The Delta Method

Variance Estimation for Transformations of Estimated Parameters

LearnByCode Project

1 Why the Delta Method?

You estimated a regression and obtained coefficient estimates $\hat{\beta}_1, \hat{\beta}_2, \ldots$ with their covariance matrix. Now someone asks: *what is the standard error of $\hat{\beta}_1/\hat{\beta}_2$?*

You cannot simply divide the standard errors—the ratio of two normal random variables is not normal, and the covariance between numerator and denominator matters. The **delta method** solves this: it gives you the asymptotic variance of *any* smooth function of your estimates, using only quantities you already have.

When do you need it? Whenever you need the standard error of a quantity that is a function of estimated coefficients:

1. **Ratio of coefficients**—e.g., the relative effect of education vs. experience on wages
2. **Elasticities** from a log-linear model—e.g., price elasticity $= \hat{\beta}_{\text{price}} \cdot \bar{p}/\bar{q}$
3. **Marginal effects** in nonlinear models—e.g., probit/logit: $\hat{\beta} \cdot \phi(X'\hat{\beta})$
4. **Long-run multipliers** in dynamic models—e.g., sum of lag coefficients divided by $(1 - \hat{\rho})$
5. **Structural parameters** from reduced-form estimates—e.g., in IV/SEM models
6. **Predicted values** at a specific point—e.g., predicted wage at education = 16, experience = 5
7. **Testing linear restrictions**—e.g., $H_0\colon \beta_1 = \beta_2$, which is $R\beta - r = 0$ with $R = (1,\ -1)$, $r = 0$

All of these are functions $g(\hat{\theta})$ of the coefficient vector. The delta method tells you:

> *If you know the variance of $\hat{\theta}$, you can get the variance of $g(\hat{\theta})$ by sandwiching with the Jacobian.*

A concrete example

Suppose OLS gives $\hat{\beta} = (3.05,\ 0.98)'$ with covariance matrix

$$\widehat{V} = \begin{pmatrix} 0.0042 & -0.0001 \\ -0.0001 & 0.0038 \end{pmatrix}.$$

You want the standard error of the ratio $g(\beta) = \beta_1/\beta_2$.

1. **Point estimate:** $\hat{\phi} = 3.05/0.98 = 3.112$.
2. **Jacobian:** $G = \dfrac{\partial g}{\partial \beta'} = \left(\dfrac{1}{\beta_2},\ -\dfrac{\beta_1}{\beta_2^2}\right)\Big|_{\hat{\beta}} = (1.020,\ -3.178)$.

3. **Variance:** $\widehat{\mathrm{Var}}(\hat{\phi}) = G\,\widehat{V}\,G' = 0.0427$.
4. **Standard error:** $\mathrm{se}(\hat{\phi}) = \sqrt{0.0427} = 0.207$.

That is all the delta method does—three lines of computation once you have the Jacobian.

2 The Result

Theorem 1 (Delta Method—Hansen (2021), Thm 6.8; Robinson (2008), Lemma 1)**.** *Let $\mu \in \mathbb{R}^k$ and $g\colon \mathbb{R}^k \to \mathbb{R}^q$. If $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \xi$, where g is continuously differentiable in a neighborhood of μ, then*

$$\sqrt{n}(g(\hat{\mu}) - g(\mu)) \xrightarrow{d} G\,\xi,$$

*where $G = \dfrac{\partial}{\partial u'}\, g(u)\Big|_{u=\mu}$ is the $q \times k$ **Jacobian matrix**.*

In particular, if $\xi \sim \mathcal{N}(0, V)$, then

$$\boxed{\sqrt{n}(g(\hat{\mu}) - g(\mu)) \xrightarrow{d} \mathcal{N}(0,\ G\,V\,G').} \tag{1}$$

Dimensions at a glance:

Object	Size	Description
$\theta_0,\ \hat{\theta}$	$k \times 1$	Parameter vector and its estimator
V	$k \times k$	Asymptotic covariance of $\sqrt{n}(\hat{\theta} - \theta_0)$
$g(\cdot)$	$\mathbb{R}^k \to \mathbb{R}^q$	Transformation of interest
$G = \partial g / \partial \theta'$	$q \times k$	Jacobian evaluated at θ_0
$G\,V\,G'$	$q \times q$	Asymptotic covariance of $\sqrt{n}(g(\hat{\theta}) - g(\theta_0))$

3 Proof Sketch

Proof (Robinson 2008, Section 5.2.1). By the mean value theorem,

$$g(\hat{\theta}) - g(\theta_0) = F(\tilde{\theta})\,(\hat{\theta} - \theta_0),$$

where $\tilde{\theta}$ lies between $\hat{\theta}$ and θ_0, and $F(\theta) = \partial g(\theta)/\partial\theta'$ is the Jacobian. Since $\hat{\theta}$ is consistent, $\tilde{\theta} \xrightarrow{p} \theta_0$, and by continuity of F:

$$F(\tilde{\theta}) \xrightarrow{p} F(\theta_0).$$

By Slutsky's theorem,

$$\sqrt{n}(g(\hat{\theta}) - g(\theta_0)) = F(\theta_0)\,\sqrt{n}(\hat{\theta} - \theta_0) + o_p(1).$$

The result follows by Cramér's theorem. □

Key ingredients: (i) a first-order Taylor expansion, (ii) consistency of $\hat{\theta}$ so the remainder vanishes, (iii) Slutsky to combine a converging matrix with a converging random vector.

4 Two Special Cases

4.1 Linear transformation: $\phi = R\theta - r$

When $g(\theta) = R\theta - r$ with R a $q \times k$ matrix and r a $q \times 1$ vector, the Jacobian is simply $G = R$ (constant), so

$$\mathrm{Var}(\hat{\phi}) = R\,V\,R'.$$

This is **exact**—no first-order approximation is involved. Common uses:

- Testing $\beta_1 = \beta_2$: $R = (1,\ -1)$, $r = 0$
- Constant returns to scale: $R = (1,\ 1)$, $r = 1$
- Exclusion restrictions: $R = (0,\ I)$, $r = 0$

See Hansen (2021), Chapter 8.1 for the connection to constrained least squares.

4.2 Nonlinear transformation: $\phi = g(\theta)$

The Jacobian $G = \partial g/\partial\theta'$ must be evaluated—analytically or numerically. The variance formula

$$\mathrm{Var}(\hat{\phi}) \approx G\,V\,G'$$

is a **first-order approximation**. Its quality depends on:

- Sample size (how close $\hat{\theta}$ is to θ_0)
- Curvature of g (higher-order Taylor terms)
- Whether G has full rank at θ_0

5 Where Does V Come From? The Smooth Function Model

The delta method theorem starts with "given that you know $V\ldots$" But where does V actually come from? In practice, most estimators are **plug-in estimators**: you replace population moments with sample moments and then apply some formula.

The *smooth function model* (Hansen 2021, Section 6.6) formalizes this. The parameter of interest is

$$\theta = g(\mu), \qquad \mu = \mathbb{E}[h(Y)],$$

where $\hat{\mu} = n^{-1}\sum_{i=1}^{n} h(Y_i)$ is the sample analogue. The plug-in estimator $\hat{\theta} = g(\hat{\mu})$ just substitutes $\hat{\mu}$ for μ.

Theorem 2 (Smooth Function Model—Hansen (2021), Thm 6.10)**.** *If $Y_i \in \mathbb{R}^m$ are i.i.d., $h\colon \mathbb{R}^m \to \mathbb{R}^k$, $\mathbb{E}\|h(Y)\|^2 < \infty$, and $G(u) = \frac{\partial}{\partial u'} g(u)$ is continuous in a neighborhood of μ, then*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0,\ G\,V\,G'),$$

where $V = \mathbb{E}\big[(h(Y) - \mu)(h(Y) - \mu)'\big]$ is $k \times k$ and $G = G(\mu)$ is $q \times k$.

Why this matters: the smooth function model tells you that the delta method applies automatically to any plug-in estimator. OLS is a special case: $\mu = \big(\mathbb{E}[X'X],\ \mathbb{E}[X'Y]\big)$, and g extracts $\beta = \mathbb{E}[X'X]^{-1}\mathbb{E}[X'Y]$. So when you compute the OLS covariance matrix, you are already using the smooth function model—the delta method is the same idea applied one more time on top.

6 Singularity: The Collinearity of Transformations

In OLS, **collinearity** means that your regressors are linearly dependent: $X'X$ is singular and you cannot invert it. The delta method has an exact analogue: if your *parameters of interest* are linearly dependent functions of the *estimated coefficients*, then $G\,V\,G'$ is singular and you cannot get independent standard errors for all of them.

Remark 3 (Robinson 2008, Remark 13)*.* If the parameters of interest ϕ are functions of fewer estimated coefficients θ than there are functions (i.e., $q > k$, or more generally the rows of G are linearly dependent), then $G\,V\,G'$ is **singular**.

Example 4*.* You estimated two coefficients $\hat{\theta}_1, \hat{\theta}_2$, but you report three quantities of interest:

$$\phi_1 = \theta_1 + \theta_2, \qquad \phi_2 = \theta_1 - \theta_2, \qquad \phi_3 = 2\theta_1 + 3\theta_2.$$

Here $\phi_3 = \frac{1}{2}(3\phi_1 + \phi_2)$—a linear combination of the other two. The 3×3 covariance matrix of $(\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3)$ has rank at most 2, just as adding a collinear regressor to an OLS model makes $X'X$ singular. The remedy is the same: drop the redundant quantity, or recognize that you have only two independent parameters.

7 Implementation

The MATLAB implementation in `DeltaMethod.m` performs the finite-sample analogue of (1):

```
para    = g(coef);                % point estimate: g(theta_hat)
F       = JacobianEst(g, coef); % numerical Jacobian: G at theta_hat
varpara = F * varcoef * F';     % sandwich: G * V_hat * G'
```

`JacobianEst.m` uses **central finite differences**:

$$\frac{\partial g}{\partial \theta_j} \approx \frac{g(\theta + h\,e_j) - g(\theta - h\,e_j)}{2h}, \qquad h = \varepsilon^{1/3} \cdot \max(|\theta_j|,\ 1),$$

where $\varepsilon \approx 2.2 \times 10^{-16}$ is machine epsilon. This balances truncation error $O(h^2)$ against roundoff error $O(\varepsilon/h)$.

Self-validation

A good test: compare the numerical Jacobian against the analytical one. For $g(\beta) = \beta_1/\beta_2$, the analytical Jacobian is $G = (1/\beta_2,\ -\beta_1/\beta_2^2)$. The numerical and analytical standard errors should agree to $\sim$10 digits. See `MinimalExample.m` for a runnable verification.

8 Assumptions

1. **Asymptotic normality** of the original estimator: $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V)$. Holds for OLS, IV, GMM, MLE under standard regularity conditions.

2. **Continuous differentiability** of g in a neighborhood of θ_0. Excludes functions like $|\theta|$ at $\theta = 0$ or indicator functions.

3. **Consistent covariance estimation**: $\widehat{V}$ is a consistent estimator of V/n.

4. **Non-degenerate Jacobian**: G has rank q for the standard errors to be finite (see Section 6 for the singular case).

References

- Hansen, B. E. (2021). *Econometrics.* Princeton University Press. Theorem 6.8, Sections 6.5–6.6, 7.10–7.11.
- Robinson, P. M. (2008). *EC484 Econometric Analysis: Lecture Notes.* LSE. Section 5.2.1, Lemma 1, Remark 13.

Online resources

- Robinson, T. S. (2024). "Delta Method," Chapter 7 in *10 Fundamental Theorems for Econometrics.* https://bookdown.org/ts_robinson1994/10EconometricTheorems/dm.html
- Stata FAQ. "Explanation of the Delta Method." https://www.stata.com/support/faqs/statistics/delta-method/
- Statlect. "Delta Method." https://www.statlect.com/asymptotic-theory/delta-method
- CRAN / modmarg. "What is the Delta Method?" https://cran.r-project.org/web/packages/modmarg/vignettes/delta-method.html