Calculus and Convergence for Estimation

PLSC 30700 — Post-Midterm Review (Lectures 9–18)

Robert Gulotty
University of Chicago

Winter 2026

This sheet collects the calculus skills and convergence concepts used in Lectures 9–18 (after the midterm). The pre-midterm calculus review is a separate document. Each section lists key definitions and results, then indicates where they appear in the course.

Abbreviations

FOC	First-Order Condition	SOC	Second-Order Condition
CLT	Central Limit Theorem	WLLN	Weak Law of Large Numbers
CMT	Continuous Mapping Theorem	CDF	Cumulative Distribution Function
MLE	Maximum Likelihood Estimation	GMM	Generalized Method of Moments
AME	Average Marginal Effect	MTE	Marginal Treatment Effect

Contents

Part I
Calculus Skills

1 Derivative Rules (Review and New Applications)

The pre-midterm review covered the power rule, product rule, and partial derivatives for OLS derivations. Post-midterm lectures apply these same rules in new settings (nonlinear models, likelihood functions, transformations of estimators) and add several new derivative techniques.

1.1 Chain Rule

If $y = f(g(x))$, then:

$$\frac{dy}{dx} = f'(g(x)) \cdot g'(x)$$

Multivariate version: if $y = f(\boldsymbol{g}(\boldsymbol{x}))$, then $\frac{\partial y}{\partial x_j} = f'(\boldsymbol{g}) \cdot \frac{\partial g}{\partial x_j}$.

Where it appears:

- **Day 9 (Probit):** Marginal effects require differentiating $\Phi(X'\beta)$:

$$\frac{\partial}{\partial x_j}\Phi(X'\beta) = \phi(X'\beta) \cdot \beta_j$$

Here $f = \Phi$ (the normal CDF) and $g = X'\beta$ (the index). Its derivative $\Phi'(\cdot) = \phi(\cdot)$ is the normal PDF.

- **Day 9 (Score):** The Probit score involves differentiating $\log \Phi(X'\beta)$:

$$\frac{\partial}{\partial \beta}\log \Phi(X'\beta) = \frac{\phi(X'\beta)}{\Phi(X'\beta)} \cdot X$$

This is the chain rule applied twice: first the log derivative ($1/u$), then the CDF derivative (ϕ).

- **Day 10 (Delta Method):** Approximating a nonlinear function $f(\hat{\beta})$ via Taylor expansion uses the chain rule to compute ∇f.

1.2 Quotient Rule

If $y = \frac{u(x)}{v(x)}$, then:

$$\frac{dy}{dx} = \frac{u'v - uv'}{v^2}$$

Equivalently, write $u/v = u \cdot v^{-1}$ and use the product rule with the power rule on v^{-1}.

Where it appears:

- **Day 10–11 (Delta Method):** The long-run elasticity $\theta = \beta_2/(1-\beta_4)$ requires the gradient:

$$\frac{\partial \theta}{\partial \beta_2} = \frac{1}{1-\beta_4}, \qquad \frac{\partial \theta}{\partial \beta_4} = \frac{\beta_2}{(1-\beta_4)^2}$$

  The second component is a quotient-rule application.

- **Day 11:** Peak experience in a log-wage equation: $\theta_3 = -50\beta_2/\beta_3$ yields gradient entries $-50/\beta_3$ and $50\beta_2/\beta_3^2$.

1.3 Logarithmic and Exponential Derivatives

$$\frac{d}{dx}\ln(x) = \frac{1}{x}, \qquad \frac{d}{dx}e^x = e^x, \qquad \frac{d}{dx}e^{f(x)} = f'(x)\, e^{f(x)}$$

For the normal PDF: $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, so $\phi'(x) = -x\,\phi(x)$.

Where it appears:

- **Day 9:** The Probit log-likelihood is $\ell(\beta) = \sum_i [Y_i \log \Phi(X_i'\beta) + (1-Y_i)\log(1-\Phi(X_i'\beta))]$. Differentiating uses $\frac{d}{du}\log u = 1/u$ and the chain rule.
- **Day 9:** The Hessian involves $\phi'(X'\beta) = -X'\beta \cdot \phi(X'\beta)$, a derivative of the normal PDF.
- **Day 17:** The airline cost regression uses a log-linear specification, relying on $\ln(x^a) = a\ln x$.

2 Partial Derivatives and Gradients

For $f : \mathbb{R}^k \to \mathbb{R}$, the **gradient** is the $k \times 1$ vector:

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_k} \end{pmatrix}$$

For $\boldsymbol{f} : \mathbb{R}^k \to \mathbb{R}^J$ (a vector-valued function), the **Jacobian** is the $J \times k$ matrix:

$$\boldsymbol{C} = \frac{\partial \boldsymbol{f}(\boldsymbol{b})}{\partial \boldsymbol{b}'} = \begin{pmatrix} \frac{\partial f_1}{\partial b_1} & \cdots & \frac{\partial f_1}{\partial b_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_J}{\partial b_1} & \cdots & \frac{\partial f_J}{\partial b_k} \end{pmatrix}$$

Where it appears:

- **Day 9:** The score vector $S_n(\beta) = \partial \ell_n / \partial \beta$ is a gradient of the log-likelihood.
- **Day 10–11 (Delta Method):** The Jacobian $\boldsymbol{C}(\boldsymbol{b})$ maps the asymptotic variance of $\hat{\boldsymbol{\beta}}$ to the asymptotic variance of $\boldsymbol{f}(\hat{\boldsymbol{\beta}})$ via $\boldsymbol{C}\,\widehat{\text{Var}}(\hat{\boldsymbol{\beta}})\,\boldsymbol{C}'$.
- **Day 15–16 (GMM):** The expected Jacobian $Q = \mathbb{E}\left[\frac{\partial g_i(\beta)}{\partial \beta'}\right]$ determines the asymptotic variance of GMM.
- **Day 17–18:** Panel data variance decomposition uses partial derivatives to separate within- and between-group variation.

3 Optimization

3.1 Unconstrained Optimization (FOC and SOC)

At a maximum $\hat{\theta}$ of $f(\theta)$:

- **FOC:** $\nabla f(\hat{\theta}) = \mathbf{0}$ (score = 0 at the optimum)
- **SOC:** $H(\hat{\theta}) = \frac{\partial^2 f}{\partial \theta \partial \theta'}$ is negative definite (Hessian confirms it is a maximum)

Where it appears:

- **Day 9:** Probit MLE: $S_n(\hat{\beta}) = 0$ (score equation). The Hessian $H_n(\hat{\beta})$ must be negative definite for a maximum.
- **Day 9:** Newton–Raphson uses both the score and Hessian: $\beta^{(t+1)} = \beta^{(t)} - H_n^{-1} S_n$.
- **Day 15:** GMM minimizes a quadratic criterion $J(\beta) = n\,\bar{g}_n(\beta)' \boldsymbol{W} \bar{g}_n(\beta)$. The FOC yields $\hat{\beta}_{GMM}$.

3.2 Constrained Optimization (Lagrange Multipliers)

To minimize $f(\boldsymbol{x})$ subject to $g(\boldsymbol{x}) = c$, form the Lagrangian:

$$\mathcal{L}(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) - \lambda \left[g(\boldsymbol{x}) - c\right]$$

FOCs: $\frac{\partial \mathcal{L}}{\partial x_j} = 0$ for all j, and $g(\boldsymbol{x}) = c$.

Where it appears:

- **Day 12:** Constrained Least Squares (CLS): minimize $(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})$ subject to $\boldsymbol{R}'\boldsymbol{\beta} = \boldsymbol{r}$, yielding the restricted estimator.
- **Day 12:** The CLS Lagrangian above is the main constrained optimization example in this course.

4 Taylor Series and Linear Approximation

First-order Taylor expansion of f around a:

$$f(b) \approx f(a) + f'(a)(b - a)$$

Multivariate version: $\boldsymbol{f}(\boldsymbol{b}) \approx \boldsymbol{f}(\boldsymbol{a}) + \boldsymbol{\Gamma}(\boldsymbol{b} - \boldsymbol{a})$, where $\boldsymbol{\Gamma} = \frac{\partial \boldsymbol{f}(\boldsymbol{a})}{\partial \boldsymbol{a}'}$ is the Jacobian at $\boldsymbol{a}$.

Where it appears:

- **Day 9 (Asymptotic Normality):** Taylor-expanding the score at β_0:

$$0 = S_n(\hat{\beta}) \approx S_n(\beta_0) + H_n(\beta_0)(\hat{\beta} - \beta_0)$$

  Rearranging gives $\sqrt{n}(\hat{\beta} - \beta_0) \approx \left[-\frac{1}{n}H_n\right]^{-1} \frac{1}{\sqrt{n}} S_n(\beta_0)$.

- **Day 10–11 (Delta Method):** If $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V})$, then for any smooth function $\boldsymbol{f}$:

$$\sqrt{n}(\boldsymbol{f}(\hat{\boldsymbol{\beta}}) - \boldsymbol{f}(\boldsymbol{\beta})) \xrightarrow{d} N(\boldsymbol{0},\ \boldsymbol{\Gamma}\boldsymbol{V}\boldsymbol{\Gamma}')$$

  The Taylor approximation is the entire justification for the delta method.

5 Integration

Definite integral: $\int_a^b f(x)\,dx$ computes the area under f from a to b.

Key rules:

- $\int_a^b c\, f(x)\,dx = c \int_a^b f(x)\,dx$ (constant factor)
- $\int_a^b x^n\,dx = \frac{x^{n+1}}{n+1}\Big|_a^b$ for $n \neq -1$ (power rule)
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$ if f is a PDF (normalization)

Where it appears:

- **Day 9 (IRT):** Integrating out the latent trait θ in the marginal likelihood:

$$L_i(a,b) = \int \prod_j \Phi(a_j\theta - b_j)^{Y_{ij}} (1 - \Phi(a_j\theta - b_j))^{1-Y_{ij}} \, \phi(\theta) \, d\theta$$

- **Day 10 (Markov's Inequality):** The proof splits $\mathbb{E}[X] = \int_0^\infty x f(x)\,dx$ into two parts at a and bounds the tail.
- **Day 14, 16 (MTE):** Treatment parameters are weighted integrals of the MTE:

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

6 Series

Geometric series: For $|\gamma| < 1$:

$$\sum_{k=0}^{\infty} \gamma^k = \frac{1}{1-\gamma}$$

Partial sum: $\sum_{k=0}^{N} \gamma^k = \frac{1-\gamma^{N+1}}{1-\gamma}$.

Where it appears:

- **Day 10:** The long-run multiplier for a lagged dependent variable model $y_t = \beta_0 + \beta_1 x_t + \gamma y_{t-1} + \varepsilon_t$. A one-unit increase in x has cumulative effect:

$$\beta_1 + \gamma\beta_1 + \gamma^2\beta_1 + \cdots = \beta_1 \sum_{k=0}^{\infty} \gamma^k = \frac{\beta_1}{1-\gamma}$$

  The long-run elasticity $\beta_1/(1-\gamma)$ is a nonlinear function of parameters — requiring the delta method for standard errors.

Part II
Convergence Concepts

7 Convergence of Random Variables

7.1 The Convergence Hierarchy

$$X_n \xrightarrow{\text{a.s.}} X \quad \Rightarrow \quad X_n \xrightarrow{p} X \quad \Rightarrow \quad X_n \xrightarrow{d} X$$

- **Almost sure (a.s.):** $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$. For almost every "path," the sequence converges.
- **In probability (p):** For all $\delta > 0$, $\mathbb{P}(|X_n - X| > \delta) \to 0$. The probability of being far from X vanishes.
- **In distribution (d):** $F_{X_n}(t) \to F_X(t)$ at all continuity points of F_X. The CDFs converge.

Each arrow is strict: the converse does not hold in general.

Where it appears:

- **Day 10:** The full hierarchy is developed. Convergence in probability is used for *consistency* (WLLN). Convergence in distribution is used for *asymptotic normality* (CLT).
- **Days 11–18:** Every asymptotic result in the course uses one of these modes.

7.2 Convergence in Probability (plim)

$X_n \xrightarrow{p} c$ means: for all $\delta > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - c| \le \delta) = 1$$

Written $\operatorname{plim} X_n = c$. Key properties:

- $\operatorname{plim}(X_n + Y_n) = \operatorname{plim} X_n + \operatorname{plim} Y_n$
- $\operatorname{plim}(X_n Y_n) = (\operatorname{plim} X_n)(\operatorname{plim} Y_n)$
- $\operatorname{plim}(X_n / Y_n) = (\operatorname{plim} X_n)/(\operatorname{plim} Y_n)$ if $\operatorname{plim} Y_n \neq 0$

Unlike expectations, plim passes through continuous functions (including division).

Where it appears:

- **Day 10:** Definition and examples. The contrast with expectation: $\mathbb{E}[X/Y] \neq \mathbb{E}[X]/\mathbb{E}[Y]$, but $\operatorname{plim}(X/Y) = (\operatorname{plim} X)/(\operatorname{plim} Y)$.
- **Day 11:** OLS consistency: $\operatorname{plim} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \operatorname{plim}\left[\left(\frac{\boldsymbol{X'X}}{n}\right)^{-1} \frac{\boldsymbol{X'e}}{n}\right] = \boldsymbol{\beta}$.
- **Day 13:** IV consistency: $\operatorname{plim} \hat{\beta}_{IV} = \beta + (\mathbb{E}[ZX'])^{-1}\mathbb{E}[Ze] = \beta$ under exogeneity.

8 Probability Inequalities

Markov's Inequality: For $X \ge 0$ and $a > 0$:

$$\mathbb{P}(X \ge a) \le \frac{\mathbb{E}[X]}{a}$$

Chebyshev's Inequality: For any X with mean μ and variance σ^2, and $k > 0$:

$$\mathbb{P}(|X - \mu| \ge k) \le \frac{\sigma^2}{k^2}$$

Chebyshev follows from Markov by setting $Y = (X - \mu)^2$ and $a = k^2$.

Where it appears:

- **Day 10:** Both inequalities are proved. Chebyshev is the key step in the WLLN proof:

$$\mathbb{P}\big(|\bar{X}_n - \mu| \ge \varepsilon\big) \le \frac{\operatorname{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0$$

- **Day 10:** Chebyshev sample size calculation: to ensure $|\bar{X}_n - \mu| < 1$ with probability 99%, need $n \ge \sigma^2/0.01$.

9 Laws of Large Numbers

Weak Law of Large Numbers (WLLN): If $X_1, \ldots, X_n$ are iid with $\mathbb{E}[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2 < \infty$, then:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$$

More generally, any sample moment converges to its population counterpart:

$$\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \xrightarrow{p} \mathbb{E}[X_i X_i']$$

Where it appears:

- **Day 10:** WLLN proved via Chebyshev. Counterexample: if observations share a common component ($X_i = Z + U_i$), then $\bar{X}_n \xrightarrow{p} Z$, a random variable, not a constant.
- **Day 11 (OLS Consistency):** $\frac{1}{n}\boldsymbol{X'X} \xrightarrow{p} Q_{XX}$ and $\frac{1}{n}\boldsymbol{X'e} \xrightarrow{p} \mathbf{0}$.
- **Day 9, 11 (MLE):** $\frac{1}{n}H_n(\beta_0) \xrightarrow{p} -\mathcal{I}(\beta_0)$ (Hessian converges to negative Fisher information).

10 Continuous Mapping Theorem and Slutsky's Theorem

Continuous Mapping Theorem (CMT): If $X_n \xrightarrow{p} c$ and g is continuous at c, then:

$$g(X_n) \xrightarrow{p} g(c)$$

Slutsky's Theorem: If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ (a constant), then:

$$X_n + Y_n \xrightarrow{d} X + c, \qquad Y_n X_n \xrightarrow{d} cX$$

Where it appears:

- **Day 10–11 (OLS):** CMT gives $\left(\frac{1}{n}\boldsymbol{X'X}\right)^{-1} \xrightarrow{p} Q_{XX}^{-1}$ (matrix inversion is continuous at nonsingular matrices). Slutsky then combines this with $\frac{1}{\sqrt{n}}\boldsymbol{X'e} \xrightarrow{d} N(\mathbf{0}, \Omega)$.
- **Day 11:** Ridge regression: $((\boldsymbol{X'X} + \lambda\boldsymbol{I})/n)^{-1} \xrightarrow{p} Q_{XX}^{-1}$ when $\lambda/n \to 0$.
- **Day 13:** IV consistency: $\text{plim}\,\hat{\beta}_{IV} = \text{plim}[(\boldsymbol{X'Z}/n)(\boldsymbol{Z'Z}/n)^{-1}(\boldsymbol{Z'X}/n)]^{-1}\cdots$ uses CMT for each component.

11 Central Limit Theorem

Lindeberg–Lévy CLT: If $X_1, \dots, X_n$ are iid with $\mathbb{E}[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2 < \infty$:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

Equivalently: $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$.

Multivariate CLT: If $\boldsymbol{W}_i$ are iid k-vectors with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$:

$$\sqrt{n}(\bar{\boldsymbol{W}}_n - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$$

Where it appears:

- **Day 10:** CLT proved intuitively via moment matching. Comparison: CLT sample size (2,862) vs. Chebyshev sample size (43,000) for the same precision.
- **Day 9, 11:** Score CLT: $\frac{1}{\sqrt{n}} S_n(\beta_0) \xrightarrow{d} N(\mathbf{0}, \mathcal{I}(\beta_0))$.
- **Day 11 (OLS):** $\frac{1}{\sqrt{n}} \sum X_i e_i \xrightarrow{d} N(\mathbf{0}, \Omega)$. Combined with Slutsky:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0},\ Q_{XX}^{-1} \Omega\, Q_{XX}^{-1})$$

- **Day 15:** GMM: $\frac{1}{\sqrt{n}} \sum g(W_i, \theta_0) \xrightarrow{d} N(\mathbf{0}, \Omega)$.

12 The Delta Method

If $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V})$ and $\boldsymbol{f} : \mathbb{R}^k \to \mathbb{R}^J$ is differentiable with Jacobian $\boldsymbol{\Gamma} = \frac{\partial \boldsymbol{f}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}$, then:

$$\sqrt{n}\Big(\boldsymbol{f}(\hat{\boldsymbol{\beta}}) - \boldsymbol{f}(\boldsymbol{\beta})\Big) \xrightarrow{d} N(\mathbf{0},\ \boldsymbol{\Gamma} \boldsymbol{V} \boldsymbol{\Gamma}')$$

Estimated standard error: $\widehat{\text{se}}(f(\hat{\beta})) = \sqrt{\boldsymbol{R}' \,\widehat{\text{Var}}(\hat{\beta})\, \boldsymbol{R}}$ where $\boldsymbol{R} = \nabla f(\hat{\beta})$.

Where it appears:

- **Day 10:** Gas consumption example: long-run elasticity $\theta = \beta_2/(1-\beta_4)$ with gradient vector $\boldsymbol{R}$. Standard error via $\sqrt{\boldsymbol{R}' \widehat{V} \boldsymbol{R}}$.
- **Day 11:** Peak experience in log-wage equation: $\theta = -50\beta_2/\beta_3$.
- **Day 9, 16:** Average marginal effects for Probit: $\text{AME}_j = \frac{1}{n} \sum \phi(X_i'\hat{\beta})\hat{\beta}_j$ with standard errors via the delta method.

13 Chi-Squared Distribution and Hypothesis Testing

If $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{I}_q)$, then $\boldsymbol{z}'\boldsymbol{z} \sim \chi^2_q$.

More generally, if $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V})$, the **Wald statistic** is:

$$W = n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\hat{\boldsymbol{V}}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \chi^2_q$$

This is a quadratic form in an asymptotically normal vector.

Where it appears:

- **Day 11–12:** Wald test for linear and nonlinear restrictions on $\boldsymbol{\beta}$.
- **Day 12:** Three equivalent tests: Wald (W), Score/LM (S), and F-test. All converge to χ^2_q under H_0.
- **Day 14:** Hausman test: $H = (\hat{\beta}_{IV} - \hat{\beta}_{OLS})'\widehat{V}^{-1}(\hat{\beta}_{IV} - \hat{\beta}_{OLS}) \xrightarrow{d} \chi^2_{k_2}$.
- **Day 15–16 (GMM):** Hansen's J-test for overidentification: $J = n\,\bar{g}_n'\hat{\Omega}^{-1}\bar{g}_n \xrightarrow{d} \chi^2_{l-k}$.

14 Summary: Concepts by Lecture Day

Lecture	**Calculus and Convergence Concepts**
Day 9 (Probit)	Chain rule (differentiating $\Phi(X'\beta)$); log derivatives (log-likelihood); product of chain rules (score and Hessian); Taylor expansion of the score (asymptotic normality sketch); integration over latent variables (IRT likelihood)
Day 10 (Asymptotics)	convergence in probability and in distribution; Markov and Chebyshev inequalities; WLLN (proof via Chebyshev); CLT; Taylor series and the delta method; geometric series (long-run multiplier); quotient rule and Jacobians for the delta method
Day 11 (Asymptotics II)	WLLN and CMT for OLS consistency; multivariate CLT and Slutsky for asymptotic normality; delta method for nonlinear functions of $\hat{\beta}$ (quotient rule, gradient vectors); Wald statistic (χ^2 from normal quadratic form)
Day 12 (Hypothesis Testing)	Constrained optimization (Lagrangian for CLS); score as gradient of log-likelihood; Wald, Score, and F tests (χ^2 convergence); test power as a function of $\sqrt{n}$
Day 13 (IV)	Probability limits of ratios (CMT/Slutsky); algebraic manipulation of variance ratios (attenuation bias)
Day 14 (2SLS)	Approximate bias via Taylor-like reasoning; no finite moments for just-identified IV; definite integrals (MTE as weighted integral); probability integral transform ($U_D \sim U[0,1]$); Hausman and Sargan χ^2 tests
Day 15 (GMM I)	FOC for quadratic criterion; Jacobian $Q = \mathbb{E}[\partial g/\partial\beta']$; sandwich collapse under optimal weighting; PSD ordering for efficiency; two-step consistency
Day 16 (GMM II)	Jacobian for Wald test; J-test (χ^2 convergence); integrals (MTE/PRTE as weighted integrals); semiparametric efficiency bound
Day 17 (Panel)	Log-linear specification (stated); within vs. between variance decomposition
Day 18 (Fixed Effects)	Variance/covariance algebra; GLS transformation ($\Sigma^{-1/2}$); Hausman test (χ^2); Nickell bias ($O(1/T)$); Arellano-Bond moment conditions; dynamic panel GMM