More about Heteroskedasticity

Tyler Ransom

Univ of Oklahoma

Feb 21, 2019

Today's plan

1. Review reading topics
   1.1 Heteroskedasticity in the LPM
   1.2 Weighted Least Squares
   1.3 Cluster-robust standard errors
2. In-class activity: More practice with robust inference

Heteroskedasticity in the LPM

The LPM

- From earlier slides: y is binary $\implies$ there *must* be heteroskedasticity
- Why? $Var(y|\mathbf{x}) = p\left(\mathbf{x}\right)\left[1 - p\left(\mathbf{x}\right)\right]$ if y is binary
- How to fix this?
- Easiest way: use heteroskedasticity-robust inference after OLS
- Alternative way: weighted least squares

Robust inference of LPM in R

- In R:

```
library(lmtest)
est <- lm(as.numeric(y) ~ x1 + ... + xk, data = df)
tidy(coeftest(est, vcov=hccm))
```

  and you're done

- Another way we can correct for heteroskedasticity is by weighting (up next)

Weighted Least Squares

Weighted Statistics

- Simple average:

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

- Weighted average:

$$\overline{Y}^w = \frac{1}{\sum_i w_i}\sum_i w_i y_i$$

where $w_i \geq 0$ for all i

Weighting: Example

- Example (from Wikipedia): Exam grades for two sections of a class:
  - Morning section: 20 students, $\overline{Y} = 80$
  - Afternoon section: 30 students, $\overline{Y} = 90$
- We can compute the weighted average of the two sections:

$$\begin{aligned}\overline{Y}^w &= \frac{1}{20+30}(20*80+30*90) \\ &= \frac{1600+2700}{50} \\ &= 86\end{aligned}$$

The larger class had a higher average, so $\overline{Y}^w > \overline{Y}$

Why is weighting useful?

- In many surveys, sample is not random
- Instead, it's **stratified random**
- Sample underrepr. groups more frequently than their population share
- e.g. oversample low-income groups to learn about SNAP particpation
- With oversamples, $\overline{Y}$ is not a great estimator of μ
- but $\overline{Y}^{w}$ is

Funhouse mirror

- Another example:

- Oversampling certain groups is like a funhouse mirror

- It distorts the sample distribution of Y

- Using weights is like “un-funhousing” the mirror

- The resulting statistics can then be used to learn about population

Weighted vs. Ordinarly Least Squares

- We can also "un-funhouse" regression estimates

- OLS solves the following problem:

$$\min_{\beta} \sum_i \left(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik}\right)^2$$

  which gives us estimates ($\hat{\beta}$'s)

- With weighted least squares (WLS), the problem becomes

$$\min_{\beta} \sum_i w_i \left(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik}\right)^2$$

  and gives us slightly different estimates ($\widetilde{\beta}$'s)

Using WLS to fix heteroskedasticity

- If we know the form of heteroskedasticity, we can weight to correct for it
- e.g. $Var\left(u|\mathbf{x}\right) = \sigma^2 h\left(\mathbf{x}\right)$ for some function $h\left(\cdot\right)$
- Key idea: weight by inverse of $h\left(\cdot\right)$, i.e. $\frac{1}{h(\cdot)}$
- This undoes the heteroskedasticity and gives us better standard errors
- This procedure is known as **Generalized Least Squares (GLS)**

It's not that easy, though

- Problem: We rarely know what $h(\cdot)$ function looks like
- So instead, we should estimate it
- This is known as **Feasible GLS** or **FGLS**
- Here we assume that

$$Var(u|\mathbf{x}) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k)$$

More on FGLS

- The feasible formula for $Var\left(u|\mathbf{x}\right) \implies$ use the squared residuals u^2:

$$\log\left(u^2\right) = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k + e$$

  where e is a mean-zero error term independent of the x's

- Then, we get the fitted values $\hat{g}_i$ from the above regression

- Run WLS on original data, with $w_i = \frac{1}{\exp(\hat{g}_i)}$

FGLS steps

1. Run a regression of y on all the x's
2. Create a new variable equal to $\log\left(u^2\right)$ for each observation
3. Regress $\log\left(u^2\right)$ on all the x's
4. Exponentiate the fitted values $\hat{h}_i = \exp\left(\hat{g}_i\right)$
5. Run the regression from (1) again, but this time use $\frac{1}{\hat{h}_i}$ as weights

FGLS in the Linear Probability Model

- Recall: $Var(y|\mathbf{x}) = p(\mathbf{x})[1 - p(\mathbf{x})]$
- So then we can just use $\hat{h}_i = \hat{y}_i(1 - \hat{y}_i)$
- **Problem:** $\hat{y}_i$ is sometimes < 0 or > 1, and we need $\frac{1}{\hat{h}_i} > 0$ for all i
- **Solution:** just use heteroskedasticity-robust SE's

Practical matters re: FGLS

- FGLS is a lot of work just to get efficient SE's
- WLS also gives $\widetilde{\beta}$'s $\neq \hat{\beta}$'s from OLS
- Which one to believe?
- If $\widetilde{\beta}$ is drastically different from $\hat{\beta}$ (i.e. completely different sign) ...

  ... you've probably got a misspecified model (i.e. $E\left(u|\mathbf{x}\right) \neq 0$)
- We'll cover this after the midterm

When should you run weighted regression?

- If you're interested in a statistic about the population (e.g. $\beta_0 = E\left(y|x\right)$)
- If you know the exact form of the heteroskedasticity (results in $\uparrow$ efficiency)
- If you have a misspecified model
- Details: Winship and Radbill (1994); Solon, Haider, and Wooldridge (2015)
- If you want a **causal effect** ($\hat{\beta}_j$), **don't worry about weighting**

Cluster-robust SEs

Clusters in hierarchical data

- Sometimes, u is correlated among observations within certain groups
- This violates the Random Sample assumption
- e.g. students in a classroom in a school in a school district
- e.g. individuals within a US state
- We call this **hierarchical data**
- In this case, regular **and** heterosk-robust SE's are **too small**

Cluster-robust SEs: t-tests

- To get the correct SE's, use **clustered SEs** a.k.a. **cluster-robust SEs**
- Use the `coef_test` function of the R package `clubSandwich`:

```
coef_test(est, vcov = "CR1", cluster = df$state)
```

- Outputs the coefficient, SE, and p-values for single hypotheses

Cluster-robust SEs: *F*-tests

- For *F*-tests:

```
Wald_test(est, c("x1","x2"), vcov = "CR1",
cluster = df$state)
```

  where `x1` and `x2` are variable names in your model.

- You can have more than two hypotheses; the above is just an example

- Outputs the *F*-stat and *p*-value for the joint hypothesis test

References I

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. 2015. “What are We Weighting For?” *Journal of Human Resources* 50 (2):301–316.

Winship, Christopher and Larry Radbill. 1994. “Sampling Weights and Regression Analysis.” *Sociological Methods & Research* 23 (2):230–257.