Generalized Least Squares and Heteroskedasticity

Walter Sosa-Escudero

Econ 507. Econometric Analysis. Spring 2009

February 4, 2009

The Classical Linear Model:

1. Linearity: $Y = X\beta + u$.
2. Strict exogeneity: $E(u|X) = 0$
3. No Multicollinearity: $\rho(X) = K$, w.p.1.
4. No heteroskedasticity/ serial correlation: $V(u|X) = \sigma^2 I_n$.

Gauss/Markov: $\hat{\beta} = (X'X)^{-1}X'Y$ is best linear unbiased.

Variance: $S^2(X'X)^{-1}$ is unbiased for $V(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$

Valid Inference: with the normality assumption, we use t and F tests.

Now we will focus on the consequences of relaxing $V(u|X) = \sigma^2 I_n$.

The Generalized Linear Model

Suppose all classical assumptions hold, but now

- $V(u|X) = \sigma^2 \Omega$ where Ω is any symmetric, positive definite $n \times n$ matrix

We are now allowing for heteroskedasticity (elements of the diagonal of Ω are not restricted to be all equal) and/or serial correlation (off-diagonal elements may now be $\neq 0$.), but we are not imposing any structure to Ω yet (besides symmetry and pd).
Plan

1. Explore consecuences on previous results of relaxing $V(u|X) = \sigma^2 I_n$.
2. Find optimal estimators and valid inference strategies for this case (GLS).

Consequences of relaxing $V(u|X) = \sigma^2 I_n$

- $\hat{\beta}$ is still linear and unbiased (why?) but the Gauss Markov Theorem does not hold anymore. We will show constructively that $\hat{\beta}$ is now inneficient by finding the BLUE for the generalized linear model.
- $V(\hat{\beta}|X)$ will now be $\sigma^2(X'X)^{-1}\Omega(X'X)^{-1}$ (check it).
- $V(\hat{\beta}|X)$ is no longer $\sigma^2(X'X)^{-1}$, and $S^2(X'X)^{-1}$ will be a biased estimator for $V(\hat{\beta})$.
- t tests no longer have the t distribution, F tests no longer valid too.

So, ignoring heteroskedasticity or serial correlation, that is, the use of $\hat{\beta}$ and $\hat{V}(\hat{\beta}|X) = S^2(X'X)^{-1}$, keeps estimation of β unbiased though ineficient, and invalidates all standard inference.

Generalized Least Squares

First we need a simple result: if Ω is $n \times n$ symmetric and pd, there is an $n \times n$ nonsingular matrix C such that

$$\Omega^{-1} = C'C$$

What does this mean, intuitively?

Consider now the following tranformed model

$$Y^* = X^*\beta + u^*$$

where $Y^* = CY$, $X^* = CX$ and $u^* = Cu$.

Now check:

1. $Y^* = X^*\beta + u^*$, so the transformed model is trivially linear.
2. $E(u^*|X) = CE(u^*|X) = 0$
3. $\rho(X^*) = \rho(CX) = K$, w.p.1. (CX is a rank preserving tranformation of X!).
4. $V(u^*|X) =$

$$\begin{aligned} V(Cu|X) = E(CuuC'|X) &= CE(uu'|X)C' \\ &= C\sigma^2\Omega C' \\ &= \sigma^2 C[\Omega^{-1}]^{-1}C' \\ &= \sigma^2 C\left[(C'C)^{-1}\right]^{-1}C \\ &= \sigma^2 I_n \end{aligned}$$

So...

Since the transformed model satisfies all the classical assumption, the Gauss-Markov Theorem hold *for the transformed model*, hence the best linear unbiased estimator is:

$$\hat{\beta}_{gls} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*$$

This the generalized least squares estimator

- Careful: the GLS estimator is an OLS estimator of a transformed 'isomorphic' model (the generalized linear model).
- It provides the BLUE under heteroskedasticity/ serial correlation.
- Now it is clear that $\hat{\beta}$ is inefficient in the generalized context (why?)
- It is important to see that the statistical properties depend on the underlying structure (they are not properties of an estimator per-se).

Note that

$$\begin{aligned}\hat{\beta}_{gls} &= (X^{**}X^{*})^{-1}X^{**}Y^{*} \\ &= (X'C'CX)^{-1}X'C'CY \\ &= (X'\Omega^{-1}X)X'\Omega^{-1}Y\end{aligned}$$

- When $\Omega = I_n$, $\hat{\beta}_{gls} = \hat{\beta}$.
- The practical implementation of $\hat{\beta}_{gls}$ requires that we know Ω (though not σ^2.)
- It is easy to check that $V(\hat{\beta}_{gls}) = \sigma^2(X^{*\prime}X^{*})^{-1}$.

Feasible GLS

Suppose there is an estimate for Ω, label it $\hat{\Omega}$. Then, replacing Ω by $\hat{\Omega}$:

$$\hat{\beta}_{fgls} = (X'\hat{\Omega}^{-1}X)X'\hat{\Omega}^{-1}Y$$

This is the feasible GLS estimator.

Is it linear and unbiased? Efficient?

Heteroskedasticity

Suppose we keep the no-serial correlation assumption, but allow for heteroskedasticity, that is:

$$V(u|X) = \sigma^2 \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$$

It is easy to see that in this case

$$C = \text{diag}(\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_n^{-1})$$

Hence

$$X^* = CX$$

is a matrix with typical element X_{ik}/σ_i^{-1}, $i = 1, \ldots, n$, $k = 1, \ldots, K$. $Y* = CY$ is constructed in the same fashion.

$\hat{\beta}_{gls}$ for this particular case is called the weighted least squares estimator.

In some cases the FGLS takes a very simple form. Consider $Y = X\beta + u$ with

$$V(u_i|X) = \sigma^2 X_{ij}^2$$

Consider the transformed model

$$Y^* = X^*\beta + u^*$$

where X^* is a matrix with typical element

$$X_{ik}^* = \frac{X_{ik}}{X_{ij}}$$

u^* and Y^* are defined in a similar fashion. Note

$$u_i^* = \frac{u_i}{X_{ij}}$$

so $V(u_i^*|X) = \sigma^2$. Hence the transformed model is homoskedastic.

- In this case the FGLS is a GLS estimator since there is no need to estimate unknown parameters.
- In any case, the implementation of a FGLS or GLS requires detailed knowledge of the structure of heteroskedasticity in order to propose a homoskedastic transformed model. This is very rarely available.
- Later on we will explore another approach: keep the OLS estimator (we will loose efficiency) but replace $S^2(X'X)^{-1}$ by some other estimator that behaves correctly even under heteroskedasticity. (we need an asymptotic framework for this).

Testing for heteroscedasticity

We will explore two strategies, with no derivations and emphasizing intuitions. Later on we will prove all results (we need an asymptotic framework for this).

a) White test

H_0 : no heteroscedasticity, H_A : there is heterocedasticity of some form.

Consider a simple case with $K = 3$:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad 1, \ldots, n$$

Steps to implement the test:

1. Estimate by OLS, save squared residuals in e^2.
2. Regress e^2 on all variables, their squares and all possible non-redundant cross-products. In our case, regress e^2 on $1, X_2, X_3, X_2^2, X_3^2, X_2 X_3$, and obtain R^2 in this auxiliar regression.
3. Under H_0, $nR^2 \sim \chi^2(p)$. $p =$ number of explanatory variables in the auxiliar regression minus one.

Reject H_0 if nR^2 is too large.

Intuition: Note that under homoskedasticity

$$E(u_i^2|X) = \sigma^2$$

The auxiliar model can be seen as trying to ‘model’ the variance of the error term. If the R^2 of this auxiliar regression were high, then we could explain the behavior of the squared residuals, providing evidence that they are not constant.

Comments:

- Valid for large samples.
- Informative if we do not reject the null (no heterocedasticity).
- When it rejects the null: there is heterocedasticity. But we do not have any information regarding what causes heterocedasticity. This will cause some trouble when trying to construct a GLS estimator, for which we need to know in a very specific way what causes heterocedasticity.

b) The Breusch-Pagan/Godfrey/Koenker test

Mechanically very similar to White's test. Checks if *certain* variables cause heterocedasticity.

Consider the linear model where all classical assumptions hold, but:

$$V(u_i|X) = h(\alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \ldots + \alpha_p Z_{pi})$$

where $h(\ )$ is *any* positive function with two derivatives.

When $\alpha_2 = \ldots = \alpha_p = 0$, $V(u_i|X) = h(\alpha_1)$, a constant!!

Then, homoscedasticity means $H_0 : \alpha_2 = \alpha_3 = \ldots = \alpha_p = 0$, and $H_A : \alpha_2 \neq 0 \vee \alpha_3 \neq 0 \vee \ldots \vee \alpha_p \neq 0$.

Steps to implement the test:

1. Estimate by OLS, and save squared residuals e_i^2.
2. Regresss e_i^2 on the Z_{ik} variables, $k = 2, \ldots, p$ and get (*ESS*). The test statistic is:

$$\frac{1}{2} ESS \sim \chi^2(p-1) \sim \chi^2(p)$$

under H_0, asymptotically. We reject if it is too large.

Comments:

- Intuition is as in the White test (a model for the variance).
- By focusing on a particular group, if we reject the null we have a better idea of what causes heterocedasticity.
- Accepting the null does not mean there isn't heterocedasticity (why?).
- Also a large sample test.
- Koenker (1980) has proposed to use nR_A^2 as a test, which is still valid if errors are non-normal.