Cornell University
Department of Economics

Econ 620
Instructor: Prof. Kiefer

Solution to Problem set # 3

1)
Recall that

$$\begin{aligned} e &= y - X\widehat{\beta} = y - X\left(X'X\right)^{-1}X'y = \left[I - X\left(X'X\right)^{-1}X'\right]y = My \\ &= M\left(X\beta + \varepsilon\right) = MX\beta + M\varepsilon = M\varepsilon \end{aligned}$$

Then,

$$E\left(e\right) = E\left(M\varepsilon\right) = ME\left(\varepsilon\right) = 0$$

since $M = \left[I - X\left(X'X\right)^{-1}X'\right]$ is non-stochastic. Hence,

$$\begin{aligned} Var\left(e\right) &= E\left[\left(e - E\left(e\right)\right)\left(e - E\left(e\right)\right)'\right] = E\left[ee'\right] \\ &= E\left[M\varepsilon\varepsilon'M'\right] = ME\left[\varepsilon\varepsilon'\right]M = \sigma^2 MIM \\ &= \sigma^2 M \end{aligned}$$

note that M is symmetric and idempotent. The variance matrix of e is an $(N \times N)$ matrix. The variance of e_j is the (j,j) element of the variance matrix, which can be picked up by

$$Var\left(e_j\right) = \sigma^2 M^{jj} = \sigma^2\left[I - X\left(X'X\right)^{-1}X'\right]^{jj} = \sigma^2\left[1 - X_j\left(X'X\right)^{-1}X'^j\right]$$

where X_j is the jth row of X and X'^j is the jth column of X'. Then,

$$\begin{aligned} Var\left(e_j\right) - \sigma^2 &= \sigma^2\left[1 - X_j\left(X'X\right)^{-1}X'^j\right] - \sigma^2 \\ &= -\sigma^2 X_j\left(X'X\right)^{-1}X'^j \\ &= -\sigma^2 X_j\left(X'X\right)^{-1}X_j' \leq 0 \end{aligned}$$

since $X_j\left(X'X\right)^{-1}X_j'$ is a quadratic form in $\left(X'X\right)^{-1}$ and we know that $\left(X'X\right)$ is positive semidefinite and hence so is $\left(X'X\right)^{-1}$.

2)

What is the operator we use to get mean deviation form? Yes, it is $A = I - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$. Then, the X matrix is now;

$$X = \begin{bmatrix} \mathbf{1} & AX_2 \end{bmatrix}$$

where $\mathbf{1}$ is an $(N \times 1)$ vector of ones and X_2 is an $(N \times (k-1))$ matrix of independent variables except for the constant term. Therefore,

$$\begin{aligned} X'X &= \begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'AX_2 \\ X_2'A\mathbf{1} & X_2'AA'X_2 \end{bmatrix} = \begin{bmatrix} N & \mathbf{0} \\ \mathbf{0} & X_2'AX_2 \end{bmatrix} \\ X'y &= \begin{bmatrix} \mathbf{1}'y \\ X_2'Ay \end{bmatrix} \end{aligned}$$

note that $\mathbf{1}'A = A\mathbf{1} = \mathbf{0}$ and again A is symmetric idempotent. Hence,

$$\begin{aligned} Var\left(\widehat{\beta}\right) &= \sigma^2 (X'X)^{-1} = \sigma^2 \begin{bmatrix} N & \mathbf{0} \\ \mathbf{0} & X_2'AX_2 \end{bmatrix}^{-1} \\ &= \sigma^2 \begin{bmatrix} \frac{1}{N} & \mathbf{0} \\ \mathbf{0} & (X_2'AX_2)^{-1} \end{bmatrix} \end{aligned}$$

We use the fact that $(X'X)$ is block diagonal. The covariance between the intercept and the slope estimator is the off-diagonal term, which is $\mathbf{0}$.

3)

1. (a) Easy!

   (b) First of all, note that

$$\begin{aligned} \widehat{\beta}_2 &= (X_2'M_1X_2)^{-1}(X_2'M_1y) \\ &= (X_2'M_1X_2)^{-1} X_2'M_1 (X_1\beta_1 + X_2\beta_2 + \varepsilon) \\ &= (X_2'M_1X_2)^{-1} X_2'M_1X_2\beta_2 + (X_2'M_1X_2)^{-1} X_2'M_1\varepsilon \\ &= \beta_2 + (X_2'M_1X_2)^{-1} X_2'M_1\varepsilon \end{aligned}$$

   where $M_1 = I - X_1(X_1'X_1)^{-1}X_1'$. The third equality come from the fact that $M_1X_1 = \mathbf{0}$. Then,

$$\begin{aligned} E\left(\widehat{\beta}_2\right) &= \beta_2 + (X_2'M_1X_2)^{-1} X_2'M_1E(\varepsilon) \\ &= \beta_2 + (X_2'M_1X_2)^{-1} X_2'M_1X_1\gamma \\ &= \beta_2 \end{aligned}$$

   again since $M_1X_1 = \mathbf{0}$.

4)

For X_n, it is obvious that

$$\text{plim} X_n = \text{plim}\left(3 - \frac{1}{n^2}\right) = 3 \tag{1}$$

On the other hand, by the central limit theorem,

$$\sqrt{n}\left(\overline{Z}_n - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right)$$

when $\overline{Z}_n = \frac{1}{N}\sum_{i=1}^{n} Z_i$ with $E\left(Z_i\right) = \mu$ and $Var\left(Z_i\right) = \sigma^2$. The CLT can also be expressed as

$$\frac{\sqrt{n}\left(\overline{Z}_n - \mu\right)}{\sigma} \xrightarrow{d} N\left(0, 1\right)$$

In our case, $E\left(Z_i\right) = 0$. Therefore,

$$Y_n = \frac{\sqrt{n}\overline{Z}_n}{\sigma} \xrightarrow{d} N\left(0, 1\right) \tag{2}$$

Moreover, recall the following theorems; If $X_n \xrightarrow{p} c$ and $Y_n \xrightarrow{d} Y$

(i) $X_n + Y_n \xrightarrow{d} c + Y$

(ii) $X_n Y_n \xrightarrow{d} cY$

(iii) If $Y_n \xrightarrow{d} Y$ and g is continuous, $g\left(Y_n\right) \xrightarrow{d} g\left(Y\right)$

(a) From (1) and (2) with (i), we have

$$X_n + Y_n \xrightarrow{d} 3 + Y$$

where $Y \sim N\left(0, 1\right)$. Then,

$$X_n + Y_n \xrightarrow{d} N\left(3, 1\right)$$

(b) From (1) and (2) with (ii), we have

$$X_n Y_n \xrightarrow{d} 3Y$$

where $Y \sim N\left(0, 1\right)$. Then,

$$X_n Y_n \xrightarrow{d} N\left(0, 9\right)$$

(c) From (2) and (iii),

$$Y_n^2 \xrightarrow{d} Y^2$$

where $Y \sim N\left(0, 1\right)$. Then,

$$Y_n^2 \xrightarrow{d} \chi^2\left(1\right)$$

5)

1. Model I;

$$y = \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 D_4 + \varepsilon$$

Data matrices are given by

$$y = \begin{bmatrix} y_1^1 \\ \cdots \\ y_n^1 \\ y_1^2 \\ \cdots \\ y_n^2 \\ y_1^3 \\ \cdots \\ y_n^3 \\ y_1^4 \\ \cdots \\ y_n^4 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where y_i^j is the observation on the dependent variable in year i, quarter j. Then,

$$\widehat{\alpha} = \left(X'X\right)^{-1} X'y = \begin{bmatrix} n & 0 & 0 & 0 \\ 0 & n & 0 & 0 \\ 0 & 0 & n & 0 \\ 0 & 0 & 0 & n \end{bmatrix}^{-1} \begin{bmatrix} n\overline{y}^1 \\ n\overline{y}^2 \\ n\overline{y}^3 \\ n\overline{y}^4 \end{bmatrix} = \begin{bmatrix} \overline{y}^1 \\ \overline{y}^2 \\ \overline{y}^3 \\ \overline{y}^4 \end{bmatrix}$$

where $\overline{y}^j$ is average value of the dependent variable in the j^{th} quarter.

Model II;

$$y = \alpha + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 D_4 + \varepsilon$$

Data matrices are given by

$$y = \begin{bmatrix} y_1^1 \\ \cdots \\ y_n^1 \\ y_1^2 \\ \cdots \\ y_n^2 \\ y_1^3 \\ \cdots \\ y_n^3 \\ y_1^4 \\ \cdots \\ y_n^4 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

Then,

$$\widehat{\alpha} = \begin{bmatrix} 4n & n & n & n \\ n & n & 0 & 0 \\ n & 0 & n & 0 \\ n & 0 & 0 & n \end{bmatrix}^{-1} \begin{bmatrix} n\left(\overline{y}^1 + \overline{y}^2 + \overline{y}^3 + \overline{y}^4\right) \\ n\overline{y}^2 \\ n\overline{y}^3 \\ n\overline{y}^4 \end{bmatrix} = \begin{bmatrix} \overline{y}^2 \\ \overline{y}^2 - \overline{y}^1 \\ \overline{y}^3 - \overline{y}^1 \\ \overline{y}^4 - \overline{y}^1 \end{bmatrix}$$

Suppose the model with other explanatory variable;

$$\begin{aligned} y &= \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 D_4 + \beta x + \varepsilon = D\alpha + \beta x + \varepsilon \\ y &= \alpha + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 D_4 + \beta x + \varepsilon = D^*\alpha^* + \beta^* x + \varepsilon \end{aligned}$$

Then,

$$\widehat{\beta} = \left(x' M_D x\right)^{-1} \left(x' M_D y\right) \text{ and } \widehat{\beta}^* = \left(x' M_{D^*} x\right)^{-1} \left(x' M_{D^*} y\right)$$

where $M_D = I - D\left(D'D\right)^{-1} D'$ and $M_{D^*} = I - D^*\left(D^{*\prime}D^*\right)^{-1} D^{*\prime}$. Then,

$$D\left(D'D\right)^{-1} D' = \begin{bmatrix} \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' & 0_{n\times n} & 0_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' & 0_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} & \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' \end{bmatrix}$$

where $\mathbf{1}_n$ is an $(n \times 1)$ vector of ones and $0_{n\times n}$ is an $(n \times n)$ matrix of zeros. On the other hand

$$D^*\left(D^{*\prime}D^*\right)^{-1} D^{*\prime} = \begin{bmatrix} \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' & 0_{n\times n} & 0_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' & 0_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} & \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} & 0_{n\times n} & \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' \end{bmatrix}$$

Therefore, $\widehat{\beta} = \widehat{\beta}^*$.

What if we run the model;

$$y = \alpha + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 D_4 + \varepsilon$$

The X matrix is given by

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The first column is the sum of the other columns. The X matrix is not of the full column rank, which results in the singularity of $(X'X)$ matrix. -Dummy trap-.

6)

1. We will be very careful in indicating which theorem we use in each step. We start from the definition of the least squares estimator;

$$\widehat{\beta} = (X'X)^{-1} X'y = \beta + (X'X)^{-1} X'\varepsilon \tag{1}$$

(a) It is much easier to see what is going on if we express the matrix expression in terms of summation. After a thoughtful moment, you notice that it is given by

$$(X'X) = \sum_{i=1}^{N} x_i x_i'$$

where x_i is a $(k \times 1)$ vector corresponding to the i^{th} observation. From the condition given in the question

$$\text{plim}\frac{1}{N} X'X = Q$$

We can conclude that

$$\text{plim}\frac{1}{N}\sum_{i=1}^{N} x_i x_i' = Q$$

The matrix notation is exactly the condition;

$$\text{plim}\frac{X'X}{N} = Q \tag{2}$$

What about $(X'\varepsilon)$? $-$ remember that $(X'\varepsilon)$ is a $(k \times 1)$ vector $-$. Again it is given by

$$\sum_{i=1}^{N} x_i \varepsilon_i$$

Let's scale the sum by N to get $\frac{1}{N}\sum_{i=1}^{N} x_i\varepsilon_i$. Note that

$$\frac{1}{N}\sum_{i=1}^{N} x_i\varepsilon_i = \frac{1}{N}\left(x_1\varepsilon_1 + x_2\varepsilon_2 + \cdots + x_N\varepsilon_N\right)$$

The term is the sample average of $x_i\varepsilon_i$, where $x_i\varepsilon_i's$ are uncorrelated random vectors with mean 0 and variance $\sigma^2 x_i x_i'$ since

$$\begin{aligned} E\left(x_i\varepsilon_i\right) &= x_i E\left(\varepsilon_i\right) = 0 \text{ since } x_i \text{ is non-stochastic.} \\ Var\left(x_i\varepsilon_i\right) &= E\left(x_i\varepsilon_i\varepsilon_i x_i'\right) = x_i x_i' E\left(\varepsilon_i^2\right) = \sigma^2 x_i x_i' \\ Cov\left(x_i\varepsilon_i, x_t\varepsilon_t\right) &= E\left[x_i\varepsilon_i\varepsilon_t x_t'\right] = x_i x_t' E\left(\varepsilon_i\varepsilon_t\right) = 0 \text{ since } i \neq t. \end{aligned}$$

Note also that Var$(\frac{X'\varepsilon}{N})$=$\frac{\sigma^2}{N}\frac{(X'X)}{N} \to 0Q = 0$. Then, from the Weak Law of Large Numbers(WLLN), we have

$$\frac{1}{N}\sum_{i=1}^{N} x_i\varepsilon_i \xrightarrow{p} \mathbf{0}$$

Then, in vector notation, we have

$$\frac{1}{N}X'\varepsilon \xrightarrow{p} \mathbf{0} \tag{3}$$

We will slightly reshape (1) to get;

$$\widehat{\beta} = \beta + \left(\frac{X'X}{N}\right)^{-1}\frac{X'\varepsilon}{N}$$

Then,

$$\begin{aligned}
\text{plim}\widehat{\beta} &= \text{plim}\beta + \text{plim}\left(\frac{X'X}{N}\right)^{-1}\frac{X'\varepsilon}{N} \text{ by (b) in question2} \\
&= \beta + \text{plim}\left(\frac{X'X}{N}\right)^{-1}\text{plim}\frac{X'\varepsilon}{N} \text{ by (a) in question2} \\
&= \beta + \left(\text{plim}\frac{X'X}{N}\right)^{-1}\text{plim}\frac{X'\varepsilon}{N} \text{ by Slutsky's theorem} \\
&= \beta + Q^{-1}\mathbf{0} \text{ by (2) and (3) and } Q \text{ is invertible} \\
&= \beta
\end{aligned}$$

i.e.

$$\widehat{\beta} \xrightarrow{p} \beta$$

In words, the least squares estimator $\widehat{\beta}$ is a consistent estimator for β.

(b) From (1), we have

$$\widehat{\beta} - \beta = (X'X)^{-1} X'\varepsilon$$

Now, we want to scale slightly differently to invoke the Central Limit Theorem(CLT);

$$\sqrt{N}\left(\widehat{\beta} - \beta\right) = \left(\frac{X'X}{N}\right)^{-1}\frac{X'\varepsilon}{\sqrt{N}} \tag{4}$$

We know that

$$\left(\frac{X'X}{N}\right)^{-1} \xrightarrow{p} Q^{-1} \tag{5}$$

from (2). Now let's take care of $\frac{X'\varepsilon}{\sqrt{N}}$. Again, $\frac{X'\varepsilon}{\sqrt{N}}$ is given by

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i\varepsilon_i = \frac{1}{\sqrt{N}}\left(x_1\varepsilon_1 + x_2\varepsilon_2 + \cdots + x_N\varepsilon_N\right)$$

As we've already seen in (a), $x_i\varepsilon_i's$ are uncorrelated random vectors with mean 0 and variance $\sigma^2 x_i x_i'$. Then, by CLT - here, we use a version of

CLT in Page 7 of the lecture note since we have different variances across observations-,

$$\left(\sum_{i=1}^{N} \sigma^2 x_i x_i'\right)^{\frac{1}{2}} \sum_{i=1}^{N} x_i \varepsilon_i \xrightarrow{d} N(\mathbf{0}, I) \tag{6}$$

where $\left(\sum_{i=1}^{N} \sigma^2 x_i x_i'\right)^{\frac{1}{2}}$ is a notation for Λ such that $\Lambda^2 = \sum_{i=1}^{N} \sigma^2 x_i x_i'$. However, we know that

$$\frac{1}{N} \sum_{i=1}^{N} \sigma^2 x_i x_i' = \sigma^2 \frac{1}{N} \sum_{i=1}^{N} x_i x_i' \xrightarrow{p} \sigma^2 Q \tag{7}$$

from part (a). Hence,

$$\left(\frac{1}{N} \sum_{i=1}^{N} \sigma^2 x_i x_i'\right)^{\frac{1}{2}} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i \varepsilon_i \xrightarrow{d} N(\mathbf{0}, I)$$

becomes

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i \varepsilon_i \xrightarrow{d} N\left(0, \sigma^2 Q\right) \tag{8}$$

Then, from (5) and (8) with (b) in question (3), we have

$$\sqrt{N}\left(\widehat{\beta} - \beta\right) = \left(\frac{X'X}{N}\right)^{-1} \frac{X'\varepsilon}{\sqrt{N}} \xrightarrow{d} N\left(\mathbf{0}, Q^{-1} Q Q^{-1}\right) = N\left(\mathbf{0}, Q^{-1}\right)$$

To appreciate the importance of this result, note that we have obtained asymptotic normality of $\widehat{\beta}_{OLS}$ WITHOUT the assumption of normality of the error term!.

(c) Note that

$$\begin{aligned}
s^2 &= \frac{e'e}{N-k} = \frac{\varepsilon' M \varepsilon}{N-k} \text{ since } e = M\varepsilon \\
&= \frac{\varepsilon'\left[I - X\left(X'X\right)^{-1} X'\right]\varepsilon}{N-k} = \frac{N}{N-k}\left[\frac{\varepsilon'\varepsilon}{N} - \frac{\varepsilon' X\left(X'X\right)^{-1} X'\varepsilon}{N}\right] \\
&= \frac{N}{N-k}\left[\frac{\varepsilon'\varepsilon}{N} - \frac{\varepsilon' X}{N}\left(\frac{X'X}{N}\right)^{-1} \frac{X'\varepsilon}{N}\right]
\end{aligned}$$

Now,

$$
\begin{aligned}
\text{plim}s^2 &= \text{plim}\frac{N}{N-k}\text{plim}\left[\frac{\varepsilon'\varepsilon}{N} - \frac{\varepsilon'X}{N}\left(\frac{X'X}{N}\right)^{-1}\frac{X'\varepsilon}{N}\right] \text{ by (a) in question2} \\
&= \text{plim}\frac{N}{N-k}\left[\text{plim}\frac{\varepsilon'\varepsilon}{N} - \text{plim}\frac{\varepsilon'X}{N}\left(\frac{X'X}{N}\right)^{-1}\frac{X'\varepsilon}{N}\right] \text{ by (b) in question2} \\
&= \text{plim}\frac{N}{N-k}\left[\text{plim}\frac{\varepsilon'\varepsilon}{N} - \text{plim}\frac{\varepsilon'X}{N}\text{plim}\left(\frac{X'X}{N}\right)^{-1}\text{plim}\frac{X'\varepsilon}{N}\right] \text{ by (a) in question2} \\
&= \text{plim}\frac{N}{N-k}\left[\text{plim}\frac{\varepsilon'\varepsilon}{N} - \text{plim}\frac{\varepsilon'X}{N}\left(\text{plim}\frac{X'X}{N}\right)^{-1}\text{plim}\frac{X'\varepsilon}{N}\right] \text{ by } Slutsky's \text{ theorem} \\
&= \left[\sigma^2 - \mathbf{0}'Q^{-1}\mathbf{0}\right] = \sigma^2
\end{aligned}
$$

since

$$
\begin{aligned}
\underset{N\to\infty}{\text{plim}}\frac{N}{N-k} &= 1, \text{plim}\frac{X'\varepsilon}{N} = \mathbf{0} \text{ by (3)} \\
\left(\text{plim}\frac{X'X}{N}\right)^{-1} &= Q^{-1} \text{ by (5)}
\end{aligned}
$$

and

$$
\frac{\varepsilon'\varepsilon}{N} = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_i^2
$$

which is again an average of $\varepsilon_i^2 {}' s$ whose mean is $E\left(\varepsilon_i^2\right) = \sigma^2$ .Here, if we assume that $\varepsilon_i^2 {}' s$ are independent, then we don't need to calculate the variance since we can use a version of WLLN in *Notes 3* on page 4 of the lecture note # 8, and therefore WLLN applies to this case. But otherwise, we need extra conditions on the $\varepsilon_i^2 {}' s$, such as that they are uncorrelated and that for all i, $E\left(\varepsilon_i^4\right) = \phi < \infty$ (i.e., second moment of the $\varepsilon_i^2 {}' s$ exists) . Then, by WLLN

$$
\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i^2 \xrightarrow{p} \sigma^2
$$

Therefore,

$$
\text{plim}\frac{\varepsilon'\varepsilon}{N} = \sigma^2
$$