# University of Rome Tor Vergata
# Department of Economics and Finance

## Static Regression - Fall 2018 - Prof. Cubadda

## Problem Set 2 - Solutions

TA Claudia De Palo
claudia.depalo@alumni.uniroma2.eu

---

## Exercise 1

Consider the testing procedure:

$$H_0 : R\beta = r$$

$$H_A : R\beta \neq r$$

where $R$ is $q \times k$ with $q \leq k$, $\beta$ is $k \times 1$ and $r$ is $q \times 1$.

The distribution of least squares estimator is the following:

$$\hat{\beta} \sim N\Big(\beta, \sigma^2 (X^{'}X)^{-1}\Big)$$

Then, we have:

$$R\hat{\beta} \sim N\Big(R\beta, \sigma^2 R(X^{'}X)^{-1}R^{'}\Big)$$

Show that:

a) Under the null hypothesis, $R\hat{\beta} - r \sim N\Big(0, \sigma^2 R(X^{'}X)^{-1}R^{'}\Big)$

b) $\dfrac{\Big(R\hat{\beta} - r\Big)^{'} \Big[s^2 R(X^{'}X)^{-1}R^{'}\Big]^{-1} \Big(R\hat{\beta} - r\Big)}{q} \sim F(q, n-k)$

*Hint: Be reminded that if a g-dimensional random variable $U$ is distributed as $N(0, \Sigma)$, then $U'\Sigma^{-1}U \sim \chi^2(g)$*

*Moreover, to show that the numerator and denominator are two independent $\chi^2$ variables, recall that you can write the above expression in terms of restricted and unrestricted residuals.*

**Solution**

We compute the first and the second moments in order to show that under the null hypothesis, $E(R\hat{\beta}-r) = 0$ and $Var(R\hat{\beta} - r) = \sigma^2 R(X'X)^{-1}R'$.

a)

$$E(R\hat{\beta} - r) = R\beta - R\beta = 0$$

$$\begin{aligned}
Var(R\hat{\beta} - r) &= E\Big[(R\hat{\beta} - r)(R\hat{\beta} - r)'\Big] \\
&= E\Big[(R\hat{\beta} - R\beta)(R\hat{\beta} - R\beta)'\Big] \\
&= E\Big[R(\hat{\beta} - \beta)(\hat{\beta} - \beta)'R'\Big] \\
&= E\left\{R\Big[(X'X)^{-1}X'\varepsilon\Big]\Big[(X'X)^{-1}X'\varepsilon\Big]'R'\right\} \\
&= E\left\{R\Big[(X'X)^{-1}X'\varepsilon\Big]\Big[\varepsilon'X(X'X)^{-1}\Big]R'\right\} \\
&= R(X'X)^{-1}X'\underbrace{E(\varepsilon\varepsilon')}_{\sigma^2}X(X'X)^{-1}R' \\
&= \sigma^2 R(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{I_k}R' \\
&= \sigma^2 R(X'X)^{-1}R'
\end{aligned}$$

b) We know that the F distribution is a ratio between two independent $\chi^2$ variables divided by their respective degrees of freedom.
In order to derive the numerator of the F distribution, we recall that if a g-dimensional random variable $U$ is distributed as $N(0, \Sigma)$, then $U'\Sigma^{-1}U \sim \chi^2(g)$. Applying this, we obtain:

$$\Big(R\hat{\beta} - r\Big)'\Big[\sigma^2 R(X'X)^{-1}R'\Big]^{-1}\Big(R\hat{\beta} - r\Big) \sim \chi^2(q)$$

where q degrees of freedom are the number of the elements of the vector $\underbrace{R\hat{\beta}}_{(q\times k)(k\times 1)}$.
In the denominator, in order to get rid of the unknown quantity $\sigma^2$ we use:

$$\frac{(n-k)s^2}{\sigma^2} = \frac{e^{*\prime}e^*}{\sigma^2} \sim \chi^2(n-k)$$

Thus, we have:

$$\frac{\left(R\hat{\beta}-r\right)'\left[\sigma^2 R(X'X)^{-1}R'\right]^{-1}\left(R\hat{\beta}-r\right)/q}{\frac{\frac{e^{*\prime}e^*}{\sigma^2}}{(n-k)}}$$

$$= \frac{\left(R\hat{\beta}-r\right)'\left[R(X'X)^{-1}R'\right]^{-1}\left(R\hat{\beta}-r\right)/q}{\frac{e^{*\prime}e^*}{(n-k)}}$$

Since $e'e - e^{*\prime}e^* = (R\hat{\beta}-r)'[R(X'X)^{-1}R'](R\hat{\beta}-r)$, then the above expression become:

$$= \frac{(e'e - e^{*\prime}e^*)/q}{e^{*\prime}e^*/(n-k)}$$

Notice that, $e'e - e^{*\prime}e^* = \varepsilon'(M - M^*)\varepsilon$ and $tr(M - M^*) = (n-k+q) - (n-k) = q$. Hence,

$$\frac{qs^2}{\sigma^2} = \frac{e'e - e^{*\prime}e^*}{\sigma^2} \sim \chi^2(q)$$

In addition, $\varepsilon'(M - M^*)\varepsilon$ and $\varepsilon' M\varepsilon$ are independent because $(M - M^*)M = 0$
Thus, we have:

$$\frac{(e'e - e^{*\prime}e^*)/q}{e^{*\prime}e^*/(n-k)} \sim F(q, n-k)$$

Rearrange the terms as follows

$$\frac{(e'e - e^{*\prime}e^*)/q}{e^{*\prime}e^*/(n-k)} = \frac{\left(R\hat{\beta}-r\right)'\left[R(X'X)^{-1}R'\right]^{-1}\left(R\hat{\beta}-r\right)/q}{\frac{e^{*\prime}e^*}{(n-k)}}$$

$$= \frac{\left(R\hat{\beta}-r\right)'\left[R(X'X)^{-1}R'\right]^{-1}\left(R\hat{\beta}-r\right)/q}{s^2}$$

$$= \frac{\left(R\hat{\beta}-r\right)'\left[s^2 R(X'X)^{-1}R'\right]^{-1}\left(R\hat{\beta}-r\right)}{q} \sim F(q, n-k)$$

# Exercise 2

Consider the population regression model

$$Y = X\beta + \varepsilon.$$

Use the Law of Large Number (LLN) and the Central Limit Theorem (CLT) to show that:

a) as $n \to \infty$, $(\hat{\beta} - \beta) \xrightarrow{p} 0 \iff E(X'\varepsilon) = 0$, namely iff $X$ and $\varepsilon$ are uncorrelated.

b) as $n \to \infty$, $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$, where $Q = plim(\frac{X'X}{n})$.

Recall that if $\{w_i\}$ is an i.i.d. sequence and Var$(w_i)< \infty$. Then

- Law of Large Number (LLN): $\frac{1}{n}\sum\limits_{i=1}^{n} w_i \xrightarrow{p} E(w_i)$
- Central Limit Theorem (CLT): $\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n}(w_i - E(w_i)) \xrightarrow{d} N(0, Var(w_i))$

**Solution**

a) Let us start from the expression for $\hat{\beta}$:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon \\ &= \beta + (X'X)^{-1}X'\varepsilon\end{aligned}$$

Hence,

$$\begin{aligned}\hat{\beta} - \beta &= (X'X)^{-1}X'\varepsilon \\ \hat{\beta} - \beta &= (X'X)^{-1}X'\varepsilon\frac{n}{n} \\ \hat{\beta} - \beta &= \Big(\frac{X'X}{n}\Big)^{-1}\frac{X'\varepsilon}{n}\end{aligned}$$

By the LLN we know that $plim\Big(\frac{X'X}{n}\Big)^{-1} = Q^{-1}$ and $plim\Big(\frac{X'\varepsilon}{n}\Big) = E(X\varepsilon)$. Therefore, $(\hat{\beta} - \beta) \xrightarrow{p} 0$, iff $E(X'\varepsilon) = 0$. In other words, $\hat{\beta}$ is a consistent estimator of $\beta$ if and only if $X$ and $\varepsilon$ are not correlated.

b)

$$\begin{aligned}\hat{\beta} - \beta &= \Big(\frac{X'X}{n}\Big)^{-1}\frac{X'\varepsilon}{n} \\ \hat{\beta} - \beta &= \Big(\frac{X'X}{n}\Big)^{-1}\frac{X'\varepsilon}{\sqrt{n}\sqrt{n}} \\ \sqrt{n}(\hat{\beta} - \beta) &= \Big(\frac{X'X}{n}\Big)^{-1}\frac{X'\varepsilon}{\sqrt{n}}\end{aligned}$$

Note that $plim\Big(\frac{X'X}{n}\Big)^{-1} = Q^{-1}$ by Law of Large Numbers and that $\frac{X'\varepsilon}{\sqrt{n}} \xrightarrow{d} N(0, D)$ by Central Limit Theorem, where $D = \sigma^2 Q$.
We have that $\sqrt{n}(\hat{\beta} - \beta)$ has a finite mean and a finite variance:

$$E\Big[\sqrt{n}(\hat{\beta} - \beta)\Big] = E\Big[\Big(\frac{X'X}{n}\Big)^{-1}\frac{X'\varepsilon}{\sqrt{n}}\Big] = 0$$

$$Var\Big[(\sqrt{n}(\hat\beta-\beta)\Big] = E\left\{\Big[\Big(\frac{X'X}{n}\Big)^{-1}\Big(\frac{X'\varepsilon}{\sqrt{n}}\Big) - \underbrace{E\Big[\Big(\frac{X'X}{n}\Big)^{-1}\frac{X'\varepsilon}{\sqrt{n}}\Big]}_{=0}\right\}\left\{\Big[\Big(\frac{X'X}{n}\Big)^{-1}\Big(\frac{X'\varepsilon}{\sqrt{n}}\Big) - \underbrace{E\Big[\Big(\frac{X'X}{n}\Big)^{-1}\frac{X'\varepsilon}{\sqrt{n}}\Big]}_{=0}\right\}'$$

$$= E\left\{\Big[\Big(\frac{X'X}{n}\Big)^{-1}\frac{X'\varepsilon}{\sqrt{n}}\Big]\Big[\Big(\frac{X'X}{n}\Big)^{-1}\frac{X'\varepsilon}{\sqrt{n}}\Big]'\right\}$$

$$= E\Big[\Big(\frac{X'X}{n}\Big)^{-1}\frac{X'\varepsilon}{\sqrt{n}}\frac{\varepsilon' X}{\sqrt{n}}\Big(\frac{X'X}{n}\Big)^{-1}\Big]$$

$$= Q^{-1}\frac{X'E(\varepsilon\varepsilon')X}{n}Q^{-1}$$

$$= Q^{-1}\sigma^2\underbrace{\frac{X'X}{n}}_{Q}Q^{-1}$$

$$= Q^{-1}\underbrace{\sigma^2 Q}_{D}Q^{-1}$$

Then, we have the following limiting distribution:

$$\sqrt{n}(\hat\beta-\beta) = \Big(\frac{X'X}{n}\Big)^{-1}\frac{X'\varepsilon}{\sqrt{n}} \quad \xrightarrow{d} \quad N(0, Q^{-1}DQ^{-1}) = N(0, Q^{-1}\sigma^2 QQ^{-1}) = N(0, \sigma^2 Q^{-1})$$

# Exercise 3

Consider the classical Gaussian linear model $Y = X\beta + \varepsilon$, with $\varepsilon \sim N_n(0, \sigma^2 I_n)$, and the null hypothesis $R\beta = r$, where $R$ is a $q \times k$ matrix of full row rank.

a) Write down the likelihood ratio (LR) test statistic.

b) Write down the Wald (W) test statistic.

c) Comment on the relation among the above test statistics and the classical F test statistic.

**Solution**
The sample log-likelihood for the classical Gaussian linear model is

$$L(\theta) = \underbrace{-\frac{n}{2}ln2\pi}_{c} - \frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}\underbrace{n^{-1}(Y-X\beta)'(Y-X\beta)}_{S_n(\beta)}, \quad \theta = (\beta, \sigma^2)$$

Let $\tilde\theta = (\tilde\beta, \tilde\sigma^2)$ be the ML estimator of the **constrained** model, that is the model under which the constraints hold. Let $\hat\theta = (\hat\beta, \hat\sigma^2)$ be the ML estimator of the **unconstrained** model.

a) The likelihood ratio (LR) test statistic is:

$$\xi^{LR} = 2(\hat L - \tilde L)$$

where $\hat{L} = L(\hat{\theta})$ is the log-likelihood evaluated at the unconstrained estimator and $\tilde{L} = L(\tilde{\theta})$ is the log-likelihood evaluated at the constrained estimator.
$\hat{\theta}$ is the maximizer over the parameter space of the log-likelihood in the unrestricted model:

$$\hat{\theta} = \underset{\theta \in \ominus}{\text{argmax}} \quad L(\theta)$$

whereas $\tilde{\theta}$ is the maximizer over the parameter space of the log-likelihood in the restricted model:

$$\tilde{\theta} = \underset{\theta \in \ominus}{\text{argmax}} \quad L(\theta) \quad s.t. \quad R\theta = r$$

For a fixed $\beta$, $L(\theta)$ attains its unique maximum at $\sigma^2 = S(\beta)$. Thus $\hat{\sigma}^2 = S(\hat{\beta})$, $\tilde{\sigma}^2 = S(\tilde{\beta})$.

$$\hat{L} = L(\hat{\theta}) = c - \frac{n}{2}\ln\hat{\sigma}^2 - \frac{n\hat{\sigma}^2}{2\hat{\sigma}^2} = \left(c - \frac{n}{2}\right) - \frac{n}{2}\ln\hat{\sigma}^2$$

$$\tilde{L} = L(\tilde{\theta}) = c - \frac{n}{2}\ln\tilde{\sigma}^2 - \frac{n\tilde{\sigma}^2}{2\tilde{\sigma}^2} = \left(c - \frac{n}{2}\right) - \frac{n}{2}\ln\tilde{\sigma}^2$$

Hence, the LR test statistic is:

$$\xi^{LR} = 2\left[\left(c - \frac{n}{2}\right) - \frac{n}{2}\ln\hat{\sigma}^2 - \left(c - \frac{n}{2}\right) + \frac{n}{2}\ln\tilde{\sigma}^2\right] = n\ln\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}$$

The LR principle leads to test that reject the null hypothesis for large values of $\xi^{LR}$.
The LR test requires computing both the unconstrained and the constrained ML estimator.

b) The Wald test only requires computing the unconstrained estimator.
Since $R\hat{\beta} - r \sim N_q(0, \sigma^2 R(X'X)^{-1}R')$ under $H_0$, where the vector $(R\hat{\beta} - r)$ indicates the extent to which the unrestricted ML estimates fit the null hypothesis, the Wald test statistic is:

$$\xi^W = \frac{(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)}{\hat{\sigma}^2}$$

We can rewrite the numerator of the Wald test statistic as the difference between the residuals sum of squares in the constrained model and the residuals sum of squares in the unconstrained model

$$\tilde{e}'\tilde{e} - \hat{e}'\hat{e} = (R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)$$

Then

$$\begin{aligned}\xi^W &= \frac{(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)}{\hat{\sigma}^2} \\ &= \frac{\tilde{e}'\tilde{e} - \hat{e}'\hat{e}}{\hat{\sigma}^2} = \frac{n\tilde{\sigma}^2 - n\hat{\sigma}^2}{\hat{\sigma}^2} \\ &= n\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2}.\end{aligned}$$

Reject the null hypothesis for large values of the test statistic.

c) Recall the F statistic under the classical Gaussian linear model:

$$F = \frac{(R\bar{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\bar{\beta} - r)/q}{s^2} \sim F(q, n-k) \quad \text{under} \quad H_0$$

Rewrite the F statistic as:

$$\begin{aligned} F = \frac{(\tilde{e}'\tilde{e} - \hat{e}'\hat{e})/q}{\hat{e}'\hat{e}/(n-k)} &= \frac{(\tilde{e}'\tilde{e} - \hat{e}'\hat{e})}{\hat{e}'\hat{e}} \frac{(n-k)}{q} \\ &= \frac{n(\tilde{\sigma}^2 - \hat{\sigma}^2)}{n\hat{\sigma}^2} \frac{(n-k)}{q} \\ &= \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \frac{(n-k)}{q} \end{aligned}$$

In order to show the equivalence to the F statistic let us rewrite the expressions for the LR and the Wald tests:

$$\begin{aligned} \xi^{LR} &= n \ln \frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \\ \xi^{W} &= n \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \end{aligned}$$

Thus, we have:

$$\begin{aligned} \xi^{LR} &= n\ ln\Big[\frac{(1 + \frac{qF}{n-k})\hat{\sigma}^2}{\hat{\sigma}^2}\Big] = n\ ln\Big(1 + \frac{qF}{n-k}\Big) \\ \xi^{W} &= n\frac{qF}{(n-k)} \end{aligned}$$

# Exercise 4

Consider the classical Gaussian linear model $Y = X\beta + \varepsilon$, with $\varepsilon \sim N_n(0, \sigma^2 I_n)$, and the null hypothesis $R\beta = r$, where $R$ is a $q \times k$ matrix of full row rank. Show that the Lagrange Multiplier (or Score) test can be written as

a) $\xi^S = NR^2$

where $R^2$ is the determination coefficient of the regression of the restricted residuals $\tilde{e}$ on $X$.

b) $\xi^S = n\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\tilde{\sigma}^2}$

**Solution**
The sample log-likelihood for the classical Gaussian linear model is

$$L(\theta) = \underbrace{-\frac{n}{2} ln 2\pi}_{c} - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\underbrace{(Y-X\beta)'(Y-X\beta)}_{S(\beta)}, \quad \theta = (\beta, \sigma^2)$$

a) The Score o Lagrange Multiplier (LM) principle leads to test that reject for large values of a statistic that is a quadratic form of the score vector, that is the gradient of the log-likelihood. The quadratic form to consider is one that uses as a weighted matrix the inverse of the information matrix.
The test statistic is:

$$\xi^S = \tilde{S}'\tilde{I}^{-1}\tilde{S}$$

where $\tilde{S} = L'(\tilde{\theta})$ and $\tilde{I} = -E[L''(\tilde{\theta})]$ are, respectively, the likelihood score and the (expected) information matrix evaluated at the constrained Maximum Likelihood (ML) estimator.
The Score or LM test only requires computing the constrained ML estimator.
Since $H_0$ leaves $\sigma^2$ unconstrained, we can focus on the derivatives with respect to $\tilde{\beta}$. This is possible since $S$ can be partitioned and the information matrix is block diagonal between $\beta$ and $\sigma^2$ with cross derivatives equal to 0.

$$S_\beta(\tilde{\theta}) = \frac{\partial \tilde{L}}{\partial \tilde{\beta}} = \frac{1}{\tilde{\sigma}^2}X'(Y - X\tilde{\beta})$$

$$\tilde{I}(\tilde{\theta})^{-1} = \begin{bmatrix} \tilde{\sigma}^2(X'X)^{-1} & 0 \\ 0 & \frac{2\tilde{\sigma}^4}{n} \end{bmatrix}$$

How do we obtain the inverse of the information matrix?
First, we compute the second derivatives of the log-likelihood function:

$$\begin{aligned} \frac{\partial^2 L}{\partial\beta\partial\beta'} &= -\frac{1}{\tilde{\sigma}^2}(X'X) \\ \frac{\partial^2 L}{\partial\sigma^2\partial\sigma^2} &= \frac{N}{2\tilde{\sigma}^4} - \frac{1}{\tilde{\sigma}^6}(Y - X\tilde{\beta})'(Y - X\tilde{\beta}) \\ \frac{\partial^2 L}{\partial\beta\partial\sigma^2} &= -\frac{1}{\tilde{\sigma}^4}(X'Y - X'X\tilde{\beta}) \end{aligned}$$

Second, we consider their expected values taken with the opposite sign:

$$\begin{aligned}
E\left(-\frac{\partial^2 L}{\partial\beta\partial\beta'}\right) &= \frac{1}{\tilde{\sigma}^2}(X'X) \\
E\left(-\frac{\partial^2 L}{\partial(\sigma^2)^2}\right) &= -\frac{N}{2\tilde{\sigma}^4} + \frac{1}{\tilde{\sigma}^6}E(\varepsilon'\varepsilon) \\
&= -\frac{N}{2\tilde{\sigma}^4} + \frac{1}{\tilde{\sigma}^6}N\tilde{\sigma}^2 = \frac{N}{2\tilde{\sigma}^4} \\
E\left(-\frac{\partial^2 L}{\partial\beta\partial\sigma^2}\right) &= \frac{1}{\tilde{\sigma}^4}(X'E(y) - X'X\tilde{\beta}) \\
&= \frac{1}{\tilde{\sigma}^4}(X'X\tilde{\beta} - X'X\tilde{\beta}) = 0
\end{aligned}$$

Therefore, the information matrix is:

$$I(\theta) = \begin{bmatrix} \frac{1}{\tilde{\sigma}^2}(X'X) & 0 \\ 0 & \frac{N}{2\tilde{\sigma}^4} \end{bmatrix}$$

Since the information matrix is diagonal, its inverse is given by:

$$I(\theta)^{-1} = \begin{bmatrix} \tilde{\sigma}^2(X'X)^{-1} & 0 \\ 0 & \frac{2\tilde{\sigma}^4}{N} \end{bmatrix}$$

Coming back to the score test statistic, we have:

$$\begin{aligned}
\xi^S = \tilde{S}'_{\tilde{\beta}}\tilde{I}^{-1}_{\tilde{\beta}\tilde{\beta}}\tilde{S}_{\tilde{\beta}} &= \frac{1}{\tilde{\sigma}^2}(Y - X\tilde{\beta})'X\tilde{\sigma}^2(X'X)^{-1}\frac{1}{\tilde{\sigma}^2}X'(Y - X\tilde{\beta}) \\
&= \frac{(Y - X\tilde{\beta})'X(X'X)^{-1}X'(Y - X\tilde{\beta})}{\tilde{\sigma}^2} \\
&= \frac{\tilde{e}X(X'X)^{-1}X'\tilde{e}}{\tilde{\sigma}^2}
\end{aligned}$$

Now, let us consider the following auxiliary regression model

$$\tilde{e} = \gamma X + errors$$

We have that the OLS estimator of $\gamma$ is equal to

$$\hat{\gamma} = (X'X)^{-1}X'\tilde{e}$$

and that the predicted values are

$$\begin{aligned}
\hat{\tilde{e}} &= X\hat{\gamma} \\
&= X(X'X)^{-1}X'\tilde{e}
\end{aligned}$$

The R-squared of the above auxiliary regression is:

$$
\begin{aligned}
R^2 &= \frac{\hat{\tilde{e}}'\hat{\tilde{e}}}{\tilde{e}'\tilde{e}} \\
&= \frac{(X(X'X)^{-1}X'\tilde{e})'(X(X'X)^{-1}X'\tilde{e})}{\tilde{e}'\tilde{e}} \\
&= \frac{\tilde{e}'X(X'X)^{-1}X'X(X'X)^{-1}X'\tilde{e}}{\tilde{e}'\tilde{e}} \\
&= \frac{\tilde{e}'X(X'X)^{-1}X'\tilde{e}}{\tilde{e}'\tilde{e}}
\end{aligned}
$$

We saw above that

$$
\xi^S = \frac{\tilde{e}X(X'X)^{-1}X'\tilde{e}}{\tilde{\sigma}^2}
$$

Given that $\tilde{\sigma}^2 = \frac{\tilde{e}'\tilde{e}}{N}$, we can rewrite the Score test statistics as

$$
\begin{aligned}
\xi^S &= \frac{\tilde{e}X(X'X)^{-1}X'\tilde{e}}{\frac{\tilde{e}'\tilde{e}}{N}} \\
&= N\frac{\tilde{e}X(X'X)^{-1}X'\tilde{e}}{\tilde{e}'\tilde{e}} \\
&= NR^2
\end{aligned}
$$

b) We saw that the Score test statistic is:

$$
\xi^S = \tilde{S}'\tilde{I}^{-1}\tilde{S}
$$

Notice that the quadratic form in the score vector is equivalent to a quadratic form in the Lagrange multiplier associated with the constraint:

$$
\xi^S = \tilde{S}'\tilde{I}^{-1}\tilde{S} = \lambda' R\tilde{I}^{-1}R'\lambda
$$

where $\lambda$ is the Lagrange multiplier associated with the constraint $R\theta = r$. The reason they are equivalent comes from the first order conditions from the constrained maximization problem. If $\tilde{\theta}$ solves this problem, then $\underbrace{L'(\tilde{\theta})}_{\tilde{S}} - R'\lambda = 0$.

Therefore

$$
S_\beta(\tilde{\theta}) = \frac{\partial \tilde{L}}{\partial \tilde{\beta}} = \frac{1}{\tilde{\sigma}^2}X'(Y - X\tilde{\beta}) = R'\lambda
$$

where $\lambda = \frac{1}{\tilde{\sigma}^2}\Big[R(X'X)^{-1}R'\Big]^{-1}(R\hat{\beta} - r)$

By computing $S = R'\lambda = \frac{1}{\tilde{\sigma}^2}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)$, we obtain:

$$\begin{aligned}
\xi^S &= \tilde{S}'_{\hat{\beta}}\tilde{I}^{-1}_{\hat{\beta}}\tilde{S}_{\hat{\beta}} \\
&= \frac{1}{\tilde{\sigma}^2}(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}R \cdot \tilde{\sigma}^2(X'X)^{-1} \cdot \frac{1}{\tilde{\sigma}^2}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \\
&= \frac{1}{\tilde{\sigma}^2}(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}\underbrace{R(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}}_{I_q}(R\hat{\beta} - r) \\
&= \frac{1}{\tilde{\sigma}^2}(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \\
&= n\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\tilde{\sigma}^2}
\end{aligned}$$

# Exercise 5

Consider the model $y = X\beta + \varepsilon$ and assume that $E(\varepsilon) = 0$ and $Var(y) = Var(\varepsilon) = E(\varepsilon\varepsilon') = V$, where $V$ is positive definite.

a) Compute $E(\hat{\beta}_{OLS})$ and $Var(\hat{\beta}_{OLS})$

b) Transform the data in order to restore the ideal conditions of the classical linear model

c) Derive $\hat{\beta}_{GLS}$

d) Compute $E(\hat{\beta}_{GLS})$ and $Var(\hat{\beta}_{GLS})$

e) Prove that $(\hat{\beta}_{GLS})$ is BLUE

**Solution**

a) We know that $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$.

$$\begin{aligned}
E(\hat{\beta}_{OLS}) &= E[(X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon] \\
&= \beta + (X'X)^{-1}X'\underbrace{E(\varepsilon)}_{0} \\
&= \beta
\end{aligned}$$

$$\begin{aligned}
Var(\hat{\beta}_{OLS}) &= E[(\hat{\beta}_{OLS} - E(\beta))(\hat{\beta}_{OLS} - E(\beta))'] \\
&= E[(\hat{\beta}_{OLS} - \beta)(\hat{\beta}_{OLS} - \beta)'] \\
&= E[((X'X)^{-1}X'\varepsilon)((X'X)^{-1}X'\varepsilon)'] \\
&= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} \\
&= (X'X)^{-1}X'VX(X'X)^{-1}
\end{aligned}$$

b) Since $V$ is positive definite, $V^{-1}$ is positive definite too. Therefore, there exists a nonsingular matrix $P$ such that $V^{-1} = P'P$.
Use $P$ to transform the model as follows:

$$Py = PX\beta + P\varepsilon$$

Since

$$E(P\varepsilon) = 0$$

and

$$\begin{aligned} Var(P\varepsilon) &= E(P\varepsilon\varepsilon' P') \\ &= PVP' \\ &= P(P'P)^{-1}P' \\ &= I \end{aligned}$$

then, the transformed model satisfies the conditions of the classical linear model.

c) The GLS estimator for $\beta$ in the model $y = X\beta + \varepsilon$ is the Least Squares estimator for $\beta$ in the model $Py = PX\beta + P\varepsilon$. Hence,

$$\begin{aligned} \hat{\beta}_{GLS} &= [(PX)'(PX)]^{-1}(PX)'Py \\ &= (X'P'PX)^{-1}X'P'Py \\ &= (X'V^{-1}X)^{-1}X'V^{-1}y \end{aligned}$$

d) The expected value of $\hat{\beta}_{GLS}$ is:

$$\begin{aligned} E(\hat{\beta}_{GLS}) &= E[(X'V^{-1}X)^{-1}X'V^{-1}y] \\ &= E[\underbrace{(X'V^{-1}X)^{-1}X'V^{-1}X}_{I}\beta + (X'V^{-1}X)^{-1}X'V^{-1}\varepsilon] \\ &= \beta + (X'V^{-1}X)^{-1}X'V^{-1}\underbrace{E(\varepsilon)}_{0} \\ &= \beta \end{aligned}$$

The variance of $\hat{\beta}_{GLS}$ is:

$$\begin{aligned}
Var(\hat{\beta}_{GLS}) &= E[\hat{\beta}_{GLS} - E(\hat{\beta}_{GLS})][\hat{\beta}_{GLS} - E(\hat{\beta}_{GLS})]' \\
&= E[\hat{\beta}_{GLS} - \beta][\hat{\beta}_{GLS} - \beta]' \\
&= E[((X'V^{-1}X)^{-1}X'V^{-1}\varepsilon)((X'V^{-1}X)^{-1}X'V^{-1}\varepsilon)'] \\
&= (X'V^{-1}X)^{-1}X'V^{-1}E(\varepsilon\varepsilon')V^{-1}X(X'V^{-1}X)^{-1} \\
&= (X'V^{-1}X)^{-1}X'V^{-1}\underbrace{VV^{-1}}_{I}X(X'V^{-1}X)^{-1} \\
&= (X'V^{-1}X)^{-1}\underbrace{X'V^{-1}X(X'V^{-1}X)^{-1}}_{I} \\
&= (X'V^{-1}X)^{-1}
\end{aligned}$$

e) In order to prove that $\hat{\beta}_{GLS}$ is BLUE, let us consider another *linear* estimator, $b$, such that

$$b = [(X'V^{-1}X)^{-1}X'V^{-1} + A]y$$

- **Unbiasedness**

$$\begin{aligned}
E(b) &= [(X'V^{-1}X)^{-1}X'V^{-1} + A]E(y) \\
&= [(X'V^{-1}X)^{-1}X'V^{-1} + A]X\beta \\
&= \underbrace{(X'V^{-1}X)^{-1}X'V^{-1}X}_{I}\beta + AX\beta \\
&= \beta + AX\beta
\end{aligned}$$

  Thus, $b$ is not biased when $AX = 0$.

- **Efficiency**

  We need to show that $Var(\hat{\beta}_{GLS}) \leq Var(b)$

$$\begin{aligned}
Var(b) &= E\{[(X'V^{-1}X)^{-1}X'V^{-1} + A]y - E(b)\}\{[(X'V^{-1}X)^{-1}X'V^{-1} + A]y - E(b)\}' \\
&= E\{[(X'V^{-1}X)^{-1}X'V^{-1} + A](X\beta + \varepsilon) - \beta\}\{[(X'V^{-1}X)^{-1}X'V^{-1} + A](X\beta + \varepsilon) - \beta\}'
\end{aligned}$$

  Notice that

$$\begin{aligned}
[(X'V^{-1}X)^{-1}X'V^{-1} + A](X\beta + \varepsilon) - \beta &= [\underbrace{(X'V^{-1}X)^{-1}X'V^{-1}X}_{I}\beta + (X'V^{-1}X)^{-1}X'V^{-1}\varepsilon + \underbrace{AX}_{0}\beta + A\varepsilon - \beta] \\
&= \beta + (X'V^{-1}X)^{-1}X'V^{-1}\varepsilon + A\varepsilon - \beta \\
&= (X'V^{-1}X)^{-1}X'V^{-1}\varepsilon + A\varepsilon
\end{aligned}$$

Therefore

$$\begin{aligned}
Var(b) &= E\{[(X'V^{-1}X)^{-1}X'V^{-1}\varepsilon + A\varepsilon][(X'V^{-1}X)^{-1}X'V^{-1}\varepsilon + A\varepsilon]'\} \\
&= [(X'V^{-1}X)^{-1}X'V^{-1}\underbrace{E(\varepsilon\varepsilon')}_{V}V^{-1}X(X'V^{-1}X)^{-1} + (X'V^{-1}X)^{-1}X'V^{-1}\underbrace{E(\varepsilon\varepsilon')}_{V}A' + \\
&+ A\underbrace{E(\varepsilon\varepsilon')}_{V}V^{-1}X(X'V^{-1}X)^{-1} + A\underbrace{E(\varepsilon\varepsilon')}_{V}A' \\
&= (X'V^{-1}X)^{-1}\underbrace{(X'V^{-1}X)^{-1}X'V^{-1}X}_{I} + (X'V^{-1}X)^{-1}\underbrace{X'A'}_{0} + \underbrace{AX}_{0}X(X'V^{-1}X)^{-1} + AVA' \\
&= \underbrace{(X'V^{-1}X)^{-1}}_{Var(\hat{\beta}_{GLS})} + AVA'
\end{aligned}$$

Hence,

$$Var(b) - Var(\hat{\beta}_{GLS}) = AVA' > 0$$

since $AVA'$ is positive definite.

# Exercise 6

Consider the model specified in the previous exercise and let $V$ be equal to $\sigma^2\Omega$

a) Derive the asymptotic distribution of $\hat{\beta}_{GLS}$, with $\Omega$ known.

b) Suppose now that $\Omega$ is unknown. Write down the expression for the Feasible GLS (FGLS) estimator, $\hat{\beta}_{FGLS}$. Is it a consistent estimator for $\beta$?

c) Prove that $plim\frac{X'(\hat{\Omega}^{-1}-\Omega^{-1})X}{N} = 0$ and $plim\frac{X'(\hat{\Omega}^{-1}-\Omega^{-1})\varepsilon}{N} = 0$ are **sufficient** conditions for $\hat{\beta}_{GLS}$ and $\hat{\beta}_{FGLS}$ to have the same asymptotic distribution.

**Solution**

a) Let us start from the expression for $\hat{\beta}_{GLS}$ we derived above and recall that now $V = \sigma^2\Omega$

$$\begin{aligned}
\hat{\beta}_{GLS} &= (X'(\sigma^2\Omega)^{-1}X)^{-1}X'(\sigma^2\Omega)^{-1}y \\
&= \Big(\frac{X'\Omega^{-1}X}{\sigma^2}\Big)^{-1}\frac{X'\Omega^{-1}y}{\sigma^2} \\
&= (X'\Omega^{-1}X)^{-1}\sigma^2\frac{X'\Omega^{-1}y}{\sigma^2} \\
&= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \\
&= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}(X\beta + \varepsilon) \\
&= \underbrace{(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X}_{I}\beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon \\
&= \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon
\end{aligned}$$

$$\sqrt{N}(\hat{\beta}_{GLS}-\beta)=\Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1}\frac{X'\Omega^{-1}\varepsilon}{\sqrt{N}}$$

By the Central Limit Theorem, $\frac{X'\Omega^{-1}\varepsilon}{\sqrt{N}} \xrightarrow{d} N(0,D)$

$$E[\sqrt{N}(\hat{\beta}_{GLS}-\beta)]=0$$

$$\begin{aligned}
Var[\sqrt{N}(\hat{\beta}_{GLS}-\beta)] &= E\Big[\Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1}\frac{X'\Omega^{-1}\varepsilon}{\sqrt{N}}\Big]\Big[\Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1}\frac{X'\Omega^{-1}\varepsilon}{\sqrt{N}}\Big]' \\
&= \Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1}\frac{X'\Omega^{-1}E(\varepsilon\varepsilon')\Omega^{-1}X}{N}\Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1} \\
&= \Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1}\frac{X'\Omega^{-1}\sigma^2\Omega\Omega^{-1}X}{N}\Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1} \\
&= \sigma^2\underbrace{\Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1}\frac{X'\Omega^{-1}X}{N}}_{I}\Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1} \\
&= \sigma^2\Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1}
\end{aligned}$$

Thus, $\sqrt{N}(\hat{\beta}_{GLS}-\beta) \xrightarrow{d} N\Big(0,\sigma^2\Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1}\Big)$

b)

$$\hat{\beta}_{FGLS}=(X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y$$

$$\hat{\beta}_{FGLS}=\beta+(X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}\varepsilon$$

$$\hat{\beta}_{FGLS}-\beta=\Big(\frac{X'\hat{\Omega}^{-1}X}{N}\Big)^{-1}\frac{X'\hat{\Omega}^{-1}\varepsilon}{N}$$

$\hat{\beta}_{FGLS}$ is a consistent estimator for $\beta$ if

$$plim\frac{X'\hat{\Omega}^{-1}X}{N}=Q$$

where $Q$ is pd and finite and

$$plim\frac{X'\hat{\Omega}^{-1}\varepsilon}{N}=0$$

c) We have that

$$\sqrt{N}(\hat{\beta}_{GLS}-\beta)-\sqrt{N}(\hat{\beta}_{FGLS}-\beta)=\Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1}\frac{X'\Omega^{-1}\varepsilon}{\sqrt{N}}-\Big(\frac{X'\hat{\Omega}^{-1}X}{N}\Big)^{-1}\frac{X'\hat{\Omega}^{-1}\varepsilon}{\sqrt{N}}$$

$$\sqrt{N}(\hat{\beta}_{GLS} - \hat{\beta}_{FGLS}) = \Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1}\frac{X'\Omega^{-1}\varepsilon}{\sqrt{N}} - \Big(\frac{X'\hat{\Omega}^{-1}X}{N}\Big)^{-1}\frac{X'\hat{\Omega}^{-1}\varepsilon}{\sqrt{N}}$$

Thus,

$$plim\sqrt{N}(\hat{\beta}_{GLS} - \hat{\beta}_{FGLS}) = plim\Bigg[\Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1}\frac{X'\Omega^{-1}\varepsilon}{\sqrt{N}} - \Big(\frac{X'\hat{\Omega}^{-1}X}{N}\Big)^{-1}\frac{X'\hat{\Omega}^{-1}\varepsilon}{\sqrt{N}}\Bigg]$$

If

$$plim\Big(\frac{X'\hat{\Omega}^{-1}X}{N}\Big)^{-1} = plim\Big(\frac{X'\Omega^{-1}X}{N}\Big)^{-1}$$

and

$$plim\frac{X'\hat{\Omega}^{-1}\varepsilon}{N} = plim\frac{X'\Omega^{-1}\varepsilon}{N}$$

then

$plim\sqrt{N}(\hat{\beta}_{GLS} - \hat{\beta}_{FGLS}) = 0 \Rightarrow \hat{\beta}_{GLS}$ and $\hat{\beta}_{FGLS}$ have the same asymptotic distribution.