## Testing for Heteroskedasticity

Homoskedasticity is a strong assumption and is rarely realistic for data sets in this course.

- This isn't a problem for applied work because we have techniques that are robust to heteroskedasticity.
- Our formula for the variance of OLS coefficients is wrong under heteroskedasticity, but heteroskedasticity-robust standard errors can be used instead.
    - Other names: Huber-White standard errors, Eicker-White, or Eicker-Huber-White
    - These researchers developed an expression for variance estimates that are unbiased without homoskedasticity.

## Testing for Heteroskedasticity (cont.)

- We can work around heteroskedasticity, but it's worth discussing how to test for it.
- Usually the best thing to do is to look at the diagnostic plots.
- R provides two plots that can help identify heteroskedasticity.

**Note: If there are more data points on one side of the plot, it may make it look like there's more variation than is actually occurring.**

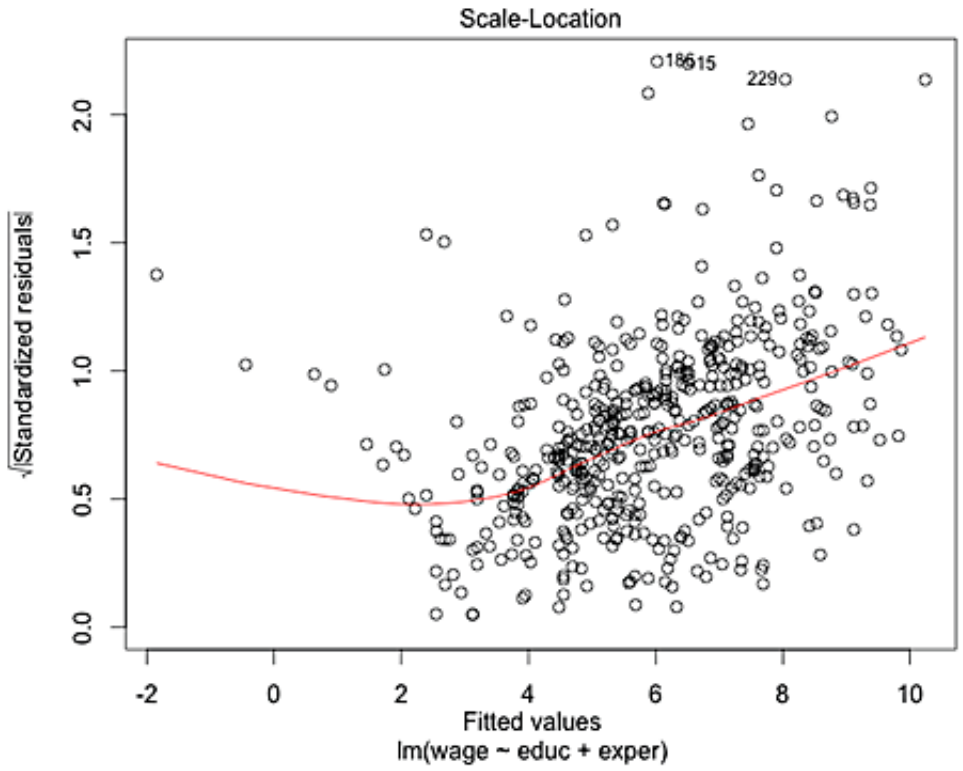
## Testing Homoskedasticity: Residuals vs. Fitted-Values Plot

- Example: Fitted wage model, using data from 1976.
  - $\text{wage} = -3.39 + 0.644\text{educ} + 0.070\text{exper} + u$
- If errors were homoskedastic, the band would be of uniform thickness.
- Remember, the red line is a smoother that approximates the mean of the residuals.
  - It appears to show a violation of zero-conditional mean.
  - But the spline does not help with heteroskedasticity because that is about the deviations from the mean.

## Scale-Location Plot

R provides something called a scale-location plot.

- This is related to the residual vs. fitted-value plot, but the residuals are transformed in two ways:
  - First, we calculate the absolute value.
    - More variance appears as points higher on the *y*-axis.
    - At this point, there is likely to be a lot of skew, making it tough to inspect visually.
  - Next, we calculate the square root.
    - This reduces skew, moving points away from the *x*-axis.
  - After these transformations, homoskedastic errors should appear as a horizontal band of points.



## Breusch-Pagan Test

- The Breusch-Pagan test is a statistical test that can be used to check for heteroskedasticity.
- The null hypothesis is that there is homoskedasticity.
- If there is a significant result, that is evidence supporting heteroskedasticity.
- Warning: As with most tests of this nature, sample size matters greatly.
    - For a few hundred data points or more, almost any amount of heteroskedasticity will appear as a significant test.
    - For small data sets, the test will rarely be significant.

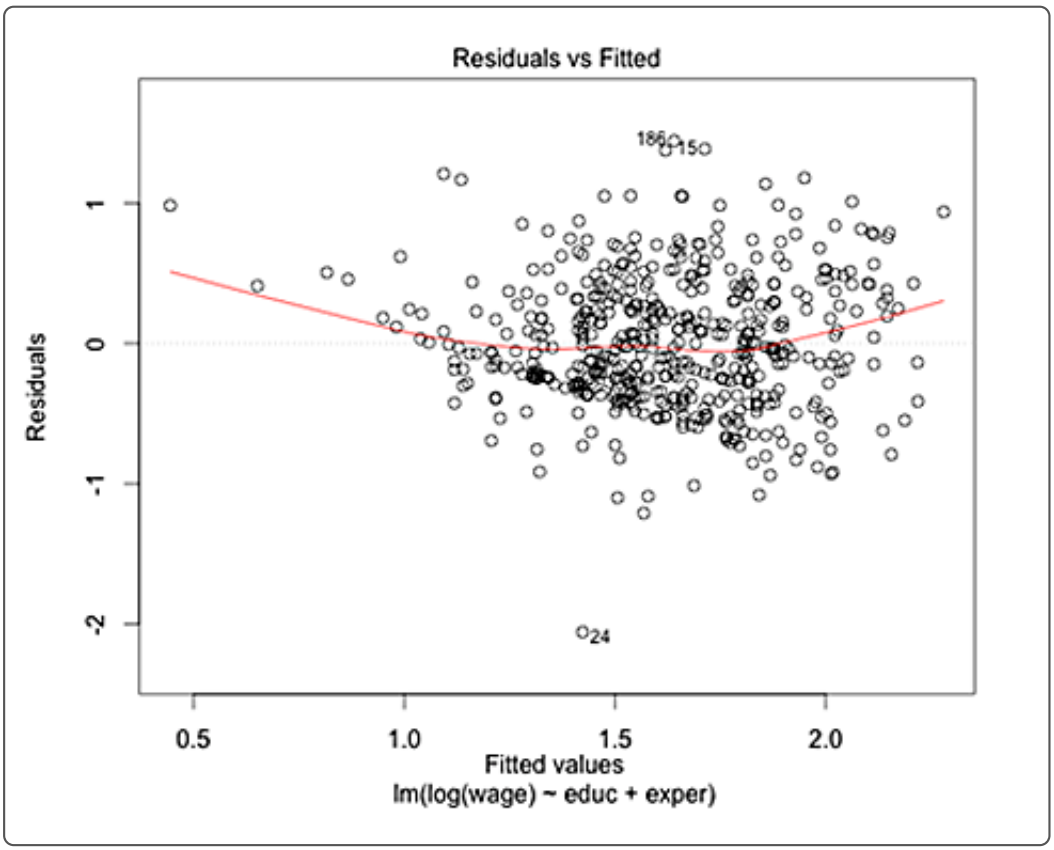
**With the Breusch-Pagan test, it's best to keep sample size in mind and to use it in conjunction with diagnostic plots.**

## Responding to Heteroskedasticity

- The simplest solution is to switch to heteroskedasticity-robust tools.
- In particular, White standard errors are robust to heteroskedasticity.
    - Many researchers recommend using these by default.
    - If there is homoskedasticity, White standard errors will be larger (more conservative) than regular standard errors, so it is slightly in your favor to use regular errors.

## Responding to Heteroskedasticity (cont.)

- At times, heteroskedasticity accompanies a violation of zero-conditional mean.
    - There may be an exponential relationship in the data, especially if $\text{var}(y)$ seems correlated with $E(y)$.
- This seemed to happen for the previous wage regression.
    - There are a lot of outliers in the positive direction, and there is also more variance on this side.
- More commonly, researchers will model the log of wage.




## Plots With Natural Log Wage

- Using the same model as previously, replacing wage with the natural log of wage.
  - $\widehat{\text{Log(wage)}} = 0.21 + 0.09\text{educ} + 0.010\text{exper}$
  - Notice that the residual vs. fitted-value plot looks more even, and the scale-location plot appears flatter.

## Log(wage) and Intuition

- Warning: In this case, it makes intuitive sense to take the log(wage).
  - Everyone has a sense for what a 10-percent increase in wage means.
  - A $100 increase could be very different for different people.
- Do not take the log of a variable just to fix heteroskedasticity if it doesn't make theoretical sense.
  - We want our specification to be guided by theory and the tests we wish to perform, not by what choices result in homoskedasticity.
  - That's why the best practice is simply to use heteroskedasticity-robust standard errors.