Lecture 14: Asymptotic Normality of the OLSE
& Instrumental Variables Estimator

Econometrics 2 , *Applying the General Theory to OLS and IV*

Instructor: Prof. S. Arvanitis | **Notes & digitisation:** 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:

- **Lectures 10–11:** Two-step consistency recipe formalised and applied to OLS.
- **Lecture 12:** Same recipe applied to the IV estimator; $\hat{\beta}_n^{IV} \xrightarrow{p} \beta_0$ for any $\omega \succ 0$.
- **Lecture 13:** LADE consistency via convexity; topology of Θ; general three-assumption framework for asymptotic normality introduced (Assumptions 1–3).

This lecture verifies Assumptions 1–5 of the general normality framework for the OLSE, derives the asymptotic sandwich distribution, and sets up the same verification for the IV estimator (completed in Lecture 15).

1 Applying the General Theory to the OLSE

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

We apply our general asymptotic theory to our examples (except for LAD). In each case, we examine conditions that imply Assumptions 1–5 of our general theory. We begin with the OLSE case.

Assuming initially that $\beta_0 \in \text{Int}(\Theta)$, we have that:

$$\frac{\partial \mu_n(\beta)}{\partial \beta} = -\frac{2X_n'X_n}{n}(\beta_0 - \beta) - \frac{2X_n'\varepsilon_n}{n}$$

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

With probability approaching 1:

$$\hat{\beta}_n - \beta_0 = -\left(\frac{\partial^2 \mu_n}{\partial \beta\, \partial \beta'}(\beta_n^*)\right)^{-1} \frac{\partial \mu_n(\beta_0)}{\partial \beta}$$

If additionally $(X_{(i)}, \varepsilon_{(i)})$ are i.i.d. in i and $\mathbb{E}(\varepsilon_{(i)}^2 X_{(i)}' X_{(i)})$ exists, then because the classical CLT is applicable we have that:

$$\sqrt{n}\, \frac{\partial \mu_n(\beta_0)}{\partial \beta} = -\frac{2X_n'\varepsilon_n}{\sqrt{n}} = -\frac{2}{\sqrt{n}} \sum_{i=1}^{n} X_{(i)}' \varepsilon_{(i)} \xrightarrow{d} Z \sim N\big(0_{p\times 1},\ 4\,\mathbb{E}\big(\varepsilon_{(i)}^2 X_{(i)}' X_{(i)}\big)\big)$$

From our general theory.

⋄ Student's Notes

The story of this lecture is simple once you see the structure. In Lecture 13 we built three assumptions for asymptotic normality of a generic M-estimator. Here we just check them one by one for OLS. The checking is almost mechanical , the conceptual heavy lifting was done in Lectures 10–13.
The equation to keep in mind throughout is the Taylor expansion from Lecture 13:

$$0 = \frac{\partial \mu_n(\beta_0)}{\partial \beta} + \frac{\partial^2 \mu_n(\beta_n^*)}{\partial \beta\, \partial \beta'}(\hat{\beta}_n - \beta_0)$$

Everything we do is figuring out what each of these two terms converges to. The score (left) goes normal via CLT. The Hessian (right) goes to a constant via LLN. Plug both in, use Slutsky, and you have normality.

The score and its connection to Lecture 11.
The cross-product $\frac{X_n'\varepsilon_n}{n}$ is the same object that appeared in Lecture 11 when proving OLS consistency. There, we showed it converges to zero under exogeneity, which killed the "tilt" term in the expansion of $M_n(\beta)$ and centered the limit function at β_0. Here we scale it up by $\sqrt{n}$ and ask what distribution it has. The CLT answers: normal, with variance $4\,\mathbb{E}(\varepsilon_{(i)}^2 X_{(i)}' X_{(i)})$.
The variance involves $\varepsilon_{(i)}^2$, not $\mathbb{E}(\varepsilon_{(i)})^2$ (which is zero). This is because the CLT applies to $X_{(i)}'\varepsilon_{(i)}$ and the variance of that product is $\mathbb{E}(X_{(i)}' X_{(i)} \varepsilon_{(i)}^2)$ by the zero-mean property plus independence across i. No homoskedasticity is assumed anywhere.

Figure: how the score distribution concentrates as n grows.
The plot below shows the distribution of the scaled score $\sqrt{n}\,\frac{\partial \mu_n(\beta_0)}{\partial \beta}$ (in the scalar $p = 1$ case) for three sample sizes. For small n it is fat-tailed and spread out. As n grows it concentrates around the limiting normal. This is the CLT in action.

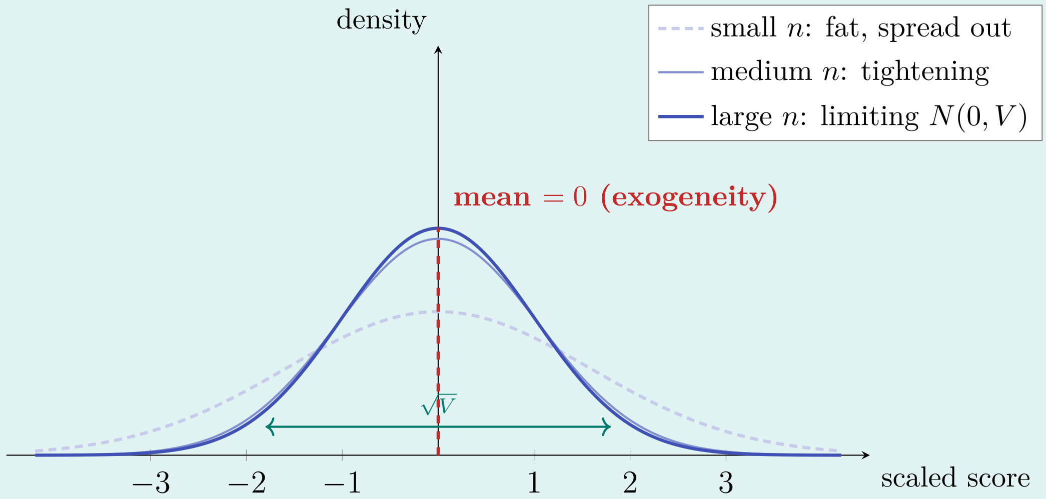

Figure 1: The scaled score $\sqrt{n}\,\partial\mu_n(\beta_0)/\partial\beta$ concentrates around the limiting $N(0, V)$ as n grows. The mean is zero by exogeneity ($\mathbb{E}(X'\varepsilon) = 0$ from Lecture 11). The spread shrinks at rate $1/\sqrt{n}$.

Why the Hessian being constant in β matters so much.
For a generic M-estimator (MLE, LAD), the Hessian $\frac{\partial^2\mu_n(\beta_n^*)}{\partial\beta^2}$ genuinely depends on β_n^*. To show it converges to the same limit as the Hessian at β_0, you need to invoke continuity of the Hessian in β plus $\beta_n^* \xrightarrow{p} \beta_0$. For OLS the Hessian is $\frac{2X'X}{n}$ everywhere , no β appears at all , so the intermediate point β_n^* is completely irrelevant and the LLN gives the limit directly. This is one of the few places where the quadratic structure of OLS makes the proof genuinely simpler, not just computationally but logically.

2 Hessian Convergence (Assumption 5)

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

Regarding Assumption 5 of our general theory: we remember that in the consistency study of the estimator, in the case where $(X_{(i)}, \varepsilon_{(i)})$ are i.i.d. and since $\mathbb{E}(X'_{(i)}X_{(i)})$ exists, Kolmogorov's Law of Large Numbers implies that:

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

Therefore, due to the Continuous Mapping Theorem:

$$\frac{2}{n}X'_nX_n \xrightarrow{p} 2\,\mathbb{E}(X'_{(i)}X_{(i)})$$

and consequently, if we denote by β_n^* the intermediate point of the Mean Value Theorem of our theory:

$$\frac{\partial^2 \mu_n(\beta_n^*)}{\partial\beta\,\partial\beta'} \xrightarrow{p} 2\,\mathbb{E}(X_{(i)}' X_{(i)}) := H(\beta_0)$$

$H(\beta_0)$ does not depend on β_0.

Assumption 5 will be satisfied ultimately as long as $\operatorname{rank}\mathbb{E}(X_{(i)}' X_{(i)}) = p$.

◇ Student's Notes

The professor's remark that $H(\beta_0)$ does not depend on β_0 is easy to miss but worth thinking about. The Hessian $\frac{2X'X}{n}$ contains no β at all , it is purely a property of the design matrix. So its limit $H = 2\,\mathbb{E}(X_{(i)}' X_{(i)})$ depends only on how the regressors are distributed, not on the true parameter value. This is special to quadratic objectives and does not hold for LAD or MLE.

The same LLN result, used twice.
In Lecture 11 the convergence $\frac{X'X}{n} \xrightarrow{p} \mathbb{E}(X_{(i)}' X_{(i)})$ was used to establish that the OLS limit function had curvature $Q_{X'X}$, making the minimum well-separated and giving consistency. Now the exact same convergence result is recycled to establish Assumption 5 for normality. The whole architecture of the course is built this way: prove the LLN results once for consistency, then reuse them for normality. The difference is what you do with them: for consistency you need the limit to be invertible so the bowl has a unique bottom; for normality you need the same invertibility so you can solve for $\hat\beta_n - \beta_0$ in the Taylor expansion.

The rank condition doing double duty.
In Lecture 11, $\operatorname{rank}(Q_{X'X}) = p$ was the identification condition: no flat directions in the limit function, no multicollinearity problem. Here, the same condition is the invertibility condition for the sandwich formula. One assumption, two jobs. If this fails, you lose both consistency and normality at once.

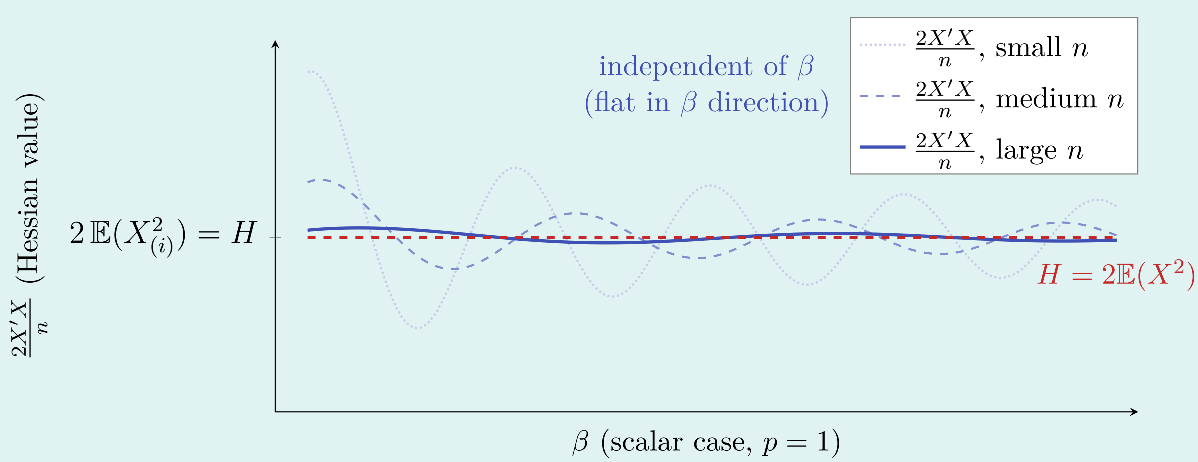

Figure 2: The Hessian $\frac{2X'X}{n}$ has no dependence on β (it is the same number regardless of where you evaluate it). As n grows, it concentrates around the limit $H = 2\mathbb{E}(X_{(i)}^2)$ by Kolmogorov's LLN. The intermediate point β_n^* is completely irrelevant.

3 The Asymptotic Distribution of the OLSE

▷ Handwritten Notes (what the professor said)

Therefore, from the above, our general theory yields:

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N\left(0_{p\times 1},\ \underbrace{\frac{1}{2}\mathbb{E}^{-1}(X_{(i)}'X_{(i)})}_{H^{-1}(\beta_0)} \cdot \underbrace{4\,\mathbb{E}(\varepsilon_{(i)}^2 X_{(i)}'X_{(i)})}_{V} \cdot \underbrace{\frac{1}{2}\mathbb{E}^{-1}(X_{(i)}'X_{(i)})}_{H^{-1}(\beta_0)}\right)$$

$$= N\big(0_{p\times 1},\ \mathbb{E}^{-1}(X_{(i)}'X_{(i)})\,\mathbb{E}(\varepsilon_{(i)}^2 X_{(i)}'X_{(i)})\,\mathbb{E}^{-1}(X_{(i)}'X_{(i)})\big)$$

Due to the above:

$$\left(\frac{1}{n}X_n'X_n\right)^{-1} \xrightarrow{P} \mathbb{E}^{-1}(X_{(i)}'X_{(i)})$$

so it will be a consistent estimator of it.

Key Result

Asymptotic Distribution of the OLSE
Under i.i.d. $(X_{(i)}, \varepsilon_{(i)})$, $\mathbb{E}(\varepsilon_{(i)}^2 X_{(i)}'X_{(i)}) < \infty$, and $\text{rank}(\mathbb{E}(X_{(i)}'X_{(i)})) = p$:

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N\Big(0_{p\times 1},\ \underbrace{\mathbb{E}^{-1}(X_{(i)}'X_{(i)})\,\mathbb{E}(\varepsilon_{(i)}^2 X_{(i)}'X_{(i)})\,\mathbb{E}^{-1}(X_{(i)}'X_{(i)})}_{\text{sandwich (heteroskedasticity-robust) variance}}\Big)$$

$$\left(\frac{X_n'X_n}{n}\right)^{-1} \xrightarrow{P} \mathbb{E}^{-1}(X_{(i)}'X_{(i)}) \quad \text{(consistent estimator of } H^{-1}(\beta_0))$$

◇ Student's Notes

This is the most important formula in the course so far and deserves a careful read.

Where the sandwich comes from.
The general template from Lecture 13 was $\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, H^{-1}VH^{-1})$. Here $H = 2\mathbb{E}(X_{(i)}'X_{(i)})$, so $H^{-1} = \frac{1}{2}\mathbb{E}^{-1}(X_{(i)}'X_{(i)})$, and $V = 4\mathbb{E}(\varepsilon_{(i)}^2 X_{(i)}'X_{(i)})$. Plugging in and cancelling:

$$H^{-1}VH^{-1} = \tfrac{1}{2}\mathbb{E}^{-1}(X'X)\cdot 4\mathbb{E}(\varepsilon^2 X'X)\cdot \tfrac{1}{2}\mathbb{E}^{-1}(X'X) = \mathbb{E}^{-1}(X'X)\,\mathbb{E}(\varepsilon^2 X'X)\,\mathbb{E}^{-1}(X'X)$$

The factors of 2 and 4 come from normalising the OLS objective as $\frac{1}{n}\|Y - X\beta\|^2$. If you used $\frac{1}{2n}\|Y - X\beta\|^2$ instead (common in other textbooks), the gradient would carry a factor of 1, the Hessian a factor of 1, and the final variance would still be the same. The sandwich is an intrinsic property of the model, not of how you normalised the objective.

Homoskedastic special case and what breaks when it fails.
If $\mathbb{E}(\varepsilon^2_{(i)} \mid X_{(i)}) = \sigma^2$, then by LIE: $\mathbb{E}(\varepsilon^2_{(i)}X'_{(i)}X_{(i)}) = \sigma^2\mathbb{E}(X'_{(i)}X_{(i)})$, and the sandwich collapses to $\sigma^2\mathbb{E}^{-1}(X'_{(i)}X_{(i)})$. This is the textbook OLS variance formula from classical econometrics courses. The whole sandwich structure exists to handle the case where ε^2 varies with X, which in practice it almost always does.

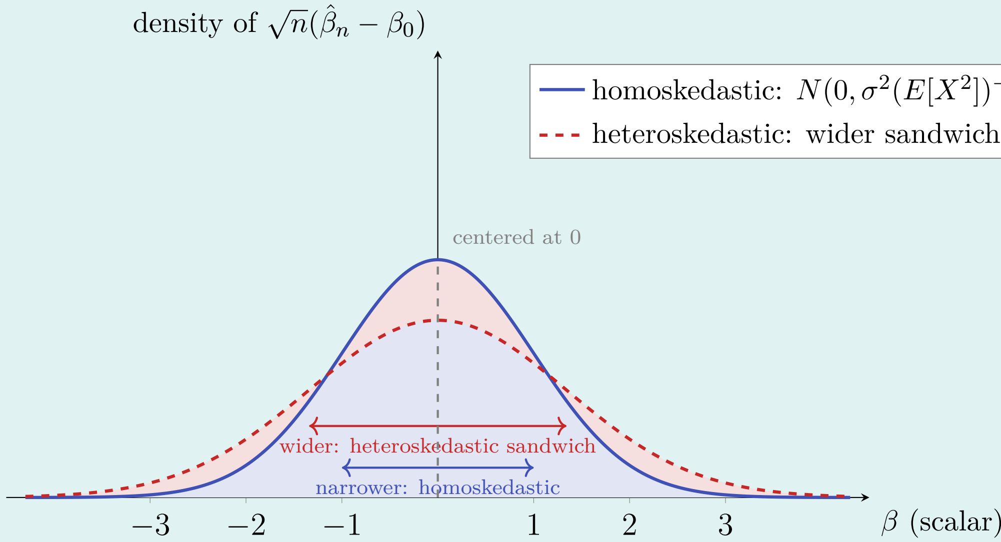

Figure 3: Under homoskedasticity (indigo, filled), the limiting distribution is $N(0, \sigma^2\mathbb{E}^{-1}(X^2))$. Under heteroskedasticity (rose, dashed), the true sandwich variance is larger. Using the homoskedastic formula when ε^2 varies with X would understate the spread and give confidence intervals that are too narrow.

The consistent estimator remark and inference.
The professor notes $\left(\frac{X'X}{n}\right)^{-1} \xrightarrow{P} \mathbb{E}^{-1}(X'_{(i)}X_{(i)})$. This is the same LLN from Lecture 11, recycled a third time. The outer bread of the sandwich is easy to estimate: just invert $X'X/n$. The inner filling $\mathbb{E}(\varepsilon^2_{(i)}X'_{(i)}X_{(i)})$ requires estimating $\varepsilon^2_{(i)}$, which you do with squared residuals $\hat{\varepsilon}^2_{(i)} = (Y_{(i)} - X_{(i)}\hat{\beta}_n)^2$. The resulting estimator $\left(\frac{X'X}{n}\right)^{-1}\left(\frac{1}{n}\sum \hat{\varepsilon}^2_{(i)}X'_{(i)}X_{(i)}\right)\left(\frac{X'X}{n}\right)^{-1}$ is the heteroskedasticity-consistent (HC) standard error that econometrics software reports. The professor is pointing at the first piece and saying it is already handled by the LLN we proved back in Lecture 11.

4 Application to the Instrumental Variables Estimator

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

Our general theory is directly applicable also to the version of the linear model with instrumental variables. Given the weighting matrix ω, the objective function is:

$$M_n(\beta) = \frac{1}{n^2}(Y_n - X_n\beta)'\, Z_n\, \omega\, Z_n'(Y_n - X_n\beta)$$

We assume that $\beta_0 \in \text{Int}(\Theta)$. The differentiability conditions are satisfied since M_n is a quadratic function of β.
We have that:

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

Evaluating the scaled score at β_0:

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

The asymptotic analysis of $(*)$ is carried out in Lecture 15.

◇ **Student's Notes**

The structure of the IV score $(*)$ is more complex than OLS but the logic is the same. The score is a product of two pieces: a LLN piece ($\frac{1}{n}\sum X_{(i)}' Z_{(i)} \xrightarrow{p} Q_{Z'X}'$) and a CLT piece ($\frac{1}{\sqrt{n}}\sum Z_{(i)}\varepsilon_{(i)} \xrightarrow{d} N$). Combining them requires Slutsky's theorem: a sequence converging in probability multiplied by a sequence converging in distribution gives convergence in distribution to the product of the limits. Lecture 15 carries this out in full.

Why the instruments appear in the CLT piece but not the regressors.
The CLT piece is $\frac{1}{\sqrt{n}}\sum Z_{(i)}\varepsilon_{(i)}$, not $\frac{1}{\sqrt{n}}\sum X_{(i)}\varepsilon_{(i)}$. The whole point of IV is that while $\mathbb{E}(X_{(i)}'\varepsilon_{(i)}) \neq 0$ (endogeneity , this is why OLS is inconsistent in this setting), we have $\mathbb{E}(Z_{(i)}'\varepsilon_{(i)}) = 0$ (instrument exogeneity, from Lectures 6 and 12). The CLT can only be applied to zero-mean sequences. So the instruments appear in the CLT piece because they are the ones orthogonal to ε. The regressors appear only in the LLN piece $\frac{1}{n}\sum X_{(i)}' Z_{(i)} \xrightarrow{p} Q_{Z'X}'$, which is just the empirical relevance condition.

The tag $(*)$ and what Lecture 15 will do with it.
The professor writes $(*)$ and stops. Lecture 15 picks up immediately from $(*)$, applies

LLN plus CLT plus Slutsky, and arrives at the full sandwich variance:

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

Note that this depends on ω. Unlike consistency (which held for any $\omega \succ 0$ as shown in Lecture 12), the asymptotic variance changes with the weighting matrix. Lecture 16 then asks which ω minimises this variance , that is the optimal GMM question , and shows that when $q = p$ the variance becomes independent of ω entirely.

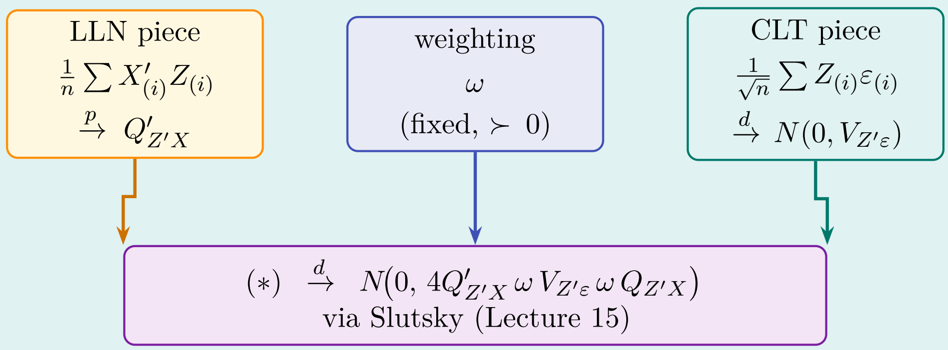

Figure 4: Structure of the IV score decomposition $(*)$. The LLN piece (amber) and CLT piece (teal) are combined via Slutsky's theorem in Lecture 15 to give the limiting normal distribution of $(*)$. The ω matrix passes through unchanged.

! Watch Out

LAD is excluded from this lecture.

The general normality framework requires the objective to be twice continuously differentiable (Assumption 2). The LAD objective $M_n(\beta) = \frac{1}{n}\sum_i |Y_i - X_i'\beta|$ has kinks wherever a residual is zero and is not C^2. Asymptotic normality of the LADE requires separate techniques not covered here.