q > p$ moment conditions but only p unknowns. The conditions cannot all be satisfied exactly in sample, so we minimise a weighted distance to zero. Different ω assign different penalties to violating different conditions, producing different estimators with different variances. The optimal ω balances these penalties in the most statistically efficient way.

The boundary point caveat.

The professor explicitly notes the ω-independence in the $p = q$ case fails when β_0 is a boundary point. The reason connects back to Assumption 0 (interior condition) from Lecture 13. At a boundary point the FOC $\frac{\partial M_n}{\partial \beta}(b_n) = 0$ need not hold, so the Taylor expansion argument breaks down and the entire sandwich formula is invalid. The limiting distribution in that case is not Gaussian at all — it may be a truncated normal or a more complex object depending on the geometry of Θ near the boundary. The ω-dependence question does not even arise because the sandwich formula does not apply.

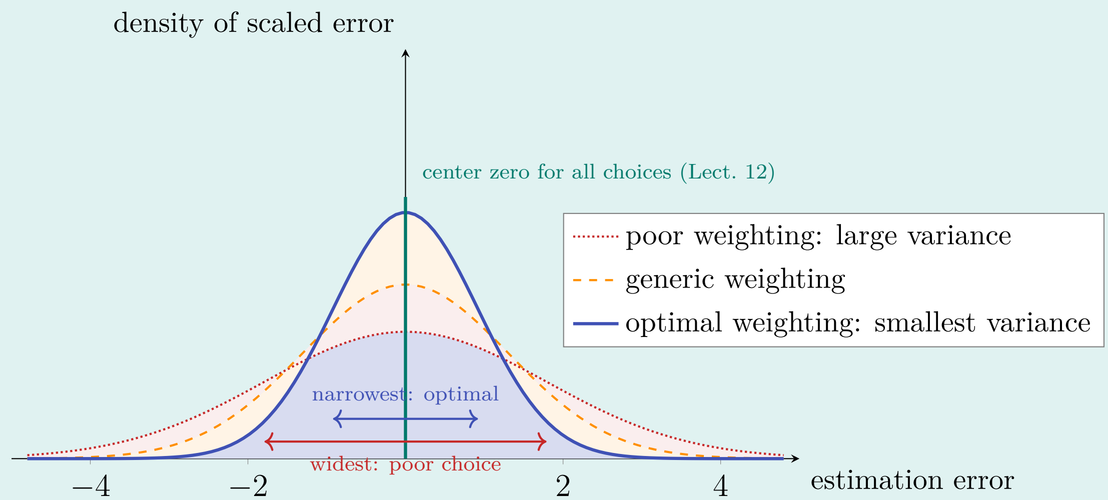

Figure 4: Limiting distributions of $\sqrt{n}(b_n - \beta_0)$ for three choices of ω. All are centered at zero since consistency holds for any $\omega \succ 0$ (Lecture 12). The spread varies strongly with ω. The optimal choice $\omega^* = V_{Z'\varepsilon}^{-1}$ (indigo, narrowest) minimises the asymptotic variance. This is the subject of Lecture 16.