Q11) a) $\hat{\beta}_1 = (X_1'\Omega^{-1}X_1)^{-1}X_1'\Omega^{-1}y$

$$= (X_1'\Omega^{-1}X_1)^{-1}X_1'\Omega^{-1}[X_1\beta_1 + X_2\beta_2 + \varepsilon]$$

$$= \beta_1 + (X_1'\Omega^{-1}X_1)^{-1}X_1'\Omega^{-1}X_2\beta_2$$

$$+ (X_1'\Omega^{-1}X_1)^{-1}X_1'\Omega^{-1}\varepsilon$$

$$E(\hat{\beta}_1) = \beta_1 + (X_1'\Omega^{-1}X_1)^{-1}X_1'\Omega^{-1}X_2\beta_2$$

So, $E(\hat{\beta}_1) \neq \beta_1$, unless $X_2\beta_2 = 0$ (which means (2) is the right model), or more generally unless $X_1'\Omega^{-1}X_2 = 0$.

b) We have applied GLS to a model that omits some relevant regressors. That is, we have applied GLS with an invalid set of linear restrictions - namely, $\beta_2 = 0$. In the case of OLS, imposing false restrictions reduces the variability of the estimator. GLS is just OLS applied to the transformed data, y^* and X^*, where $y^* = Py$, $X^* = PX$, and $P'P = \Omega^{-1}$. So, the same result will apply here. The variability of the GLS estimator of β_1 will be less in (2) than in (1).

c) We estimate the model:

$$y = X_1\beta_1 + X_2\beta_2 + u$$

$$= (X_1, X_2)\begin{pmatrix}\beta_1 \\ \beta_2\end{pmatrix} + u$$

$$= X\beta + u$$

The GLS estimator will be:

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

$$= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}(X_1\beta_1 + \varepsilon)$$

Now, noting that $X_1 = (X_1, X_2)\begin{pmatrix} I \\ 0\end{pmatrix} = XS$, we have:

$$\hat{\beta}_G = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}(XS\beta_1 + \varepsilon)$$

$$= S\beta_1 + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon$$

$$E(\hat{\beta}_G) = S\beta_1 + 0$$

$$= \begin{pmatrix} I \\ 0\end{pmatrix}\beta_1 = \begin{pmatrix}\beta_1 \\ 0\end{pmatrix}$$

So, $E\begin{bmatrix}\hat{\beta}_{1G} \\ \hat{\beta}_{2G}\end{bmatrix} = \begin{pmatrix}\beta_1 \\ 0\end{pmatrix} = \begin{matrix}\beta_1 \\ \beta_2\end{matrix}$ (since $\beta_2 = 0$).

So, the estimator is unbiased for both β_1 and β_2.

d) They will be consistent. Here we have failed to impose the (valid) restriction that $\beta_2 = 0$. In the OLS case this will lead to a loss of efficiency, relative to what could have been achieved by estimating (3) instead of (4). The same is true here, with GLS, as GLS is just OLS with the transformed data. So, our estimator in this case will be inefficient, both in finite samples and asymptotically.

12. a) $L(\lambda | y) = \dfrac{\lambda^{\Sigma y_i}}{e^{n\lambda} \prod_{i=1}^{n} y_i!}$

b) $\log L = \Sigma y_i \log \lambda - n\lambda - \text{constant}$

$$\frac{\partial \log L}{\partial \lambda} = \frac{\Sigma y_i}{\lambda} - n = 0 \Rightarrow \tilde{\lambda} = \bar{y}.$$

c) Need $E(y_i) = \lambda$.

$$E(\tilde{\lambda}) = E\left(\frac{\Sigma y_i}{n}\right) = \frac{n\lambda}{n} = \lambda.$$

So $\tilde{\lambda}$ is unbiased for λ.

d) In MLE, the variance of the maximum likelihood estimator $\tilde{\theta}$ for θ, is given by:

$$\text{var}(\tilde{\theta}) = -E[H]^{-1}.$$

$$H = \frac{\partial^2 \log L}{\partial \lambda^2} = -\frac{\Sigma y_i}{\lambda^2}.$$

$$-E[H] = \frac{\Sigma E(y_i)}{\lambda^2} = \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda}$$

So, $\text{var}(\tilde{\theta}) = \dfrac{\lambda}{n}$. An estimator for this variance is:

$\widetilde{\text{var}}(\tilde{\theta}) = \dfrac{\tilde{\lambda}}{n}$, where we make use of the <u>invariance</u> property.

e) See pg. 10, lecture 13.

13. a) A linear relationship implies the ~~coefficient~~ coefficient of RY^2 is zero. The p-value for testing this hypothesis is reported as 0.3118. We ~~reje~~ fail to reject the null hypothesis at the 10% significance level. That is, we could use a linear relationship.

For the t-test to be valid we need:

- zero mean for errors
- spherical disturbances
- normality of errors
- non-random X

b) Here we need to test the hypothesis that the coefficient on the dummy variable is zero. The p-value is 0.07. So, we would reject H_0 at the 10% sig. level, but not at the 5% level. Same assumptions needed as above.

c) $$s^2 = \frac{e'e}{n-k} = 0.105660^2 = 0.011164035$$

So, $$e'e = s^2(n-k) = (32-4)s^2 = 0.31259298$$

and $$R^2 = 1 - \frac{e'e}{\hat{var}(y)\times(n-1)} = 1 - \frac{0.31259298}{(0.117046)^2(31)} = 0.26395$$

The model "explains" just over 26% of the sample variability of the dependent variable.

d) The interval is:

This critical value was provided during the exam.

$$(4.98 \times 10^{-5}) \pm (3.79 \times 10^{-5}) \times 2.048$$

$$= [0.0000498 \pm 0.00007762]$$

If we constructed many such intervals in this way, 95% of them would cover the true value of β_2, although there is no guarantee that this particular one does.