Heteroskedasticity-Robust Inference

Tyler Ransom

Univ of Oklahoma

Feb 19, 2019

Today's plan

1. Review reading topics
   1.1 The Lagrange Multiplier test
   1.2 Consequences of Heteroskedasticity
   1.3 Heteroskedasticity-Robust Inference after OLS Estimation
   1.4 Testing for Heteroskedasticity
2. In-class activity: Practice testing and correcting for heteroskedasticity

The Lagrange Multiplier (LM) test

The LM test

1. Aside from the F test, you may come across the LM test
2. Slightly different way to test joint hypotheses
3. The LM test statistic has a χ^2 distribution with $df = q$
4. Otherwise, it's pretty much the same as F test
5. LM test a.k.a. "n R-squared test"'
6. See today's in-class lab for further details

Consequences of heteroskedasticity for OLS

Review: Gauss-Markov Assumptions

1. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$
2. random sampling from the population
3. no perfect collinearity in the sample
4. $E(u|\mathbf{x}) = E(u) = 0$ (exogenous explanatory variables)
5. $Var(u|\mathbf{x}) = Var(u) = \sigma^2$ (homoskedasticity)

Properties of OLS

- Under these five assumptions, OLS has lots of nice properties
- OLS is BLUE and asymptotically efficient
- If we add normality (CLM), the tests are exact for any sample size
- Without normality, usual OLS test are asymptotically justified
- But what if we act as if we know nothing about

$$Var(u|\mathbf{x})?$$

What happens when $Var(u|\mathbf{x}) \neq Var(u)$?

- OLS is **still unbiased and consistent** under A(1)-(4)
- But it's no longer BLUE
- Usual standard errors are no longer valid
- t stats, F stats, and CIs cannot be trusted
- Need to adjust the SEs to make them valid
- Continue to use OLS; but do **heteroskedasticity-robust inference**

Heteroskedasticity-robust inference

Correcting SEs for Heteroskedasticity

- SEs, test statistics can be modified to be valid
- Can conduct hypoth. tests without worrying A(5)'s validity
- Most regression packages include an option to compute **heteroskedasticity-robust standard errors**
- These then produce **heteroskedasticity-robust** t **statistics**
- and **heteroskedasticity-robust confidence intervals**

How to do this in R

- Easiest way to do this in R is with `lmtest` package

```
library(lmtest)
tidy(coeftest(est, vcov=hccm))
```

- Result will be slightly different than typical `tidy(est)` output

- Typically (but not always), **robust SEs** larger than regular SEs

- Resulting *t* tests are valid

- "hccm" stands for "Heteroskedasticity Corrected Covariance Matrix"

How to do robust *F* test in R

- To do a robust *F* test, use the `car` package

```
library(car)
library(lmtest)

tidy(linearHypothesis(est, c('x1=0','x2=0'),
vcov=hccm))

# or, in piped form:
est %>%
linearHypothesis(c('x1=0','x2=0', vcov=hccm)) %>%
tidy
```

Why bother with default SEs at all?

1. Tradition (not necessarily a good answer)

2. Robust stats and CIs only have asymptotic justification …

   … even if the full set of CLM assumptions hold

- Typically, researchers report the robust standard errors

- Especially with large sample size

Example results

- Using `college` data from `wooldridge` package:

$$\begin{aligned}
\widehat{lwage} &= \underset{\substack{(.0720)\\[.0754]}}{1.6492} - \underset{\substack{(.0318)\\[.0325]}}{.2202}\ female + \underset{\substack{(.0058)\\[.0060]}}{.0521}\ exper + \underset{\substack{(.0066)\\[.0068]}}{.0762}\ coll \\
n &= 750,\ R^2 = .302,\ \overline{R}^2 = .299
\end{aligned}$$

Testing for Heteroskedasticity

Some history

- Before the discovery of heteroskedasticity-robust inference:
- Workflow was to first test for it and then,
- if it was found, abandon OLS for weighted least squares
- Nowadays, there is less of a case for even testing for heteroskedasticity.

Why test for it?

We may want to:

1. know if we need to report robust standard errors
2. know if we can improve over OLS (possible if there's heterosk.)
3. determine if variance in y about its mean changes with the values of the x's

Testing for heteroskedasticity

- In order to test for heteroskedasticity, we maintain

$$
\begin{aligned}
y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u \\
E(u|\mathbf{x}) &= 0,
\end{aligned}
$$

  which are A(1) and A(4), respectively

- also assume random sampling A(2)

- and of course rule out perfect collinearity A(3)

Testing for heteroskedasticity (cont'd)

- If $E(u|\mathbf{x}) = 0$ then

$$Var(u|\mathbf{x}) = E(u^2|\mathbf{x}).$$

- Therefore, A(5) can be written

$$E(u^2|\mathbf{x}) = \sigma^2 = E(u^2),$$

The null hypothesis

- A(5) as a testable null hypothesis is then:

$$H_0 : E(u^2 | x_1, x_2, ..., x_k) = \sigma^2 \text{ (constant)}$$

- We can formulate this as a regression equation with F test:

$$\begin{aligned} u^2 &= \delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k + v \\ E(v | x_1, ..., x_k) &= 0 \end{aligned}$$

and then test whether all slope coefficients are zero:

$$H_0 : \delta_1 = \delta_2 = ... = \delta_k = 0$$

More on the null hypothesis

- The previous equation is an odd looking regression model
- dependent variable is u^2, the *squared* error
- But it satisfies A(1)-A(4)
- Under the null $H_0 : \delta_1 = \delta_2 = ... = \delta_k = 0$, the intercept must be σ^2: $\delta_0 = \sigma^2$
- Under the null, it makes sense to assume that v is independent of the x_j
- Thus, it satisfies A(1)-A(5), so use original F test

Some complications

1. u^2 can't be normally distributed
   - In fact $u^2 \sim \chi^2$ if $u \sim N$
   - We'll have to appeal to Central Limit Theorem
2. We don't actually observe u!
   - Will need to use residuals $\hat{u}$ instead

The Breusch-Pagan (BP) test

The **Breusch-Pagan test** is the process described previously. Steps:

1. Estimate your regression by OLS
2. Saving the residuals, $\hat{u}_i$ and compute their squares $\hat{u}_i^2$
3. Regress $\hat{u}_i^2$ on all x's
4. Compute the default overall F test
5. If p-value is sufficiently small, reject H_0 : homoskedasticity

You'll get to practice this in today's lab

Performing the *BP* test in R

- The code to do the *BP* test in R is below:

```
library(lmtest)
tidy(bptest(est))
```

- Note: An alternative to the *BP* test is the White test

- You'll practice using both in the lab today