POLS 509: The Linear Model, *Lecture # 9*

Eric Reinhardt

Mar 28, 2005

Heteroscedasticity.

Mar 23: Heteroscedasticity. The problem & diagnosis. GLS and robust SEs.

§ Fox, 301-309, 320-321, 326-328.

§ Wooldridge, 248-278.

§ Optional: Greene (4th ed.), 499-524.

§ Optional: George W. Downs and David M. Rocke, "Interpreting Heteroscedasticity," American Journal of Political Science 23:4 (November 1979), 816-828.

1 Outline

1) The concept of heteroscedasticity. Sources of heteroscedasticity.

2) Consequences for OLS estimates.

3) Detection and diagnosis. Graphical methods. Goldfeld-Quandt Test. Breusch-Pagan Test. White Test.

4) What to do about heteroscedasticity. GLS/WLS/FGLS. Robust SEs. Heteroscedastic Regression (MLE).

2 Heteroscedasticity

[Greene 455-6] Consider a variant of the linear regression model, such that

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U},$$

expressed in matrix terms, so that $\mathbf{Y}$ has order $n \times 1$, $\mathbf{X}$, $n \times k$, $\boldsymbol{\beta}$, $k \times 1$, and $\mathbf{U}$, naturally, $n \times 1$. As per the CLRM, we assume

$$E[\mathbf{U}|\mathbf{X}] = \mathbf{0},$$

but also, generalizing from the CLRM formulation,

$$E[\mathbf{U}\mathbf{U}'|\mathbf{X}] = \sigma^2\boldsymbol{\Omega} \text{ instead of just } E[\mathbf{U}\mathbf{U}'|\mathbf{X}] = \sigma^2\mathbf{I},$$

where $\boldsymbol{\Omega}$ is a symmetric (and hence square) matrix of $n \times n$ order.

Of course, $\boldsymbol{\Omega} = \mathbf{I}$ will satisfy the classical regression assumption as a special case. More generally, however,

$$\sigma^2\boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}.$$

If $\sigma_i^2 \neq \sigma_j^2$ for any two observations i, j, then we have nonconstant error variance, that is, *heteroscedasticity.*

Note right away: heteroscedasticity can come in an infinite variety of forms.

Autocorrelation, as we'll see next week, is another special case of $\boldsymbol{\Omega}$, for example,

$$\sigma^2\boldsymbol{\Omega} = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ & & \ddots & \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix}.$$

Here, the correlation between the errors of any two observations k and $k - j$ is ρ_j.

3 Consequences

What effect does a generalized heteroscedastic $\boldsymbol{\Omega}$ have on the OLS regression estimates?

3.1 Still Unbiased

Assume a heteroscedastic $\mathbf{\Omega}$ as above. The OLS estimator $\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$, or, if we substitute $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}$ in for the $\mathbf{Y}$ here and re-arrange, then $\widehat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{U})$. All we need for $E\left[\widehat{\beta}\right] = \beta$ is thus to assume, as per the CLRM, that $E[\mathbf{U}|\mathbf{X}] = \mathbf{0}$, which zeroes out the rightmost part of the last expression. The $\mathbf{\Omega}$, however, is about the error variances, not the error means, so it does not invalidate this necessary assumption. Hence, despite heteroscedasticity, the OLS estimates are unbiased.

3.2 But Standard Errors Are Wrong

The Variance-Covariance matrix of the beta estimates for the regular OLS model is

$$VarCov\left(\widehat{\beta}\right) = (\mathbf{X}'\mathbf{X})^{-1}\left(\mathbf{X}'\sigma^2\mathbf{\Omega}\mathbf{X}\right)(\mathbf{X}'\mathbf{X})^{-1},$$

as we showed before. In the case of homoscedasticity, $\mathbf{\Omega} = \mathbf{I}$, which can then be pulled out as a scalar, allowing the $(\mathbf{X}'\mathbf{X})^{-1}$ to cancel out the internal $(\mathbf{X}'\mathbf{X})$, which leaves the standard OLS $VarCov\left(\widehat{\beta}\right) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. However, with real heteroscedasticity, that simplification does not hold.

NOTE: The amount of bias is

$$VarCov\left(\widehat{\boldsymbol{\beta}}\right) - VarCov(\boldsymbol{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$$

This quantity, as Greene p. 219 notes, hinges on

$$\mathbf{\Delta} = \frac{1}{n}\sum_{i=1}^{n}(1 - w_i)\,\mathbf{x}_i\mathbf{x}_i^{'},$$

where w is parts of the function expressed as some generic weight. The key point here is that the amount of bias in the variance-covariance matrix (and hence the estimated SEs) is related to $\mathbf{x}_i\mathbf{x}_i^{'}$. If the form of heteroscedasticity here is some function of the $\mathbf{X}$ variables, then the bias is larger. If not, then in sufficiently large samples, the estimated variance-covariance matrix will not be biased, even though it is still inefficient.

Furthermore, it is important to note that, while heteroscedasticity does not affect R^2 or adjusted R^2, it does mean that the standard regression F

statistic no longer follows the F distribution, just as it means the t statistic doesn't follow the t distribution (because the SE in its denominator is biased).

4 Diagnostics

4.1 Graph the Residuals

After the regression, plot the fitted residuals $(Y - \widehat{Y})$ against the regression predictions $(\widehat{Y})$, e.g., with Stata's `rvfplot` command. Inspect visually for wider vertical spreads of dots above one range of $\widehat{Y}$ values than above another range of $\widehat{Y}$ values.

Also, plot the residuals against individual independent variables (`predict r, res; scatter r` x) suspected to be a source of heteroscedasticity.

Advantages: (a) gives a good impression of the extent and nature of heteroscedasticity as a function of plotted variables; (b) can indicate heteroscedasticity of particular functional forms.

Disadvantages: (a) does not allow hypothesis testing about the presence of heteroscedasticity, doesn't yield a "yes-no" answer; (b) you must stipulate each form of heteroscedasticity to be examined — does not provide any generic answers regarding all forms of heteroscedasticity.

4.2 Goldfeld-Quandt Test

(Greene p. 223)

Start by stipulating a division of your sample into two groups. You make the division in the way that you suspect will yield the starkest possible evidence of heteroscedasticity, with one group having lower error variance than the other, say, by ranking observations by their values of a particular variable X_j, and splitting those with the highest values from those with the lowest. Now you have two subsamples with n_1 and n_2 observations, respectively.

Your null hypothesis is homoscedasticity, equal error variances in the two groups. If true, then residual sums of squares of the two groups should be equal, and the ratio of the two, weighted by degrees of freedom, can serve as a test statistic following an F distribution:

$$F\left(n_1-k,n_2-k\right)=\frac{\widehat{\mathbf{U}}_1'\widehat{\mathbf{U}}_1/\left(n_1-k\right)}{\widehat{\mathbf{U}}_2'\widehat{\mathbf{U}}_2/\left(n_2-k\right)}.$$

Make sure to put the group with the higher error variance on top of this expression, so that the resulting F statistic will exceed 1 if unequal.

If there really is heteroscedasticity, as believed, in the grouping variable X_j, then we can show that more starkly if we exclude the middle values of X_j and only compare two groups taken from the extremes. Cutting the sample into thirds is a common approach here, but the place(s) to divide the sample is really arbitrary. The point is that, if we are right, then this technique may expose a higher F ratio and thus make our test of heteroscedasticity more likely to reject the null if the null is really false — which gives our test greater **power**. However, the tradeoff here is to reduce the size of n_1 and n_2, which *reduces* the power of our test. Hence the lack of any real convention about whether to exclude any observations from the test.

Advantages: (a) valid for small samples, not just for asymptotically large ones; (b) more powerful than other hypothesis tests (e.g., the White test) of heteroscedasticity; (c) gives a clear sense of directions for "treatment," since form of heteroscedasticity is so clearly identified.

Disadvantages: (a) requires that you stipulate *ex ante* the form and source of heteroscedasticity to be tested for; (b) doesn't address all generic forms of heteroscedasticity, only one at a time.

4.3 Breusch-Pagan Test

(Greene p. 223-5; Wooldridge p. 266-8; other variants described there)

Assume that the error variances follow a particular hypothesized function

$$\sigma_i^2=\sigma^2\times f\left(\alpha_0+\boldsymbol{\alpha}'\mathbf{z}_i\right),$$

where $\mathbf{z}_i$ is a vector of some set of m independent variables presumed to be sources of heteroscedasticity. If $\boldsymbol{\alpha}=\mathbf{0}$, then we have homoscedasticity. So let's test that!

1. Estimate your original regression, and compute the fitted residuals $\widehat{u}_i$. (predict u, res)

2. Square those residuals, creating $\widehat{u}_i^2$. (gen u2=u^2)

3. Divide those squared residuals by $\widehat{\mathbf{U}}'\widehat{\mathbf{U}}/N$ (one estimate of σ^2), yielding $\frac{\widehat{u}_i^2}{SSR/N}$.

4. Now run an auxiliary regression of $\frac{\widehat{u}_i^2}{SSR/N}$ on $\mathbf{z}_i$, including a constant.

5. Multiply the model or explained sum of squares from this auxiliary regression by one-half.

6. This is the **Breusch-Pagan Test Statistic**, which is a kind of Lagrange Multiplier statistic, asymptotically following a χ^2 distribution with m degrees of freedom, which you use to reject the null of homoscedasticity as appropriate.

The intuition: if our selected subset of variables explains a sufficient amount of the variance in the estimated errors, then we have heteroscedasticity. (Type `bpagan xvar1 xvar2` in Stata, naming the *xvar*s desired, after running your original OLS regression.)

Advantages: (a) can address a variety of sources of heteroscedasticity at once, albeit only those named in the auxiliary regression; (b) hence can bear on broader, if not full infinite, variety of generic heteroscedasticity; (c).gives a useful picture of the form of heteroscedasticity and hence points to potential corrections.

Disadvantages: (a) valid only in sufficiently large samples, unlike the GQ Test; (b) less powerful than GQ, although more powerful than the White test below.

4.4 White Test

(Greene p. 222-3; Wooldridge p. 268-70)

Assume homoscedasticity. Then the error variance should be uncorrelated with all the explanatory variables, the squares of all the explanatory variables, and the all cross-products of pairs of explanatory variables. These are the elements in the matrix $\mathbf{X}'\mathbf{\Omega}\mathbf{X}$, which is the kernel of $VarCov(\widehat{\boldsymbol{\beta}})$ whose properties are the subject of the violation, if you recall.

1. Estimate your original OLS regression, and compute the squared fitted residuals, $\widehat{u}_i^2$.

2. Regress $\widehat{u}_i^2$ on all of your original explanatory variables $\mathbf{X}$, plus the squares of the variables in $\mathbf{X}$ (not counting the constant), plus the cross-products of pairs of all variables in $\mathbf{X}$ (not counting the constant). This is a regression with $l = k + (k-1) + \frac{(k-1)!}{2(k-3)!}$ variables on the RHS. E.g., for an original regression with $k = 4$ (three substantive independent variables plus the constant), this yields $4 + 3 + (3 \times 2)/(2 \times 1) = 10$ variables in the auxiliary regression.

3. For the LM version of the statistic, we multiply $N \times R^2_{aux}$, which is asymptotically distributed according to a chi-square distribution with $l - 1$ degrees of freedom. If this yields $p < 0.05$, we reject the null of homoscedasticity.

4. In Stata, use `whitetst` following your original `regress` command.

Advantages: the only test for all generic forms of heteroscedasticity.

Disadvantages: (a) valid only in sufficiently large samples; (b) not "constructive," since it leaves the form of heteroscedasticity unknown; (c) not "specific," i.e., high "false positive" rate, since it may pick up model misspecification and other errors besides heteroscedasticity; (d) less powerful than the other tests, since it uses so many extra degrees of freedom.

5 Corrections

5.1 Weighted Least Squares

Procedures:

1. Stipulate the nature of the heteroscedasticity, where the variances are inversely proportional to a weight w_i.

2. `regress y x1 x2...xk [aw=w]`

Advantages: like OLS, BLUE.

Disadvantages: you must stipulate the exact weights, i.e., the form of heteroscedasticity to correct for; we don't often know this in practice.

5.2 Feasible Generalized Least Squares

Procedures:

1. Obtain a consistent estimate of the variance of errors, by regressing fitted $\widehat{u}_i^2$ on the independent variables in $\mathbf{X}$, and generating the predicted error variance for each observation, as the squared residuals $w = \widehat{\widehat{u}}_i^2$ as fitted from this auxiliary regression.

2. Estimate the original regression as per WLS, weighting in Stata by $1/\widehat{\widehat{u}}_i^2$: `regress y x1 x2...xk [aw=1/w]`

Advantages: does NOT require that you stipulate the precise form of heteroscedasticity, unlike WLS.

Disadvantages: not small-sample unbiased, nor efficient in small samples; is only consistent and efficient asymptotically, in sufficiently large samples, because of it relies on only an estimate of the error variance.

5.3 Heteroscedasticity-Consistent Robust Standard Errors

The robust SE for an individual variable x_j's coefficient estimate $\widehat{\beta}_j$ is

$$Var(\widehat{\beta}_j) = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2 \widehat{u}_i^2}{\left[\sum_{i=1}^{n} (x_i - \overline{x})^2\right]^2},$$

so obviously the $SE(\widehat{\beta}_j) = \sqrt{\text{that}}$.

This is based on an estimator for the population σ_i^2 which is valid only as $n \to \infty$, so it is "just" consistent, not valid in small samples.

Advantages: (a) like White's test, does *not* require that you know the form of heteroscedasticity; corrects for all forms; (b) the correction of choice in current applied Political Science studies, for better or worse.

Disadvantages: (a) consistent – asymptotically unbiased – but **potentially biased in insufficiently large samples**, hence bringing its derived ts and Fs into question in such circumstances; (b) **less efficient** than OLS SEs if errors are indeed homoscedastic, inflating SEs to even an analytically costly degree in certain circumstances.