Generalized Least Squares

Up until now, we have assumed that

$$Euu' = \sigma^2 I.$$

Now we generalize to let

$$Euu' = \Omega$$

where Ω is restricted to be a positive definite, symmetric matrix.

1 Common Examples

1.1 Autoregressive Models

1.1.1 AR(1)

$$\begin{aligned} u_t &= \rho u_{t-1} + e_t; \\ |\rho| &< 1; \\ e_t &\sim iid\left(0, \sigma^2\right). \end{aligned}$$

$$\begin{aligned} Varu_t &= Eu_t^2 = E\left(\rho u_{t-1} + e_t\right)^2 \\ &= E\left[\rho^2 u_{t-1}^2 + 2\rho u_{t-1} e_t + e_t^2\right] \\ &= \rho^2 E u_{t-1}^2 + 2\rho E u_{t-1} e_t + E e_t^2 \\ &= \rho^2 E u_{t-1}^2 + E e_t^2 \\ &= \rho^2 E u_t^2 + \sigma^2 \quad \text{(stationarity)} \end{aligned}$$

$$\Rightarrow Varu_t = \frac{\sigma^2}{1-\rho^2}.$$

$$\begin{aligned} Cov\left(u_t, u_{t-1}\right) &= Eu_t u_{t-1} = E\left(\rho u_{t-1} + e_t\right) u_{t-1} \\ &= \rho E u_{t-1}^2 + E e_t u_{t-1} = \rho E u_t^2 \\ &= \frac{\rho\sigma^2}{1-\rho^2}. \end{aligned}$$

$$\begin{aligned} Cov\left(u_t, u_{t-2}\right) &= Eu_t u_{t-2} = E\left(\rho u_{t-1} + e_t\right) u_{t-2} \\ &= \rho E u_{t-1} u_{t-2} + E e_t u_{t-2} = \rho E u_t u_{t-1} \\ &= \frac{\rho^2\sigma^2}{1-\rho^2}. \end{aligned}$$

$$Cov\left(u_t, u_{t-n}\right) = \frac{\rho^n \sigma^2}{1-\rho^2}.$$

$\Rightarrow$

$$\Omega = \frac{\sigma^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{pmatrix}$$

1.1.2 AR(2)

$$\begin{aligned} u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t; \\ e_t &\sim iid\left(0, \sigma_e^2\right). \end{aligned}$$

$$\begin{aligned} Var u_t &= \sigma_0 = E u_t^2 = E\left[\rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t\right]^2 \qquad (1) \\ &= \rho_1^2 E u_{t-1}^2 + \rho_2^2 E u_{t-2}^2 + E e_t^2 \\ &\quad + 2\rho_1\rho_2 E u_{t-1} u_{t-2} + 2\rho_1 E u_{t-1} e_t + 2\rho_2 E u_{t-2} e_t \\ &= \rho_1^2 \sigma_0 + \rho_2^2 \sigma_0 + \sigma_e^2 + 2\rho_1\rho_2\sigma_1 \end{aligned}$$

where

$$\sigma_n = Cov\left(u_t, u_{t-n}\right).$$

$$\begin{aligned} \sigma_1 &= E u_t u_{t-1} = E\left[\rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t\right] u_{t-1} \qquad (2) \\ &= \rho_1 E u_{t-1}^2 + \rho_2 E u_{t-2} u_{t-1} + E e_t u_{t-1} \\ &= \rho_1 \sigma_0 + \rho_2 \sigma_1. \end{aligned}$$

We can write equations (1) and (2) in matrix form as

$$\begin{bmatrix} 1-\rho_1^2-\rho_2^2 & -2\rho_1\rho_2 \\ -\rho_1 & 1-\rho_2 \end{bmatrix}\begin{pmatrix} \sigma_0 \\ \sigma_1 \end{pmatrix} = \begin{pmatrix} \sigma_e^2 \\ 0 \end{pmatrix} \qquad (3)$$

and solve for

$$\begin{pmatrix} \sigma_0 \\ \sigma_1 \end{pmatrix} = \begin{bmatrix} 1-\rho_1^2-\rho_2^2 & -2\rho_1\rho_2 \\ -\rho_1 & 1-\rho_2 \end{bmatrix}^{-1}\begin{pmatrix} \sigma_e^2 \\ 0 \end{pmatrix}.$$

The condition for stationarity is that the eigenvalues of

$$\begin{bmatrix} 1-\rho_1^2-\rho_2^2 & -2\rho_1\rho_2 \\ -\rho_1 & 1-\rho_2 \end{bmatrix}^{-1}$$

are greater than one. Equation (3) is called the Yule-Walker equations.

Students should work out the Yule-Walker equations for an AR(3).

1.2 Moving Average

1.2.1 MA(1)

$$\begin{aligned} u_t &= \rho_0 e_t + \rho_1 e_{t-1}; \\ e_t &\sim iid\left(0, \sigma_e^2\right). \end{aligned}$$

$$\begin{aligned} Eu_t^2 &= E\left[\rho_0 e_t + \rho_1 e_{t-1}\right]^2 \\ &= \rho_0^2 E e_t^2 + \rho_1^2 E e_{t-1}^2 + 2\rho_0\rho_1 E e_t e_{t-1} \\ &= \rho_0^2 \sigma_e^2 + \rho_1^2 \sigma_e^2; \end{aligned}$$

$$Eu_t u_{t-1} = E\left[\rho_0 e_t + \rho_1 e_{t-1}\right]\left[\rho_0 e_{t-1} + \rho_1 e_{t-2}\right] = \rho_0 \rho_1 \sigma_e^2.$$

$$Eu_t u_{t-n} = E\left[\rho_0 e_t + \rho_1 e_{t-1}\right]\left[\rho_0 e_{t-n} + \rho_1 e_{t-n-1}\right] = 0$$

if $n > 1$. Thus

$$\Omega = \sigma_e^2 \begin{bmatrix} \rho_0^2 + \rho_1^2 & \rho_0\rho_1 & 0 & \cdots & 0 \\ \rho_0\rho_1 & \rho_0^2 + \rho_1^2 & \rho_0\rho_1 & \cdots & 0 \\ 0 & \rho_0\rho_1 & \rho_0^2 + \rho_1^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \rho_0^2 + \rho_1^2 \end{bmatrix}$$

1.2.2 MA(n)

$$\begin{aligned} u_t &= \sum_{i=0}^{n} \rho_i e_{t-i} \\ e_t &\sim iid\left(0, \sigma_e^2\right). \end{aligned}$$

$$\begin{aligned} Eu_t^2 &= E\left[\sum_{i=0}^{n} \rho_i e_{t-i}\right]^2 = \sigma_e^2 \sum_{i=0}^{n} \rho_i^2; \\ Eu_t u_{t-1} &= E\left[\sum_{i=0}^{n} \rho_i e_{t-i}\right]\left[\sum_{i=0}^{n} \rho_i e_{t-i-1}\right] \\ &= E\left[\sum_{i=0}^{n} \rho_i e_{t-i}\right]\left[\sum_{i=1}^{n+1} \rho_{i-1} e_{t-i}\right] = \sigma_e^2 \sum_{i=1}^{n} \rho_i \rho_{i-1}; \\ Eu_t u_{t-k} &= \begin{cases} \sigma_e^2 \sum_{i=k}^{n} \rho_i \rho_{i-1} & \text{if } k \leq n \\ 0 & \text{if } k > n \end{cases}. \end{aligned}$$

1.3 ARMA

$$\begin{aligned} u_t &= \sum_{i=1}^{m} \alpha_i u_{t-i} + \sum_{i=0}^{n} \rho_i e_{t-i} \\ e_t &\sim iid\left(0, \sigma_e^2\right) \end{aligned}$$

is an ARMA(m,n) process. Work out the Yule-Walker equations.

1.4 Heteroskedasticity

$$\Omega = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 \end{pmatrix}.$$

1.5 Random Coefficients

Consider the model

$$y_t = X_t \beta_t + u_t$$

with

$$\beta_t \sim iid\left(\overline{\beta}, \Omega_\beta\right).$$

Then we can write the model as

$$\begin{aligned} y_t &= X_t \overline{\beta} + X_t\left(\beta_t - \overline{\beta}\right) + u_t \\ &= X_t \overline{\beta} + e_t \end{aligned}$$

where

$$e_t = X_t\left(\beta_t - \overline{\beta}\right) + u_t.$$

$$\begin{aligned} Ee' &= E\left[X\left(\beta - \overline{\beta}\right) + u\right]\left[X\left(\beta - \overline{\beta}\right) + u\right]' \\ &= X\Omega_\beta X' + \sigma_u^2 I. \end{aligned}$$

1.6 Random Effects

Consider the model

$$\begin{aligned} y_{it} &= X_{it}\beta + u_i + e_{it}, \quad i = 1, .., n; \quad t = 1, .., T \\ u_i &\sim iid\left(0, \sigma_u^2\right); \quad e_{it} \sim iid\left(0, \sigma_e^2\right) \end{aligned}$$

(explain panel data; give examples). Define

$$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}; \quad X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{iT} \end{pmatrix}; \quad e_i = \begin{pmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{iT} \end{pmatrix}; \quad \iota = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Then the model can be written as

$$y_i = X_i\beta + \iota u_i + e_i.$$

Now define

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}; \quad X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}; \quad e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}; \quad u = \begin{pmatrix} \iota u_1 \\ \iota u_2 \\ \vdots \\ \iota u_n \end{pmatrix}.$$

Then the model can be written as

$$y = X\beta + v$$

with

$$v = u + e.$$

$$\begin{aligned} \Omega_v &= Evv' = E\left(u+e\right)\left(u+e\right)' \\ &= \begin{pmatrix} A & 0_T & \cdots & 0_T \\ 0_T & A & \cdots & 0_T \\ \vdots & \vdots & \ddots & \vdots \\ 0_T & 0_T & \cdots & A \end{pmatrix} \end{aligned}$$

where

$$A_{T\times T} = \begin{pmatrix} \sigma_u^2 + \sigma_e^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_e^2 & \cdots & \sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 + \sigma_e^2 \end{pmatrix}.$$

2 GLS and OLS

Consider the model

$$y = X\beta + u, \quad u \sim \left(0, \Omega\right).$$

$$\widehat{\beta}_{OLS} = (X'X)^{-1} X'y = \beta + (X'X)^{-1} X'u$$

$$E\widehat{\beta}_{OLS} = \beta + E(X'X)^{-1} X'u = \beta + (X'X)^{-1} X'Eu = \beta.$$

$$\begin{aligned} D\left(\widehat{\beta}_{OLS}\right) &= E(X'X)^{-1} X'uu'X(X'X)^{-1} \\ &= (X'X)^{-1} X'Euu'X(X'X)^{-1} \\ &= (X'X)^{-1} X'\Omega X(X'X)^{-1} \neq \sigma_u^2 (X'X)^{-1} \end{aligned}$$

(especially because there is no such object as σ_u^2). Therefore, $\widehat{\beta}_{OLS}$ is unbiased but the standard errors are wrong. We could correct the standard errors.

Alternatively, let

$$R'R = \Omega^{-1},$$

and consider

$$Ry = RX\beta + Ru.$$

Note that

$$\begin{aligned} ERu &= REu = 0; \\ ERuu'R' &= REuu'R' = R\Omega R' = I. \end{aligned}$$

Define

$$\begin{aligned} \widehat{\beta}_{GLS} &= \left[(RX)'(RX)\right]^{-1} \left[(RX)'(Ry)\right] \\ &= \left[X'R'RX\right]^{-1} \left[X'R'Ry\right] \\ &= \left[X'\Omega^{-1}X\right]^{-1} \left[X'\Omega^{-1}y\right]. \end{aligned}$$

Note that

$$\begin{aligned} \widehat{\beta}_{GLS} &= \left[X'\Omega^{-1}X\right]^{-1} \left[X'\Omega^{-1}y\right] \\ &= \left[X'\Omega^{-1}X\right]^{-1} \left[X'\Omega^{-1}(X\beta + u)\right] \\ &= \left[X'\Omega^{-1}X\right]^{-1} \left[X'\Omega^{-1}X\beta\right] + \left[X'\Omega^{-1}X\right]^{-1} \left[X'\Omega^{-1}u\right] \\ &= \beta + \left[X'\Omega^{-1}X\right]^{-1} \left[X'\Omega^{-1}u\right] \end{aligned}$$

$\Rightarrow$

$$\begin{aligned} E\widehat{\beta}_{GLS} &= \beta + E\left[X'\Omega^{-1}X\right]^{-1} \left[X'\Omega^{-1}u\right] \\ &= \beta + E\left[X'\Omega^{-1}X\right]^{-1} \left[X'\Omega^{-1}Eu\right] \\ &= \beta; \end{aligned}$$

$$
\begin{aligned}
D\left(\widehat{\beta}_{GLS}\right) &= E\left[X'\Omega^{-1}X\right]^{-1}\left[X'\Omega^{-1}u\right]\left[u'\Omega^{-1}X\right]\left[X'\Omega^{-1}X\right]^{-1} \\
&= \left[X'\Omega^{-1}X\right]^{-1}\left[X'\Omega^{-1}Euu'\Omega^{-1}X\right]\left[X'\Omega^{-1}X\right]^{-1} \\
&= \left[X'\Omega^{-1}X\right]^{-1}\left[X'\Omega^{-1}\Omega\Omega^{-1}X\right]\left[X'\Omega^{-1}X\right]^{-1} \\
&= \left[X'\Omega^{-1}X\right]^{-1}\left[X'\Omega^{-1}X\right]\left[X'\Omega^{-1}X\right]^{-1} \\
&= \left[X'\Omega^{-1}X\right]^{-1}.
\end{aligned}
$$

Note that, if

$$\Omega = \sigma^2 I,$$

then

$$
\begin{aligned}
\widehat{\beta}_{GLS} &= \left[X'\Omega^{-1}X\right]^{-1}\left[X'\Omega^{-1}y\right] \\
&= \left[X'\sigma^{-2}IX\right]^{-1}\left[X'\sigma^{-2}Iy\right] \\
&= \left[X'X\right]^{-1}\left[X'y\right] = \widehat{\beta}_{OLS},
\end{aligned}
$$

and

$$D\left(\widehat{\beta}_{GLS}\right) = \left[X'\sigma^{-2}IX\right]^{-1} = \sigma^2\left[X'X\right]^{-1} = D\left(\widehat{\beta}_{OLS}\right).$$

3 Realities of Data

3.1 Testing for Heteroskedasticity and Other Deviations from $\Omega = \sigma^2 I$

3.1.1 Durbin-Watson test statistic

$$
\begin{aligned}
DW &= \frac{\sum_{t=2}^{T}\left(\widehat{u}_t - \widehat{u}_{t-1}\right)^2}{\sum_{t=2}^{T}\widehat{u}_{t-1}^2} \\
&= \frac{\sum_{t=2}^{T}\left(\widehat{u}_t^2 - 2\widehat{u}_t\widehat{u}_{t-1} + \widehat{u}_{t-1}^2\right)}{\sum_{t=2}^{T}\widehat{u}_{t-1}^2}.
\end{aligned}
$$

Consider the error structure associated with an AR(1) process.

$$
\begin{aligned}
plimDW &= \frac{plim\frac{1}{T-1}\sum_{t=2}^{T}\left(\widehat{u}_t^2 - 2\widehat{u}_t\widehat{u}_{t-1} + \widehat{u}_{t-1}^2\right)}{plim\frac{1}{T-1}\sum_{t=2}^{T}\widehat{u}_{t-1}^2} \\
&= \frac{2\sigma_0 - 2\sigma_1}{\sigma_0} = 2\left(1-\rho\right).
\end{aligned}
$$

$$
\begin{array}{ll}
\text{If } \rho = 0, & plimDW = 2 \\
\text{If } \rho \approx 1, & plimDW \approx 0 \\
\text{If } \rho \approx -1, & plimDW \approx 4
\end{array}.
$$

The distribution of the DW test can be found in tables. The DW test is a special case of a Lagrange-Multiplier test statistic (to be learned later).

3.1.2 White Heteroskedasticity Test Statistic

Consider

$$\Omega = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_T^2 \end{pmatrix}$$

and

$$H_0 : \sigma_1^2 = \sigma_2^2 = .. = \sigma_T^2 \text{ vs } H_0 : \sigma_1^2 \neq \sigma_2^2 \neq .. \neq \sigma_T^2.$$

Let $\widehat{u}$ be the OLS residuals. Consider regressing

$$\widehat{u}_t^2 = \alpha_0 + \sum_{j \neq k} X_{tj} X_{tk} \alpha_{jk} + e_t.$$

Under H_0, $\alpha_{jk} = 0 \; \forall j, k$. One can show that

$$TR^2 \sim \chi^2_{k(k+1)/2}.$$

3.2 Estimation when Ω is not Known

Consider the model

$$\begin{aligned} y &= X\beta + u \\ u &\sim (0, \Omega) \end{aligned}$$

and Ω is unknown. Let $\widehat{u}$ be the OLS residuals. The goal is to use the residuals to construct a consistent estimate of Ω. In general, Ω has $T(T+1)/2$ free parameters, and there are only T residuals (with only $T-(n+1)$ degrees of freedom). So, to make any progress, we will have to put a lot of structure on Ω (White heteroskedasticity corrected estimators are exceptions).

Examples:

1. AR(1): For an AR(1) model, there are only 2 parameters: ρ and σ_e^2. Consider

$$\widehat{\rho} = \frac{\sum_{t=2}^T \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=2}^T \widehat{u}_{t-1}^2}.$$

$$plim\ \widehat{\rho} = \frac{plim \frac{1}{T-1} \sum_{t=2}^T \widehat{u}_t \widehat{u}_{t-1}}{plim \frac{1}{T-1} \sum_{t=2}^T \widehat{u}_{t-1}^2} = \frac{\rho\sigma_e^2 / \left(1-\rho^2\right)}{\sigma_e^2 / \left(1-\rho^2\right)} = \rho.$$

   Consider

$$\widehat{\sigma}_e^2 = \left(1 - \widehat{\rho}^2\right) \frac{1}{T-1} \sum_{t=2}^T \widehat{u}_{t-1}^2.$$

$$plim \widehat{\sigma}_e^2 = plim \left(1 - \widehat{\rho}^2\right) \frac{\sigma_e^2}{(1-\rho^2)} = \left(1-\rho^2\right) \frac{\sigma_e^2}{(1-\rho^2)} = \sigma_e^2.$$

We can plug $\widehat{\rho}$ and $\widehat{\sigma}_e^2$ into

$$\Omega = \frac{\sigma_e^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{pmatrix}$$

to get a consistent estimate of Ω, let's say $\widehat{\Omega}$. Now plug $\widehat{\Omega}$ into the GLS estimator for Ω to get

$$\widehat{\beta}_{GLS} = \left[X'\widehat{\Omega}^{-1}X\right]^{-1} \left[X'\widehat{\Omega}^{-1}y\right].$$

We can no longer talk about $E\widehat{\beta}_{GLS}$. But

$$\begin{aligned} \widehat{\beta}_{GLS} &= \left[X'\widehat{\Omega}^{-1}X\right]^{-1} \left[X'\widehat{\Omega}^{-1}(X\beta+u)\right] \\ &= \left[X'\widehat{\Omega}^{-1}X\right]^{-1} \left[X'\widehat{\Omega}^{-1}X\beta\right] + \left[X'\widehat{\Omega}^{-1}X\right]^{-1} \left[X'\widehat{\Omega}^{-1}u\right] \\ &= \beta + \left[X'\widehat{\Omega}^{-1}X\right]^{-1} \left[X'\widehat{\Omega}^{-1}u\right]; \end{aligned}$$

$$\begin{aligned} plim \widehat{\beta}_{GLS} &= \beta + plim \left[\frac{X'\widehat{\Omega}^{-1}X}{T}\right]^{-1} \left[\frac{X'\widehat{\Omega}^{-1}u}{T}\right] \\ &= \beta + \left[plim \frac{X'\widehat{\Omega}^{-1}X}{T}\right]^{-1} \left[plim \frac{X'\widehat{\Omega}^{-1}u}{T}\right] = \beta; \end{aligned}$$

and

$$\sqrt{T}\left(\widehat{\beta}_{GLS} - \beta\right) \sim N(0, V)$$

with

$$V = \left[plim \frac{X'\widehat{\Omega}^{-1}X}{T}\right]^{-1} = \left[plim \frac{X'\Omega^{-1}X}{T}\right]^{-1}.$$

2. MA(1): For an MA(1) model, there are three parameters: ρ_0, ρ_1, and σ_e^2. Without loss of generality, we can set $\rho_0 = 1$ (later we will understand this as an identification issue). What are good consistent estimates for ρ_1 and σ_e^2?

3. Random Effects: For a random effects model, there are two parameters: σ_u^2 and σ_e^2. Let $\widehat{v}_{it}$ be OLS residuals. Consider

$$s_1^2 = \frac{1}{nT} \sum_{i,t} \widehat{v}_{it}^2.$$

Since

$$Ev_{it}^2 = \sigma_u^2 + \sigma_e^2,$$

$$plim\ s_1^2 = \sigma_u^2 + \sigma_e^2. \tag{4}$$

Consider

$$s_2^2 = \frac{1}{n} \sum_{i} \widehat{v}_{i.}^2$$

where

$$\widehat{v}_{i.} = \frac{1}{T} \sum_{t} \widehat{v}_{it}.$$

Since

$$Ev_{i.}^2 = \sigma_u^2 + \frac{1}{T}\sigma_e^2,$$

$$plim\ s_2^2 = \sigma_u^2 + \frac{1}{T}\sigma_e^2. \tag{5}$$

We can solve equations (4) and (5) together to get consistent estimates of σ_u^2 and σ_e^2.

4. General Ω: Assume we can parameterize Ω in terms of a small number of parameters α, and let $\Omega(\alpha)$ represent Ω as a function of α. Let

$$\widehat{u} = (I - P_X)\, y$$

be OLS residuals. Define

$$\widehat{\alpha} = \arg\min_{\alpha} \left[D^* \left(\Omega(\alpha) \right) - D^* \left(\frac{1}{T} \widehat{u}\widehat{u}' \right) \right]$$

where $D^*(\cdot)$ takes the independent elements of the argument. Later we will show that $\widehat{\alpha}$ is a consistent estimate of α. Plug $\widehat{\alpha}$ into $\Omega(\alpha)$ to get $\widehat{\Omega}$, a consistent estimate of Ω. Plug $\widehat{\Omega}$ into the GLS estimator.

3.3 White Heteroskedasticity Correction

Consider the model

$$\begin{aligned} y_t &= X_t \beta + u_t \\ u_t &\sim ind\left(0, \sigma_t^2\right). \end{aligned}$$

We know that

$$D\left(\widehat{\beta}_{OLS}\right) = \left(X'X\right)^{-1} X'\Omega X \left(X'X\right)^{-1}$$

where $\Omega_{tt} = \sigma_t^2$ and $\Omega_{ts} = 0$ for $t \neq s$. Consider $X'\widehat{\Omega}X$ where $\widehat{\Omega}_{tt} = \widehat{u}_t^2$ and $\widehat{\Omega}_{ts} = 0$ for $t \neq s$.

$$\Rightarrow plim \frac{X'\widehat{\Omega}X}{T} = \frac{X'\Omega X}{T}$$

even though $plim\widehat{\Omega}$ doesn't exist.

$$\Rightarrow plim \left(\frac{X'X}{T}\right)^{-1} \frac{X'\widehat{\Omega}X}{T} \left(\frac{X'X}{T}\right)^{-1} = \left(\frac{X'X}{T}\right)^{-1} \frac{X'\Omega X}{T} \left(\frac{X'X}{T}\right)^{-1}.$$

Newey-West generalizes this by letting

$$\widehat{\Omega}_{ts} = \frac{1}{T - |j-k|} \sum_t \widehat{u}_t \widehat{u}_s.$$