Weighted Least Squares (WLS) Estimation Given Heteroscedasticity

Econometric Methods, ECON 370

We have learned that our OLS estimator remains unbiased in the face of heteroskedasticity. To address the problem the variance of the parameters are no longer B.L.U.E, we know that all we need to do is to use robust standard errors to ensure our inferences are not affected.

We can however also adopt another method of estimation in the face of heteroskedasticity call the Weighted Least Squares Method, which yields greater efficiency then OLS, if the form of the variance is correctly specified.

1 When Heteroskedasticity is Known up to a Multiplicative Constant

Let $\mathbf{x} = x_1, x_2, ..., x_k$, and

$$Var(\epsilon|\mathbf{x}) = \sigma^2 h(\mathbf{x})$$

where $h(\mathbf{x}) > 0$, is some function of the independent variables that determines the particular heteroskedasticity. And assume that $h(.)$ is known. Then for each observation in the population we can express the variance of the parameter in the regression equation as,

$$\sigma_i^2 = Var(\epsilon|\mathbf{x}_i) = \sigma^2 h(\mathbf{x}_i) = \sigma^2 h_i$$

Where h_i varies with each observation since $\mathbf{x}_i$ varies with each observation. Given this structure, we could then transform the OLS regression equation so that the error term would become homoskedastic and satisfy the Gauss-Markov assumptions, or the OLS assumptions. How can we do that? Will it really work?

Consider the transformation where we divide the error term by $\sqrt{h_i}$. That is the new error becomes $\frac{\epsilon_i}{h_i}$. What happens to variance with this transformation?

$$Var(\frac{\epsilon_i}{\sqrt{h_i}}|\mathbf{x}_i) = E(\frac{\epsilon_i^2}{h_i}|\mathbf{x}_i) = \frac{\sigma^2 h_i}{h_i} = \sigma^2$$

Viola! The new error term is homoskedastic. This then suggests that we could divide the entire original OLS regression by $\sqrt{h_i}$, so that we get

$$\frac{y_i}{\sqrt{h_i}} = \frac{\beta_0}{\sqrt{h_i}} + \frac{\beta_1}{\sqrt{h_i}}x_{1,i} + \frac{\beta_2}{\sqrt{h_i}}x_{2,i} + ... + \frac{\beta_k}{\sqrt{h_i}}x_{k,i} + \frac{\epsilon_i}{\sqrt{h_i}}$$

or more simply as

$$y_i^* = \beta_0 x_{0,i}^* + \beta_1 x_{1,i}^* + \beta_2 x_{2,i}^* + ... + \beta_k x_{k,i}^* + \epsilon_i^*$$

Although this new regression equation looks peculiar, and seem to complicate interpretation, you have to keep in mind that this transformation is performed so as to achieve efficiency (obtain the correct variance). At the end of the day, when we interpret the results, we are concerned with the form of the original, untransformed regression, and not the weighted least squares regression equation. Lastly, note that the WLS equation is still linear in the parameters, and that it satisfies all of the Gauss-Markov Assumptions of the OLS as long as besides heteroskedasticity, all other assumptions are not violated.

Note that the two estimators, that from the original OLS, and that from the transformed regression equation which is a Generalized Least Squares (GLS) Estimator. However, what we actually need to do after transformation is to perform the same OLS procedure. Although in the interpretation part of the analysis, we are using the linear OLS, you have to keep in mind that the estimators, standard errors, t statistics, F statistic are all from the transformed equation. As a final note, although the goodness of fit measure is useful for calculation of the F statistic, it is not informative to us here since as noted before, we are only concerned with the original regression relationship.

To see why all we need to do is to perform the OLS procedure is to recall that when finding the estimators, we are in effect finding the estimator that minimizes the square of the error terms.

$$min \sum (\epsilon_i^*)^2 = min \sum \left(y_i^* - d_0 x_{0,i}^* - d_1 x_{1,i}^* - d_2 x_{2,i}^* ... - d_k x_{k,i}^* \right)^2$$

where d_j is just the sample version of the population parameters δ_j for $j \in \{1, 2, ..., k\}$. The GLS estimators that minimizes the above problem are called weighted least squares (WLS) estimators. The name is derived from the fact that the estimators are found by minimizing the weighted, $\frac{1}{h_i}$, sum of the the square of the residuals/errors. The idea behind this method is to give less weight to observations with the greatest variance (note that the weight is the inverse of h_i). You should be able to see now that OLS is a special case of the WLS since it gives all observations the same weight. You should also realize why we said earlier that the estimators are not the same. To see the reason why the OLS estimator is not the same as the WLS estimator, first note that the OLS estimator in the case of the simple regression is,

$$\widehat{\beta}_1 = \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2}$$

whereas the estimator from the WLS is,

$$\tilde{\beta}_1 = \frac{\sum (x_i^* - \overline{x^*}) y_i}{\sum (x_i^* - \overline{x^*})^2} = \frac{\sum (\frac{x_i}{h_i} - \overline{x^*}) \frac{y_i}{h_i}}{\sum (\frac{x_i}{h_i} - \overline{x^*})^2}$$

Consequently, we cannot use the OLS estimates in making inference. This however is not the efficient GLS procedure which weights $\widehat{\epsilon}^2$ by the conditional variance for each observation, that is $Var(\epsilon_i | \mathbf{x}_i)$. The command to do this in STATA is vwls. Note lastly that the WLS estimator can be defined for any positive weight.

The above discussion predicates on us knowing the relationship between the variance of the errors and the independent variable. First of all, this need not be so. If we don't, but use an arbitrary weight, just like the OLS estimator, the WLS estimator remains unbiased and consistent. However, just like the OLS estimator, the variances are no longer correct. And just as in the OLS case, we can always use a robust variance estimator. This has led to some criticism that if we had the robust variances for OLS, what is the need in performing a WLS. Although correct, we must realize that in the face of strong heteroskedasticity, it is always better to account for it than not, even though the manner in which we account for it may be incorrect.

Read your text, pages 289-290 on an example of when the weights needed for WLS occurs naturally

2 When the Heteroskedasticity Function must be Estimated: Feasible GLS

As mentioned earlier, usually the exact form heteroskedasticity is not known, and we have to estimate $h(\mathbf{x}_i$ using the data first, which would then give $\widehat{h(\mathbf{x}_i}$, and the use of the latter yields the **Feasible GLS (FGLS) Estimator**.

A particularly flexible manner in which we can account for the unknown functional form of heteroskedasticity is to apply

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

Of course the use of the exponential function is not unique, and depends on your imagination. Why do we use the exponential function in the first place? Principally, recall that we need h_i to be strictly positive, and the fact of the matter is that the predicted values under linear functional form does not guarantee us that, whereas exponential function does recalling the range of that functional form.

The next question is how do we estimate it? Realize that if we perform a log transformation, we would get in return a log-levels model noting first that the estimator of the variance of error is the expected value of the square of the errors. That is we get,

$$\epsilon^2 = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + \nu$$

which can be easily estimated using OLS, and where of course e has a mean of 0, and is independent of $\mathbf{x}$ by assumption. Where do we get ϵ. Well, simple, perform a OLS on the original (very first model that started this whole fracas) first to get it. Upon estimation of the above, we next have to produce $\widehat{h}_i$ which is just,

$$\widehat{h}_i = exp(\widehat{g}_i) = exp(\widehat{\alpha}_0 + \widehat{\delta}_1 x_1 + \widehat{\delta}_2 x_2 + ... + \widehat{\delta}_k x_k)$$

We will now summarize the entire procedure,

1. Run OLS for $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \epsilon$.
2. Create $log(\widehat{\epsilon}^2$ by squaring $\widehat{\epsilon}$ obtained from step 1.
3. Run OLS on $\epsilon^2 = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + \nu$ and obtain $\widehat{h}_i$.
4. Estimate $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \epsilon$ again, but this time using the weight $\frac{1}{\widehat{h}_i}$ and WLS.

Although the FGLS procedure is no longer unbiased (since we are using $\widehat{h}_i$ instead of h_i), it is consistent and asymptotically more efficient than OLS. Also, if we suspect that the model of variance used above yields heteroskedasticity in ν, we could always use the robust variance and its consequent test statistics in the final regression.

Another alternative to the functional form suggested is that we could include the square of the independent variables just as was suggested in the Breusch-Pagan test using the predicted dependent variable, and its squared value. As a final note, you cannot test for heteroskedasticity with the WLS estimation, though you can account for it using robust variances.

When testing the usual multiple hypothesis using the F test, we must apply the WLS estimation for all the models in the sense that we have to use the same weights (Weights are from the unrestricted model. Why? Recall first what the null hypothesis is.)

In practice, as you will find out, the OLS and WLS estimators will produce different estimates (due not only to the formulation) due to sampling error. It is fine if for statistically significant parameters, the differences pertain to slight differences in magnitude. If however, the magnitudes differ by large amount, or that the sign changes by for statistically significant and not significant parameters in the OLS, or the parameters that were not significant becomes significant (or the other way around), we should be concerned as it is be a signal that there are violations of other OLS assumptions such as $E(\epsilon|\mathbf{x}) \neq 0$ (which would bring about bias and inconsistency in both the OLS and WLS estimators).