University of Glasgow

Data Science and Machine Learning in Finance (ACCFIN 5246)

Problem Set 2 – Spring 2025

Hormoz Ramian

Question 1 The model is:

$$y_i \quad = \quad \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + u_i \tag{1}$$

At each step, state any additional assumption you need to use:

(1.1) Derive the OLS estimators without using vectors/matrix notations.

(1.2) Show that OLS estimator is unbiased.

(1.3) Assume that,

$$\widehat{\beta} = \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' + \lambda \boldsymbol{I}_k \right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{x}_i y_i \right)$$

Discuss the behaviour of this estimator $\widehat{\boldsymbol{\beta}}$ as sample size increases to a very large number $n \to \infty$.

(1.4) Comment on the previous part. In particular, can you think of a case where $\widehat{\beta}$ takes the form above, and what would be the main purpose of such regression?

Question 2 Consider the regression model:

$$y \quad = \quad X\beta_0 + u$$

where y is $T \times 1$, X is $T \times k$ and $rank(X) = k$, β_0 is the $k \times 1$ parameter vector, and $u \sim N(0_T, \sigma_0^2 I_T)$ where σ_0^2 is unknown but a positive constant.

(2.1) Using this result, propose a decision rule to test:

$$\begin{aligned} H_0 &: R\beta_0 = r \\ H_A &: R\beta_0 \neq r \end{aligned}$$

where R and r are respectively a $q \times k$ matrix and a $q \times 1$ vector of constants. Define the test-statistic associated with this hypothesis testing in terms of R, r, etc. What would constitute a Type I error in this context and what is the probability of a Type I error associated with your decision rule?

(2.2) Define the p-value of the test in previous part.

Question 3 The model is:

$$y_i \quad = \quad \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + u_i$$

for $i = 1, \ldots, N$ and we wish to test the null hypothesis: $H_0 : \beta_1 = \beta_2 = 0$.

(3.1) What is the alternative hypothesis? Re-write the regression model, and the null hypothesis in terms of notations used in the lecture (R, r, etc.), indicating the size of each variable. Using the null hypothesis, what are the numerical values for elements in R, r, etc.

(3.2) What is the test statistic and its distribution when the variance of the error term is unknown?

(3.3) Represent elements[1] in $(X'X)^{-1} = \{c_{jk}\}$. What is $[R(X'X)^{-1}R']^{-1}$ in terms of c_{jk} elements?

(3.4) What is the test-statistic in terms of c_{jk} elements?

(3.5) Suppose the test conclusion is to reject the null, comment on this conclusion.

(3.6) Suppose the test conclusion is to fail-to-reject the null, comment on this conclusion.

[1] e.g. $\left(\begin{smallmatrix} c_{11} & c_{12} & \cdots \\ c_{21} & \cdots & \cdots \end{smallmatrix}\right)$ depending on the size of $(X'X)^{-1}$

Question 4 Consider the probability density function, $f(x;\theta) = \lambda e^{-\lambda x}$. Find the MLE of λ and its variance (assuming that the sample is i.i.d.).

Question 5 Consider a simple linear regression model with non-stochastic regressors and $i = 1, \ldots, n$:

$$y_i = \alpha + \beta x_i + u_i \tag{2}$$

$$u_i \sim i.i.d\ \mathcal{N}(0, \sigma^2) \tag{3}$$

(5.1) Define the ML estimator for α and β.
(5.2) Clearly stating any assumption you need, derive the ML estimators for α and β.
(5.3) Is this estimator BLUE?

ANSWERS

Problem 1
(1.1)

$$y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + u_i$$

The objective function is $RSS = \sum_{i\in N} u_i^2$ and the first order conditions (foc) are:

$$\begin{aligned}
\partial RSS/\partial\beta_1\big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} &= -2\sum_{i\in N}\Big(y_i - [\widehat{\beta}_1 + \widehat{\beta}_2 x_{1i} + \widehat{\beta}_3 x_{2i}]\Big)\\
\partial RSS/\partial\beta_2\big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} &= -2\sum_{i\in N} x_{1i}\Big(y_i - [\widehat{\beta}_1 + \widehat{\beta}_2 x_{1i} + \widehat{\beta}_3 x_{2i}]\Big)\\
\partial RSS/\partial\beta_3\big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} &= -2\sum_{i\in N} x_{2i}\Big(y_i - [\widehat{\beta}_1 + \widehat{\beta}_2 x_{1i} + \widehat{\beta}_3 x_{2i}]\Big)
\end{aligned}$$

where $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3)'$. Setting the foc's equal to zero determines the following linear system which determines solution to the OLS problem:

$$\begin{aligned}
0 &= -2\sum_{i\in N}\Big(y_i - [\widehat{\beta}_1 + \widehat{\beta}_2 x_{1i} + \widehat{\beta}_3 x_{2i}]\Big) && (4)\\
0 &= -2\sum_{i\in N} x_{1i}\Big(y_i - [\widehat{\beta}_1 + \widehat{\beta}_2 x_{1i} + \widehat{\beta}_3 x_{2i}]\Big) && (5)\\
0 &= -2\sum_{i\in N} x_{2i}\Big(y_i - [\widehat{\beta}_1 + \widehat{\beta}_2 x_{1i} + \widehat{\beta}_3 x_{2i}]\Big) && (6)
\end{aligned}$$

The first equations gives:

$$\begin{aligned}
\sum_{i\in N} y_i &= \sum_{i\in N}\widehat{\beta}_1 + \sum_{i\in N}\widehat{\beta}_2 x_{1i} + \sum_{i\in N}\widehat{\beta}_3 x_{2i}\\
&= N\widehat{\beta}_1 + \widehat{\beta}_2\sum_{i\in N} x_{1i} + \widehat{\beta}_3\sum_{i\in N} x_{2i}
\end{aligned}$$

Dividing throughout by N and defining $\overline{x}_1 = \sum_{i\in N} x_{1i}/N$ and $\overline{x}_2 = \sum_{i\in N} x_{2i}/N$ yields:

$$\widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2\overline{x}_1 - \widehat{\beta}_3\overline{x}_2$$

which together with equations (5) and (6) gives:

$$\begin{aligned}
0 &= \sum_{i\in N} x_{1i}y_i - \sum_{i\in N} x_{1i}(\overline{Y} - \widehat{\beta}_2\overline{x}_1 - \widehat{\beta}_3\overline{x}_2) + \widehat{\beta}_2\sum_{i\in N} x_{1i}^2 + \widehat{\beta}_3\sum_{i\in N} x_{1i}x_{2i} && (7)\\
0 &= \sum_{i\in N} x_{2i}y_i - \sum_{i\in N} x_{2i}(\overline{Y} - \widehat{\beta}_2\overline{x}_1 - \widehat{\beta}_3\overline{x}_2) + \widehat{\beta}_2\sum_{i\in N} x_{2i}x_{1i} + \widehat{\beta}_3\sum_{i\in N} x_{2i}^2 && (8)
\end{aligned}$$

Re-arranging gives:

$$\widehat{\beta}_2 = \frac{\operatorname{cov}(x_1,y)\operatorname{var}(x_2) - \operatorname{cov}(x_2,y)\operatorname{cov}(x_1,x_2)}{\operatorname{var}(x_1)\operatorname{var}(x_2) - [\operatorname{cov}(x_2,x_1)]^2}$$
$$\widehat{\beta}_3 = \frac{\operatorname{cov}(x_2,y)\operatorname{var}(x_1) - \operatorname{cov}(x_2,y)\operatorname{cov}(x_1,x_2)}{\operatorname{var}(x_1)\operatorname{var}(x_2) - [\operatorname{cov}(x_1,x_2)]^2}$$

There is another intuitive way to approach this question. To start, we need to know that the regression line goes through the sample mean $(\overline{x}_1, \overline{x}_2, \overline{y})$. This is easy to verify, by substituting values for $(\overline{x}_1, \overline{x}_2, \overline{y})$ into *any* regression line and show that it satisfies the equality. This is a helpful lemma because we can now run the following regression:

$$\tilde{y}_i = \beta_2\tilde{x}_{1i} + \beta_3\tilde{x}_{2i} + u_i$$

where $\tilde{x}_{1i} = x_{1i} - \overline{x}_1$, $\tilde{x}_{2i} = x_{2i} - \overline{x}_2$ and $\tilde{y}_i = y_i - \overline{y}$. In fact, instead of regressing y_i on an intercept, x_1 and x_2, we regress $y_i - \overline{y}$ on $x_{1i} - \overline{x}_1$ and $x_{2i} - \overline{x}_2$ (with no intercept) which is to say, we de-mean the data (subtract average values of each variable) before running the regression. As a result, this regression goes through the origin, and has no intercept which helps us to reduce one parameter from the model and solve the foc system only for two parameters. This transformation leaves the slope parameters intact:

$$\begin{aligned} \partial RSS/\partial\beta_2\big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} &= -2\sum_{i\in N}\tilde{x}_{1i}\left(\tilde{y}_i + \widehat{\beta}_2\tilde{x}_{1i} + \widehat{\beta}_3\tilde{x}_{2i}\right) \\ \partial RSS/\partial\beta_3\big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} &= -2\sum_{i\in N}\tilde{x}_{2i}\left(\tilde{y}_i + \widehat{\beta}_2\tilde{x}_{1i} + \widehat{\beta}_3\tilde{x}_{2i}\right) \end{aligned}$$

Re-arrange:

$$\begin{aligned} 0 &= \sum_{i\in N}\tilde{x}_{1i}\tilde{y}_i + \widehat{\beta}_2\sum_{i\in N}\tilde{x}_{1i}^2 + \widehat{\beta}_3\sum_{i\in N}\tilde{x}_{1i}\tilde{x}_{2i} \\ 0 &= \sum_{i\in N}\tilde{x}_{2i}\tilde{y}_i + \widehat{\beta}_2\sum_{i\in N}\tilde{x}_{2i}\tilde{x}_{1i} + \widehat{\beta}_3\sum_{i\in N}\tilde{x}_{2i}^2 \end{aligned}$$

Re-arrange:

$$\begin{aligned} 0 &= \operatorname{cov}(\tilde{x}_1,\tilde{y}) + \widehat{\beta}_2\operatorname{var}(\tilde{x}_1) + \widehat{\beta}_3\operatorname{cov}(\tilde{x}_1,\tilde{x}_2) \\ 0 &= \operatorname{cov}(\tilde{x}_2,\tilde{y}) + \widehat{\beta}_2\operatorname{cov}(\tilde{x}_2,\tilde{x}_1) + \widehat{\beta}_3\operatorname{var}(\tilde{x}_2) \end{aligned}$$

Re-write in terms of $\widehat{\beta}_2$:

$$\begin{aligned} \widehat{\beta}_2 &= -(\operatorname{cov}(\tilde{x}_1,\tilde{y}) + \widehat{\beta}_3\operatorname{cov}(\tilde{x}_1,\tilde{x}_{2i}))/\operatorname{var}(\tilde{x}_1) \\ \widehat{\beta}_2 &= -(\operatorname{cov}(\tilde{x}_2,\tilde{y}) + \widehat{\beta}_3\operatorname{var}(\tilde{x}_2))/\operatorname{cov}(\tilde{x}_2,\tilde{x}_1) \end{aligned}$$

Equate the right hand sides:

$$\frac{\operatorname{cov}(\tilde{x}_1,\tilde{y}) + \widehat{\beta}_3\operatorname{cov}(\tilde{x}_1,\tilde{x}_2)}{\operatorname{var}(\tilde{x}_1)} = \frac{\operatorname{cov}(\tilde{x}_2,\tilde{y}) + \widehat{\beta}_3\operatorname{var}(\tilde{x}_2)}{\operatorname{cov}(\tilde{x}_2,\tilde{x}_1)}$$

Therefore:

$$\operatorname{cov}(\tilde{x}_2,\tilde{x}_1)\operatorname{cov}(\tilde{x}_1,\tilde{y}) + \widehat{\beta}_3[\operatorname{cov}(\tilde{x}_2,\tilde{x}_1)]^2 = \operatorname{var}(\tilde{x}_1)\operatorname{cov}(\tilde{x}_2,\tilde{y}) + \widehat{\beta}_3\operatorname{var}(\tilde{x}_1)\operatorname{var}(\tilde{x}_2)$$

Therefore:

$$\begin{aligned} \widehat{\beta}_2 &= \frac{\operatorname{var}(\tilde{x}_2)\operatorname{cov}(\tilde{x}_1,\tilde{y}) - \operatorname{cov}(\tilde{x}_1,\tilde{x}_2)\operatorname{cov}(\tilde{x}_2,\tilde{y})}{\operatorname{var}(\tilde{x}_2)\operatorname{var}(\tilde{x}_1) - [\operatorname{cov}(\tilde{x}_2,\tilde{x}_1)]^2} \\ \widehat{\beta}_3 &= \frac{\operatorname{var}(\tilde{x}_1)\operatorname{cov}(\tilde{x}_2,\tilde{y}) - \operatorname{cov}(\tilde{x}_2,\tilde{x}_1)\operatorname{cov}(\tilde{x}_1,\tilde{y})}{\operatorname{var}(\tilde{x}_1)\operatorname{var}(\tilde{x}_2) - [\operatorname{cov}(\tilde{x}_2,\tilde{x}_1)]^2} \end{aligned}$$

which are the slope estimators. We need to complete one more step, which is to find the original regression intercept. However, once we have $\widehat{\beta}_2$ and $\widehat{\beta}_3$ then we can use them to uncover the intercept: $\widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2\overline{x}_1 - \widehat{\beta}_3\overline{x}_2$.

(1.2) We use the population model $\tilde{y}_i = \beta_2\tilde{x}_{1i} + \beta_3\tilde{x}_{2i} + u_i$ together with the following:

$$\begin{aligned} \mathbb{E}\widehat{\beta}_2 &= \frac{\operatorname{var}(\tilde{x}_2)\operatorname{cov}(\tilde{x}_1,\tilde{y}) - \operatorname{cov}(\tilde{x}_1,\tilde{x}_2)\operatorname{cov}(\tilde{x}_2,\tilde{y})}{\operatorname{var}(\tilde{x}_2)\operatorname{var}(\tilde{x}_1) - [\operatorname{cov}(\tilde{x}_2,\tilde{x}_1)]^2} \\ \mathbb{E}\widehat{\beta}_3 &= \frac{\operatorname{var}(\tilde{x}_1)\operatorname{cov}(\tilde{x}_2,\tilde{y}) - \operatorname{cov}(\tilde{x}_2,\tilde{x}_1)\operatorname{cov}(\tilde{x}_1,\tilde{y})}{\operatorname{var}(\tilde{x}_1)\operatorname{var}(\tilde{x}_2) - [\operatorname{cov}(\tilde{x}_2,\tilde{x}_1)]^2} \end{aligned}$$

Noting that by (strict) exogeneity assumption $\mathrm{cov}(\tilde{x}_1, u) = 0$ and $\mathrm{cov}(\tilde{x}_2, u) = 0$ then:

$$\begin{aligned}
\mathbb{E}\widehat{\beta}_2 &= \frac{\mathrm{var}(\tilde{x}_2)\,\mathrm{cov}(\tilde{x}_1, \beta_2\tilde{x}_1 + \beta_3\tilde{x}_2 + u) - \mathrm{cov}(\tilde{x}_1, \tilde{x}_2)\,\mathrm{cov}(\tilde{x}_2, \beta_2\tilde{x}_1 + \beta_3\tilde{x}_2 + u)}{\mathrm{var}(\tilde{x}_2)\,\mathrm{var}(\tilde{x}_1) - [\mathrm{cov}(\tilde{x}_2, \tilde{x}_1)]^2} \\
&= \frac{\beta_2\,\mathrm{var}(\tilde{x}_2)\,\mathrm{var}(\tilde{x}_1) + \beta_3\,\mathrm{var}(\tilde{x}_2)\,\mathrm{cov}(\tilde{x}_1, \tilde{x}_2) - \beta_2[\mathrm{cov}(\tilde{x}_1, \tilde{x}_2)]^2 - \beta_3\,\mathrm{cov}(\tilde{x}_1, \tilde{x}_2)\,\mathrm{var}(\tilde{x}_2)}{\mathrm{var}(\tilde{x}_2)\,\mathrm{var}(\tilde{x}_1) - [\mathrm{cov}(\tilde{x}_2, \tilde{x}_1)]^2} \\
&= \frac{\beta_2\{\mathrm{var}(\tilde{x}_2)\,\mathrm{var}(\tilde{x}_1) - [\mathrm{cov}(\tilde{x}_1, \tilde{x}_2)]^2\} + \beta_3\{\mathrm{var}(\tilde{x}_2)\,\mathrm{cov}(\tilde{x}_1, \tilde{x}_2) - \mathrm{cov}(\tilde{x}_1, \tilde{x}_2)\,\mathrm{var}(\tilde{x}_2)\}}{\mathrm{var}(\tilde{x}_2)\,\mathrm{var}(\tilde{x}_1) - [\mathrm{cov}(\tilde{x}_2, \tilde{x}_1)]^2} \\
&= \beta_2
\end{aligned}$$

The similar derivation holds for $\mathbb{E}\widehat{\beta}_3 = \beta_3$.

(1.3) Assuming the $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{y}$ $(n \times 1)$ and $\boldsymbol{X}$ $(n \times k)$ and the property that $\mathbb{E}(\boldsymbol{x}_i e_i) = 0$. The ridge regression estimator:

$$\widehat{\beta} = \left(\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' + \lambda\boldsymbol{I}_k\right)^{-1}\left(\sum_{i=1}^{n} \boldsymbol{x}_i y_i\right)$$

Find probability limit of $\widehat{\boldsymbol{\beta}}$ as $n \to \infty$

$$\begin{aligned}
\widehat{\beta} &= \left(\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' + \lambda\boldsymbol{I}_k\right)^{-1}\left(\sum_{i=1}^{n} \boldsymbol{x}_i y_i\right) \\
&= \left(\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' + \lambda\boldsymbol{I}_k\right)^{-1}\left(\sum_{i=1}^{n} \boldsymbol{x}_i(\boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{e})\right) \\
&= \left(\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' + \lambda\boldsymbol{I}_k\right)^{-1}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{\beta} + \left(\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' + \lambda\boldsymbol{I}_k\right)^{-1}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{e}
\end{aligned}$$

Dividing and multiplying by $\dfrac{1}{n}$ yields:

$$\widehat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' + \frac{1}{n}\lambda\boldsymbol{I}_k\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\right)\boldsymbol{\beta} + \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' + \frac{1}{n}\lambda\boldsymbol{I}_k\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{e}\right)$$

Let $\widehat{\boldsymbol{Q}}_{xx\lambda}^{-1} = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' + \frac{1}{n}\lambda\boldsymbol{I}_k\right)^{-1}$, $\widehat{\boldsymbol{Q}}_{xx} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'$ and $\widehat{\boldsymbol{Q}}_{xe} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{e}$. First, because $\mathbb{E}(\boldsymbol{x}_i e_i) = 0$, we can show that:

$$\widehat{\boldsymbol{Q}}_{xe} \xrightarrow{p} \boldsymbol{0}$$

therefore, we can simplify the following as:

$$\widehat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' + \frac{1}{n}\lambda\boldsymbol{I}_k\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\right)\boldsymbol{\beta} + \boldsymbol{0}$$

In the first term, we can show that asymptotically $\frac{1}{n}\lambda\boldsymbol{I}_k$ disappears and we have:

$$\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' + \frac{1}{n}\lambda\boldsymbol{I}_k\right)^{-1} \xrightarrow{p} \boldsymbol{Q}_{xx}^{-1} \text{ because } \frac{1}{n}\lambda\boldsymbol{I}_k \to \boldsymbol{0}_k \text{ as } n \to \infty$$

Re-writing the expression for $\widehat{\beta}$ yields,

$$\begin{aligned}
\widehat{\beta} &\xrightarrow{p} \boldsymbol{Q}_{xx}^{-1}\boldsymbol{Q}_{xx}\boldsymbol{\beta} \\
&\xrightarrow{p} \boldsymbol{\beta}
\end{aligned}$$

which shows that the ridge regression estimator is a consistent estimator for $\boldsymbol{\beta}$. Intuitively, this result holds since the impact of particular form of constraint on the regression model, becomes less important as the sample size increases. Adding the term $\lambda \boldsymbol{I}_k$ is particularly important when the sample size is small which enables the inverse form to be well-conditioned.

(1.4) The main reason to introduce this additional term is that the inverse form $\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1}$ only exists if $\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)$ is positive-definite and full rank. However, even if we have a full rank matrix, still this may be ill-conditioned, indicating that the number observations are very close to the number of explanatory variables. In this case the usual OLS estimator is poorly estimated. In particular, confidence intervals are very large and estimates largely depend on individual observation perturbation. The ridge regression estimator overcomes this issue via introducing the additional term. As we see in the derivations above the impact fades as sample size increases.

Question 2

(2.1) The test-statistic is:

$$F\text{-statistic} \quad = \quad \mathcal{F} \quad = \quad \frac{(R\widehat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - r)}{q\widehat{\sigma}^2}$$

when the variance term in the denominator is unknown[2] then $\widehat{\sigma}^2 = \text{SSE}/(T-k)$ and $\widehat{\beta} = (X'X)^{-1}X'y$ and the decision rule is: Reject $H_0 : R\beta_0 = r$ at $100\alpha\%$ significance level if,

$$\mathcal{F} > \text{critical-value}[F_{q,T-k}(1-\alpha)] \tag{9}$$

where $F_{q,T-k}(1-\alpha)$ is the $100(1-\alpha)$th percentile of the F distribution with q and $T-k$ degrees of freedom. A Type I error occurs if $\mathcal{F} >$ critical-value$[F_{q,T-k}(1-\alpha)]$ but H_0 is true, that is, we reject $R\beta_0 = r$ when it actually holds, given the confidence level $1-\alpha$. The probability of type I error is α.

(2.2) The associated probability $1 - F_{q,T-k}(F\text{-statistic})$ is the p-value: tail area under the distribution defined by the test statistic.

Question 3

(3.1) $H_A : \beta_1 \neq 0$ or $\beta_2 \neq 0$. The alternative should include all possible cases that are complement to the null.

$$\underset{N\times 1}{\boldsymbol{Y}} = \underset{N\times 3}{\boldsymbol{X}}\underset{3\times 1}{\boldsymbol{\beta}} + \underset{N\times 1}{\boldsymbol{u}}$$

$$H_0 : \underset{2\times 3}{\boldsymbol{R}}\underset{3\times 1}{\boldsymbol{\beta}} = \underset{2\times 1}{\boldsymbol{r}}$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

(3.2) With two linear restrictions, the statistic $\mathcal{F}$, compares to:

$$\mathcal{F} \quad = \quad \frac{\left[(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r})\right]}{2s^2} \sim F_{2,N-3}(1-\alpha)$$

[2]When σ^2 is known, we use the Wald-statistic: $W = (R\widehat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - r)/(q\sigma^2) \sim \chi^2_q$. The distribution changes to χ^2_q because the only random variable is the quadratic form of $\widehat{\beta}$. When σ^2 is unknown, we need to estimate it and we have an additional χ^2-distributed random variable in the denominator $\widehat{\sigma}^2$ instead of σ^2.

(3.3)

$$\begin{aligned}
\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}' &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}' = \begin{bmatrix} c_{22} & c_{23} \\ c_{32} & c_{33} \end{bmatrix} \\
\left(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right)^{-1} &= \frac{1}{c_{22}c_{33} - c_{23}c_{32}} \begin{bmatrix} c_{33} & -c_{32} \\ -c_{23} & c_{22} \end{bmatrix} = \frac{1}{c_{22}c_{33} - c_{23}^2} \begin{bmatrix} c_{33} & -c_{23} \\ -c_{23} & c_{22} \end{bmatrix}
\end{aligned}$$

where in the last part, we use the symmetry property of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ to simply $c_{32} = c_{23}$

(3.4)

$$\mathcal{F} = \frac{\left[(\widehat{\beta}_2, \widehat{\beta}_3)'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\widehat{\beta}_2, \widehat{\beta}_3)\right]/2}{s^2} = \frac{\widehat{\beta}_2^2 c_{33} - 2\widehat{\beta}_2\widehat{\beta}_3 c_{23} + \widehat{\beta}_3^2 c_{22}}{2s^2(c_{22}c_{33} - c_{23}^2)} \sim F_{2,N-3}(1-\alpha)$$

(3.5) The rejection region, is the interval under distribution $F_{2,N-3}(1-\alpha)$ tail such that the area under the curve is equal to α, pinning down a specific critical value c. The test conclusion is to reject the null if statistic $\mathcal{F}$ is greater than the critical value:

$$F\text{-statistic} > c$$

Or, to alternatively say that p-value(F-statistic) $< \alpha$. The test rejects the joint hypothesis in the null, at 5% significance level: parameters of interest are simultaneously distinguishable from zero.

(3.6) There is inadequate evidence to reject the null: when testing both parameters, we are unable to statistically distinguish at least one of them from zero, at the 5% significance level.

Question 4 The likelihood function $\mathcal{L}(\lambda; x)$, under the independence assumption, is:

$$\begin{aligned}
\mathcal{L}(\lambda; x) &= f(x_1, x_2, \ldots, x_T; \lambda) \\
&= f(x_1; \lambda) \times f(x_1; \lambda) \times, \ldots, \times f(x_T; \lambda) = \prod_{i=1}^{T} f(x_i; \lambda)
\end{aligned}$$

Taking logs gives the log-likelihood function $\ell(\lambda; x)$:

$$\begin{aligned}
\ell(\lambda; x) &= \log \prod_{i=1}^{T} f(x_i; \lambda) = \sum_{i=1}^{T} \log f(x_i; \lambda) \\
&= \sum_{i=1}^{T} \log \lambda e^{-\lambda x} = \sum_{i=1}^{T} \log \lambda - \sum_{i=1}^{T} \lambda x_i \\
&= T \log \lambda - \lambda \sum_{i=1}^{T} x_i
\end{aligned}$$

Taking derivative with respect to the parameter give:

$$0 = \left.\frac{\partial \ell(\lambda; x)}{\partial \lambda}\right|_{\lambda = \widehat{\lambda}_{ML}} = \frac{T}{\widehat{\lambda}_{ML}} - \sum_{i=i}^{T} x_i$$

giving the ML estimator,

$$\widehat{\lambda}_{ML} = \left(\frac{1}{T}\sum_{i=1}^{T} x_i\right)^{-1} \tag{10}$$

To derive the asymptotic variance, we form the second derivative of the log-likelihood function with respect to λ to obtain the hessian. Then the Fisher's information, its large sample counterpart and asymptotic variance of the ML estimator are:

$$\begin{aligned}\mathcal{H} &= -\frac{T}{\lambda^2} \\ I &= -\mathbb{E}\mathcal{H} \;=\; \frac{T}{\lambda^2} \\ \mathcal{I} &= \frac{1}{T}\text{plim}\frac{T}{\lambda^2} \;=\; \frac{1}{\lambda^2} \\ \text{AsyVar}(\widehat{\lambda}) &= \frac{1}{\mathcal{I}} \;=\; \lambda^2\end{aligned}$$

Question 5

(5.1) We wish to maximize the joint density of random variables given the parameters, which under independence, simplifies to the following product:

$$\mathcal{L} \;=\; \max_{\boldsymbol{\theta}\in\Theta} \prod_{i=1}^{n} f(y_i;\boldsymbol{\theta}) \tag{11}$$

The ML estimator $\widehat{\boldsymbol{\theta}}_{ML}$ or for short $\widehat{\boldsymbol{\theta}}$ is:

$$\widehat{\boldsymbol{\theta}} \;=\; \arg\max_{\boldsymbol{\theta}\in\Theta} \mathcal{L}(\boldsymbol{\theta}) \tag{12}$$

where $\Theta \subset \mathbb{R}^2$ is the two-dimensional parameter space, where each parameter belongs to the real line. Moreover, since logarithmic transformation is a monotonic transformation, which retains the parameter estimates unchanged, we can re-write the previous equation as:

$$\begin{aligned}\widehat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta}\in\Theta} \mathcal{L}(\boldsymbol{\theta}) \\ &= \arg\max_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}) \qquad (13)\end{aligned}$$

(5.2) The probability density function for Normal distribution is:

$$\widehat{\boldsymbol{\theta}} \;=\; \arg\max_{\boldsymbol{\theta}\in\Theta} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(y_i-\alpha-\beta x_i)^2}{\sigma^2}\right\} \tag{14}$$

$$= \arg\max_{\boldsymbol{\theta}\in\Theta} \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} \prod_{i=1}^{n} \exp\left\{-\frac{1}{2}\frac{(y_i-\alpha-\beta x_i)^2}{\sigma^2}\right\} \tag{15}$$

$$= \arg\max_{\boldsymbol{\theta}\in\Theta} -\frac{n}{2}\log(2\pi\sigma^2) + \sum_{i=i}^{n} \log\left(\exp\left\{-\frac{1}{2}\frac{(y_i-\alpha-\beta x_i)^2}{\sigma^2}\right\}\right) \tag{16}$$

Because $\frac{n}{2}\log(2\pi)$ in the first term is a constant and can be separated from the rest of the objective function, we can drop it from the optimization. In other words, $\frac{n}{2}\log(2\pi)$ is only a constant shift and in inconsequential to the value of $\widehat{\boldsymbol{\theta}}$:

$$\widehat{\boldsymbol{\theta}} \;=\; \arg\max_{\boldsymbol{\theta}\in\Theta} -\cancel{\frac{n}{2}\log(2\pi)} - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2 \tag{17}$$

Furthermore, since the question has give the true value of σ^2, we do not estimate it and as a result the objective function reduces to:

$$\widehat{\boldsymbol{\theta}} \;=\; \arg\max_{\boldsymbol{\theta}\in\Theta} -\cancel{\frac{n}{2}\log(\sigma^2)} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2 \tag{18}$$

$$= \arg\max_{\boldsymbol{\theta}\in\Theta} -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2 \tag{19}$$

Differentiating with respect to each parameter gives the first order conditions, or the score functions:

$$\begin{aligned}
\left.\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} &= -\frac{1}{2\sigma^2}\frac{\partial}{\partial \alpha}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2 \\
\left.\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} &= -\frac{1}{2\sigma^2}\frac{\partial}{\partial \beta}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2
\end{aligned}$$

Setting the equations equal to zero:

$$0 = \sum_{i=1}^{n}(y_i-\widehat{\alpha}-\widehat{\beta}x_i) \tag{20}$$

$$0 = \sum_{i=1}^{n}x_i(y_i-\widehat{\alpha}-\widehat{\beta}x_i) \tag{21}$$

Solving the system above gives:

$$\widehat{\alpha} = \overline{y}-\widehat{\beta}\overline{x} \tag{22}$$

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2} \tag{23}$$

The estimates are identical to those from the OLS.

(5.3) We start by differentiating the objective function twice, or differentiating the first order conditions once, with respect to the parameter to construct the hessian matrix $\mathcal{H}(\boldsymbol{\theta}) = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$:

$$\begin{aligned}
\mathcal{H}(\boldsymbol{\theta}) &= \begin{bmatrix} \partial^2\ell(\boldsymbol{\theta})/\partial\alpha^2 & \partial^2\ell(\boldsymbol{\theta})/\partial\alpha\partial\beta \\ \partial^2\ell(\boldsymbol{\theta})/\partial\beta\partial\alpha & \partial^2\ell(\boldsymbol{\theta})/\partial\beta^2 \end{bmatrix} \\
&= \begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i \\ -\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i & -\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i^2 \end{bmatrix}
\end{aligned} \tag{24}$$

The information matrix $I(\boldsymbol{\theta})$ is defined as $I(\boldsymbol{\theta}) = -\mathbb{E}(\mathcal{H})$ which can be simplified to $I(\boldsymbol{\theta}) = -\mathcal{H}$ under non-stochastic regressors assumption. Therefore,

$$\begin{aligned}
I(\boldsymbol{\theta}) &= -\mathbb{E}(\mathcal{H}) \\
&= \begin{bmatrix} \frac{n}{\sigma^2} & \frac{1}{\sigma^2}\sum_{i=1}^{n}x_i \\ \frac{1}{\sigma^2}\sum_{i=1}^{n}x_i & \frac{1}{\sigma^2}\sum_{i=1}^{n}x_i^2 \end{bmatrix} \\
&= \frac{n}{\sigma^2}\begin{bmatrix} 1 & \frac{1}{n}\sum_{i=1}^{n}x_i \\ \frac{1}{n}\sum_{i=1}^{n}x_i & \frac{1}{n}\sum_{i=1}^{n}x_i^2 \end{bmatrix}
\end{aligned}$$

then:

$$\begin{aligned}
\mathcal{I}(\boldsymbol{\theta}) &= \text{plim}_{n\to\infty}\frac{1}{n}I(\boldsymbol{\theta}) \\
&= \frac{1}{\sigma^2}\begin{bmatrix} 1 & \frac{1}{n}\sum_{i=1}^{n}x_i \\ \frac{1}{n}\sum_{i=1}^{n}x_i & \frac{1}{n}\sum_{i=1}^{n}x_i^2 \end{bmatrix} \\
&= \frac{1}{\sigma^2}\frac{\boldsymbol{X}'\boldsymbol{X}}{n} = \frac{1}{\sigma^2}Q_{\boldsymbol{X}}
\end{aligned}$$

where $Q_{\boldsymbol{X}} = \frac{\boldsymbol{X}'\boldsymbol{X}}{n}$ and we have $\mathcal{I}^{-1}(\boldsymbol{\theta}) = \sigma^2 Q_{\boldsymbol{X}}^{-1}$. We conclude that the asymptotic covariance matrix of the MLE coincides with the asymptotic covariance matrix of the OLS estimators (The correlation between parameters is only equal to zero iff $\sum_{i=1}^{n}x_i = 0$). The ML estimators obtained in equations (22) and (23), $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha},\widehat{\beta})'$ are identical to those of the OLS estimation and therefore unbiased $\mathbb{E}\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}$ also note that expressing the estimator as an average:

$$\widehat{\beta} \;=\; \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2} \;=\; \beta + \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})u_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2} \tag{25}$$

noting that

$$\frac{1}{n}\sum_{i=1}^{n} x_i u_i \;\xrightarrow{p}\; 0$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i \;\xrightarrow{d}\; \mathcal{N}(0,\sigma^2 Q_{xx})$$

where $Q_{xx} = \text{plim}\frac{1}{n}\sum_{i=1}^{n} x_i^2$ because, under i.i.d. assumption:

$$\begin{aligned}\text{var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(x_i-\overline{x})u_i\right] \;&=\; \frac{1}{n}\text{var}\left[\sum_{i=1}^{n}(x_i-\overline{x})u_i\right] \;\overset{\text{i.i.d.}}{=}\; \frac{1}{n}\sum_{i=1}^{n}\text{var}((x_i-\overline{x})u_i) \\ &=\; \sigma^2\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\right)\end{aligned} \tag{26}$$

re-arranging equation (25) gives:

$$\sqrt{n}(\widehat{\beta}-\beta) \;=\; \left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(x_i-\overline{x})u_i \tag{27}$$

now taking limit of the random variable $\sqrt{n}(\widehat{\beta}-\beta)$, as the sample size goes to infinity $\lim_{n\to\infty}\sqrt{n}(\widehat{\beta}-\beta)$ yields the following aymptotic results:

$$\sqrt{n}(\widehat{\beta}-\beta) \;\xrightarrow{d}\; \mathcal{N}(0,\mathcal{I}^{-1}) \tag{28}$$

first, we use normal distribution by assuming the Central Limit Theorem as the estimator can be constructed as an average (under some regularity conditions, mainly finite variance, the averages converge in distribution to a normal distribution), second, using the unbiasedness property, the limiting distribution is centered at zero, and third the asymptotic variance is determined by the results in equation (26). repeating for $\widehat{\alpha}$, and collecting both parameter in $\boldsymbol{\theta} = \begin{bmatrix}\alpha & \beta\end{bmatrix}'$ then,

$$\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta} \tag{29}$$

and then,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \;\xrightarrow{d}\; \mathcal{N}(\mathbf{0},\mathcal{I}^{-1}) \tag{30}$$

shows asymptotic normality of the ML estimator.