

## 8 Heteroskedasticity and serial correlation. HAC standard errors.

The homoskedasticity and no serial correlation assumption (GM3) can be violated in three ways:

- Heteroskedasticity only (B) -  $Var(\varepsilon|X)$  is diagonal with unequal elements along the diagonal.
- Serial correlation only (C) -  $Var(\varepsilon|X)$  has non-zero off-diagonal elements, but all diagonal elements are the same.
- Heteroskedasticity and serial correlation (D) -  $Var(\varepsilon|X)$  is a general non-diagonal matrix with unequal elements along the diagonal.

$$A = \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad B = \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad C = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix} \quad D = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 2 & \rho \\ \rho^2 & \rho & 3 \end{bmatrix}$$

### 8.1 Heteroskedasticity

Under heteroskedasticity OLS is still consistent and asymptotically normal, although is no longer efficient and has a different asymptotic covariance matrix. Thus the default standard errors will be wrong. Recall the large sample OLS assumptions, now consider the weaker assumptions OLS2' and OLS3'

(OLS0)  $(y_i, x_i)$  is an i.i.d. sequence

(OLS1)  $E(x_i x_i')$  is finite non-singular

(OLS2)  $E(y_i | x_i) = x_i' \beta$

(OLS3)  $Var(y_i | x_i) = \sigma^2$

(OLS4)  $E\varepsilon_i^4 < \infty, \quad E\|x_i\|^4 < \infty$

(OLS2')  $E(\varepsilon_i x_i) = 0$

(OLS3')  $Var(\varepsilon_i x_i) = V < \infty$  and is non-singular

**Theorem 8.1.1.** Under OLS0,1,2',3',4

1.  $\hat{\beta}_{OLS} \xrightarrow{p} \beta$  (OLS is consistent)
2.  $\sqrt{n}(\hat{\beta}_{OLS} - \beta) \xrightarrow{d} N(0, (E(x_i x_i'))^{-1} V (E(x_i x_i'))^{-1})$

**Proof.** 1. We only require OLS0,1,2' for consistency

$$\begin{aligned} \hat{\beta}_{OLS} &= \beta + (X'X)^{-1} X' \varepsilon \\ &= \beta + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n x_i \varepsilon_i \\ &\xrightarrow{p} \beta + [E(x_i x_i')]^{-1} E(\varepsilon_i x_i) \\ &= \beta \end{aligned}$$

2.

$$\sqrt{n}(\hat{\beta}_{OLS} - \beta) = \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \varepsilon_i$$

Using the CLT:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \varepsilon_i \xrightarrow{d} N(0, V)$$

Using the CMT:

$$\begin{aligned} \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \varepsilon_i &\xrightarrow{d} [\mathbb{E}(x_i x_i')]^{-1} N(0, V) \\ &\Rightarrow \sqrt{n}(\hat{\beta}_{OLS} - \beta) \xrightarrow{d} N(0, (\mathbb{E}(x_i x_i'))^{-1} V (\mathbb{E}(x_i x_i'))^{-1}) \end{aligned}$$

□

When the errors are homoskedastic the variance is as in previous lectures:

$$\mathbb{E}[X'X]^{-1} \mathbb{E}[X'X \varepsilon_i^2] \mathbb{E}[X'X]^{-1} = \mathbb{E}[X'X]^{-1} \sigma^2 \mathbb{E}[X'X] \mathbb{E}[X'X]^{-1} = \sigma^2 \mathbb{E}[X'X]^{-1}$$

The classic covariance matrix estimator can be highly biased if homoskedasticity fails, we now consider how to construct covariance matrix estimators which do not require homoskedasticity.

If  $\varepsilon_i$  were known, we could have estimated  $V$  as follows:

$$\frac{1}{n} \sum_{i=1}^n x_i x_i' \hat{\varepsilon}_i^2 \xrightarrow{p} V$$

Of course  $\varepsilon_i$  is unknown, but since  $\hat{\beta}_{OLS}$  remains consistent we can use the observed residuals  $\hat{\varepsilon}_i = Y_i - x_i' \hat{\beta}_{OLS}$ :

$$\hat{V} = \frac{1}{n} \sum_{i=1}^n x_i x_i' \hat{\varepsilon}_i^2$$

To show this is a consistent estimator:

$$\begin{aligned} \hat{V} &= \frac{1}{n} \sum_{i=1}^n x_i x_i' \hat{\varepsilon}_i^2 \\ &= \frac{1}{n} \sum_{i=1}^n x_i x_i' \left( \varepsilon_i - x_i' (\hat{\beta}_{OLS} - \beta) \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n x_i x_i' \left( \varepsilon_i^2 - 2\varepsilon_i x_i' (\hat{\beta}_{OLS} - \beta) + \left( x_i' (\hat{\beta}_{OLS} - \beta) \right)^2 \right) \\ &= \frac{1}{n} \sum_{i=1}^n x_i x_i' \varepsilon_i^2 - \frac{2}{n} \sum_{i=1}^n (x_i x_i') \varepsilon_i x_i' (\hat{\beta}_{OLS} - \beta) + \frac{1}{n} \sum_{i=1}^n x_i x_i' \left( x_i' (\hat{\beta}_{OLS} - \beta) \right)^2 \\ &\xrightarrow{p} V \quad \text{since } \hat{\beta}_{OLS} \xrightarrow{p} \beta \end{aligned}$$

**Definition 8.1.1: White's heteroskedasticity robust covariance matrix**

$$\widehat{Var}(\hat{\beta}_{OLS}) = (X'X)^{-1} \left( \sum_{i=1}^n x_i x_i' \hat{\varepsilon}_i^2 \right) (X'X)^{-1}$$

**Note:-**

Whilst this estimator is consistent, it is biased in finite samples. To see this, suppose the actual covariance matrix of the population regression residuals is given by  $\mathbb{E}[\varepsilon\varepsilon'|X] = \Phi = \text{diag}(\phi_i)$ . The covariance matrix of the OLS estimator is then

$$V = (X'X)^{-1}(X'\Phi X)(X'X)^{-1}$$

Denote the  $i$ -th column of the residual maker matrix  $M$  by  $m_i$  then  $\hat{\varepsilon}_i = m_i'\varepsilon$ .

$$\Rightarrow \mathbb{E}[\hat{\varepsilon}_i^2] = \mathbb{E}[m_i'\varepsilon\varepsilon'm_i] = m_i'\Phi m_i$$

Notice that  $m_i$  is the  $i$ -th column of the identity matrix (denoted as  $e_i$ ) minus the  $i$ -th column of the projection matrix  $X(X'X)^{-1}X'$  ( $p_i$ ). Hence  $m_i = e_i - p_i$  and

$$\mathbb{E}[\hat{\varepsilon}_i^2] = (e_i - h_i)'\Phi(e_i - h_i) = \phi_i - 2\phi_i h_{ii} + h_i'\Phi h_i$$

where  $h_{ii}$  is the  $i$ -th diagonal element of the projection matrix. Because this matrix is symmetric and idempotent,  $h_{ii} = h_i'h_i$  so:

$$\begin{aligned} \mathbb{E}(\hat{V} - V) &= (X'X)^{-1}(X'\Phi X)(X'X)^{-1} - (X'X)^{-1}(X'\hat{\Phi}X)(X'X)^{-1} \\ &= (X'X)^{-1}(X'(\Phi - \hat{\Phi})X)(X'X)^{-1} \\ &= (X'X)^{-1}(X'\text{diag}(\phi_i - (\phi_i - 2\phi_i h_{ii} + h_i'\Phi h_i))X)(X'X)^{-1} \\ &= (X'X)^{-1}(X'\text{diag}(h_i'(\Phi - 2\phi_i I)h_i)X)(X'X)^{-1} \end{aligned}$$

Whilst  $\hat{V}$  is biased, here we can see that it is also consistent. Notice that  $\hat{\Phi}$  is not consistent for  $\Phi$ , since there are more elements to estimate as the sample gets large. However,  $\hat{\varepsilon}_i$  is consistent for  $\varepsilon_i$ . We know

$$X'\hat{\Phi}X = \frac{1}{n} \sum_{i=1}^n x_i x_i' \hat{\varepsilon}_i^2$$

and since  $\text{plim } \hat{\varepsilon}_i^2 = \phi_i$  we get  $\text{plim } X'\hat{\Phi}X = X'\Phi X$ .

In summary,  $\hat{V}$  is biased since  $\hat{\varepsilon}_i^2$  is a biased estimate of  $\varepsilon_i^2$ .

## 8.2 Serial correlation (and heteroskedasticity)

As with heteroskedasticity, OLS remains consistent and asymptotically normal, but the default standard errors are wrong. This cannot happen if the data are i.i.d. - if OLS0 holds it must be the case that  $\Omega = \text{Var}(\varepsilon|X)$  is diagonal. If the data are dependent, then  $\Omega$  is typically no longer diagonal.

### Definition 8.2.1: Strict Stationarity

A sequence of random variables  $\{Z_t\}_{t=-\infty}^{\infty}$  is strictly stationary if, for any finite nonnegative integer  $m$ ,

$$f_{Z_t, Z_{t+1}, \dots, Z_{t+m}}(x_0, x_1, \dots, x_m) = f_{Z_s, Z_{s+1}, \dots, Z_{s+m}}(x_0, x_1, \dots, x_m)$$

which is to say that the joint distribution,  $f$ , does not depend on the index,  $t$ .

Strict stationarity implies that the (marginal) distribution of  $Z_t$  does not vary over time. It also implies that the bivariate distributions of  $(Z_t, Z_{t+1})$  and multivariate distributions of  $(Z_t, \dots, Z_{t+m})$  are stable over time.

**Theorem 8.2.1.** If  $Z_t$  is i.i.d., then it is strictly stationary

**Proof.** Let  $F$  denote the joint distribution function, then:

$$\begin{aligned} F(x_{n+1}, \dots, x_{n+m}) &= F(x_{n+1}) \cdot \dots \cdot F(x_{n+m}) \\ &= F(x_{n+k+1}) \cdot \dots \cdot F(x_{n+k+m}) \\ &= F(x_{n+k+1}, \dots, x_{n+k+m}) \end{aligned}$$

Lines 1 and 3 follow from the fact that the joint distribution function of a set of mutually independent variables is equal to the product of their marginal distribution functions. On line 2 we have used the fact that all the terms of the sequence have the same distribution.  $\square$

### Definition 8.2.2: Covariance stationarity

A sequence of random variables  $\{Z_t\}_{t=-\infty}^{\infty}$  is covariance (weakly) stationary if just the first two moments do not depend on  $t$ , e.g.

$$\begin{aligned} \mathbb{E}Z_1 &= \mathbb{E}Z_2 = \dots \\ \text{Var}(Z_1) &= \text{Var}(Z_2) = \dots \\ \text{Cov}(Z_1, Z_{1+m}) &= \text{Cov}(Z_2, Z_{2+m}) = \dots \end{aligned}$$

A strictly stationary process is covariance-stationary as long as the variance and covariances are finite.

Consider a new set of OLS assumptions:

- (SC0)  $\{(y_t, x_t)\}_{t=1}^T$  is strictly stationary
- (SC1)  $\{(x_t x_t')\}$  satisfies LLN:  $\frac{1}{T} \sum x_t x_t' \xrightarrow{p} \mathbb{E}(x_t x_t') < \infty$ , positive definite
- (SC2)  $\{(x_t \varepsilon_t)\}$  satisfies LLN:  $\frac{1}{T} \sum x_t \varepsilon_t \xrightarrow{p} \mathbb{E}(x_t \varepsilon_t) = 0$
- (SC3)  $\{(x_t \varepsilon_t)\}$  satisfies CLT:  $\frac{1}{\sqrt{T}} \sum x_t \varepsilon_t \xrightarrow{d} N(0, V)$ , where

$$V = \mathbb{E}(\varepsilon_t^2 x_t x_t') + \sum_{l=1}^{\infty} (\mathbb{E}(\varepsilon_t \varepsilon_{t-l} x_t x_{t-l}') + \mathbb{E}(\varepsilon_t \varepsilon_{t-l} x_{t-l} x_t'))$$

These assumptions further generalise our GM/OLS conditions, such that if the data were independent, we would have  $V = \mathbb{E}(\varepsilon_t^2 x_t x_t')$  as in OLS3'.

**Theorem 8.2.2.** Under SC0,1,2,3

1.  $\hat{\beta}_{OLS} \xrightarrow{p} \beta$  (OLS is consistent)
2.  $\sqrt{T} (\hat{\beta}_{OLS} - \beta) \xrightarrow{d} N(0, (\mathbb{E}(x_t x_t'))^{-1} V (\mathbb{E}(x_t x_t'))^{-1})$

The proof is identical to the heteroskedastic case in Theorem 8.2.1.

### Newey-West Method

Under the SC assumptions, the conventional covariance matrix estimators are inconsistent as they do not capture the serial dependence in  $x_t \varepsilon_t$ . To consistently estimate the covariance matrix, we need a different estimator. The appropriate class of estimators are called Heteroskedasticity and Autocorrelation Consistent (HAC) covariance matrix estimators.

Define  $V_T$  as follows:

$$\begin{aligned}
V_T &\equiv Var \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T x_t \varepsilon_t \right) \\
&= \mathbb{E} \left[ \frac{1}{T} \left( \sum_{t=1}^T x_t \varepsilon_t \right) \left( \sum_{t=1}^T x_t \varepsilon_t \right)' \right] \\
&= \mathbb{E} \left[ \frac{1}{T} (x_1 \varepsilon_1 + x_2 \varepsilon_2 + \cdots + x_T \varepsilon_T) (x_1' \varepsilon_1 + x_2' \varepsilon_2 + \cdots + x_T' \varepsilon_T)' \right] \\
&= \mathbb{E} \left[ \underbrace{\frac{1}{T} \sum_{t=1}^T \varepsilon_t^2 x_t x_t'}_{\text{variance at } t} + \underbrace{\frac{1}{T} \sum_{\ell=1}^{T-1} \sum_{t=\ell+1}^T (\varepsilon_t \varepsilon_{t-\ell} x_t x_{t-\ell}' + \varepsilon_t \varepsilon_{t-\ell} x_{t-\ell} x_t')}_{\text{all covariances with all other time periods}} \right] \\
&= \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\varepsilon_t^2 x_t x_t'] + \frac{1}{T} \sum_{\ell=1}^{T-1} \sum_{t=\ell+1}^T (\mathbb{E}(\varepsilon_t \varepsilon_{t-\ell} x_t x_{t-\ell}') + \mathbb{E}(\varepsilon_t \varepsilon_{t-\ell} x_{t-\ell} x_t')) \\
&= \mathbb{E}[\varepsilon_t^2 x_t x_t'] + \sum_{\ell=1}^{T-1} \frac{T-\ell}{T} (\mathbb{E}(\varepsilon_t \varepsilon_{t-\ell} x_t x_{t-\ell}') + \mathbb{E}(\varepsilon_t \varepsilon_{t-\ell} x_{t-\ell} x_t')) \quad \text{Using SC0}
\end{aligned}$$

As  $T$  get large,  $V_T \approx V$ . Since we have  $T$  data points, we can only estimate  $G < T$  autocovariances of  $x_t \varepsilon_t$ , where  $G$  is the truncation lag. Newey and West propose the following procedure:

1. Choose  $G$  such that:  $G = O(T^\alpha)$  for  $0 < \alpha < 1/4$
2. Estimate autocovariances of  $x_t \varepsilon_t$  of order  $\ell$  by

$$\hat{\Gamma}_\ell = \frac{1}{T} \sum_{t=\ell+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-\ell} x_t x_{t-\ell}'$$

3. Estimate  $V$  by

$$\hat{V}_{nw} = \hat{\Gamma}_0 + \sum_{\ell=1}^G \frac{G+1-\ell}{G+1} (\hat{\Gamma}_\ell + \hat{\Gamma}_\ell')$$

If we know a priori that autocovariances are zero in population beyond a certain finite lag  $q$ , we can consistently estimate  $V$  with

$$\hat{V} = \hat{\Gamma}_0 + \sum_{\ell=1}^q (\hat{\Gamma}_\ell + \hat{\Gamma}_\ell')$$

However in the case where we do not know  $q$  (which is potentially infinite), we can use the weighted sum suggested by Newey and West. For example, for  $q(n) = 3$

$$\hat{V}_{NW} = \hat{\Gamma}_0 + \frac{2}{3}(\hat{\Gamma}_1 + \hat{\Gamma}_1') + \frac{1}{3}(\hat{\Gamma}_2 + \hat{\Gamma}_2')$$

The weighting term ensures  $\hat{V}_{nw}$  is positive semi-definite. We can see the similarities between this and our expression for  $V_T$  earlier, giving some intuition for its consistency.

$$\begin{aligned}
V_T &= \mathbb{E}[\varepsilon_t^2 x_t x_t'] + \sum_{\ell=1}^{T-1} \frac{T-\ell}{T} \left[ \mathbb{E}(\varepsilon_t \varepsilon_{t-\ell} x_t x_{t-\ell}') + \mathbb{E}(\varepsilon_t \varepsilon_{t-\ell} x_{t-\ell} x_t') \right] \\
\hat{V}_{nw} &= \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2 x_t x_t' + \sum_{\ell=1}^G \frac{G+1-\ell}{G+1} \left[ \frac{1}{T} \sum_{t=\ell+1}^T (\varepsilon_t \varepsilon_{t-\ell} x_t x_{t-\ell}') + \frac{1}{T} \sum_{t=\ell+1}^T (\varepsilon_t \varepsilon_{t-\ell} x_{t-\ell} x_t') \right]
\end{aligned}$$

Now we can estimate the covariance matrix of  $\hat{\beta}_{OLS}$  as

$$\frac{1}{T} \left[ \frac{1}{T} \sum_{t=1}^T x_t x_t' \right]^{-1} \hat{V}_{nw} \left[ \frac{1}{T} \sum_{t=1}^T x_t x_t' \right]^{-1}$$

**Lemma 8.2.1.** The matrix of sample covariances for any process is positive semi-definite.

**Proof.** Let  $z_1, \dots, z_T$  be any sequence of  $T$  numbers, and let  $P$  be a  $m \times m$  matrix of sample covariances:

$$P = \begin{bmatrix} \frac{1}{T} \sum_{t=1}^T z_t^2 & \frac{1}{T} \sum_{t=2}^T z_t z_{t-1} & \cdots & \frac{1}{T} \sum_{t=m+1}^T z_t z_{t-m} \\ \frac{1}{T} \sum_{t=2}^T z_t z_{t-1} & \frac{1}{T} \sum_{t=1}^T z_t^2 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \frac{1}{T} \sum_{t=2}^T z_t z_{t-1} \\ \frac{1}{T} \sum_{t=m+1}^T z_t z_{t-m} & \cdots & \frac{1}{T} \sum_{t=2}^T z_t z_{t-1} & \frac{1}{T} \sum_{t=1}^T z_t^2 \end{bmatrix}$$

Consider the  $m \times (2T-1)$  matrix:

$$Z = \begin{bmatrix} z_1 & z_2 & \cdots & z_m & \cdots & z_T & 0 & \cdots & 0 \\ 0 & z_1 & z_2 & \cdots & z_m & \cdots & z_T & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & z_1 & z_2 & \cdots & \cdots & \cdots & z_T \end{bmatrix}$$

$$\begin{aligned} \frac{1}{T} Z Z' &= \frac{1}{T} \underbrace{\begin{bmatrix} z_1 & z_2 & \cdots & z_m & \cdots & z_T & 0 & \cdots & 0 \\ 0 & z_1 & z_2 & \cdots & z_m & \cdots & z_T & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & z_1 & z_2 & \cdots & \cdots & \cdots & z_T \end{bmatrix}}_{m \times (2T-1)} \underbrace{\begin{bmatrix} z_1 & 0 & \cdots & 0 \\ z_2 & z_1 & \cdots & \vdots \\ \vdots & z_2 & \ddots & 0 \\ z_m & \vdots & \ddots & z_1 \\ \vdots & z_m & \ddots & z_2 \\ z_T & \vdots & \ddots & \vdots \\ 0 & z_T & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & z_T \end{bmatrix}}_{(2T-1) \times m} \\ &= \frac{1}{T} \begin{bmatrix} \sum_{t=1}^T z_t^2 & \sum_{t=2}^T z_t z_{t-1} & \cdots & \sum_{t=m+1}^T z_t z_{t-m} \\ \sum_{t=2}^T z_t z_{t-1} & \sum_{t=1}^T z_t^2 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \sum_{t=2}^T z_t z_{t-1} \\ \sum_{t=m+1}^T z_t z_{t-m} & \cdots & \sum_{t=2}^T z_t z_{t-1} & \sum_{t=1}^T z_t^2 \end{bmatrix} = P_{m \times m} \end{aligned}$$

Thus  $P$  is p.s.d. since for any vector  $v$ ,  $v' Z Z' v = u' u = \sum_{i=1}^m u_i^2 \geq 0$  □

**Theorem 8.2.3.**  $\hat{V}_{nw}$  is positive semi-definite

**Proof.** Let  $c$  be any deterministic  $k$ -dimensional vector, we aim to show  $c'\hat{V}_{nw}c \geq 0$ . Consider the  $G + 1$  matrix

$$P = \begin{bmatrix} c'\hat{\Gamma}_0c & c'\hat{\Gamma}_1c & \ddots & c'\hat{\Gamma}_Gc \\ c'\hat{\Gamma}_1c & c'\hat{\Gamma}_0c & \ddots & \ddots \\ \ddots & \ddots & \ddots & c'\hat{\Gamma}_1c \\ c'\hat{\Gamma}_Gc & \ddots & c'\hat{\Gamma}_1c & c'\hat{\Gamma}_0c \end{bmatrix}$$

If  $i$  is a  $G + 1$ -dimensional vector of ones, then we have

$$\begin{aligned} i'Pi &= \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} c'\hat{\Gamma}_0c & c'\hat{\Gamma}_1c & \ddots & c'\hat{\Gamma}_Gc \\ c'\hat{\Gamma}'_1c & c'\hat{\Gamma}_0c & \ddots & \ddots \\ \ddots & \ddots & \ddots & c'\hat{\Gamma}_1c \\ c'\hat{\Gamma}'_Gc & \ddots & c'\hat{\Gamma}'_1c & c'\hat{\Gamma}_0c \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} \sum_{\ell=0}^G c'\hat{\Gamma}_\ell c \\ \sum_{\ell=1}^1 c'\hat{\Gamma}'_\ell c + \sum_{\ell=0}^{G-1} c'\hat{\Gamma}_\ell c \\ \vdots \\ \sum_{\ell=0}^G c'\hat{\Gamma}'_\ell c \end{bmatrix} \quad m\text{-th row} = \sum_{\ell=1}^m c'\hat{\Gamma}'_\ell c + \sum_{\ell=0}^{G-m} c'\hat{\Gamma}_\ell c \\ &= \sum_{\ell=0}^G c'\hat{\Gamma}_\ell c + \sum_{\ell=1}^1 c'\hat{\Gamma}'_\ell c + \sum_{\ell=0}^{G-1} c'\hat{\Gamma}_\ell c + \dots + \sum_{\ell=0}^G c'\hat{\Gamma}'_\ell c \\ &= (G+1)c'\hat{\Gamma}_0c + G(c'\hat{\Gamma}'_1c + c'\hat{\Gamma}_1c) + (G-1)(c'\hat{\Gamma}'_2c + c'\hat{\Gamma}_2c) + \dots \\ &= (G+1)c'\hat{\Gamma}_0c + \sum_{\ell=1}^G (G+1-\ell)(c'\hat{\Gamma}'_\ell c + c'\hat{\Gamma}_\ell c) \\ &\Rightarrow \frac{1}{G+1}i'Pi = c'\hat{\Gamma}_0c + \sum_{\ell=1}^G \frac{G+1-\ell}{G+1}(c'\hat{\Gamma}'_\ell c + c'\hat{\Gamma}_\ell c) \\ &= c'\hat{V}_{nw}c \end{aligned}$$

Hence, it is sufficient to show that  $P$  is positive semi-definite. However,  $P$  is the matrix of sample covariances of the process  $z_t = c'x_t\hat{\varepsilon}_t$  with autocovariances:

$$\mathbb{E}[c'x_t\varepsilon_t\varepsilon_{t-j}x'_{t-j}c] = c'\mathbb{E}[\varepsilon_t\varepsilon_{t-j}x_tx'_{t-j}]c = c'\Gamma_jc \quad \forall j \in \mathbb{Z}$$

The matrix of sample covariances for any process is positive semi-definite, thus  $\hat{V}_{nw}$  is p.s.d.  $\square$

**Note:-**

The population covariance matrix is always positive semi-definite, so it's desirable for its estimate to also be positive semi-definite. Thus in a time series context we define sample covariances as:

$$\frac{1}{T} \sum_{t=|i-j|+1}^T z_t z_{t-|i-j|} \quad \text{rather than as} \quad \frac{1}{T-|i-j|} \sum_{t=|i-j|+1}^T z_t z_{t-|i-j|}$$

Even though the former is biased and the latter unbiased, had we used the latter we might get an estimate that is not positive semi-definite.