Lecture 11: Asymptotic Identification

Econometrics 2 — *Linear Model Example & Mild Misspecification*

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

Semester: Spring 2026

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

◇ **Teal boxes = Student's Notes**

Recall from Lectures 4 & 10

In **Lecture 4** we introduced extremum estimators $\hat{\beta}_n \in \arg\min_{\beta\in\Theta} M_n(\beta)$ and noted that consistency requires the sample objective M_n to "converge to" a limit that is uniquely minimised at β_0. In **Lecture 10** this was formalised as the *two-step consistency recipe*:

1. **Locally uniform convergence in probability:** $\sup_{\beta^*\in O_\beta} |\mu_n(\beta^*) - \mu(\beta^*)| \xrightarrow{p} 0$ for every β.
2. **Asymptotic identification:** the limit function $\mu(\beta)$ has a *well-separated* unique minimum at β_0.

This lecture applies both steps to the linear model, derives what the limit function looks like, and analyses what happens when the true β_0 lies *outside* the parameter space Θ (mild misspecification).

1 The Condition of Asymptotic Identification

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

The "well-separated minimum" requirement we discussed is also called the **condition of asymptotic identification**. Let us see how it is specified in our main example: the semi-parametric linear model.

Definition: Asymptotic Identification / Well-Separated Minimum

Let $\mu : \Theta \to \mathbb{R}$ be the limit of the sample objective μ_n. The parameter β_0 is **asymptotically identified** if there is a unique global minimum at β_0 and it is **well-separated**:

for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\inf_{\beta\,:\,\|\beta-\beta_0\|>\varepsilon} \mu(\beta) \;\geq\; \mu(\beta_0) + \delta.$$

◇ **Student's Notes**

Why "well-separated" is stronger than "unique minimum":
A function can have a unique global minimum yet fail the well-separated condition—for example, if it asymptotes toward the minimum value without ever quite reaching it, or if it has a very flat approach from one side.

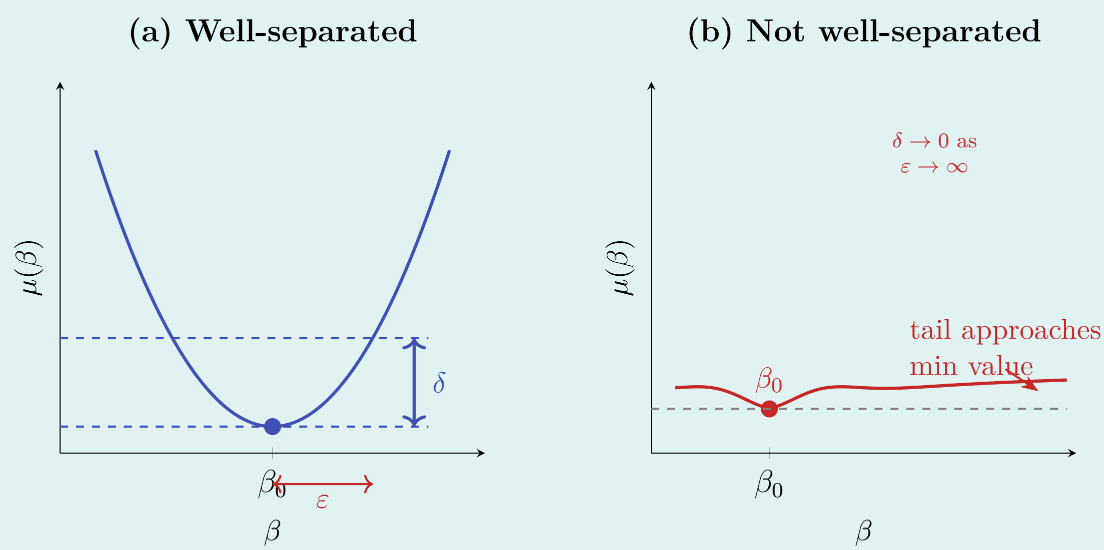

Figure 1: (a) A well-separated minimum: any point more than ε away from β_0 is at least δ higher. (b) A function with a unique minimum at β_0 but whose tail asymptotes toward the minimum value, so no uniform δ gap exists for large ε.

Connection to Lecture 10: The well-separated condition is exactly what allows us to conclude $\arg\min M_n \xrightarrow{p} \arg\min \mu = \beta_0$ once we have uniform convergence $M_n \xrightarrow{p} \mu$. Without it, the argmin can "wander" along flat regions of μ even as $n \to \infty$—precisely the phenomenon observed in the *degenerate LAD design* (Lecture 9, where the non-unique median creates flat regions and the two LP solvers disagree).

2 Applying the Theory: The Linear Model

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

Consider the standard objective function for the linear model (introduced in Lectures 2–3):

$$M_n(\beta) = \frac{1}{n}(Y_n - X_n\beta)'(Y_n - X_n\beta)$$

Substitute the true DGP $Y_n = X_n\beta_0 + \varepsilon_n$:

$$\begin{aligned} M_n(\beta) &= \frac{1}{n}\big(\varepsilon_n + X_n(\beta_0 - \beta)\big)'\big(\varepsilon_n + X_n(\beta_0 - \beta)\big) \\ &= \underbrace{\frac{1}{n}\varepsilon_n'\varepsilon_n}_{\text{const. w.r.t. } \beta} + \underbrace{\frac{2}{n}(\beta_0 - \beta)'X_n'\varepsilon_n}_{\text{cross term}} + \underbrace{(\beta_0 - \beta)'\frac{X_n'X_n}{n}(\beta_0 - \beta)}_{\text{quadratic term}} \end{aligned}$$

Since the first term does not depend on β, it does not affect the argmin. Define the **reduced objective**:

$$\widetilde{M}_n(\beta) = \frac{2}{n}(\beta_0 - \beta)'X_n'\varepsilon_n + (\beta_0 - \beta)'\frac{X_n'X_n}{n}(\beta_0 - \beta)$$

We study the asymptotic behaviour of its two key components: $\dfrac{X_n'X_n}{n}$ and $\dfrac{X_n'\varepsilon_n}{n}$.

◇ **Student's Notes**

Why we can drop $\frac{1}{n}\varepsilon_n'\varepsilon_n$:
This term is a random constant *with respect to* β—it does not depend on our choice of parameter at all. Adding a constant to an objective function shifts the minimum value but *does not move the minimiser*:

$$\arg\min_{\beta} M_n(\beta) = \arg\min_{\beta} \widetilde{M}_n(\beta).$$

This mirrors exactly what happened in the *population* identification proof in **Lecture 3**: the "noise" term $\frac{1}{n}\operatorname{tr}(I_n) = 1$ also dropped out of the argmin.

The two objects to track:

Object	Role in $\widetilde{M}_n$	What we need
$\dfrac{X_n'X_n}{n}$	Curvature of the bowl	$\xrightarrow{p} Q_{X'X}$, rank p
$\dfrac{X_n'\varepsilon_n}{n}$	Tilt of the bowl	$\xrightarrow{p} 0_{p\times 1}$ (exogeneity)

Both convergences follow from the **Law of Large Numbers**, as shown next.

3 Asymptotic Behaviour of the Two Components

3.1 The Gram Matrix: $X_n'X_n/n$

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

We assume:

$$\frac{X_n'X_n}{n} \xrightarrow{p} Q_{X'X} \qquad \text{with } \operatorname{rank}(Q_{X'X}) = p.$$

How to verify this:

$$\frac{X_n'X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_{(i)}'X_{(i)}$$

If the rows $X_{(i)}$ are i.i.d. and $\mathbb{E}[X_{(1)}'X_{(1)}]$ exists, Kolmogorov's LLN gives:

$$\frac{1}{n}\sum_{i=1}^{n} X_{(i)}'X_{(i)} \xrightarrow{p} \mathbb{E}[X_{(1)}'X_{(1)}] =: Q_{X'X}.$$

Furthermore, $Q_{X'X}$ has rank p if no regressor is a linear combination of the others (no multicollinearity in the population).

◇ **Student's Notes**

Why rank p matters:
$Q_{X'X}$ plays the same role asymptotically that $\frac{X_n'X_n}{n}$ played in the finite-sample identification proof (**Lecture 3**). There, we needed $\operatorname{rank}(X_n) = p$ to ensure positive definiteness of $\frac{X'X}{n}$. Here, we need $\operatorname{rank}(Q_{X'X}) = p$ to ensure positive definiteness of the *limit* matrix—this is the **asymptotic identification** condition.

Dictionary: finite-sample vs. asymptotic identification

Finite-sample (Lect. 2–3)	Asymptotic (Lect. 11)
$\operatorname{rank}(X_n) = p$	$\operatorname{rank}(Q_{X'X}) = p$
$X_n'X_n/n$ positive definite	$Q_{X'X}$ positive definite
Bowl has no flat directions	Limit bowl has no flat directions
β_0 = unique minimiser of M_n^*	β_0 = unique, well-separated minimiser of M

Connection to Kolmogorov's LLN:
For i.i.d. $\{W_i\}$ with $\mathbb{E}[|W_1|] < \infty$: $\frac{1}{n}\sum_{i=1}^{n} W_i \xrightarrow{a.s.} \mathbb{E}[W_1]$. Applying element-wise to $W_i = X_{(i)}'X_{(i)}$ gives the result. Almost-sure convergence implies convergence in probability, so we get the $\xrightarrow{p}$ statement we need.

3.2 The Cross Product: $X_n'\varepsilon_n/n$

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

We assume:

$$\frac{1}{n}X_n'\varepsilon_n \xrightarrow{p} 0_{p\times 1}.$$

How to verify this:

$$\frac{1}{n}X_n'\varepsilon_n = \frac{1}{n}\sum_{i=1}^{n} X_{(i)}'\varepsilon_{(i)}.$$

If pairs $(X_{(i)}, \varepsilon_{(i)})$ are i.i.d. and $\mathbb{E}[X_{(1)}'\varepsilon_{(1)}]$ exists, then by Kolmogorov's LLN:

$$\frac{1}{n}\sum_{i=1}^{n} X_{(i)}'\varepsilon_{(i)} \xrightarrow{p} \mathbb{E}[X_{(1)}'\varepsilon_{(1)}] = 0_{p\times 1}.$$

The last equality holds because strict exogeneity $\mathbb{E}[\varepsilon_{(1)}|X_{(1)}] = 0$ implies $\mathbb{E}[X_{(1)}'\varepsilon_{(1)}] = 0_{p\times 1}$ by the Law of Iterated Expectations.

◇ **Student's Notes**

The Law of Iterated Expectations (LIE) step in detail:

$$\mathbb{E}\big[X_{(1)}'\varepsilon_{(1)}\big] = \mathbb{E}\Big[\mathbb{E}\big[X_{(1)}'\varepsilon_{(1)} \,\big|\, X_{(1)}\big]\Big] = \mathbb{E}\Big[X_{(1)}' \underbrace{\mathbb{E}\big[\varepsilon_{(1)} \,\big|\, X_{(1)}\big]}_{=0}\Big] = 0.$$

This is exactly the same mechanism used in **Lecture 3** to kill the cross term when taking conditional expectations. Here, instead of conditioning on the full design matrix X_n, we apply LIE marginally to one observation pair.

Intuition—the "tilt" that vanishes:
The cross term $(\beta_0 - \beta)'\frac{X_n'\varepsilon_n}{n}$ tilts the sample bowl away from the true minimum.

$$\text{Exogeneity} \implies \text{tilt} \xrightarrow{p} 0 \implies \text{sample bowl aligns with population bowl} \implies \hat{\beta}_n \xrightarrow{p} \beta_0.$$

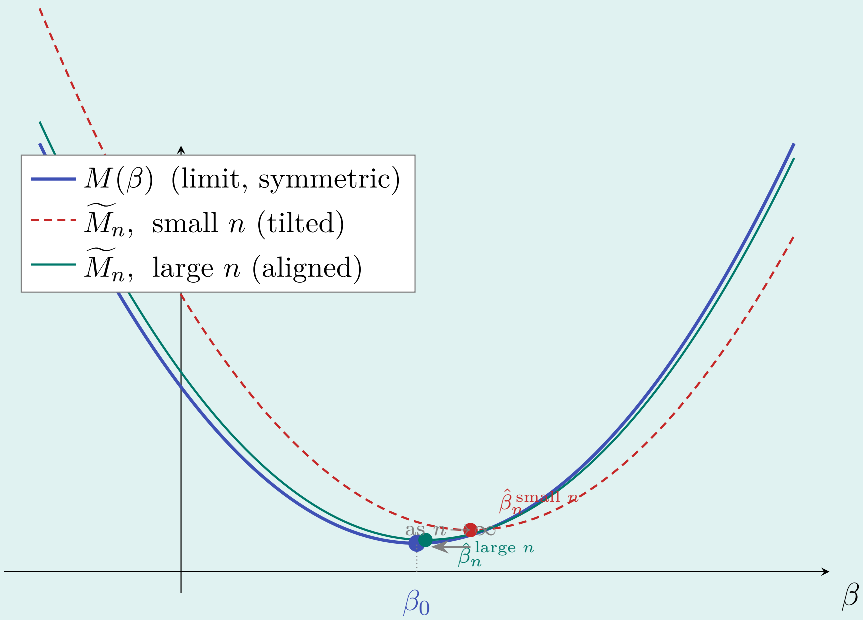

Figure 2: The cross term tilts the sample objective (rose dashed, small n). As $n \to \infty$ the tilt vanishes (teal, large n) and the sample bowl aligns with the symmetric limit $M(\beta)$ (indigo). Consequently the sample minimiser converges to β_0.

4 The Limit Function and Asymptotic Identification

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

Given convergences (a) and (b), our reduced objective converges:

$$\begin{aligned}\widetilde{M}_n(\beta) &= (\beta_0 - \beta)' \frac{X_n' X_n}{n} (\beta_0 - \beta) + 2(\beta_0 - \beta)' \frac{X_n' \varepsilon_n}{n} \\ &\xrightarrow{p} (\beta_0 - \beta)' \, Q_{X'X} \, (\beta_0 - \beta) + 2(\beta_0 - \beta)' \cdot 0_{p \times 1} \\ &= (\beta_0 - \beta)' \, Q_{X'X} \, (\beta_0 - \beta) \;\; =: \;\; M(\beta).\end{aligned}$$

Since $Q_{X'X}$ is positive definite (full rank p), $M(\beta)$ is a **strictly convex** quadratic form.
Second-order condition: $\dfrac{\partial^2 M}{\partial \beta \, \partial \beta'} = 2 \, Q_{X'X} \succ 0$.
$\min_\beta M(\beta) = 0$, attained *uniquely* at $\beta = \beta_0$. The minimum is therefore well-separated.

Key Result

Limit Function for the Linear Model
Under i.i.d. data, exogeneity, and $\operatorname{rank}(Q_{X'X}) = p$:

$$\widetilde{M}_n(\beta) \xrightarrow{p} M(\beta) = (\beta_0 - \beta)'\, Q_{X'X}\, (\beta_0 - \beta)$$

$M(\beta)$ is:

- **Non-negative:** $M(\beta) \geq 0$ for all β
- **Strictly convex:** Hessian $= 2\,Q_{X'X} \succ 0$
- **Uniquely minimised at β_0:** $M(\beta_0) = 0 < M(\beta)$ for all $\beta \neq \beta_0$
- **Well-separated:** the gap satisfies $\delta \geq \lambda_{\min}(Q_{X'X})\,\varepsilon^2 > 0$

Therefore: $\hat{\beta}_n \xrightarrow{p} \beta_0$ *(consistency of OLS).*

■ Proof / Derivation

Why the minimum is well-separated (quantitative bound)
For any β with $\|\beta - \beta_0\| \geq \varepsilon$:

$$M(\beta) = (\beta_0 - \beta)'\, Q_{X'X}\, (\beta_0 - \beta) \;\geq\; \lambda_{\min}(Q_{X'X})\, \|\beta - \beta_0\|^2 \;\geq\; \underbrace{\lambda_{\min}(Q_{X'X})\,\varepsilon^2}_{=:\,\delta} \;>\; 0,$$

where the first inequality uses the Rayleigh quotient bound and the second uses $\|\beta - \beta_0\| \geq \varepsilon$. Since $\operatorname{rank}(Q_{X'X}) = p$, all eigenvalues are strictly positive, so $\lambda_{\min}(Q_{X'X}) > 0$ and $\delta > 0$. □

◇ Student's Notes

The full consistency argument at a glance:

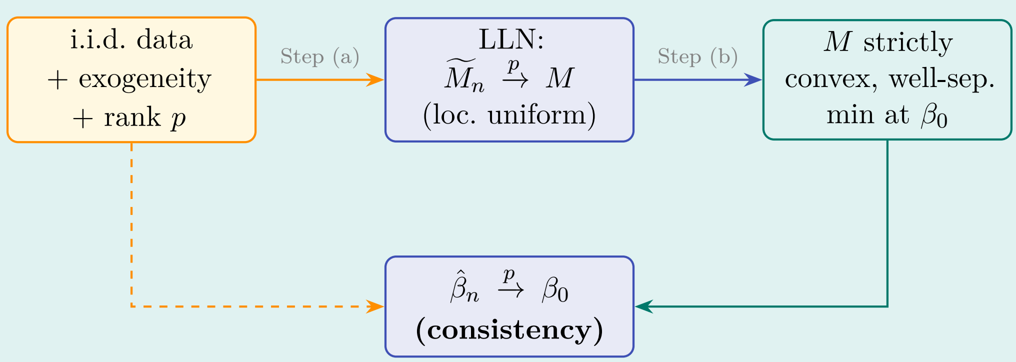

Figure 3: Logical chain from assumptions to consistency. Steps (a) and (b) are the two-step recipe from Lecture 10. The dashed arrow indicates that the assumptions feed into both steps.

Comparison with Lectures 3–4 (population identification):

Aspect	Lecture 3 (Population)	Lecture 11 (Asymptotic)
Objective analysed	$M_n^*(\beta) = \mathbb{E}[M_n \mid X_n]$	Limit $M(\beta)$ of $\widetilde{M}_n$
Curvature matrix	$X_n' X_n / n$ (finite sample)	$Q_{X'X} = \mathbb{E}[X'X]$ (population)
Key condition	$\text{rank}(X_n) = p$	$\text{rank}(Q_{X'X}) = p$
Cross-term killed by	Cond. exogeneity: $\mathbb{E}[\varepsilon \mid X] = 0$	LLN + LIE + exogeneity
Noise term	$\frac{1}{n}\text{tr}(I) = \sigma^2$ (constant)	Drops out of argmin
Conclusion	Unique min at β_0	Unique *and well-separated* min at β_0

5 Mild Misspecification: $\beta_0 \notin \Theta$

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

What happens if $\Theta \subset \mathbb{R}^p$ is a convex subset but the true parameter β_0 does **not** lie in Θ? (This is "mild" misspecification—the DGP $Y = X\beta_0 + \varepsilon$ is correct, but our parameter space excludes the truth.)

Because we never assumed $\beta_0 \in \Theta$ in the derivation of $\widetilde{M}_n(\beta)$, the convergence $\widetilde{M}_n \xrightarrow{p} M$ still holds on all of Θ.

If Θ is additionally **closed and convex**, then M has a unique, well-separated minimiser *within* Θ; call it $\beta^* \in \Theta$, $\beta^* \neq \beta_0$.

The estimator therefore converges: $\hat{\beta}_n \xrightarrow{p} \beta^*$ (not β_0).

The geometry: β^* is the point in Θ at **minimum Mahalanobis distance** from β_0, where the distance is measured by $Q_{X'X}$:

$$\beta^* = \arg\min_{\beta \in \Theta} \, (\beta_0 - \beta)' \, Q_{X'X} \, (\beta_0 - \beta).$$

◇ **Student's Notes**

Why the derivation still works without $\beta_0 \in \Theta$:

Look at the expansion:

$$\widetilde{M}_n(\beta) = (\beta_0 - \beta)' \tfrac{X'X}{n} (\beta_0 - \beta) + 2(\beta_0 - \beta)' \tfrac{X'\varepsilon}{n}.$$

Neither term requires $\beta_0 \in \Theta$; β_0 appears only as a fixed point in $\mathbb{R}^p$. The LLN applies to $X'X/n$ and $X'\varepsilon/n$ regardless of where β_0 sits relative to Θ.

Geometric picture (generalising Lecture 4, Figure 3):

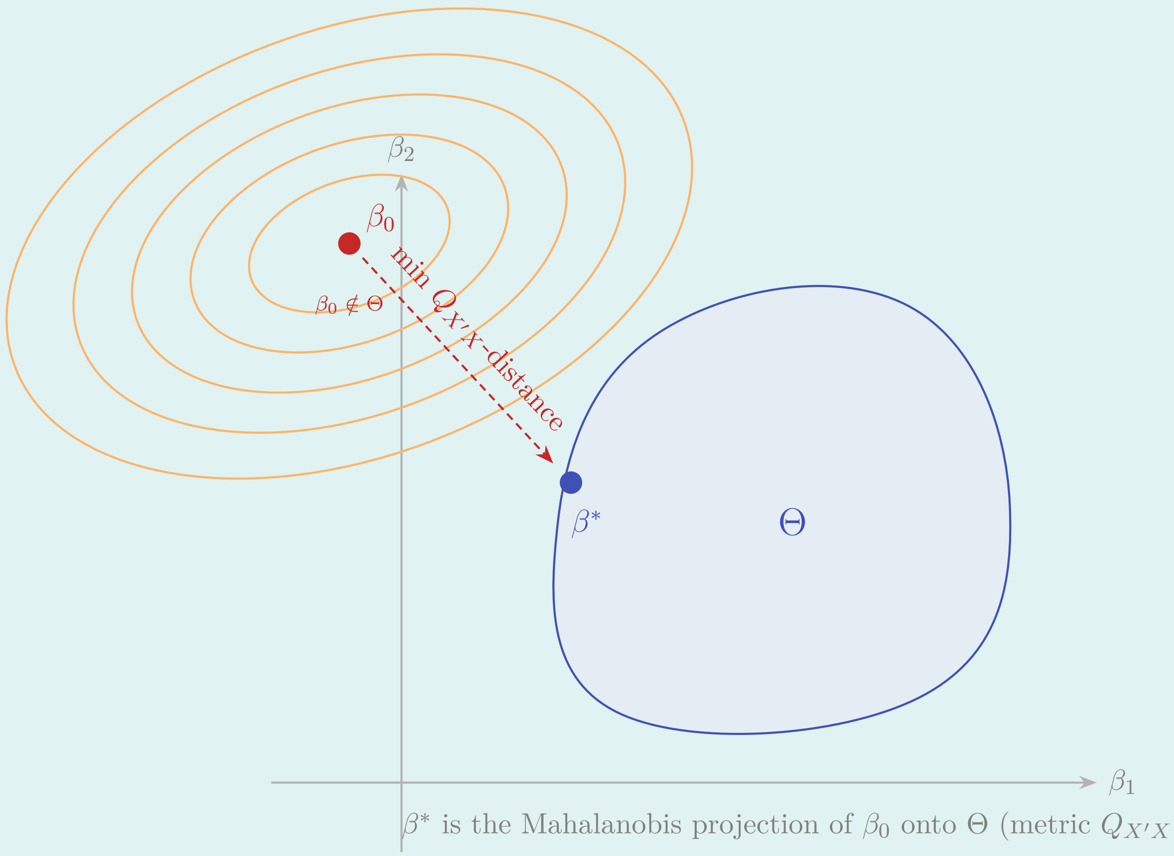

*Figure 4: Mild misspecification. The ellipses are level sets of $M(\beta) = (\beta_0 - \beta)' Q_{X'X} (\beta_0 - \beta)$ centred at β_0. The constrained minimiser β^* is the tangency point of the smallest ellipse with the boundary of Θ—the Mahalanobis projection. This generalises the finite-sample picture first introduced in **Lecture 4** (and revisited with the asymptotic metric $Q_{X'X}$ in **Lecture 7** for constrained IV).*

Key takeaways:

Situation	Consequence
$\beta_0 \in \Theta$ (well-specified)	$\hat{\beta}_n \xrightarrow{p} \beta_0$; **consistent**
$\beta_0 \notin \Theta$ (mildly misspecified)	$\hat{\beta}_n \xrightarrow{p} \beta^* \neq \beta_0$; **inconsistent**
Definition of β^*	$\arg\min_{\beta \in \Theta} (\beta_0 - \beta)' Q_{X'X} (\beta_0 - \beta)$
Metric	Mahalanobis distance with weight $Q_{X'X}$
Does $\widetilde{M}_n \to M$ hold?	Yes—derivation never used $\beta_0 \in \Theta$

Key Result

Pseudo-True Value Under Mild Misspecification
When $\beta_0 \notin \Theta$ but Θ is closed and convex:

$$\boxed{\hat{\beta}_n \xrightarrow{p} \beta^* = \arg\min_{\beta \in \Theta} (\beta_0 - \beta)' Q_{X'X} (\beta_0 - \beta) \neq \beta_0}$$

β^* is called the **pseudo-true value**. It is the point in Θ closest to β_0 under the $Q_{X'X}$-weighted Mahalanobis distance.

◇ Student's Notes

Scalar case ($p = 1$):
When $\Theta = [\underline{\beta}, \bar{\beta}]$ and $\beta_0 < \underline{\beta}$, the projection is simply $\beta^* = \underline{\beta}$.

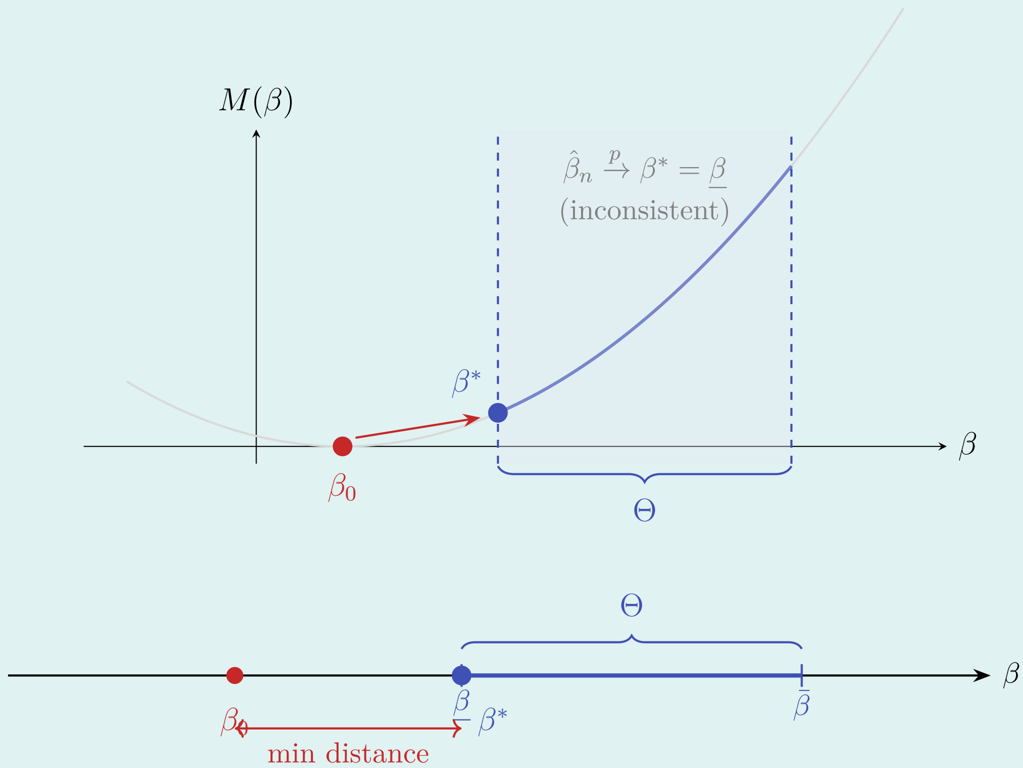

Figure 5: Scalar mild misspecification. ***Upper:*** *the parabola $M(\beta)$ has its unconstrained minimum at β_0 (rose dot, outside Θ). The constrained minimum β^* is at the left boundary of Θ (indigo dot).* ***Lower:*** *number-line view showing that β^* is the nearest point in Θ to β_0. This picture is consistent with the one first introduced in* ***Lecture 4****.*

! Watch Out

“Mild” misspecification $\neq$ harmless

“Mild” only means the DGP is still truly linear ($Y = X\beta_0 + \varepsilon$); it does *not* mean the consequences are minor:

- The estimator is **inconsistent**: $\hat{\beta}_n \not\to \beta_0$ no matter how large n is.
- Hypothesis tests and confidence intervals built around β^* have **no interpretation** in terms of the true β_0.
- The **bias** $\beta^* - \beta_0$ **does not shrink** with n; collecting more data cannot cure a misspecified parameter space.

Practical lesson: Always check whether your constraints are compatible with economic theory. If theory says $\beta_0 > 0$ but data strongly suggest $\beta_0 < 0$, the constraint $\Theta = [0, \infty)$ may be creating an inconsistency that no amount of data can resolve.

★ Intuition

Connecting Lectures 3, 4, 7, and 11:
The "projection onto Θ" idea appears throughout the course, each time with a richer structure:

Lecture	Setting	Metric used for projection
3	Population identification	$X_n'X_n/n$ (finite-sample)
4	Mild misspecification (finite)	$X_n'X_n/n$ (finite-sample)
7	Constrained IV ($q = p$)	$A_n = \frac{1}{n^2}X'ZWZ'X$ (instrument-based)
11	**Asymptotic misspecification**	$\boldsymbol{Q_{X'X}}$ **(population Gram matrix)**

In every case the structure is the same: a quadratic bowl centred at the truth, and a constrained space that may or may not contain the truth. The only thing that changes is *which matrix* defines the bowl's shape.

Quick-Reference Summary

◇ Student's Notes

Lecture 11 at a glance:

Topic	What was accomplished
Asymptotic identification	Defined the well-separated minimum condition; explained why it is stronger than uniqueness alone
Gram matrix convergence	$X_n'X_n/n \xrightarrow{p} Q_{X'X}$ via LLN under i.i.d. rows with rank $= p$
Cross product convergence	$X_n'\varepsilon_n/n \xrightarrow{p} 0$ via LLN + LIE + exogeneity
Limit function	$M(\beta) = (\beta_0 - \beta)'Q_{X'X}(\beta_0 - \beta)$; strictly convex; well-separated min at β_0
Consistency	Two-step recipe (Lect. 10) gives $\hat{\beta}_n \xrightarrow{p} \beta_0$
Mild misspecification	$\beta_0 \notin \Theta \Rightarrow \hat{\beta}_n \xrightarrow{p} \beta^*$ ($Q_{X'X}$-projection of β_0 onto Θ)

Key formulas:

Object	Expression / Convergence
Reduced objective	$\widetilde{M}_n(\beta) = (\beta_0-\beta)'\frac{X'X}{n}(\beta_0-\beta) + 2(\beta_0-\beta)'\frac{X'\varepsilon}{n}$
Gram matrix	$X_n'X_n/n \xrightarrow{p} Q_{X'X} = \mathbb{E}[X_{(1)}'X_{(1)}]$
Cross product	$X_n'\varepsilon_n/n \xrightarrow{p} 0$ (exogeneity + LLN)
Limit function	$M(\beta) = (\beta_0-\beta)'Q_{X'X}(\beta_0-\beta)$
Pseudo-true value	$\beta^* = \arg\min_{\beta\in\Theta}(\beta_0-\beta)'Q_{X'X}(\beta_0-\beta)$

Figures in this lecture:

Fig.	Content
1	Well-separated vs. non-well-separated minimum (two panels)
2	Cross-term tilt: sample bowl converges to limit bowl as $n\to\infty$
3	Logical chain: assumptions $\to$ LLN $\to$ well-separation $\to$ consistency
4	Mahalanobis projection of β_0 onto Θ (2-D, elliptical contours)
5	Scalar misspecification: parabola + number line (two panels)

Lectures 2–11 narrative arc (updated):

Lectures	Theme
2–3	Model setup; population identification (M_n^* unique min at β_0)
4–5	Analogy principle; extremum estimators; three asymptotic properties
6–7	Endogeneity; IV/GMM; just-identified simplification
8	LAD: median-based estimation
9–10	LP reformulation of LAD; two-step consistency recipe
11	**Asymptotic identification in the linear model; mild misspecification & pseudo-true values**