PhD Econometrics 1: Study Questions Week 2
Imperial College London

Hormoz Ramian

Solutions

Question 1

(1.1) No, hypothesis testing and its conclusion are only concerned with the alternative hypothesis which is the interest of statistical inference. In general, failure to reject the null does not provide evidence that the null is actually statistically (given a confidence level) true.

(1.2) No, in general, a matrix is a projection matrix if it is an idempotent matrix but not the reverse, as projection matrix is only a special case of idempotency. If an idempotent matrix is symmetric, then it is also a projection matrix. A projection matrix is also known as an orthogonal projection matrix. In econometrics, we mostly work with orthogonal projection matrices and the term *orthogonal* is often dropped. However, in a more general setting we also have a definition for non-orthogonal projection matrices. Both orthogonal and non-orthogonal projection matrices are idempotent, but an orthogonal projection matrix is also necessarily *symmetric*:

$$\boldsymbol{P} = \boldsymbol{P}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{P}' = \boldsymbol{P}'\boldsymbol{P}$$

A non-orthogonal projection matrix is not necessarily symmetric.

(1.3) No, invertibility and idempotency are two separate properties. A matrix can be idempotent but not full rank.

(1.4) No, invertibility and projection properties need not coincide. A matrix can be projection but not full rank.

(1.5) Both statements are equivalent in terms of measurements. The question wishes to establish that 'percentage change' and 'percentage points change' are different ways of measuring changes (in this part, the text measures changes correctly using both methods). Naturally, each group reported the measure that made its position sound most favorable to non-econometricians!

Question 2

(2.1) To estimate the parameter $\theta = \alpha\gamma$,

$$y_i \quad = \quad \alpha + \gamma x_i + u_i$$

and obtain OLS estimates $\widehat{\alpha}$ and $\widehat{\gamma}$ that are unbiased and efficient together with $\text{cov}([\widehat{\alpha}\ \widehat{\gamma}]') = \sigma^2([\mathbf{1}\ \boldsymbol{x}]'[\mathbf{1}\ \boldsymbol{x}])^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

(2.2) Denote $\boldsymbol{\beta} = (\alpha \quad \gamma)'$ that obtained by using OLS. In order to implement inference, we need to use the delta method to obtain correct standard error for $\widehat{\theta} = \widehat{\alpha}\widehat{\gamma}$ or simply $\theta \equiv f(\boldsymbol{\beta}) = \alpha\gamma$:

$$\sqrt{n}(f(\widehat{\boldsymbol{\beta}}) - f(\boldsymbol{\beta})) \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial f}{\partial \boldsymbol{\beta}}\sigma^2\left(\frac{\boldsymbol{X}'\boldsymbol{X}}{n}\right)^{-1}\frac{\partial f}{\partial \boldsymbol{\beta}'}\right) \tag{1}$$

where $\frac{\partial f}{\partial \boldsymbol{\beta}} = [\frac{\partial f}{\partial \alpha}\ \frac{\partial f}{\partial \gamma}] = [\gamma\ \alpha]$, then:

$$\sqrt{n}(f(\widehat{\boldsymbol{\beta}}) - f(\boldsymbol{\beta})) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2[\gamma\ \alpha]\left(\frac{\boldsymbol{X}'\boldsymbol{X}}{n}\right)^{-1}[\gamma\ \alpha]'\right) \tag{2}$$

where the asymptotic variance of $\widehat{\theta}$ is now used which provides us with the test-statistic:

$$t = \frac{\widehat{\theta} - 0}{s(\widehat{\theta})} \tag{3}$$

where $s(\widehat{\theta}) = n^{-\frac{1}{2}}\sqrt{\widehat{\boldsymbol{\beta}}' \left(\widehat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\right)\widehat{\boldsymbol{\beta}}}$ that can be used together with critical values associated with a given confidence level to perform the test.

Question 3

(3.1) Given the regression model $y_i = x_i\beta + e_i$, with $x_i \in \mathbb{R}$ and that $\mathbb{E}(e_i|x_i) = 0$ and also $\sigma_i^2 = \mathbb{E}(e_i^2|x_i)$, then find $\mathbb{E}\left((\widehat{\beta} - \beta)^3|\boldsymbol{X}\right)$:

$$\begin{aligned}
\mathbb{E}\left((\widehat{\beta} - \beta)^3|\boldsymbol{X}\right) &= \mathbb{E}\left(\widehat{\beta}^3 - 3\widehat{\beta}^2\beta + 3\widehat{\beta}\beta^2 - \beta^3|\boldsymbol{X}\right) \\
&= \mathbb{E}\left(\widehat{\beta}^3|\boldsymbol{X}\right) - 3\beta\mathbb{E}\left(\widehat{\beta}^2|\boldsymbol{X}\right) + 3\beta^2\mathbb{E}\left(\widehat{\beta}|\boldsymbol{X}\right) - \beta^3 \\
&= \mathbb{E}\left(\widehat{\beta}^3|\boldsymbol{X}\right) - 3\beta\Bigg[\mathbb{E}\left(\widehat{\beta}^2|\boldsymbol{X}\right) - \beta\ \underbrace{\mathbb{E}\left(\widehat{\beta}|\boldsymbol{X}\right)}_{\substack{\widehat{\beta}\text{ is unbiased:}\\ \mathbb{E}(\widehat{\beta}|\boldsymbol{X}) = \beta}}\ \Bigg] - \beta^3 \\
&= \mathbb{E}\left(\widehat{\beta}^3|\boldsymbol{X}\right) - 3\beta\underbrace{\left[\mathbb{E}\left(\widehat{\beta}^2|\boldsymbol{X}\right) - \beta \times \beta\right]}_{\text{var}(A)=\mathbb{E}(A^2)-(\mathbb{E}A)^2} - \beta^3 \\
&= \underbrace{\mathbb{E}\left(\widehat{\beta}^3|\boldsymbol{X}\right)}_{(A)} - 3\beta\underbrace{\mathbb{E}\left[(\widehat{\beta} - \beta)^2|\boldsymbol{X}\right]}_{(B)} - \beta^3 \qquad \text{(Eqaution 1)}
\end{aligned}$$

I simplify expressions (A) and (B) below. First, expression (A) yields:

$$\begin{aligned}
\mathbb{E}\left(\widehat{\beta}^3|\boldsymbol{X}\right) &= \mathbb{E}\left[\left(\beta + \frac{\sum_{i=1}^n x_i e_i}{\sum_{i=1}^n x_i^2}\right)^3 \Bigg| \boldsymbol{X}\right] \\
&= \mathbb{E}\left[\beta^3 + 3\beta^2\frac{\sum_{i=1}^n x_i e_i}{\sum_{i=1}^n x_i^2} + 3\beta\left(\frac{\sum_{i=1}^n x_i e_i}{\sum_{i=1}^n x_i^2}\right)^2 + \left(\frac{\sum_{i=1}^n x_i e_i}{\sum_{i=1}^n x_i^2}\right)^3 \Bigg| \boldsymbol{X}\right] \\
&= \beta^3 + 3\beta^2\frac{\sum_{i=1}^n x_i \mathbb{E}[e_i|\boldsymbol{X}]}{\sum_{i=1}^n x_i^2} + 3\beta\mathbb{E}\left[\left(\frac{\sum_{i=1}^n x_i e_i}{\sum_{i=1}^n x_i^2}\right)^2 \Bigg| \boldsymbol{X}\right] + \mathbb{E}\left[\left(\frac{\sum_{i=1}^n x_i e_i}{\sum_{i=1}^n x_i^2}\right)^3 \Bigg| \boldsymbol{X}\right]
\end{aligned}$$

Noting that $\mathbb{E}[e_i|\boldsymbol{X}] = 0$. Expanding the last two terms in the expression above contains cross-products e_ie_j, $e_i^2e_j$ and $e_ie_j^2$ $(\forall i \neq j)$ that, in expectation, are equal to zero under the assumption that error terms are linearly independent ($\mathbb{E}(e_ie_j|\boldsymbol{X}) = \mathbb{E}(e_i|\boldsymbol{X})\mathbb{E}(e_j|\boldsymbol{X}) = 0$ and $\mathbb{E}(e_i^2e_j|\boldsymbol{X}) = \mathbb{E}(e_i^2|\boldsymbol{X})\mathbb{E}(e_j|\boldsymbol{X}) = 0$). This simplifies the squared and cubic summations as below:

$$\begin{aligned}
&= \beta^3 + 3\beta\mathbb{E}\left[\frac{\sum_{i=1}^n x_i^2 e_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2}\Bigg|\boldsymbol{X}\right] + \mathbb{E}\left[\frac{\sum_{i=1}^n x_i^3 e_i^3}{\left(\sum_{i=1}^n x_i^2\right)^3}\Bigg|\boldsymbol{X}\right] \\
&= \beta^3 + 3\beta\frac{\sum_{i=1}^n x_i^2\mathbb{E}\left[e_i^2|\boldsymbol{X}\right]}{\left(\sum_{i=1}^n x_i^2\right)^2} + \frac{\sum_{i=1}^n x_i^3\mathbb{E}\left[e_i^3|\boldsymbol{X}\right]}{\left(\sum_{i=1}^n x_i^2\right)^3} \\
&= \beta^3 + 3\beta\frac{\sum_{i=1}^n x_i^2\sigma_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2} + \frac{\sum_{i=1}^n x_i^3\mu_{3i}}{\left(\sum_{i=1}^n x_i^2\right)^3}
\end{aligned}$$

Second, expression (B) yields:

$$\begin{aligned}
\mathbb{E}\left[(\widehat{\beta}-\beta)^2|\boldsymbol{X}\right] &= \mathbb{E}\left[\left(\beta+\frac{\sum_{i=1}^n x_i e_i}{\sum_{i=1}^n x_i^2}-\beta\right)^2\middle|\boldsymbol{X}\right] \\
&= \mathbb{E}\left[\frac{\sum_{i=1}^n x_i^2 e_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2}\middle|\boldsymbol{X}\right] \\
&= \frac{\sum_{i=1}^n x_i^2 \mathbb{E}\left[e_i^2|\boldsymbol{X}\right]}{\left(\sum_{i=1}^n x_i^2\right)^2} \\
&= \frac{\sum_{i=1}^n x_i^2 \sigma_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2}
\end{aligned}$$

Substituting simplified versions of expressions (A) and (B) into 'Equation 1' yield:

$$\begin{aligned}
\mathbb{E}\left((\widehat{\beta}-\beta)^3|\boldsymbol{X}\right) &= \beta^3+3\beta\frac{\sum_{i=1}^n x_i^2\sigma_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2}+\frac{\sum_{i=1}^n x_i^3\mu_{3i}}{\left(\sum_{i=1}^n x_i^2\right)^3}-3\beta\frac{\sum_{i=1}^n x_i^2\sigma_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2}-\beta^3 \\
&= \frac{\sum_{i=1}^n x_i^3\mu_{3i}}{\left(\sum_{i=1}^n x_i^2\right)^3}
\end{aligned}$$

(3.2) This statistic measures skewness of the estimator that shows asymmetry in the distribution.

Question 4

(4.1)

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

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

$$\begin{aligned}
\partial RSS/\partial\beta_1\big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} &= -2\sum_{i\in N}\left(y_i-[\widehat{\beta}_1+\widehat{\beta}_2x_{1i}+\widehat{\beta}_3x_{2i}]\right) \\
\partial RSS/\partial\beta_2\big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} &= -2\sum_{i\in N}x_{1i}\left(y_i-[\widehat{\beta}_1+\widehat{\beta}_2x_{1i}+\widehat{\beta}_3x_{2i}]\right) \\
\partial RSS/\partial\beta_3\big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} &= -2\sum_{i\in N}x_{2i}\left(y_i-[\widehat{\beta}_1+\widehat{\beta}_2x_{1i}+\widehat{\beta}_3x_{2i}]\right)
\end{aligned}$$

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:

$$0 = -2\sum_{i\in N}\left(y_i-[\widehat{\beta}_1+\widehat{\beta}_2x_{1i}+\widehat{\beta}_3x_{2i}]\right) \tag{4}$$

$$0 = -2\sum_{i\in N}x_{1i}\left(y_i-[\widehat{\beta}_1+\widehat{\beta}_2x_{1i}+\widehat{\beta}_3x_{2i}]\right) \tag{5}$$

$$0 = -2\sum_{i\in N}x_{2i}\left(y_i-[\widehat{\beta}_1+\widehat{\beta}_2x_{1i}+\widehat{\beta}_3x_{2i}]\right) \tag{6}$$

The first equations gives:

$$\begin{aligned}
\sum_{i\in N}y_i &= \sum_{i\in N}\widehat{\beta}_1+\sum_{i\in N}\widehat{\beta}_2x_{1i}+\sum_{i\in N}\widehat{\beta}_3x_{2i} \\
&= N\widehat{\beta}_1+\widehat{\beta}_2\sum_{i\in N}x_{1i}+\widehat{\beta}_3\sum_{i\in N}x_{2i}
\end{aligned}$$

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:

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

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

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:

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

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

Re-arrange:

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

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

Re-arrange:

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

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

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

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

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

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

(4.2) 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 $\operatorname{cov}(\tilde{x}_1,u)=0$ and $\operatorname{cov}(\tilde{x}_2,u)=0$ then:

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

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

(4.3) 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 $\frac{1}{n}$ 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.

(4.4) 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 5

(5.1) and (5.2): The actual outcome variable is observable with noise v_i around it which gives the following expression in terms of observables (y_i, x_i):

$$y_i \;=\; \alpha + \beta x_i + (u_i - v_i) \tag{9}$$

because two error terms are mutually exclusive, $\text{cov}(v_i, u_j) = 0\ \forall_{i,j}$, and that both errors have zero expectations[1], we apply the GLS estimator:

$$\widehat{\boldsymbol{\gamma}} \;=\; [\widehat{\alpha}, \widehat{\beta}]' \;=\; (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Y} \tag{10}$$

where $\boldsymbol{X} = [\boldsymbol{1}\quad \boldsymbol{x}]$ and noting that the variance structure is $\mathbb{E}[(\boldsymbol{u}-\boldsymbol{v})(\boldsymbol{u}-\boldsymbol{v})'|\boldsymbol{X}] = \boldsymbol{\Omega} = \sigma_u^2 diag(z_i) + \sigma_v^2\boldsymbol{I} \neq (\sigma_u^2+\sigma_v^2)\boldsymbol{I}$ since z_i's vary across i and $diag(.)$ is a $n\times n$ square matrix with z_i's on the main diagonal and zeros elsewhere (n is the sample size). We can write $\widehat{\boldsymbol{\gamma}} \to_p \boldsymbol{\gamma}$ because $\widehat{\boldsymbol{\gamma}} = \boldsymbol{\gamma} + (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}(\boldsymbol{u}-\boldsymbol{v})$ includes terms with the following properties $\frac{\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{u}}{N} \to_p \mathbb{E}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{u}] = \boldsymbol{0}_{k\times 1}$ and $\frac{\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{v}}{N} \to_p \mathbb{E}[\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{v}] = \boldsymbol{0}_{k\times 1}$ ($\widehat{\boldsymbol{\gamma}}$ is consistent). Conditions are GM and that z_i's should be known.

[1]Ideally we would need zero conditional expectation which is a stronger condition.

(5.3)

$$\begin{aligned}
\mathrm{var}(\boldsymbol{e}|\boldsymbol{X}) &= \mathrm{var}(\boldsymbol{y} - \boldsymbol{X}\widehat{\gamma}|\boldsymbol{X}) \\
&= \mathrm{var}(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Y}|\boldsymbol{X}) \\
&= \mathrm{var}(\{\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\}\boldsymbol{Y}|\boldsymbol{X}) \\
&= \mathrm{var}(\boldsymbol{M}_{\mathrm{GLS}}\boldsymbol{Y}|\boldsymbol{X}) \\
&= \boldsymbol{M}_{\mathrm{GLS}}\,\mathrm{var}(\widehat{\boldsymbol{u}} - \widehat{\boldsymbol{v}}|\boldsymbol{X})\boldsymbol{M}_{\mathrm{GLS}} \\
&= \boldsymbol{M}_{\mathrm{GLS}}(\sigma_u^2.diag(z_i) + \sigma_v^2\boldsymbol{I})\boldsymbol{M}_{\mathrm{GLS}}
\end{aligned}$$

Question 6

(6.1) The test-statistic is:

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

when the variance term in the denominator is unknown[2] then $\widehat{\sigma}^2 = \mathrm{SSE}/(T-k)$ 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,

$$\mathcal{F} > \text{critical-value}[F_{q,T-k}(1-\alpha)] \tag{11}$$

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.

(6.2) 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 7

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

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

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

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

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

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

[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.

(7.3)

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

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

(7.4)

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

(7.5) 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.

(7.6) 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.