Heteroskedasticity-Consistent Covariance Matrices (HACCM)

1 Introduction

All the testing procedures we have used thus far use, implicitly, if not explicitly, standard errors or estimated covariance matrices. If we are to make reliable inferences about the value of parameters, these estimates should be reliable. Previously, we assumed that under the OLS assumptions, that the error terms of the regression were IID. This assumption is needed to show that $s^2(\mathbf{X}^\top \mathbf{X})^{-1}$, the usual estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}$, is consistent in the sense that as $n \to \infty$, the probability limit of the estimator for the covariance matrix tends to a limiting covariance matrix. However, even without the IID assumption, it is possible to obtain a consistent estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}$.

Let us focus our attention to:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},\ \mathbb{E}[\mathbf{u}] = 0,\ \mathbb{E}[\mathbf{u}\mathbf{u}^\top] = \boldsymbol{\Omega}, \tag{1}$$

where $\boldsymbol{\Omega}$, the error covariance matrix, is an $n \times n$ matrix with t^{th} diagonal element equal to ω_t^2 and all the off-diagonal elements equal to 0:

$$\boldsymbol{\Omega} = \begin{bmatrix} \omega_1^2 & 0 & \cdots & 0 \\ 0 & \omega_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \omega_n^2 \end{bmatrix}.$$

Since X is assumed to be exogenous, the expectations in (1) can be treated as conditional on X. Conditional on X, then, implies the error terms in (1) are uncorrelated and have mean 0, but they do not have the same variance for all observations. These error terms are said to be heteroskedastic, or to exhibit heteroskedasticity. If the error terms all have the same variance, then they are said to be homoskedastic, or to exhibit homoskedasticity.

The assumption that X is exogenous is fairly strong, but it is often reasonable for cross sectional data. We make it largely for simplicity, since we would obtain essentially the same asymptotic results if we replaced it with the weaker assumption that X is predetermined: $\mathbb{E}(u_t|X_t) = 0$. When the data are generated by a DGP that belongs to (1) with $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, the exogeneity assumption implies that $\hat{\boldsymbol{\beta}}$ is unbiased.

Whatever the form of the error covariance matrix $\mathbf{\Omega}$, the covariance matrix of the OLS estimator $\hat{\boldsymbol{\beta}}$ can be derived as follows:

$$\begin{aligned}
\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\
\implies (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{u} \\
\mathbb{E}\left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top\right] &= (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbb{E}[\mathbf{u}\mathbf{u}^\top|\mathbf{X}]\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \\
&= (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{\Omega}\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}. \qquad (2)
\end{aligned}$$

This form of covariance matrix is often called a sandwich covariance matrix, for the obvious reason that $\mathbf{X}^\top\mathbf{\Omega}\mathbf{X}$ is sandwiched between the two instances of the matrix $(\mathbf{X}^\top\mathbf{X})^{-1}$. Covariance matrices of an inefficient estimator very often takes this sandwich form.

We can see why the OLS estimator is inefficient when there is heteroskedasticity by noting that observations with low variance convey more information about the parameters than observations with high variance.

Assume we don't know the diagonal elements of $\mathbf{\Omega}$, in fact it's likely that we don't know them. Furthermore there are n of them, so we cannot hope to estimate the ω_t^2 terms without making further assumptions. The situation seems hopeless. However, even though we cannot evaluate the sandwich covariance matrix, we can estimate it without having to attempt the impossible task of estimating $\mathbf{\Omega}$ consistently.

2 Asymptotic theory

For the purposes of asymptotic theory, we wish to consider the covariance matrix, not of $\hat{\boldsymbol{\beta}}$, but rather of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. This is simply the probability limit of n times the matrix (2). We then take the limit of each of the factors, and we find the asymptotic covariance of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is[1]:

$$\left(\lim_{n\to\infty}\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1}\left(\lim_{n\to\infty}\frac{1}{n}\mathbf{X}^\top\mathbf{\Omega}\mathbf{X}\right)\left(\lim_{n\to\infty}\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1} \qquad (3)$$

[1]We could also have done this by the following

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \beta) = \left(\frac{\mathbf{X}^\top\mathbf{X}}{n}\right)^{-1}\frac{1}{n}\mathbf{X}^\top\mathbf{u} \overset{a}{\sim} N(0, \mathbf{S}_{\mathbf{X}^\top\mathbf{X}}^{-1}\mathbf{S}_{\mathbf{X}^\top\mathbf{\Omega}\mathbf{X}}\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}^{-1}),$$

which is equivalent to what we derived in (3).

We know that that the two factors to the left and the right, tend to a finite, deterministic, positive definite matrix $\mathbf{S}_{\mathbf{X}^\top \mathbf{X}}^{-1}$. To estimate the limit, we can simply use the matrix $n^{-1}\mathbf{X}^\top \mathbf{X}$ itself. The middle factor is difficult to estimate however (which is the filling in the sandwich). White (1980) showed that, under certain conditions, including the existence of the limit, this matrix can be estimated consistently by:

$$\frac{1}{n}\mathbf{X}^\top \hat{\mathbf{\Omega}}\mathbf{X}, \tag{4}$$

where $\hat{\mathbf{\Omega}}$ is an *inconsistent* estimator of $\mathbf{\Omega}$. As we will see, there are several admissible versions of $\hat{\mathbf{\Omega}}$. The simplest version, is a diagonal matrix with t^{th} diagonal element equal to $\hat{u}_t^2$, the t^{th} squared OLS residual.

The middle factor of (3) is a $k \times k$ symmetric matrix. Therefore it has exactly $\frac{1}{2}(k^2 + k)$ distinct elements. Since this number is independent of sample size, the matrix can be estimated consistently. Its ij^{th} element is:

$$\lim_{n\to\infty}(\frac{1}{n}\sum_{t-1}^{n}\omega_t^2 X_{ti}X_{tj}). \tag{5}$$

This is to be estimated by the ij^{th} element of (4), which for the simplest version of $\hat{\mathbf{\Omega}}$, is:

$$\frac{1}{n}\sum_{t=1}^{n}\hat{u}_t^2 X_{ti}X_{tj}. \tag{6}$$

Because $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$, $\hat{u}_t$ is consistent for u_t, and $\hat{u}_t^2$ is therefore consistent for u_t^2. Thus, asymptotically, (6) is equal to:

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}u_t^2 X_{ti}X_{tj} &= \frac{1}{n}\sum_{t=1}^{n}(\omega_t^2 + v_t)X_{ti}X_{tj} && (7)\\ &= \frac{1}{n}\sum_{t=1}^{n}\omega_t^2 X_{ti}X_{tj} + \frac{1}{n}\sum_{t=1}^{n}v_t X_{ti}X_{tj}, && (8)\end{aligned}$$

where v_t is defined as u_t^2 minus its mean of ω_t^2. Under suitable assumptions about the X_{ti} and the ω_t^2, we can apply a LLN to the second term in (8). Since v_t has mean 0, this term by construction converges to 0, while the first term converges to (5).

The above argument shows that (7) tends in probability to expression (5). Because the former is asymptotically equivalent to (6), that expression also tends in probability to (5). Thus, we can use the matrix (4), of which typical element is (6), to estimate $\lim(n^{-1}\mathbf{X}^\top \mathbf{\Omega}\mathbf{X})$ consistently, and the matrix

$$(n^{-1}\mathbf{X}^\top \mathbf{X})^{-1} n^{-1}\mathbf{X}^\top \hat{\mathbf{\Omega}}\mathbf{X}(n^{-1}\mathbf{X}^\top \mathbf{X})^{-1} \tag{9}$$

to estimate expression (3) consistently. We drop the factors of n^{-1} and use:

$$\widehat{\mathrm{Var}}_h(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \hat{\mathbf{\Omega}}\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \tag{10}$$

to directly estimate the covariance matrix of the OLS estimator. The sandwich estimator (10) is an example of a heteroskedasticity-consistent covariance matrix estimator (HCCME).

We could further this by saying the OLS estimator is root-n consistent and asymptotically normal, with (10) being a consistent estimator of its covariance matrix. Recall that an estimator is root-n consistent if it satisfies the following properties:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\frac{\mathbf{X}^\top \mathbf{X}}{n}\right)^{-1} \frac{1}{\sqrt{n}}\mathbf{X}^\top \mathbf{u} \overset{a}{\sim} N(0, \mathbf{S}^{-1}_{\mathbf{X}^\top \mathbf{X}} S_{\mathbf{X}^\top \mathbf{\Omega}\mathbf{X}} S^{-1}_{\mathbf{X}^\top \mathbf{X}}).$$

i.e. Since multiplying the estimation error by $\sqrt{n}$ gives rise to an expression of zero mean and finite covariance matrix, it follows that the estimation error itself tends to zero at the same rate as $n^{-\frac{1}{2}}$. This property is expressed as saying that $\hat{\boldsymbol{\beta}}$ is root-n consistent.