Lecture 3: Identification in Linear Models

Econometrics — *continuation from Lecture 2*

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

Semester: Spring 2026

▷ **Amber boxes = Handwritten Notes (professor's words)**

◇ **Teal boxes = Student's Notes**

Recall from Lecture 2

In Lecture 2 we set up the semi-parametric linear model $Y_n = X_n b_0 + \epsilon_n$ with three assumptions (exogeneity, spherical errors, full rank) and proposed the expected conditional squared-error loss

$$U_n^*(b) \;:=\; \mathbb{E}\left[\frac{1}{n}(Y_n - X_n b)'(Y_n - X_n b) \,\middle|\, X_n\right]$$

as our objective function. **This lecture proves that U_n^* is uniquely minimised at the true parameter** b_0—i.e. the identification result.

1 The Core Intuition

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

We want to uncover the true relationship, b_0, that generated our data. If we guess a parameter b, how wrong is our guess on average? The mathematical proof below demonstrates that our "average squared mistake" forms a perfect, multi-dimensional *bowl*.

Because of our statistical assumptions, the absolute bottom of this bowl occurs exactly when our guess matches reality ($b = b_0$). Keeping this "bowl" shape in mind helps anchor the algebra: every step is simply moving toward proving that the loss function is a parabola pointing upwards.

◇ **Student's Notes**

Think of the scalar analogy first. If $p = 1$ the expected loss becomes

$$U_n^*(b) = \underbrace{\frac{\sum x_i^2}{n}}_{>0}(b_0 - b)^2 + 1,$$

a literal upward-opening parabola in b with vertex at b_0. The full proof generalises this to p dimensions: the parabola becomes an *elliptic paraboloid* because $\frac{X'X}{n}$ is positive definite.

Road-map of the proof:

1. Substitute the DGP into the residual.
2. Expand the quadratic and take the conditional expectation.
3. Show the cross-term dies (exogeneity) and the constant term equals 1 (spherical errors).
4. Conclude uniqueness via positive definiteness of $X'X/n$.

2 Model Setup (Recap from Lecture 2)

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

Our data generating process is defined as:

$$Y_n = X_n b_0 + \epsilon_n$$

The critical assumptions are:

- **Strict Exogeneity:** $\mathbb{E}(\epsilon_n \mid X_n) = 0_{n\times 1}$ *(The errors are pure noise.)*
- **Spherical Errors:** $\mathrm{Var}(\epsilon_n \mid X_n) = I_{n\times n}$ *(The noise is uniform and uncorrelated.)*
- **Full Rank (Identification):** $\mathrm{rank}(X_n) = p$ *(Enough unique info to pinpoint the parameters.)*

◇ **Student's Notes**

Dimension check. Y_n is $n \times 1$, X_n is $n \times p$, b_0 is $p \times 1$, ϵ_n is $n \times 1$.

- *Strict exogeneity* is stronger than just $\mathrm{Cov}(X, \epsilon) = 0$; it says **no column** of X_n carries information about **any** ϵ_i.
- *Spherical errors* bundles two things: homoskedasticity ($\mathrm{Var}(\epsilon_i|X) = 1$ for every i)

and no serial correlation ($\text{Cov}(\epsilon_i, \epsilon_j|X) = 0$ for $i \neq j$). When these fail, we move to GLS.

- *Full rank* requires $n \geq p$ (at least as many observations as parameters) **and** no exact multicollinearity. This is what makes $X'X$ invertible.

! Watch Out

The spherical-errors assumption here sets the variance to $\sigma^2 I$ with $\sigma^2 = 1$ for notational simplicity. In the general case $\text{Var}(\epsilon|X) = \sigma^2 I$, every formula carries a σ^2 factor, but the identification argument is identical.

3 The Objective Function (The "Mistake" Function)

▷ Handwritten Notes (what the professor said)

We define our expected loss function as:

$$U_n^*(b) \;:=\; \mathbb{E}\left[\frac{1}{n}(Y_n - X_n b)'(Y_n - X_n b) \;\middle|\; X_n\right]$$

First, substitute the true model into our guess to see the error gap:

$$Y_n - X_n b = X_n b_0 + \epsilon_n - X_n b = X_n(b_0 - b) + \epsilon_n$$

Next, expand the quadratic form algebraically:

$$\begin{aligned}&\big(X_n(b_0 - b) + \epsilon_n\big)'\big(X_n(b_0 - b) + \epsilon_n\big)\\ &\quad = \underbrace{(b_0 - b)'X_n'X_n(b_0 - b)}_{\text{Term 1}} + \underbrace{2(b_0 - b)'X_n'\epsilon_n}_{\text{Term 2}} + \underbrace{\epsilon_n'\epsilon_n}_{\text{Term 3}}\end{aligned}$$

◇ Student's Notes

This is just the identity $(a + b)'(a + b) = a'a + 2a'b + b'b$ with

$$a = X_n(b_0 - b), \qquad b = \epsilon_n.$$

Each of the three terms will be handled separately under the conditional expectation in the next section.

4 Applying the Expectation (Filtering the Noise)

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

Now, we apply the conditional expectation operator $\frac{1}{n}\mathbb{E}(\cdot \mid X_n)$ to each of the three terms. This is where we mathematically "average out" the noise to reveal the underlying structure.

4.1 Term 1: The Deterministic Signal

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

Since b, b_0, and X_n are treated as constants under the condition, they pass through unchanged:

$$\frac{1}{n}\mathbb{E}\big[(b_0 - b)'X_n'X_n(b_0 - b) \mid X_n\big] \;=\; (b_0 - b)'\left(\frac{X_n'X_n}{n}\right)(b_0 - b)$$

◇ **Student's Notes**

Conditioning on X_n means we treat the entire design matrix as *known*. The only random object inside the expectation is ϵ_n, and this term contains no ϵ_n at all—so the expectation is the identity operation here.

The matrix $\frac{X_n'X_n}{n}$ is the **sample second-moment matrix** of the regressors. It will play the role of the curvature of our bowl.

4.2 Term 2: The Cross-Product (Zeroing Out)

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

Because the errors are pure noise $\big(\mathbb{E}(\epsilon_n \mid X_n) = 0\big)$, this term vanishes. Recognizing that the expected value of the noise is zero immediately simplifies the expression:

$$\frac{1}{n}\mathbb{E}\big[2(b_0 - b)'X_n'\epsilon_n \mid X_n\big] = \frac{2}{n}(b_0 - b)'X_n' \underbrace{\mathbb{E}(\epsilon_n \mid X_n)}_{=0} = 0$$

◇ **Student's Notes**

This is the **crucial** step powered by strict exogeneity. If exogeneity failed (e.g. an omitted variable that correlates with both X and ϵ), this cross-term would *not* vanish and the minimum of U_n^* would *not* be at b_0—we would have **bias**.

Linearity of expectation lets us pull the non-random $(b_0 - b)'X_n'$ out, leaving only $\mathbb{E}(\epsilon_n|X_n)$ inside.

4.3 Term 3: The Trace Trick (The Error Variance)

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

For a scalar value like $\epsilon_n'\epsilon_n$, the value equals its trace. Because the trace is a linear operator, we swap the trace and expectation to isolate the variance matrix:

$$\frac{1}{n}\,\mathbb{E}[\epsilon_n'\epsilon_n \mid X_n] = \frac{1}{n}\,\mathbb{E}\big[\mathrm{tr}(\epsilon_n\epsilon_n') \mid X_n\big] = \frac{1}{n}\,\mathrm{tr}\big(\mathbb{E}[\epsilon_n\epsilon_n' \mid X_n]\big)$$

Since $\mathrm{Var}(\epsilon_n \mid X_n) = I_{n\times n}$, the trace of an $n \times n$ identity matrix is simply n:

$$\frac{1}{n}\,\mathrm{tr}(I_{n\times n}) = \frac{n}{n} = 1$$

◇ **Student's Notes**

Why the trace trick? $\epsilon_n'\epsilon_n$ is a 1×1 scalar, while $\epsilon_n\epsilon_n'$ is an $n \times n$ matrix. We *want* the matrix form because we know $\mathbb{E}[\epsilon_n\epsilon_n'|X_n] = \mathrm{Var}(\epsilon_n|X_n)$ (since $\mathbb{E}[\epsilon_n|X_n] = 0$). The cyclic property of trace bridges the two:

$$\epsilon_n'\epsilon_n = \mathrm{tr}(\epsilon_n'\epsilon_n) = \mathrm{tr}(\epsilon_n\epsilon_n').$$

The first equality holds because any scalar equals its own trace. The second uses $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

General case: if $\mathrm{Var}(\epsilon|X) = \sigma^2 I$, this term becomes σ^2 instead of 1. It shifts the bowl *up* but does not change the location of the minimum.

5 The Unique Minimum

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

Reassembling the filtered function gives us:

$$\boxed{U_n^*(b) \;=\; (b_0 - b)'\left(\frac{X_n'X_n}{n}\right)(b_0 - b) \;+\; 1}$$

Because X_n has full rank, the matrix $A = \frac{X_n'X_n}{n}$ is **positive definite**. This mathematically guarantees our "bowl" shape.

For any positive definite matrix A, the quadratic form $y'Ay \geq 0$, and it *strictly* equals 0 if and only if $y = 0_{px1}$.

Let $y = (b_0 - b)$. The minimum possible value of this entire function is uniquely achieved when:

$$b_0 - b = 0 \quad\Longrightarrow\quad b = b_0$$

Therefore, minimizing this objective function cleanly and uniquely identifies the true parameter b_0:

$$\arg\min_{b \in \Theta} U_n^*(b) = b_0$$

Key Result

$$\arg\min_{b \in \Theta} U_n^*(b) \;=\; b_0$$

The expected conditional loss is **uniquely minimised** at the true parameter vector. This is the **identification** result promised at the end of Lecture 2: b_0 is the *only* value in the parameter space that minimises the population objective, so it is in principle recoverable from data.

◇ Student's Notes

Unpacking positive definiteness: A symmetric matrix A is positive definite iff all its eigenvalues $\lambda_1, \ldots, \lambda_p > 0$. Since $A = \frac{1}{n}X'X$ and $\text{rank}(X) = p$, the columns of X are linearly independent, which forces every eigenvalue of $X'X$ (and hence A) to be strictly positive.

Geometrically, positive definiteness means the level sets $\{b : U_n^*(b) = c\}$ are **ellipsoids** centred at b_0; there are no flat directions, so the minimum is unique.

Why this matters going forward:

1. *Identification* is a **population** concept—it says the truth *can* in principle be found; it does not say how fast or accurately a finite sample does so.
2. The *estimation* step replaces U_n^* with the sample analogue $\hat{U}_n(b) = \frac{1}{n}(Y - Xb)'(Y - Xb)$ and minimises it, giving the OLS estimator $\hat{b} = (X'X)^{-1}X'Y$.
3. *Consistency* then asks whether $\hat{b} \to b_0$ as $n \to \infty$—this is where the law of large numbers (for $X'X/n$ and $X'\epsilon/n$) enters.

Quick-Reference Summary

◇ Student's Notes

Term	Expression	Role
Signal	$(b_0 - b)'\frac{X'X}{n}(b_0 - b)$	Bowl curvature; $= 0$ iff $b = b_0$
Cross	$\frac{2}{n}(b_0 - b)'X'\epsilon$	Killed by exogeneity
Noise	$\frac{1}{n}\epsilon'\epsilon$	$\to 1$ (or σ^2); constant shift

Lectures 2–3 narrative arc:

Lecture	What was accomplished
Lect. 2	Defined parametric vs. semi-parametric models; chose the objective function
Lect. 3	Proved the objective has a unique minimum at b_0 (identification)
Next	Replace population objective with sample analogue $\Rightarrow$ OLS estimator