# Heteroskedasticity

BS1802 Statistics and Econometrics

*Jiahua Wu*

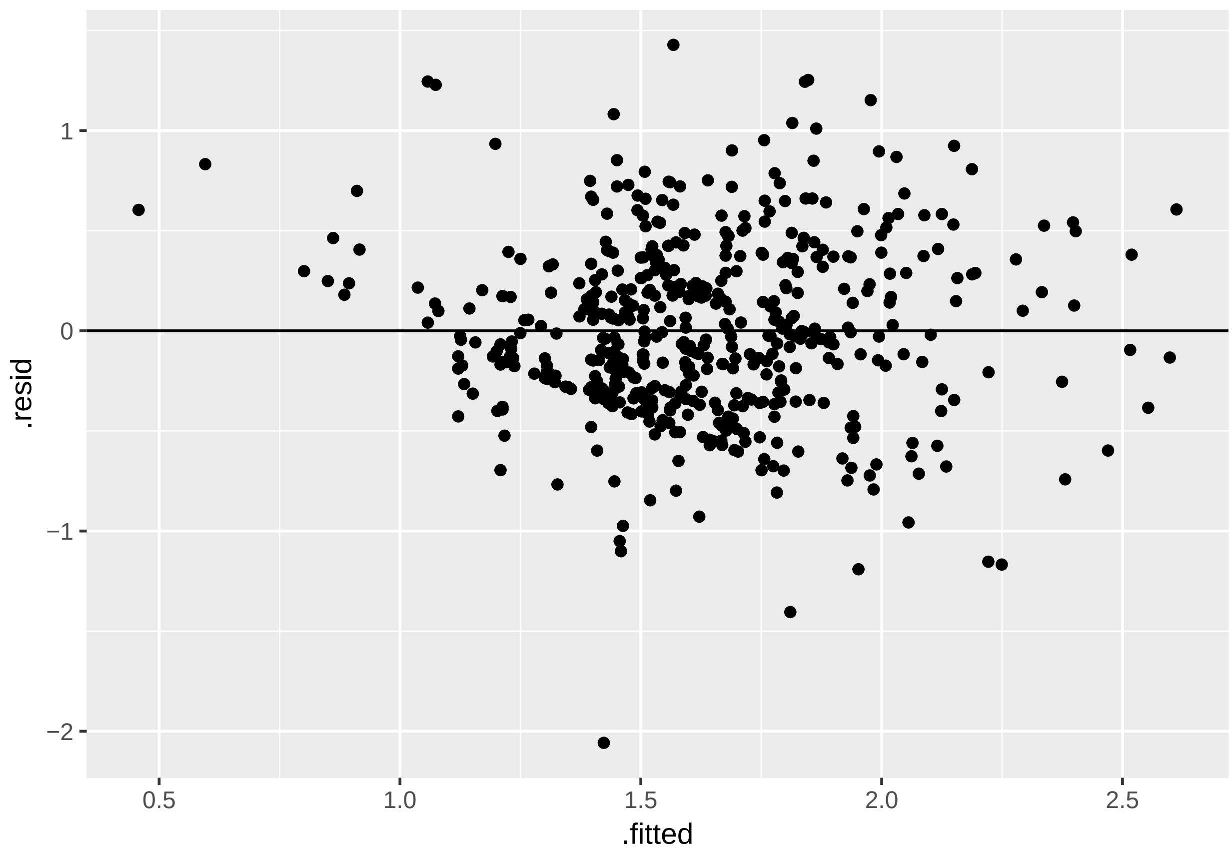
## Tests for heteroskedasticity

In the class, we discuss two formal tests for heteroskedasticity, i.e., Breusch-Pagan test and White test. We use the example on page 13 to discuss how to run the tests in R. The two tests are implemented in the *lmtest* package.

Before we run formal tests, it is proven to be quite useful to check out the residual plot, and get some intuitive idea whether the variance in the residual is constant or not.

```
load("wage1.RData")
wage.m1 <- lm(log(wage) ~ educ + exper + tenure, data)

# residual plot
ggplot(wage.m1, aes(.fitted, .resid)) + geom_point() + geom_hline(yintercept = 0)
```




We plot residuals against fitted values using *ggplot*. Judging from the plot, there exhibits some weak evidence of heteroskedasticity. The variation in residuals is smaller with either very small fitted value or very large fitted value. But also, we have fewer observations with very small fitted values or very large fitted values. So this might contribute to the smaller variances.

Next we run the two formal tests. These two tests are implemented by the function *bptest* in the *lmtest* package. By default, *bptest* regresses $\hat{u}^2$ on all the independent variables, which is based on Breusch-Pagan test. If we want to use White test, we need to specify the independent variables to be the fitted value and squared of the fitted value in the second step. The implementation of *bptest* is based on an alternative test, rather than F test as discussed in the slides. So we only care about the $p$-value in the output. A smaller $p$-value indicates stronger evidence of heteroskedasticity.

```
library(lmtest)
# BP test
bptest(wage.m1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  wage.m1
## BP = 10.761, df = 3, p-value = 0.01309
```

```
# White test
fitted.wage <- wage.m1$fitted.values
bptest(wage.m1, ~ fitted.wage + I(fitted.wage^2))
```

```
##
##  studentized Breusch-Pagan test
##
## data:  wage.m1
## BP = 6.6953, df = 2, p-value = 0.03517
```

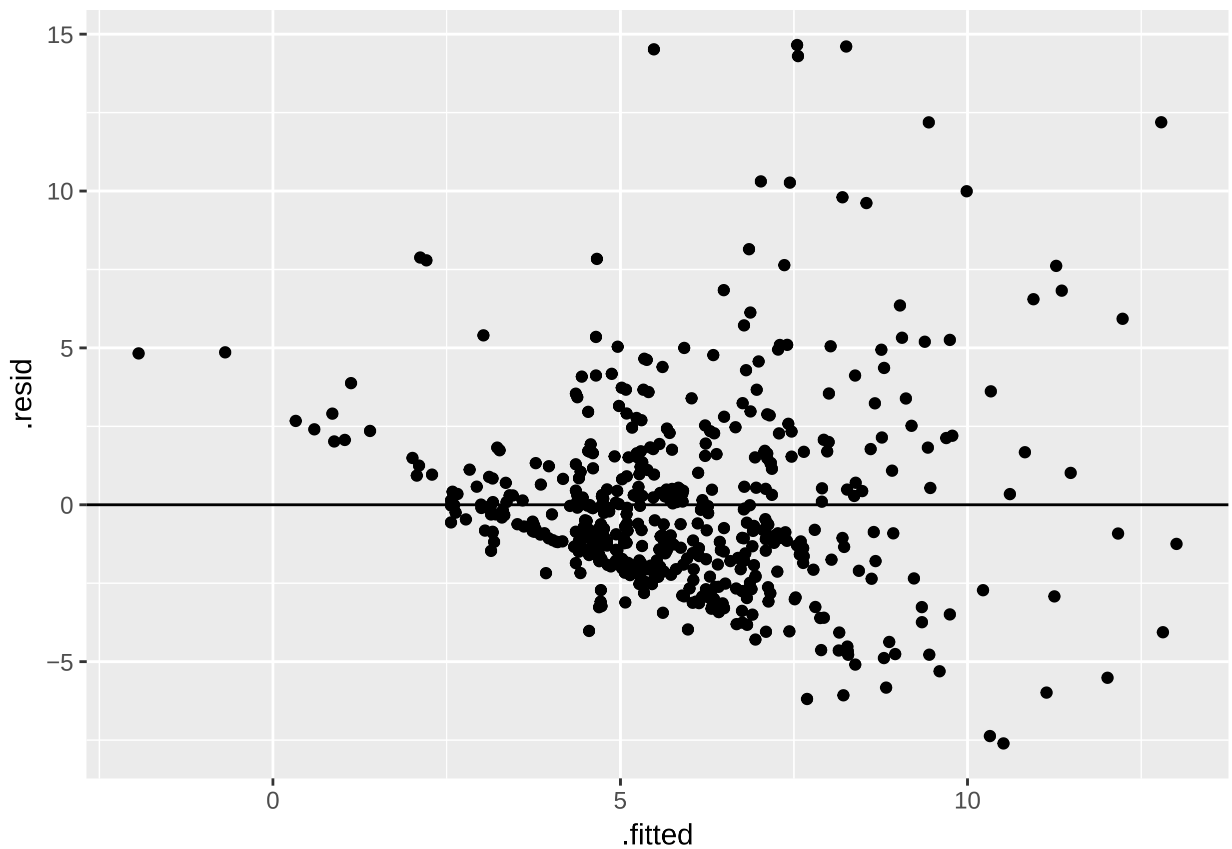
We can code manually for Breusch-Pagan test using the code below. We can see that $p$-values from *bptest* and manual coding are consistent.

```
# Running BP test manually
res.sq <- wage.m1$residuals^2
aux <- lm(res.sq ~ educ + exper + tenure, data)
aux.rsq <- summary(aux)$r.squared
aux.F <- (aux.rsq/3) / ((1 - aux.rsq) / wage.m1$df.residual)
1 - pf(aux.F, 3, wage.m1$df.residual)
```

```
## [1] 0.01285541
```

Now, let us consider an alternative model, where we use *wage* as the dependent variable. We can see that both residual plots and BP tests suggest stronger evidence of heteroskedasticity. In general, log transformation can mitigate the issue of heteroskedasticity.

```
wage.m2 <- lm(wage ~ educ + exper + tenure, data)
ggplot(wage.m2, aes(.fitted, .resid)) + geom_point() + geom_hline(yintercept = 0)
```




```r
bptest(wage.m2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  wage.m2
## BP = 43.096, df = 3, p-value = 2.349e-09
```

### Heteroskedasticity-robust inference

We can derive robust standard errors using *sandwich* package in R. There are multiple ways to adjust standard errors to make it robust against heteroskedasticity. For details, please check out "sandwich.pdf" posted on the Hub (btw, it is not examinable). The code for robust t test and robust F test are shown below.

```r
library(sandwich)

# calculate robust variance and covariance matrix
vcov.robust <- vcovHC(wage.m1, "HC1")

# t test
coeftest(wage.m1, vcov = vcov.robust)
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value  Pr(>|t|)
```

```
## (Intercept) 0.2843596  0.1117069  2.5456   0.01120 *
## educ        0.0920290  0.0079212 11.6181 < 2.2e-16 ***
## exper       0.0041211  0.0017459  2.3605   0.01862 *
## tenure      0.0220672  0.0037820  5.8348 9.461e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# F test
linearHypothesis(wage.m1, "educ - exper = 0", white.adjust = "hc1")
```

```
## Linear hypothesis test
##
## Hypothesis:
## educ - exper = 0
##
## Model 1: restricted model
## Model 2: log(wage) ~ educ + exper + tenure
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df      F    Pr(>F)
## 1    523
## 2    522  1 143.46 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```