Robust Standard Errors

Gov 2001: Quantitative Social Science Methods I

Week 12, Lecture 23

Spring 2026

For Today

Required Reading

- Aronow & Miller, §4.1.4 (pp. 151–156)
- Angrist & Pischke, §8.1 (pp. 219–221)

Today: What happens when variance isn't constant, and how to fix inference.

Roadmap

1. Homoskedasticity vs. heteroskedasticity
2. Consequences of heteroskedasticity
3. Robust standard errors
4. When to use robust SEs
5. Practical guidance

Part I: Homoskedasticity vs. Heteroskedasticity

Recall: The Variance of $\hat{\beta}$

Under the **Gauss-Markov assumptions**:

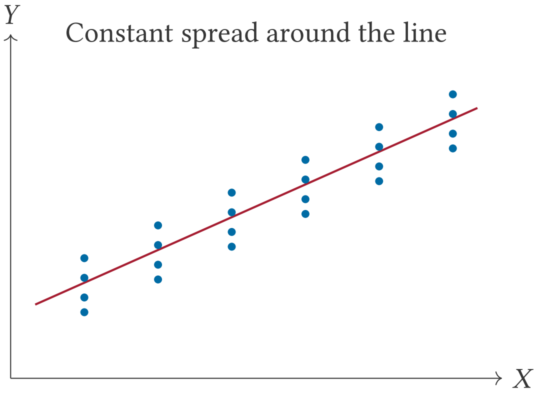
$$\text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

This requires **homoskedasticity**:

$$\text{Var}(\varepsilon_i|X_i) = \sigma^2 \quad \text{for all } i$$

The error variance is the **same** for all observations.

Homoskedasticity: Visual



Error variance is the same at all values of X.

Heteroskedasticity

Heteroskedasticity: Error variance varies with X.

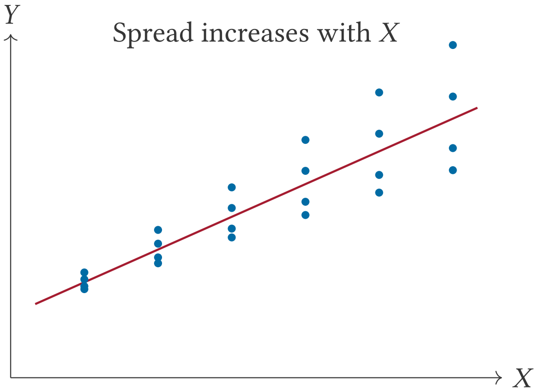
$$\mathrm{Var}(\varepsilon_i | X_i) = \sigma_i^2 \neq \sigma^2$$

Common patterns:

- Variance increases with X (e.g., income and spending)
- Variance decreases with X
- Variance differs by group

"Hetero" = different, "skedastic" = scatter.

Heteroskedasticity: Visual



Error variance is larger for larger values of X.

Real-World Examples

Income and consumption:
Rich households have more variable spending than poor households.

Firm size and growth:
Large firms have more variable growth rates than small firms.

Education and wages:
Returns to education may be more variable for college graduates.

Heteroskedasticity is the rule, not the exception.

Part II: Consequences of Heteroskedasticity

Good News: OLS Is Still Unbiased

Unbiasedness only requires:

$$\mathbb{E}[\varepsilon_i | X_i] = 0$$

This doesn't involve the variance!

Result:

- OLS is still unbiased under heteroskedasticity
- OLS is still consistent
- Point estimates are fine

Bad News: Standard Errors Are Wrong

The usual variance formula:

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Assumes $\text{Var}(\varepsilon_i|X) = \sigma^2$ for all i.

Under heteroskedasticity:

- This formula is **wrong**
- Standard errors could be too large or too small
- t-statistics are wrong
- Confidence intervals have wrong coverage
- Hypothesis tests have wrong size

Direction of Standard Error Bias

No general rule for direction:

- If high-variance observations have high leverage (extreme X):
  Classical SEs tend to be **too small**
  $\Rightarrow$ Over-reject null hypotheses

- If high-variance observations have low leverage:
  Classical SEs tend to be **too large**
  $\Rightarrow$ Under-reject null hypotheses

Either way, inference is unreliable.

OLS Is No Longer BLUE

Under heteroskedasticity:

- OLS is still unbiased
- OLS is still linear
- But OLS is **not efficient**

Gauss-Markov doesn't apply.
There exist other linear unbiased estimators with lower variance.

(Weighted Least Squares can be more efficient if you know the variance structure.)

Part III: Robust Standard Errors

The Robust Standard Error Idea

Problem: We're using the wrong variance formula.

Solution: Use a formula that's valid under heteroskedasticity.

Key insight: We can estimate the variance without assuming homoskedasticity.

- Use each residual $\hat{e}_i^2$ as an estimate of σ_i^2
- This accounts for varying error variances

The General Variance Formula

Without assuming homoskedasticity:

$$\text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

where $\boldsymbol{\Omega} = \text{Var}(\boldsymbol{\varepsilon}|\mathbf{X})$ is the variance-covariance matrix of errors.

Under heteroskedasticity:

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

Diagonal, but not $\sigma^2\mathbf{I}$.

The Sandwich Estimator

Structure: “Bread-Meat-Bread”

$$\underbrace{(\mathbf{X}'\mathbf{X})^{-1}}_{\text{Bread}}\underbrace{\mathbf{X}'\hat{\Omega}\mathbf{X}}_{\text{Meat}}\underbrace{(\mathbf{X}'\mathbf{X})^{-1}}_{\text{Bread}}$$

Eicker-Huber-White (EHW) estimator:

$$\hat{\Omega} = \text{diag}(\hat{e}_1^2, \hat{e}_2^2, \ldots, \hat{e}_n^2)$$

Use squared residuals to estimate each σ_i^2.

The Robust Variance Formula (HC0)

$$\widehat{\mathrm{Var}}_{HC0}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{i=1}^{n} \hat{e}_i^2 \mathbf{x}_i \mathbf{x}_i'\right)(\mathbf{X}'\mathbf{X})^{-1}$$

where $\mathbf{x}_i$ is the i-th row of $\mathbf{X}$ (as a column vector).

Key property: Consistent for $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ even under heteroskedasticity.

Standard errors from this matrix are **robust standard errors**.

Small-Sample Corrections

HC0 can be biased in small samples (residuals underestimate true errors).

Common corrections:

- **HC1:** Multiply by $n/(n-k-1)$
  Simple degrees-of-freedom correction
- **HC2:** Weight by $1/(1-h_{ii})$
  Accounts for leverage (h_{ii} = diagonal of hat matrix)
- **HC3:** Weight by $1/(1-h_{ii})^2$
  More conservative; jackknife-like

Most software uses HC1 or HC2 by default.

Part IV: When to Use Robust SEs

The Modern Consensus

Angrist & Pischke's advice:

"Always use robust standard errors."

Reasoning:

- Heteroskedasticity is common in practice
- You rarely know the true variance structure
- Robust SEs are valid either way
- The "cost" of robust SEs is small (slightly less efficient if homoskedastic)

When in doubt, use robust standard errors.

When Classical SEs Might Be Fine

Robust SEs aren't magic. Classical SEs may be preferable when:

- You have good reason to believe homoskedasticity holds
- Sample size is small (robust SEs can be unreliable)
- You're doing simulation studies with known error structure
- Computational simplicity matters

But in most applied work: **default to robust**.

Testing for Heteroskedasticity

Should you test before deciding?

Classic tests:

- Breusch-Pagan test
- White test

Modern view: Don't bother testing.

- Tests have low power against some alternatives
- Pre-testing affects final inference
- Just use robust SEs by default

Part V: Practical Guidance

Using Robust SEs in Practice

R:

- `lm_robust()` from `estimatr` package
- `coeftest(model, vcov = vcovHC(model))` from `sandwich`

Stata:

- `reg y x, robust`
- `reg y x, vce(robust)`

Python:

- `statsmodels`: `model.fit(cov_type='HC1')`

Reporting in Papers

Always state what SEs you're using:

"Standard errors are heteroskedasticity-robust (HC1)."
"Robust standard errors in parentheses."

Don't:

- Switch between classical and robust without explanation
- Use whichever gives smaller SEs
- Forget to mention what you're using

When Robust SEs Aren't Enough

Robust SEs handle **heteroskedasticity**.

They do **NOT** handle:

- Autocorrelation (errors correlated across observations)
- Clustering (groups of correlated observations)
- Omitted variable bias
- Measurement error
- Model misspecification

Next lecture: Clustered standard errors for when observations are correlated within groups.

Part VI: Robust SEs in R

Classical vs. Robust SEs in R

First, fit a model with classical SEs:

```r
# Standard OLS (classical SEs)
model <- lm(y ~ x, data = mydata)

# View classical standard errors
summary(model)$coefficients
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    2.145     0.3821   5.614  4.2e-08
## x              0.892     0.0234  38.120  < 2e-16
```

These SEs assume homoskedasticity.

Using estimatr for Robust SEs

The `lm_robust()` function makes robust SEs easy:

```
# Install if needed: install.packages("estimatr")
library(estimatr)

# Robust SEs (HC2 by default)
model_robust <- lm_robust(y ~ x, data = mydata)

# View results
summary(model_robust)
##              Estimate  Std. Error t value  Pr(>|t|)
## (Intercept)     2.145     0.4102    5.229  3.1e-07
## x               0.892     0.0312   28.590  < 2e-16
```

Notice: Same coefficient, different (larger) standard error.

Choosing the HC Type

You can specify which HC estimator to use:

```r
# HC0: Original Eicker-Huber-White
lm_robust(y ~ x, data = mydata, se_type = "HC0")

# HC1: Stata default (degrees-of-freedom correction)
lm_robust(y ~ x, data = mydata, se_type = "HC1")

# HC2: R default (leverage adjustment)
lm_robust(y ~ x, data = mydata, se_type = "HC2")

# HC3: Most conservative (jackknife-like)
lm_robust(y ~ x, data = mydata, se_type = "HC3")
```

HC2 is a good default; HC1 matches Stata's robust.

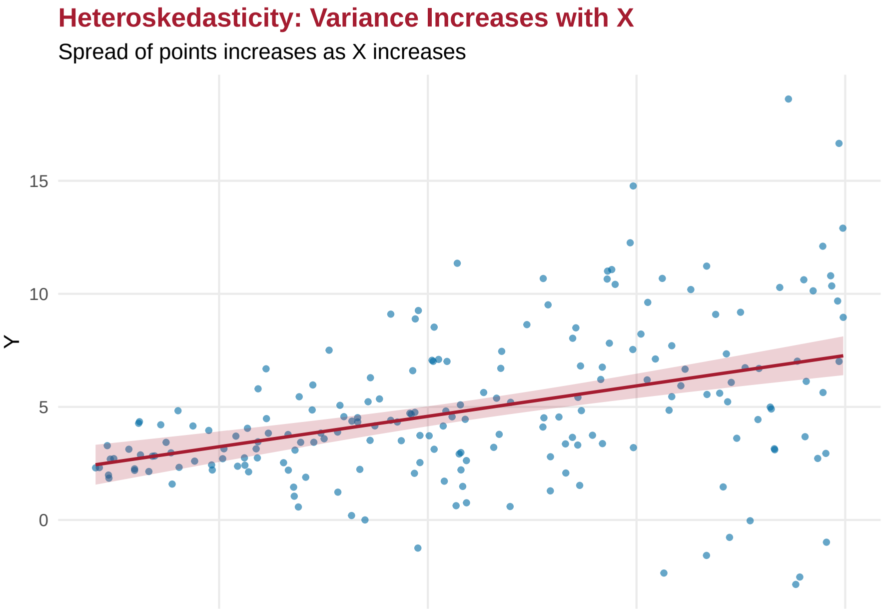
Alternative: The sandwich Package

For more flexibility, use sandwich + lmtest:

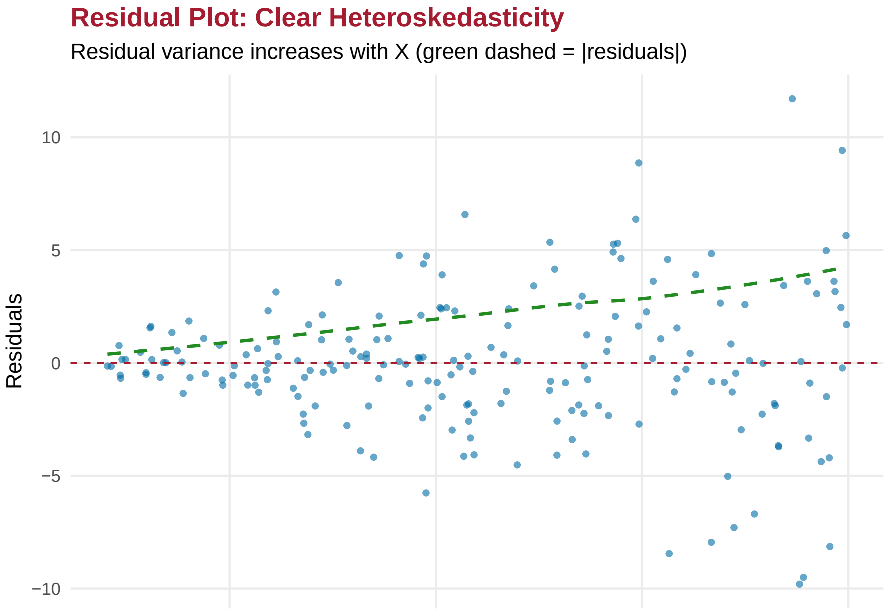
```
library(sandwich)
library(lmtest)

# Fit standard model
model <- lm(y ~ x, data = mydata)

# Get robust SEs with coeftest()
coeftest(model, vcov = vcovHC(model, type = "HC1"))

# Or extract the variance-covariance matrix
vcovHC(model, type = "HC2")
```

This approach is useful when you need the vcov matrix for other tests.

Visualizing Heteroskedasticity



Detecting Heteroskedasticity



Summary

Key points:

1. **Heteroskedasticity**: Error variance varies with X
2. Under heteroskedasticity:
   - OLS is still unbiased and consistent
   - But classical SEs are wrong
3. **Robust SEs**: Valid whether or not there's heteroskedasticity
4. **Modern advice**: Use robust SEs by default

Looking Ahead

Next lecture: Clustered Standard Errors

- When observations are correlated within groups
- Cluster-robust variance estimator
- When and at what level to cluster
- Design-based vs. sampling-based inference

Heteroskedasticity means $\mathrm{Var}(\varepsilon|X)$ isn't constant.

OLS point estimates are still fine.

But classical standard errors are wrong.

Use robust (HC) standard errors by default.