Generalised Least Squares

Generalised Linear Regression Model
We extend the linear regression model by allowing a more general form for the variance/covariance matrix of the error term. Specifically, we consider the (data generating) model

$$y = X\beta + u, \ \Sigma \equiv E\left[uu'|X\right] = \sigma^2 V\left(X\right) \neq \sigma^2 I_n;$$

where $V\left(X\right)$ is a symmetric and positive definite matrix and a function of X, i.e. non-spherical errors. We write V for convenience.

Example:

Under heteroskedasticity with no autocorrelation,

$$V = \begin{bmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & \vdots & h_n \end{bmatrix} \neq I_n.$$

Under autocorrelation with homoskedasticity,

$$V = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & & \ddots & \vdots \\ \rho_{n-1} & \cdots & \rho_2 & \rho_1 & 1 \end{bmatrix} \neq I_n.$$

Estimation: GLS Estimator

Consequences for Least Squares estimator: Not BLUE

Under strict exogeneity $E[u|X] = 0$, the LS estimator $b = (X'X)^{-1} X'y$ is unbiased and consistent.

The (conditional) variance/covariance matrix of the LS estimator b is

$$
\begin{aligned}
Var[b|X] &= E\left[(b - E[b|X])(b - E[b|X])' |X\right] \\
&= E\left[(b-\beta)(b-\beta)' |X\right] \\
&= E\left[\left((X'X)^{-1} X'u\right)\left((X'X)^{-1} X'u\right)' |X\right] \\
&= E\left[(X'X)^{-1} X'uu'X (X'X)^{-1} |X\right] \\
&= (X'X)^{-1} X'E[uu'|X] X (X'X)^{-1} \\
\implies Var[b|X] &= \sigma^2 (X'X)^{-1} X'VX (X'X)^{-1}.
\end{aligned}
$$

Notice that if $V = I_n \implies Var\left[b|X\right] = \sigma^2 \left(X'X\right)^{-1}$. But in general this is not the case. Therefore, using $se\left(b_k\right) = \sqrt{s^2 \left[\left(X'X\right)^{-1}\right]_{kk}}$ is not valid; the 'usual' *se* is biased and inconsistent. Similarly, using the 'usual' *se* implies that the t-statistic does not, under normality, follow t-distribution or does not converge to the normal distribution asymptotically. The same result holds for the F-statistic and Wald-statistic.

LS is Not BLUE:

Since $V \neq I_n$ violates the GM theorem premises; Hence, b it's not BLUE.

Efficient Estimation: GLS Estimator

One of the requirements of the GM theorem is that of spherical errors. So, to obtain the BLUE, we transform the model to one with spherical errors.

Fact: If V is a symmetric and positive definite matrix, then there exists a matrix $C_{n\times n}$ such that $V^{-1} = C'C$ (note: C is not unique).

Note: In fact, since V is a positive definite matrix, we have that $V = P\Lambda P'$, where $PP' = I$ and columns of P contain the characteristic vectors of V. Then, let $C' = P\Lambda^{-1/2}$, so that $C'C = P\Lambda^{-1}P' = V^{-1}$.

Perform the Linear Transformation

$$\begin{aligned} y &\to \tilde{y} = Cy, \\ X &\to \tilde{X} = CX, \\ u &\to \tilde{u} = Cu, \end{aligned}$$

so that we have the linear model

$$Cy = CX\beta + Cu \implies \tilde{y} = \tilde{X}\beta + \tilde{u}.$$

Remark: Since C is a non-singular, $\tilde{X}$ and X have the same information. Hence, $E\left[t|\tilde{X}\right] = E\left[t|X\right]$ for any t.

Remark: By construction $V^{-1} = C'C$, therefore

$$\begin{aligned} &(C')^{-1} V^{-1} = (C')^{-1} C'C \iff (C')^{-1} V^{-1} = C \\ \iff &(C')^{-1} V^{-1} C^{-1} = I_n \\ \iff &(CVC')^{-1} = I_n \\ \iff &CVC' = I_n \end{aligned}$$

Therefore, the variance-covariance matrix of $\tilde{u}$ is

$$\begin{aligned} E\left[\tilde{u}\tilde{u}'|\tilde{X}\right] &= E\left[\tilde{u}\tilde{u}'|X\right] \\ &= E\left[Cu\left(Cu\right)'|X\right] \\ &= E\left[Cuu'C'|X\right] \\ &= CE\left[uu'|X\right]C' \\ &= \sigma^2 CVC' \\ \iff\quad E\left[\tilde{u}\tilde{u}'|X\right] &= \sigma^2 I_n; \end{aligned}$$

that is, $\tilde{u}$ is spherical.
Also, strict exogeneity holds in the transformed variables as well

$$E\left[\tilde{u}|\tilde{X}\right] = E\left[\tilde{u}|X\right] = CE\left[u|X\right] = 0$$

Fact:[**GLS estimator is BLUE**] The transformed model satisfies the GM conditions. So, LS on the transformed model gives the BLUE. This estimator is called the **GLS estimator**.

Proof: GM requirements are Linearity of the model, no multicollinearity, strict exogeneity and spherical error. The transformed model satisfies all the requirements. Then, we know that in this case LS on the transformed model gives the BLUE. Hence, GLS is BLUE.

The GLS estimator (i.e. LS estimator on the transformed model) is

$$\begin{aligned} \hat{\beta}_{GLS} &= \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{y} \\ &= \left(X'C'CX\right)^{-1}X'C'Cy \\ \iff \hat{\beta}_{GLS} &= \left(X'V^{-1}X\right)^{-1}X'V^{-1}y. \end{aligned} \tag{1}$$

As we said GLS estimator is BLUE. The GLS estimator puts different weights on observations, whereas the OLS puts the same weight. This will be most apparent when we consider the case of heteroskedasticity and the Weighted Least Squares (WLS) estimator, where the weight is inversely related to the skedasticity of the error.

(Conditional) Variance of $\hat{\beta}_{GLS}$ differs from that of b[1]

$$\begin{aligned} Var\left[\hat{\beta}_{GLS}|X\right] &= \sigma^2\left(\tilde{X}'\tilde{X}\right)^{-1} \\ &= \sigma^2\left(X'C'CX\right)^{-1} \\ &= \sigma^2\left(X'V^{-1}X\right)^{-1} \end{aligned}$$

[1] Note that since V is symmetric so is V^{-1}. Also, recall that $V^{-1} = C'C$.

It can be seen that under strict exogeneity b and $\hat{\beta}_{GLS}$ are both unbiased. But, $\hat{\beta}_{GLS}$ is relatively more efficient in that it has lower variance in the matrix form. The gain in efficiency is achieved by exploiting heteroskedasticity and correlation between observations in the error term, which operationally, is to insert the inverse of a (a matrix proportional to) $Var[u|X]$ in the OLS formula [i.e., $\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1} X'V^{-1}y = \left(X'Var[u|X]^{-1}X\right)^{-1} X'Var[u|X]^{-1}y$].

Interpretation of estimate:
The interpretation is with respect tot the original model. That is, $\hat{\beta}_{GLS}$ is an estimate of β; i.e. the effect of X on the mean value of y.

Similarly, in order to estimate the GLS residuals, we use $\hat{\beta}_{GLS}$ and the original model (i.e. model of interest):
$\hat{u}_{GLS} = y - X\hat{\beta}_{GLS}$.

LS residuals are estimated in the usual way, $\hat{u} = y - Xb$.

Remark:[**Unknown V matrix**] If we have to estimate V then we are concerned with the Feasible (estimated) GLS (FGLS) estimator, which replaces V in $\hat{\beta}_{GLS}$ with $\hat{V}$.

In this case, the LS may perform better in finite samples than the feasible GLS. Also, earlier results for GLS, i.e. unbiasedness and BLUE, may not hold. To see this, suppose $\hat{V} = m(X, Y)$, then

$$E\left[\hat{\beta}_{FGLS}|X\right] = E\left[\left(X'm(X,Y)^{-1}X\right)^{-1}X'm(X,Y)^{-1}y|X\right],$$

then we cannot pass the expectation inside so unbiasedness does not hold.

Algebra of GLS: Normal equations

GLS estimator is the LS in the transformed model. The estimator minimizes a "*weighted*" *sum of squares*:

$$\min \tilde{u}'\tilde{u} = u'C'Cu = u'V^{-1}u = (y - X\beta)' V^{-1} (y - X\beta).$$

FOCs:

$$\tilde{X}'\hat{\tilde{u}} = 0, \tag{2}$$

Using the fact that $\hat{\tilde{u}} = C\left(y - X\hat{\beta}_{GLS}\right) = C\hat{u}_{GLS}$ we can re-write the

$$\begin{aligned} \tilde{X}'\hat{\tilde{u}} &= X'C'C\hat{u}_{GLS} = 0 \\ \implies X'V^{-1}\hat{u}_{GLS} &= 0 \end{aligned} \tag{3}$$

Remark: see that $X'\hat{u}_{GLS} \neq 0$

Note that if we also assume normality of the error term, $u|X \sim N(0, \sigma^2 V)$, then the ML estimator is found by maximizing the likelihood function

$$\max L = \frac{1}{(2\pi)^{N/2} |\sigma^2 V|^{1/n}} \exp\left\{ -\frac{1}{2\pi\sigma^2} (y - X\beta)' V^{-1} (y - X\beta) \right\},$$

which gives the same FOCs as above;

$$\tilde{X}' \hat{\tilde{u}} = 0 \iff X' V^{-1} \hat{u}_{GLS} = 0$$

Finite Sample Properties of GLS estimator

Below we list the finite sample properties of the GLS estimator of the linear model, $y = X\beta + u$ with $\Sigma \equiv E[uu'|X] = \sigma^2 V$.

Then, the estimator $\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1} X'V^{-1}y$ is

1) unbiased: $E\left[\hat{\beta}_{GLS}|X\right] = \beta$ (under $E[u|X] = 0$)
2) variance is $Var\left[\hat{\beta}_{GLS}|X\right] = \sigma^2 (X'V^{-1}X)^{-1}$
3) it's BLUE.

Asymptotic Properties of GLS estimator

We can re-write $\hat{\beta}_{GLS} = \beta + \left(X'V^{-1}X\right)^{-1} X'V^{-1}u$. Assume that

$$\tilde{Q} = p \lim_{n\to\infty} \left(\frac{X'V^{-1}X}{n}\right)$$

is finite positive definite matrix. Then, $\hat{\beta}_{GLS}$ is consistent,

$$\text{bias}\left(\hat{\beta}_{GLS}; \beta\right) = 0 \;\&\; Var\left[\hat{\beta}_{GLS}|X\right] = \frac{\sigma^2}{n}\left(\frac{X'V^{-1}X}{n}\right)^{-1} \to 0$$

$$\Longrightarrow \hat{\beta}_{GLS} \xrightarrow{m.s.} \beta \implies \hat{\beta}_{GLS} \xrightarrow{p} \beta.$$

Also, $\hat{\beta}_{GLS}$ is asymptotically normally distributed with the stated variance above.

LS Estimator in the GL Regression

Most of the times we don't know $\Sigma = \sigma^2 V$ and we have to assume a certain structure and to estimate - which entails the risk of misspecification. The LS estimator may outperform the feasible GLS. Most importantly, as we will see, *LS does not need the assumption of a certain structure of* V. So, it's worthwhile to consider the LS estimator in the generalised linear model

$$y = X\beta + u; \; E\left[u|X\right] = 0 \; \& \; E\left[uu'|X\right] = \sigma^2 V \equiv \Sigma.$$

We know that the OLS estimator is

$$b = \left(X'X\right)^{-1} X'y = \beta + \left(X'X\right)^{-1} X'u.$$

Properties of LS estimator

Finite sample results:

1) Unbiased: $E[b|X] = \beta$, under strict exogeneity.
2) Covariance matrix changes to:
   $Var[b|X] = (X'X)^{-1} X'\Sigma X (X'X)^{-1} \neq \sigma^2 (X'X)^{-1}$.
3) $u|X \sim N(0, \Sigma) \implies$
   $b|X \sim N\left(\beta, (X'X)^{-1} X'\Sigma X (X'X)^{-1}\right)$
4) It's not BLUE.

Asymptotic results:

1) CAN: Consistent and Asymptotically Normal.
2) But, not efficient.

Consistency

Suppose $\Sigma = \sigma^2 V$, then

$Var[b|X] = \sigma^2 (X'X)^{-1} X'VX (X'X)^{-1}$. We are going to show that b is a consistent estimator for β.

Assume that the probability limits

$$Q = p \lim_{n\to\infty} \left(\frac{X'X}{n} \right) \;\&\; S = p \lim_{n\to\infty} \left(\frac{X'VX}{n} \right)$$

are finite and positive definite matrices. Also, impose strict exogeneity. Then,

$$E[b|X] = \beta + E\left[(X'X)^{-1} X'u|X \right] = \beta + (X'X)^{-1} X'E[u|X] = \beta$$

$$\implies \text{bias}(b;\beta) = 0$$

and

$$Var[b|X] = \frac{\sigma^2}{n} \left(\frac{X'X}{n} \right)^{-1} \frac{X'VX}{n} \left(\frac{X'X}{n} \right)^{-1} \overset{a}{\sim} \frac{\sigma^2}{n} Q^{-1} S Q^{-1} \longrightarrow 0$$

We thus have

$$b \xrightarrow{m.s.} \beta \implies b \xrightarrow{p} \beta.$$

Asymptotic Normality

We know that

$$\sqrt{n}\,(b-\beta) = \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'u}{\sqrt{n}}\right) = \left(\frac{X'X}{n}\right)^{-1}\left(\frac{\sum_{i=1}^{n}\mathbf{x}_i u_i}{\sqrt{n}}\right)$$

has the same limiting distribution as

$$Q^{-1}\left(\frac{\sum_{i=1}^{n}\mathbf{x}_i u_i}{\sqrt{n}}\right).$$

Conditional Heteroskedasticity:

Assume conditional heteroskedasticity, but no autocorrelation and independent observations; i.e.
$\sigma_i^2 \equiv E\left[u_i^2|X\right] = \sigma^2 v_i\left(X\right) = \{indep\} = \sigma^2 v_i\left(\mathbf{x}_i\right) \equiv \sigma^2 v_{ii}$ and $E\left[u_i u_j|X\right] = 0$ for $i \neq j$. So, we have that

$$\begin{aligned} Var\left[\sum_{i=1}^{n} \mathbf{x}_i u_i\right] &= \sum_{i=1}^{n} Var\left[\mathbf{x}_i u_i\right] = \sum_{i=1}^{n} E\left[\mathbf{x}_i u_i \left(\mathbf{x}_i u_i\right)'\right] \\ &= \sum_{i=1}^{n} E\left[\mathbf{x}_i \mathbf{x}_i' u_i^2\right] = \sum_{i=1}^{n} E\left[\mathbf{x}_i \mathbf{x}_i' E\left[u_i^2|X\right]\right] \\ &= \sigma^2 \sum_{i=1}^{n} E\left[v_{ii} \mathbf{x}_i \mathbf{x}_i'\right] \equiv \sigma^2 \sum_{i=1}^{n} Q_i \end{aligned}$$

The Asymptotic results are the same as before; In our analysis of the classical model, the heterogeneity of the variances arose because of the regressors, but we still achieved the limiting normal distribution. All that has changed here is that the variance of u varies across observations *as well*. Therefore, *the proof of asymptotic normality from before is general enough to include this model without modification*. As long as X is well-behaved and the diagonal elements of V are finite and well-behaved, the LS estimator is asymptotically normally distributed with the stated covariance matrix.

Apply the Lindeberg-Feller CLT with our new definition of $Q_i = E\left[v_{ii}\mathbf{x}_i\mathbf{x}_i'\right]$. That is, as long as the variances of u_i are finite and are not dominated by any single term, so that the conditions of the Lindeberg-Feller CLT apply to $\sqrt{n}\left(\frac{\sum_{i=1}^{n}\mathbf{x}_i u_i}{n}\right)$, then we have

$$\sqrt{n}\left(\frac{\sum_{i=1}^{n}\mathbf{x}_i u_i}{n}\right) \xrightarrow{d} N\left(0, \sigma^2 S\right),$$

where

$$S = p\lim_{n\to\infty}\left(\frac{X'VX}{n}\right)$$

It then follows that

$$\sqrt{n}\,(b-\beta) \xrightarrow{d} N\left(0, Q^{-1}\sigma^2 SQ^{-1}\right)$$

Since, $AV\left(\sqrt{n}b\right) = Q^{-1}\sigma^2 SQ^{-1}$, the variance for finite sample of size n is

$$AV\,(b) = \frac{\sigma^2}{n} Q^{-1} SQ^{-1}$$

Then as long as V is known, then a consistent estimator is

$$\widehat{AV}\,(b) = \frac{s^2}{n}\left(\frac{X'X}{n}\right)^{-1}\frac{X'VX}{n}\left(\frac{X'X}{n}\right)^{-1}$$

Unknown V

Consider again the conditional heteroskedasticity case with no serial correlation and independent observations, but suppose that V is unknown. Then, how do we estimate $AV\left(\sqrt{n}b\right)$? we need to estimate V.

Example: Consider the simple regression model, $y = \beta_0 + \beta_1 x_i + u$ and suppose that $v_{ii} = x_i^2$ then $\sigma_i^2 \equiv Var\left[u_i | x_i\right] = \sigma^2 x_i^2$, and,

$$E\left[uu' | X\right] = \sigma^2 \begin{bmatrix} x_1^2 & 0 & \cdots & 0 \\ 0 & x_2^2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & x_n^2 \end{bmatrix}.$$

In this case, the solution is easy; use s^2 and we have the estimator

$$\widehat{AV}\left(\sqrt{n}b\right) = s^2 \left(\frac{X'X}{n}\right)^{-1} \frac{X'\text{diad}\left\{x_1^2, \ldots, x_n^2\right\} X}{n} \left(\frac{X'X}{n}\right)^{-1}$$
$$= s^2 \left(\frac{\sum_{i=1}^n x_i^2}{n}\right)^{-1} \frac{\sum_{i=1}^n x_i^4}{n} \left(\frac{\sum_{i=1}^n x_i^2}{n}\right)^{-1}.$$

Now suppose instead that

$$E\left[uu'|X\right] = \sigma^2 \begin{bmatrix} x_1^\theta & 0 & \cdots & 0 \\ 0 & x_2^\theta & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & x_n^\theta \end{bmatrix}.$$

Then, we need to estimate θ.

Why use LS? White's Standard Errors

Although GLS is efficient, we still use LS. Specifically, if we don't know V then feasible GLS requires specification of V, but this entails the risk of specification error.

On the other hand, the use of LS does not require specification of V. With LS we just need a consistent estimator of $X'VX$. This is sufficient for valid asymptotic analysis with the LS.

Therefore, by using LS we trade efficiency for no specification error.

To see the issue more clearly, suppose V is a full matrix, then we have $\frac{n(n+1)}{2}$ variance and covariance parameters that we need to estimate. We also have k beta parameters. So, with n observations we have $\frac{n(n+1)}{2} + k$ unknowns. We have a dimensionality problem.

Under LS, we can find a consistent estimator of $X'VX$ which is robust to any form of V. But with GLS, the variance is $\sigma^2 \left(X'V^{-1}X\right)^{-1}$, but cannot find a consistent and robust estimator for $\left(X'V^{-1}X\right)^{-1}$. So, since the standard errors are robust, in practice we use LS estimator, even though it's not BLUE. We will expand on this now.

We have the model

$$y = X\beta + \varepsilon; \quad E\left[\varepsilon\varepsilon'|X\right] = \sigma^2 V = \sigma^2 \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nn} \end{bmatrix},$$

or

$$\sigma^2 V \equiv \Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{1n} & \cdots & \sigma_n^2 \end{bmatrix} \implies \sigma_i^2 = \sigma^2 v_i\left(X\right) = \sigma^2 v_{ii}.$$

The LS estimator is

$$b = (X'X)^{-1} X'y$$

with

$$\sqrt{n}\,(b-\beta) \xrightarrow{d} N\left(0, \underbrace{\sigma^2 Q^{-1} S Q^{-1}}_{AV\left(\sqrt{n}b\right)}\right),$$

where $Q = p\lim_{n\to\infty}\left(\frac{X'X}{n}\right)$ and $S = p\lim_{n\to\infty}\left(\frac{X'VX}{n}\right)$. For a finite sample $b \overset{a}{\sim} N\left(\beta, \underbrace{\frac{\sigma^2}{n} Q^{-1} S Q^{-1}}_{AV(b)}\right)$.

Conditional Heteroskedasticity
Assume that

$$V = \text{diag}\left\{v_1\left(\mathbf{x}_1\right), v_2\left(\mathbf{x}_2\right), \ldots, v_n\left(\mathbf{x}_n\right)\right\} = \text{diag}\left\{v_{11}, v_{22}, \ldots, v_{nn}\right\};$$

i.e., conditional heteroskedasticity with no autocorrelation and independent observations.

Suppose $V = \text{diag}\left\{v_{11}, v_{22}, \ldots, v_{nn}\right\}$ is known, then a consistent estimator of $AV\left(b\right)$ is

$$\frac{\sigma^2}{n}\left(\frac{X'X}{n}\right)^{-1}\left(\frac{\sum_{i=1}^{n} v_{ii}\mathbf{x}_i\mathbf{x}_i'}{n}\right)\left(\frac{X'X}{n}\right)^{-1},$$

but note that if we know V, then we can also use GLS.

If V is not known, then there are two ways to proceed:

1) White's approach: Use the LS estimator but be agnostic about the form of heteroskedasticity.
2) Use feasible GLS. Know the structure of V and thus estimate V. But we have n date points and we want to estimate $n + K$ parameters (K β and n σ_i^2). This is not possible. So, we need to impose some structure on V by allowing to be a function of few parameters. We will come back to this later.

Agnostic about V: White's (1980) standard errors

Under this approach, we are using the LS estimator for β, but we are going to adjust the standard errors to account for the possible heteroskedasticity. But, we are not going to go look for a consistent estimator of σ_i^2. In fact, White (1980) proposed to look at

$$S_0 = \frac{1}{n} \sum_{i=1}^{n} \hat{u}_i^2 \mathbf{x}_i \mathbf{x}_i',$$

where $\hat{u}_i$ is the LS residual; i.e. $\hat{u}_i^2$ used in lieu of $\sigma_i^2 = \sigma^2 v_{ii}$. White showed that

$$p \lim_{n \to \infty} \left(\frac{\sum_{i=1}^{n} \hat{u}_i^2 \mathbf{x}_i \mathbf{x}_i'}{n} - \frac{\sum_{i=1}^{n} \sigma_i^2 \mathbf{x}_i \mathbf{x}_i'}{n} \right) = 0.$$

Note: we are ***not*** using n residuals to estimate each of σ_i^2s. But we use $\left(\sum_{i=1}^n \hat{u}_i^2 \mathbf{x}_i \mathbf{x}_i'\right)/n$ as an estimator for $\left(\sum_{i=1}^n \sigma_i^2 \mathbf{x}_i \mathbf{x}_i'\right)/n$. We made no assumption about the form of the heteroskedasticity (i.e. σ_i^2 could take any form). That is, this procedure works for any kind of heteroskedasticity; i.e., it even works if we have homoskedasticity.

$$\widehat{AV}(b) = \frac{1}{n}\left(\frac{X'X}{n}\right)^{-1} \frac{\sum_{i=1}^n \hat{u}_i^2 \mathbf{x}_i \mathbf{x}_i'}{n} \left(\frac{X'X}{n}\right)^{-1}$$
$$\equiv \frac{1}{n}\left(\frac{X'X}{n}\right)^{-1} S_0 \left(\frac{X'X}{n}\right)^{-1},$$

is a consistent estimator of

$$AV(b) = \frac{1}{n} Q^{-1} S Q^{-1},$$

The estimator $\widehat{AV}(b)$ is the White Heteroskedasticity Consistent Covariance Estimator (HCCE). HCCE does not assume any information or knowledge of σ_i^2's. That is, HCCE is a robust estimator under ***any*** form of heteroskedasticity. So, HCCE is ***valid asymptotically*** even if there is homoskedasticity.

The square root of diagonal elements of $\hat{AV}(b)$ are the White SE's, or, the ***heteroskedasticity-Robust-Standard Errors***. Given that HCCE is asymptotically valid, then inference based on White SE's is ***asymptotically valid***; i.e. robust t-ratios are asymptotically normally distributed.

Example: Consider the model $y = \beta_1 + \beta_2 x_i + u_i$, then the LS estimator for β_2 is

$$\begin{aligned} b_2 &= \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \\ &= \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right) y_i}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \\ &= \beta_2 + \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right) u_i}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}, \end{aligned}$$

then under conditional heteroskedasticity with no autocorrelation and independent observations,

$$Var\left[b_2\right] = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 Var\left(u_i | x_i\right)}{\left[\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right]^2} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sigma_i^2}{\left[\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right]^2},$$

White's approach replaces σ_i^2 with the squared LS residual $\hat{u}_i^2$ and thus have the consistent estimator

$$\widehat{Var}\left[b_2\right] = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \hat{u}_i^2}{\left[\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right]^2}.$$

But notice that we are ***not*** saying that $\hat{u}_i^2$ is a consistent estimator for σ_i^2. What we are saying, it's that $\widehat{Var}\left[b_2\right]$ is a consistent estimator for $Var\left[b_2\right]$; that is, $\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \hat{u}_i^2$ is a consistent estimator for $\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sigma_i^2$.

The White standard errors' are

$$\text{White } se\left(b_2\right) = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \hat{u}_i^2}{\left[\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right]^2}}.$$

Notice that under conditional homoskedasticity $\sigma_i^2 = \sigma^2$ for all i, and we would have

$$Var\left[b_2\right] = \sigma^2 \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{\left[\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right]^2} = \frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}.$$

Why it doesn't work with GLS?
Recall the GLS estimator is $\hat{\beta}_{GLS} = \left(X'V^{-1}X\right)^{-1} X'V^{-1}y$ with $Var\left[\hat{\beta}_{GLS}\right] = \sigma^2 \left(X'V^{-1}X\right)^{-1}$. We don't know V and so we estimate it by $\hat{V}$. But note that, under conditional heteroskedasticity, $\sum_{i=1}^{n} \hat{u}_i^2 \mathbf{x}_i \mathbf{x}_i' / n$ is a consistent estimator for $X'VX/n$. But a consistent estimator of $X'V^{-1}X/n$ is not available.