

POLS/CS&SS 503:  
Advanced Quantitative Political Methodology

# RESIDUALS IN OLS: NON-NORMALITY, HETEROSKEDASTICITY

April 28, 2015

Jeffrey B. Arnold



# Overview

Non-Normal Errors

Heteroskedasticity

Non-Normal Errors

Heteroskedasticity

# Non-normal errors

Suppose that errors  $\epsilon$  have  $E(\epsilon) = 0$ , but not normal distribution

- $\mathbf{b}$  still unbiased
- $V(\mathbf{b})$  incorrect in small samples
- Still BLUE, but not MVUE

# How to diagnose non-normal errors?

## Graphical methods

- Plot studentized residuals against theoretical  $t$ -quantiles.
  - Studentized residuals (approximately standardized to have std dev of 1)

$$E_i^* \approx \frac{E_i}{S_E}$$

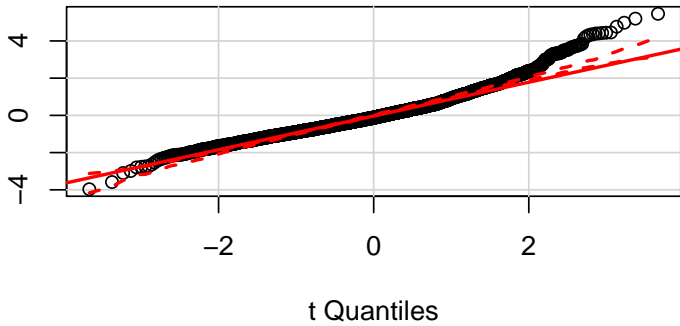
- Density plot
- Box-plot

# SLID

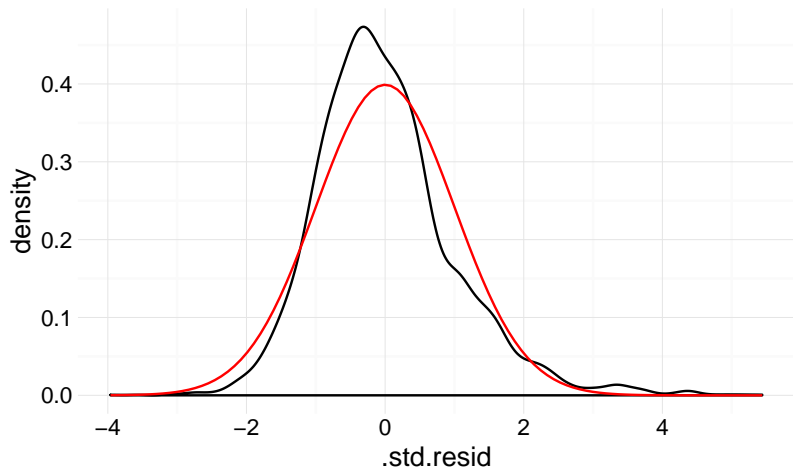
- Example in Fox
- Survey of Labour Income and Dynamics (Canada)
- 3,997 employed individuals between 16–65 residing in Ontario

```
##
## Call:
## lm(formula = wages ~ sex + age + education, data = SLID)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26.111  -4.328  -0.792   3.243  35.892
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.905243   0.607771  -13.01  <2e-16 ***
## sexMale      3.465251   0.208494   16.62  <2e-16 ***
## age          0.255101   0.008634   29.55  <2e-16 ***
## education    0.918735   0.034514   26.62  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.602 on 4010 degrees of freedom
## (3411 observations deleted due to missingness)
## Multiple R-squared:  0.2972, ^IAdjusted R-squared:  0.2967
## F-statistic: 565.3 on 3 and 4010 DF,  p-value: < 2.2e-16
```

Studentized Residuals(mod\_slid)







Non-Normal Errors

Heteroskedasticity

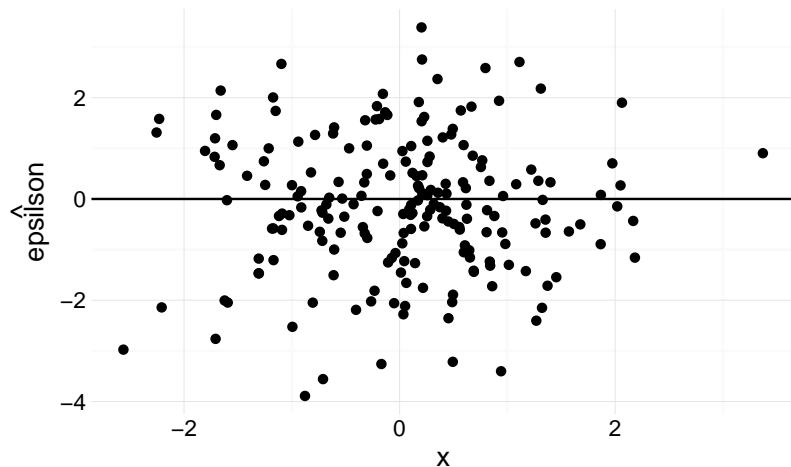
## Homoskedasticity

- equal variance
- $V(\epsilon_i) = \sigma^2$  for all obs

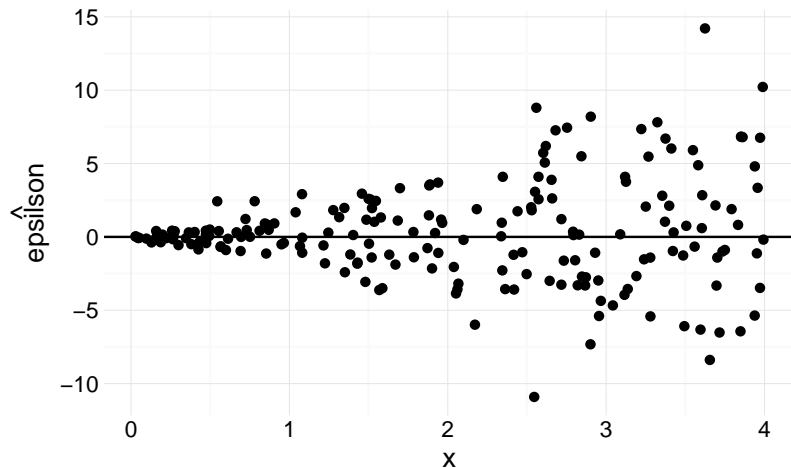
## Heteroskedasticity

- unequal variance
- Some  $V(\epsilon_i) \neq V(\epsilon_j)$
- In both cases, errors are uncorrelated  $C(\epsilon_i, \epsilon_j) = 0$  if  $i \neq j$ .

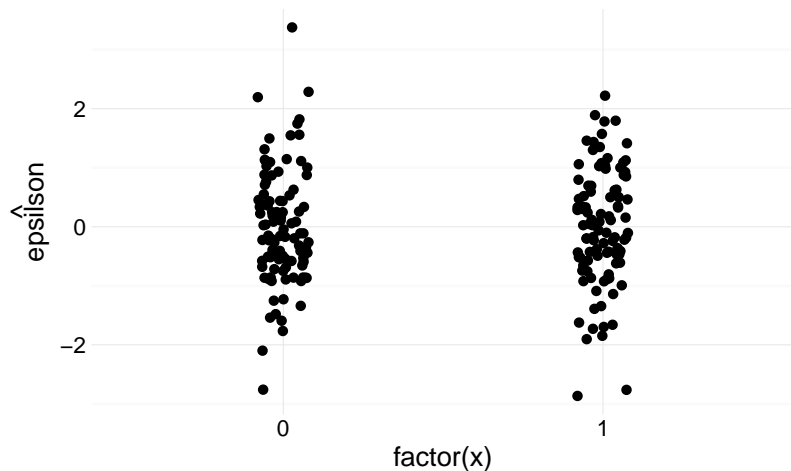
# Homoskedasticity for a continuous X



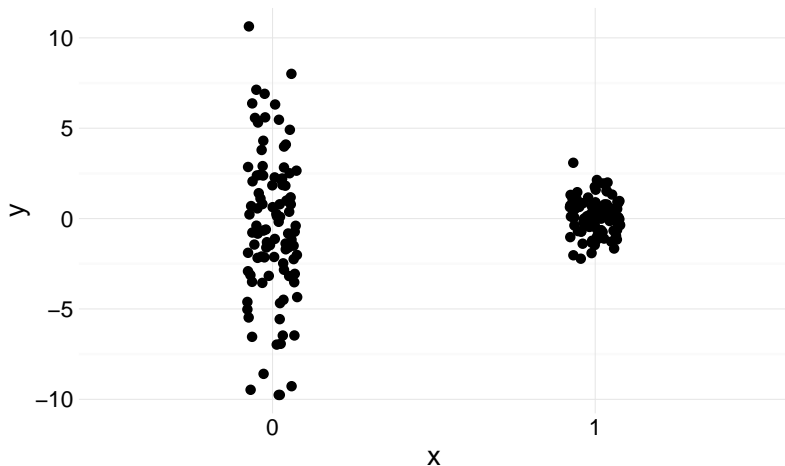
# Heteroskedasticity for a continuous $X$



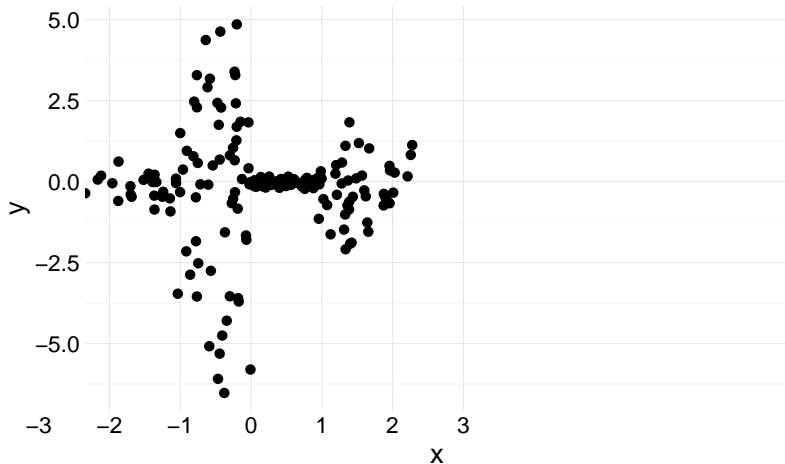
# Homoskedasticity with Binary Explanatory Variables



# Heteroskedasticity with Binary Explanatory Variables



# Unusual Heteroskedasticity





# Diagnosing heteroskedasticity?

## Diagnosing

- Plot  $E$  or  $|E|^2$  against  $\hat{Y}$  or  $X$
- R function `residualPlots`

## Tests

- All tests of the form regress residuals on functions of  $X$   
(Breusch-Pagan, White) `car::ncvTest`
- Are robust standard errors different from classic standard errors?

# What does heteroskedasticity do?

Violates some of the Gauss-Markov Assumptions.

- Point estimate is still unbiased:  $E(\mathbf{b}) = \beta$
- But, variance wrong:  $V(\mathbf{b}) \neq \sigma_\epsilon^2(\mathbf{X}'\mathbf{X})^{-1}$
- And OLS is not BLUE or MVUE

# What to do about Heteroskedasticity

## Weighted Least Squares

If you know the form of the heteroskedasticity

## Heteroskedasticity consistent standard errors

If you don't.

# Weighted Least Squares

Like OLS, but weight each observation

$$\hat{\beta}_{WLS} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$$

where  $\mathbf{W}$  is a diagonal matrix with  $\text{diag}(\mathbf{W}) = (w_1^2, w_2^2, \dots, w_n^2)$

This minimizes the weighted sum of squares

$$\hat{\beta}_{WLS} = \arg \min_{\beta} \sum w_i^2 (y_i - \mathbf{x}_i \beta)^2$$

Note  $\hat{\beta}_{WLS} \neq \hat{\beta}_{OLS}$ , but both are unbiased if form of heteroskedasticity known.

# Where do the weights in WLS come from?

Weights are such that

$$y_i \sim N(\mathbf{X}\beta, \sigma_\epsilon^2/w_i^2)$$

## Example

You have a survey and are using average values from counties. What weights should you use? What is the justification?

What if you had no idea what  $\sigma_i^2$  was?

What would you use as an estimates?

## , robust

- Use  $\mathbf{b}$  from OLS, only correct the  $V(\mathbf{b})$
- Since  $E(\epsilon_i) = 0$ ,  $V(\epsilon_i) = \sigma_i^2 = E(\epsilon_i^2)$
- So use  $E_i^2$  as estimate of  $\sigma_i^2$
- Then

$$\tilde{V}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

- Where  $\hat{\Sigma} = \text{diag}(E_1^2, \dots, E_n^2)$
- R functions `hccm` in **car**.

# Thoughts on Heteroskedasticity

- Affects standard errors, not bias
- Non-constant error variance only an issue when ratio of largest to smallest variance is  $\geq 4$  (Fox)
- If using robust standard errors, always compare them to classical standard errors
- Angrist and Pischke suggest using max of robust and classical standard errors
- Tests tell you if it is a problem, visualization needed to get ideas how to fix it
- **MOST IMPORTANT:** problems with residuals point to misspecification issues



# Residuals related to specifications of the model

Diagnostics for Simple Linear Regression