

### 8.4.2 Unknown Heteroskedasticity: Feasible GLS

What if we do not know the exact form of the heteroskedasticity? Instead of using heteroskedasticity robust standard errors, one could use a version of the GLS estimator called the Feasible GLS.

Estimate conditional variance of the error term from data:

$$Var(u|\mathbf{x}) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k)$$

We could have used a linear model of conditional variance but they cause predicted variance to be negative.

How do we estimate this model?

$$u^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k) v$$

We transform this model by taking log to estimate the following model:

$$\log(u^2) = \alpha_0 + \delta_1 x_1 + \cdots + \delta_k x_k + e$$

We can now use the fitted values from the above regression  $\hat{h}_i = \exp(\hat{g}_i)$  to obtain weights  $\frac{1}{\hat{h}_i}$  on the WLS regression.

Having to estimate  $h_i$  using the same data means that the FGLS estimator is no longer unbiased (so it cannot be BLUE, either). Nevertheless, the FGLS estimator is consistent and asymptotically more efficient than OLS. This means that if you have a large sample, the FGLS estimator would do a better job than OLS in precision of estimates.

F-stat after WLS: The only caution we must take is that the weights should be identical in both the restricted and unrestricted regression.

Difference between OLS and FGLS: they tend to give different estimates, but we would not be worried unless the results are overturned. Its best to run both models to get a sense of the magnitude and direction of the estimates.

- ▶ If they differ, then our model is not clearly identified i.e. the other gauss markov assumptions are not met. Always check whether  $E(y|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$  i.e. the MLR.4 is satisfied or not.
- ▶ There could be functional form mis-specification too.

### 8.4.3 What if we assume wrong Heteroskedasticity Function

The key issue here is whether the misspecification of  $h(x)$  causes bias or inconsistency in the WLS estimator.

Answer: It doesn't if MLR.4 assumption is valid.

The logic is that if  $E(u|\mathbf{x}) = 0$ , then any functional transformation of the error is also zero and weighting is a functional transformation.

Implications of incorrect heteroskedasticity function:

1. Standard errors are incorrect even for large samples. *Solution*: obtain Heteroskedasticity consistent standard errors for the WLS too!
2. WLS is inefficient in comparison to OLS. *However, practically* - it is often better to use a wrong form of heteroskedasticity and apply WLS than to ignore heteroskedasticity altogether in estimation and use OLS.

### 8.4.4 Prediction Intervals with Heteroskedasticity

1. Prediction intervals get affected by heteroskedasticity but not the predictions themselves. In order to obtain the prediction, we extend the methods explored earlier by estimating the following regression under WLS

$$y_i = \theta_0 + \beta_1(x_{i1} - x_1^0) + \cdots + \beta_k(x_{ik} - x_k^0) + u_i$$

From this we get  $\hat{y}^0 = \hat{\theta}_0$  which is the in-sample prediction with its standard error as  $se(\hat{\theta}_0)$

2. For out of sample prediction, we need  $sd(u^0)$ . Under our assumption of variance  $Var(u^0|\mathbf{x} = \mathbf{x}^0) = \sigma^2 h(\mathbf{x}_0)$  we can obtain a standard error by replacing the population variance with its estimated sample counterpart.
3. *FGLS and prediction intervals*: IF we have to estimate variance function as in FGLS, then we cannot obtain exact intervals. The next best option is to use  $\hat{h}(\mathbf{x}_0)$

4. JW proceeds with explaining how to obtain prediction intervals for a log model of the following form:

$$\log(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

where  $u|\mathbf{x} \sim N[0, \exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k)]$

- (a) The CEF (Conditional expectation function) is

$$E(y|\mathbf{x}) = \exp(\beta_0 + \mathbf{x}\boldsymbol{\beta} + \exp(\delta_0 + \mathbf{x}\boldsymbol{\delta})/2)$$

which is estimated using WLS with estimated variance as in FGLS.

- (b) The estimated variance under FGLS is  $\hat{\sigma}^2 \exp(\hat{g}_i) = \hat{\sigma}^2 \exp(\hat{h}_i)$  and the fitted value is

$$\hat{y}_i = \exp(\widehat{\log(y)}_i + \hat{\sigma}^2 \exp(\hat{g}_i))$$

Prediction intervals can be obtained by using the strategy adopted previously for level models with adjustments for log dependent variables.

## 8.5 Linear Probability Model: Revisited

1. **LPM model with heteroskedasticity:** simplest solution is to compute heteroskedasticity robust standard errors. Recall: this is when we don't know the form of heteroskedasticity.

(a) Recall the response probability:

$$p(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

(b) For each observation  $i$ ,  $Var(y_i|x_i)$  is estimated by OLS fitted values as follows:

$$Var(y_i|x_i) = \hat{y}_i(1 - \hat{y}_i) = \hat{h}_i$$

However, when the fitted values are negative or greater than 1 the variance would be negative which does not help. In such instances, just abandon WLS and use heteroskedasticity robust standard error.