

# Heteroskedasticity

---

ECONOMETRICS (ECON 360)

BEN VAN KAMMEN, PHD

# Introduction

---

For pedagogical reasons, OLS is presented initially under strong simplifying assumptions.

- One of these is homoskedastic errors, i.e., with constant variance,  $\sigma^2$ .
- Along with the other Gauss-Markov assumptions, this gives OLS the desirable properties of unbiasedness, consistency, and has minimal variance among the class of linear estimators, i.e., it is BLUE.
- If the homoskedasticity assumption fails (with the other assumptions remaining intact), however, the last property ceases to apply, and OLS is no longer the most efficient estimator.

This “heteroskedasticity,” or variance that depends on  $x$ , is the norm in empirical economics—not the exception.

So this chapter describes what can be done to detect heteroskedasticity and what can be done about it once detected.

# Outline

---

Consequences of Heteroskedasticity for OLS.

Heteroskedasticity-Robust Inference after OLS Estimation.

Testing for Heteroskedasticity.

Weighted Least Squares Estimation.

# Consequences of heteroskedasticity for OLS

---

OLS estimators do not need the homoskedasticity assumption to be unbiased and consistent.

It is required to have the standard errors that justify inference using  $t$  and  $F$  statistics, though.

- These tests are not valid under heteroskedasticity, i.e.,

when  $Var(u|x_1, \dots, x_k) = \sigma^2$  is violated.

Lastly OLS loses its efficiency properties under heteroskedasticity.

- It is possible to find more efficient (less variance) estimators than OLS, if one knows the form of the heteroskedasticity.

# Heteroskedasticity-robust inference after OLS estimation

---

Large samples are wonderful: they can justify OLS estimation, even with heteroskedasticity, as long as the bias in the standard errors is corrected.

- The efficiency loss is tolerable with large samples because the standard errors will be small enough to make valid inferences anyway.

So then the problem is reduced to one of correcting the bias in the variance of the OLS estimators.

Fortunately some very smart people (Halbert White, Friedhelm Eicker, and Peter J. Huber) have devised a consistent estimator for the variance of the OLS estimator under heteroskedasticity.

What is more remarkable: this estimator is valid even if one doesn't know the form of the heteroskedasticity.

# Heteroskedasticity-robust inference (continued)

---

It is deceptively easy to have software compute heteroskedasticity-robust standard errors.

In STATA it just requires an option at the end of a “reg” command. Typing “, **vce(robust)**” after your regressors will run the regression and calculate robust standard errors.

**reg yvar {xvar1 xvar2 . . . xvark}, vce(robust)** is the whole syntax.

# Heteroskedasticity-robust inference (continued)

---

As we have demonstrated before, the simple OLS estimator,  $\hat{\beta}_1$ , is expressed:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

The estimate of its variance, then, is the square of the second term.

$$\text{Var}(\hat{\beta}_1) = E(\hat{\beta}_1 - \beta_1)^2 = E\left(\frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma_i^2}{SST_x^2},$$

where the last equality holds if the errors are at least independent (for the cross products to cancel out).

# Heteroskedasticity-robust inference (continued)

---

This is the same expression for variance from Chapter 2, with one notable exception: the subscript “i” on the variance.

- This means the variance will not simplify by canceling an “ $SST_x$ ”s because you can’t pull  $\sigma_i^2$  through the sum operator like when it was a constant!

Professor White’s contribution\* was showing that the sample analog of the variance:

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2 \hat{u}_i^2}{SST_x^2}$$

consistently estimates the variance of  $\hat{\beta}_1$  as expressed above.

\*White, Halbert. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48: 817-838.

# Heteroskedasticity-robust inference (continued)

---

Retracing White's footsteps is definitely beyond the scope and level of technicality for this class.

- The interested student may read White's paper or a graduate-level text on the subject.

The simple OLS case generalizes to multiple regression, in which the estimators are expressed in terms of their residual variance (not explained by other regressors).

The “partialling-out” expression for  $\hat{\beta}_j$  is:

$$\hat{\beta}_j = \beta_j + \frac{\sum_{i=1}^n \hat{r}_{ij} u_i}{\sum_{i=1}^n \hat{r}_{ij}^2},$$

where  $\hat{r}_{i1}$  is the residual from regressing  $x_j$  on all the other regressors.

# Heteroskedasticity-robust inference (concluded)

---

The expression for heteroskedasticity-robust variance, then, is:

$$\text{Var}(\hat{\beta}_j) = \frac{\sum_{i=1}^n \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}, \text{ where } SSR_j = \sum_{i=1}^n \hat{r}_{ij}^2.$$

The square root is the heteroskedasticity-robust standard error.

- Sometimes just “robust standard error” for short.
- The degrees of freedom are corrected for by multiplying the variance by  $(\frac{n}{n-k-1})$  before taking the square root.

With the robust standard error in hand, it is possible to conduct inference using the familiar methods of  $t$  and  $F$  tests, but it is worth reminding you that the tests are only valid asymptotically, i.e., in large samples.

# Testing for heteroskedasticity

---

Even though heteroskedasticity is commonplace in empirical work, we would like to know when Assumption MLR.5 *does* hold because in those instances inference can be conducted with an *exact*  $t$  distribution, instead of just an asymptotic one.

- Also if there is heteroskedasticity, one may prefer a different estimator with improved efficiency, so detecting it is necessary for making an informed decision.

Detecting heteroskedasticity amounts to testing whether the hypothesis,

$$E(u^2 | x_1, \dots, x_k) = E(u^2) = \sigma^2,$$

(that the error variance is independent of the regressors) is rejected or not.

# Testing for heteroskedasticity (continued)

---

Keeping with the use of regression analysis to tell when two variables are related, regressing an estimate of the variance on the regressors suggests itself as a test:

$$\widehat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k + v.$$

If the magnitude of the errors is independent of the regressors, then it should be the case that:

$$\delta_0 = \delta_1 = \delta_2 = \dots = \delta_k = 0; \text{ call this } H_0.$$

This can be readily tested with an  $F$  statistic (test of the overall model's explanatory power).

- In the R Squared form, this is:

$$F = \frac{R_u^2}{(1 - R_u^2)} * \frac{n - k - 1}{k}.$$

# Testing for heteroskedasticity (concluded)

---

There are more exotic tests that can be performed to detect heteroskedasticity,

- i.e., White's test that adds interactions and quadratics to the procedure outlined above.

We will not examine all of the variations in detail; this is left as an exercise for the interested student.

Suffice it to say that there are regression-based tests for heteroskedasticity that are intuitive and easy to implement (see computer exercise C4, for example).

# Weighted least squares estimation

---

If the particular form of heteroskedasticity can be discerned,

- i.e., functionally how variance depends on  $x$ ,

the regression model can be made homoskedastic with the application of weights.

E.g., if the variance of the error, conditional on  $x$  (all regressors), is simply a constant multiplied by a known function (“ $h$ ”) of  $x$ , a new error can be written by dividing the heteroskedastic error ( $u$ ) by  $h$ .

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + u_i; \text{Var}(u|\mathbf{x}) = \sigma^2 h(\mathbf{x}), \text{ so}$$

$$\text{Var}\left(\frac{u}{h(\mathbf{x})} \mid \mathbf{x}\right) = \sigma^2, \text{ s. d.}\left(\frac{u}{h(\mathbf{x})} \mid \mathbf{x}\right) = \sigma, \text{ and } E\left(\frac{u}{h(\mathbf{x})} \mid \mathbf{x}\right) = 0.$$

# Weighted least squares (continued)

---

To make the regression homoskedastic, just divide through by  $\sqrt{h_i} = \sqrt{h(\mathbf{x}|i)}$ .

$$\frac{y_i}{\sqrt{h_i}} = \frac{\beta_0}{\sqrt{h_i}} + \beta_1 \frac{x_{i1}}{\sqrt{h_i}} + \dots + \beta_k \frac{x_{ik}}{\sqrt{h_i}} + \frac{u_i}{\sqrt{h_i}}.$$

If the original model satisfied all the Gauss-Markov assumptions except for homoskedasticity, the new transformed model satisfies all of them, along with their desirable properties (BLUE).

Estimating the parameters in the preceding line by OLS produces what are called the generalized least squares (GLS) estimators.

# Weighted least squares (continued)

---

So where does the “weighting” come in? When you estimate these parameters by OLS, you’re minimizing

$$\sum_{i=1}^n \frac{\hat{u}_i^2}{h_i},$$

the sum of squared residuals, weighted by  $1/h_i$ .

Since  $h$  is what increases the variance of  $u$ , the weights assigned are the inverses of how “noisy” each observation is.

- This is how weighted least squares improves on the efficiency of regular OLS, which simply weights all observations equally.

The GLS estimates will differ from regular OLS, but the interpretation of the coefficients still comes from the original model.

# The heteroskedasticity function must be estimated: Feasible GLS

---

The crucial step in performing weighted least squares is figuring out what the function,  $h$  is.

When the weighting function is modeled and estimated from data, the resulting estimated weights ( $\hat{h}_i$ ) are the bases for feasible GLS (FGLS).

In a spirit similar to testing for heteroskedasticity, one could estimate the regression

$$\ln(\hat{u}^2) = \text{constant} + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k + e, \text{ by OLS, where}$$

$\hat{u}^2$  are the (y on x) OLS residuals squared.

The fitted values from the regression of (logs of) squared residuals equal the logs of the weights, under the assumptions that

$$h(\mathbf{x}) = \exp(\delta_0 + \delta_1 x_1 + \dots + \delta_k x_k); \text{Var}(u|\mathbf{x}) = \sigma^2 h(\mathbf{x}).$$

# Feasible GLS (continued)

---

So,

$$\hat{h}_i = \exp(\widehat{constant} + \hat{\delta}_1 x_{i1} + \dots + \hat{\delta}_k x_{ik}),$$

are the weights for squared residuals in an FGLS procedure (you just have to invert them). So FGLS minimizes

$$\sum_{i=1}^n \frac{\hat{u}_i^2}{\hat{h}_i}.$$

Since the weights have to be estimated from the data, FGLS is no longer either unbiased or BLUE.

- But it is asymptotically more efficient than regular OLS, and it is consistent.

# FGLS “cookbook”

---

1. Regress  $y$  on all  $k$  regressors in the model.
2. Obtain the residuals,  $\hat{u}$ .
3. Square the residuals and take the log of each of the squares.
4. Regress the logs of the squares on all  $k$  regressors.
  - See p.288 for an alternative method that uses the fitted values of  $y$  instead.
5. Obtain the fitted values.
6. Exponentiate the fitted values:  $e$  raised to the fitted value.
7. Estimate the original model using the inverse of the  $\hat{h}_i$  as weights.

# FGLS (concluded)

---

If the FGLS estimates differ substantially, e.g., a statistically significant coefficient changes sign, from OLS, it is probably because the functional form of the model is mis-specified or a relevant variable has been omitted.

- This would cause OLS and FGLS estimators to have different probability limits.

Furthermore the specification of the heteroskedasticity as,  $Var(u|\mathbf{x}) = \sigma^2 h(\mathbf{x})$ , could be wrong because the model of  $h(\mathbf{x})$  is mis-specified.

FGLS is still unbiased and consistent if this is the case (assuming the underlying model satisfies the first 4 Gauss-Markov assumptions).

- The standard errors computed under a mis-specified  $h(\mathbf{x})$  will not yield valid inferences, though.
- Once again, though, robust standard errors come to the rescue if weighted least squares fails to solve the heteroskedasticity problem.

Even if a WLS procedure mis-specifies the weighting function, though, it is an improvement (in terms of smaller robust standard errors) over ignoring heteroskedasticity and estimating by OLS.

# Prediction and prediction intervals with heteroskedasticity

---

The construction of prediction intervals follows a procedure analogous to what we are familiar with.

- There are some technical modifications to the standard errors used in constructing the intervals, but I prefer not to devote time to scrutinizing them here.
- There are topics of more general interest to which to proceed, so they are left to the interested student.

# Conclusion

---

Heteroskedasticity is the rule—not the exception—in empirical economics.

Robust inference can be performed by having software compute robust (“White”) standard errors.

Heteroskedasticity can be detected by one of several regression-based tests.

If the form of heteroskedasticity is known, a weighted least squares estimator has better efficiency properties than OLS.