Econometrics

The Delta Method

Lasha Chochua

2026

Why Do We Need the Delta Method?

- OLS gives us $\hat{\beta}$ – we do care about $\hat{\beta}$ itself.
- But we often also care about **functions** of β:
  - a single coefficient, like β_1
  - a **ratio** of coefficients, like β_j/β_l
  - a **sum**, like $100\beta_2 + 20\beta_3$
  - an **optimum**, like the experience level that maximizes wages
- Call this object of interest $\theta = r(\beta)$.
- **Question:** if $\hat{\beta}$ has a standard error, what is the standard error of $r(\hat{\beta})$?

The Plug-in Estimator

- The natural estimator is just:

$$\hat{\theta} = r(\hat{\beta}).$$

- Good news: if $\hat{\beta}$ is close to β, then $r(\hat{\beta})$ is close to $r(\beta)$ (as long as r is continuous).
- Bad news: knowing $\text{Var}(\hat{\beta})$ does **not** immediately tell us $\text{Var}(r(\hat{\beta}))$, because r can be nonlinear.
- **The Delta Method** is the trick that solves this.

Two Objects: r and R – Don't Confuse Them!

- **Lowercase** r is the **function** itself. It takes $\beta \in \mathbb{R}^k$ and returns a scalar. Examples: $r(\beta) = 100\beta_1$, or $r(\beta) = -50\beta_2/\beta_3$.
- **Uppercase** R is the **gradient** of r – its vector of partial derivatives, also called the **Jacobian**:

$$R = \frac{\partial r(\beta)}{\partial \beta} = \begin{pmatrix} \partial r/\partial \beta_1 \\ \partial r/\partial \beta_2 \\ \vdots \\ \partial r/\partial \beta_k \end{pmatrix}.$$

- **Shape matters:** if β has k entries, then R is $k \times 1$.
- **Mnemonic:** r is the recipe, R tells you the sensitivity at each ingredient.

The Core Idea – Taylor Expansion

- Suppose r is smooth. Expand $r(\hat{\beta})$ around the true β:

$$r(\hat{\beta}) \approx r(\beta) + R'(\hat{\beta} - \beta).$$

- **Locally**, a nonlinear function looks linear. R is the slope.
- So $\hat{\theta} - \theta$ is approximately a **linear combination** of $\hat{\beta} - \beta$ – which is what makes the variance calculation easy.

From Taylor to Variance

- Subtract $r(\beta)$ from both sides:

$$\hat{\theta} - \theta \approx R'(\hat{\beta} - \beta).$$

- Take variances on both sides:

$$\text{Var}(\hat{\theta}) \approx R' \, \text{Var}(\hat{\beta}) \, R.$$

- **That's the Delta Method formula.** Everything else is just applying it.
- Standard error:

$$s(\hat{\theta}) = \sqrt{\hat{R}' \, \hat{V}_{\hat{\beta}} \, \hat{R}},$$

where we plug in $\hat{\beta}$ wherever β appears in R.

Delta Method in One Sentence

Delta Method

If $\hat{\beta}$ is approximately normal with variance $V_{\hat{\beta}}$, and r is smooth, then $\hat{\theta} = r(\hat{\beta})$ is approximately normal with variance $R' \, V_{\hat{\beta}} \, R$, where R is the gradient of r at β.

- Think of R as a **currency converter**: it translates uncertainty about β into uncertainty about θ.

Log Wage Regression – Example

- We estimate:

$$\log(wage) = \beta_1\, educ + \beta_2\, exper + \beta_3\, \frac{exper^2}{100} + \beta_4 + e.$$

- Three questions we care about:
  - $\theta_1 = 100\beta_1$ – percentage return to one extra year of education
  - $\theta_2 = 100\beta_2 + 20\beta_3$ – percentage return to experience at $exper = 10$
  - $\theta_3 = -50\,\beta_2/\beta_3$ – experience level that maximizes expected log wages
- θ_1, θ_2 are **linear** in β. θ_3 is **nonlinear** – this is where the Delta Method really earns its keep.

Estimates and Covariance Matrix

- Point estimates: $\hat{\beta}_1 = 0.118$, $\hat{\beta}_2 = 0.016$, $\hat{\beta}_3 = -0.022$.
- Heteroskedasticity-robust (HC2) covariance matrix:

$$\hat{V}_{\hat{\beta}} = \begin{pmatrix} 0.632 & 0.131 & -0.143 & -11.1 \\ 0.131 & 0.390 & -0.731 & -6.25 \\ -0.143 & -0.731 & 1.48 & 9.43 \\ -11.1 & -6.25 & 9.43 & 246 \end{pmatrix} \times 10^{-4}.$$

θ_1: Return to Education (Linear Case)

- Function: $r(\beta) = 100\beta_1$. Differentiate w.r.t. each of the four coefficients:

$$R = \begin{pmatrix} 100 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (4 \times 1).$$

- Because only the **first** slot of R is nonzero, only the **top-left entry** of $\hat{V}_{\hat{\beta}}$ – that is, 0.632×10^{-4} – survives the sandwich $R'\hat{V}_{\hat{\beta}}R$. Everything else gets multiplied by zero.
- Estimate: $\hat{\theta}_1 = 100 \times 0.118 = 11.8$.
- Standard error: $s(\hat{\theta}_1) = \sqrt{100^2 \times 0.632 \times 10^{-4}} = 0.8$.
- **Interpretation:** one extra year of schooling raises wages by about 11.8%, give or take 0.8 percentage points. Tight estimate!

θ_2: Return to Experience at 10 Years

- Function: $r(\beta) = 100\beta_2 + 20\beta_3$. Gradient:

$$R = \begin{pmatrix} 0 \\ 100 \\ 20 \\ 0 \end{pmatrix} \quad (4 \times 1).$$

- **Active block:** slots 2 and 3 are nonzero, so the sandwich uses the **middle** 2×2 **block** of $\widehat{V}_{\hat{\beta}}$ – rows/columns 2 and 3:

$$\text{block} = \begin{pmatrix} 0.390 & -0.731 \\ -0.731 & 1.48 \end{pmatrix} \times 10^{-4}.$$

- The off-diagonal -0.731 **must** be included – ignoring it is a common student mistake.

θ_2: Return to Experience at 10 Years (Cont.)

- Estimate: $\hat{\theta}_2 = 100(0.016) + 20(-0.022) = 1.6 - 0.44 = 1.16$.
- Standard error uses only the 2×2 submatrix for (β_2, β_3):

$$s(\hat{\theta}_2)^2 = \begin{pmatrix} 100 & 20 \end{pmatrix} \begin{pmatrix} 0.390 & -0.731 \\ -0.731 & 1.48 \end{pmatrix} \begin{pmatrix} 100 \\ 20 \end{pmatrix} \times 10^{-4}.$$

- Step by step: inner product $\begin{pmatrix} 0.390(100) + (-0.731)(20) \\ -0.731(100) + 1.48(20) \end{pmatrix} = \begin{pmatrix} 24.38 \\ -43.5 \end{pmatrix}$, then $100(24.38) + 20(-43.5) = 1568$, so $s(\hat{\theta}_2)^2 = 0.1568$ and $s(\hat{\theta}_2) \approx 0.40$.
- **Lesson:** when the transformation is linear, the off-diagonal (covariance) terms matter – ignoring them would give the wrong SE.

θ_3: The Nonlinear Case

- Function: $r(\beta) = -50\,\beta_2/\beta_3$. Gradient (still a 4×1 vector!):

$$R = \begin{pmatrix} 0 \\ -50/\beta_3 \\ 50\beta_2/\beta_3^2 \\ 0 \end{pmatrix}.$$

- Same **active block** as θ_2: slots 2 and 3, so the sandwich again uses the middle 2×2 submatrix of $\hat{V}_{\hat{\beta}}$.
- **New feature:** R depends on β itself – we plug in $\hat{\beta}$ to get $\hat{R}$.
- Estimate: $\hat{\theta}_3 = -50 \times 0.016/(-0.022) \approx 36.4$ years.

θ_3: Standard Error

- With $\hat{\beta}_2 = 0.016$, $\hat{\beta}_3 = -0.022$:

$$\hat{R}_2 = -50/(-0.022) \approx 2273, \qquad \hat{R}_3 = 50(0.016)/(-0.022)^2 \approx 1653.$$

- Plug into the sandwich using the same middle 2×2 block of $\hat{V}_{\hat{\beta}}$ as for θ_2:

$$s(\hat{\theta}_3) = \sqrt{\hat{R}'\hat{V}_{\hat{\beta}}\hat{R}} \approx 7.5.$$

- **Interpretation:** the peak of the wage-experience profile is estimated at about 36 years, with a standard error of about 7.5 years. A 95% CI runs roughly from 22 to 51 years – a wide band.
- **Takeaway:** nonlinear transformations – especially **ratios** – often inflate uncertainty. The Delta Method exposes this honestly, while a naive "just report the point estimate'' would hide it.

Confidence Intervals From the Delta Method

- Once we have $\hat{\theta}$ and $s(\hat{\theta})$, a 95% CI is the usual:

$$\hat{\theta} \pm 1.96 \cdot s(\hat{\theta}).$$

- For our three parameters:
  - θ_1: $11.8 \pm 1.96(0.8) = [10.2,\ 13.3]$ – tight
  - θ_2: $1.16 \pm 1.96(0.40) \approx [0.38,\ 1.94]$ – positive but wider
  - θ_3: $36 \pm 1.96(7) \approx [22,\ 50]$ – imprecise

What to Remember

- **Recipe for the Delta Method:**
  1. Write down the function of interest $\theta = r(\beta)$.
  2. Compute its gradient $R = \partial r / \partial \beta$.
  3. Plug in $\hat{\beta}$ to get $\hat{R}$.
  4. Standard error: $s(\hat{\theta}) = \sqrt{\hat{R}' \hat{V}_{\hat{\beta}} \hat{R}}$.
  5. Build the CI as usual: $\hat{\theta} \pm 1.96\, s(\hat{\theta})$.
- **Why it works:** locally, every smooth function is linear. Linear functions of normal variables are normal.
- **Warning:** the approximation is **local**. If $\hat{\beta}$ is noisy or r is very curved near β, the Delta Method can be misleading – bootstrap is the usual backup.