University of Glasgow

Data Science and Machine Learning in Finance (ACCFIN 5246)

Problem Set 1 – Spring 2025

Hormoz Ramian

Question 1 Consider two investment assets X and Y. The observed asset returns generated by X have been historically low but always non-negative and equal to 0%, +1% and +2%, whereas observed returns generated by Y have been more volatile and equal to −5%, +2% and +8%. Furthermore, the following table provides further information about their joint investment performances with a summary that informs the likelihoods associated with each observed return outcomes across each other:

		X		
		0	+1	+2
	−5	10%	20%	10%
Y	+2	20%	0%	15%
	+8	10%	0%	15%

Table 1: *Each value within the table is a probability and values across the top row and left column are returns in percentage points.*

Each entry shows the probability of the corresponding X and Y values occurring jointly. For example: $P(Y = 5, X = 1) = 0.2$,

(1.1) Compute the marginal distributions for X and Y. In other words, compute the unconditional probabilities for each value of X: $\mathbb{P}(X = 0)$, $\mathbb{P}(X = 1)$, $\mathbb{P}(X = 2)$, and do the same for Y: With these in hand, compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(1.2) Compute the conditional distribution of Y given X, that is, $\mathbb{P}(Y = y|X = x)$ for all possible values of X and Y. Recall that:

$$\mathbb{P}(Y = y|X = x) \quad = \quad \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)} \tag{1}$$

(1.3) Compute the conditional expectation of $\mathbb{E}(Y|X = x)$ for all three values of X.

(1.4) Compute $\mathbb{E}\big[\mathbb{E}[Y|X]\big]$ and compare with $\mathbb{E}[Y]$.

Question 2

(2.1) "Failure to reject H_0 means the null hypothesis is true", true or false? If true, explain why? If false, explain why.

(2.2) Is the statement, "A matrix is a projection matrix iff it is an idempotent matrix", true? If so, explain why? If not, explain when this can be true.

(2.3) "An idempotent matrix is always invertible", true or false?

(2.4) "A projection matrix is always invertible", true or false?

(2.5) In March 1994, Michigan voters approved a sales tax increase from 4% to 6%. In political advertisements, supporters of the measure referred to this as a two percentage point increase, or an increase of two cents on the dollar. Opponents of the tax increase called it a 50% increase in the sales tax rate. Explain which way of measuring the increase in the sales tax is more accurate.

Question 3 Let y_i represent the share price of a stock in the S&P, and x_i be a dummy variable equal to 1 if stock i is classified in the financial sector and 0 otherwise. Suppose we see N stocks total, and N_x of these stocks are in the financial sector (which of course means that $N - N_x$ are in other sectors).

(3.1) Write $\overline{x}$ (the average of x_i) in terms of N_x and N.

(3.2) Suppose we run an OLS regression of y_i on a constant and x_i:

$$y_i \quad = \quad \beta_0 + \beta_1 x_i + \nu_i \tag{2}$$

Show that $\widehat{\beta}^{OLS}$ is quantitatively equal to the difference in means between the two sectors:

$$\widehat{\beta}^{OLS} \quad = \quad \left[\sum_{i \in \text{Financial Sector}} \frac{y_i}{N_x} \right] - \left[\sum_{i \in \text{Non-financial Sector}} \frac{y_i}{N - N_x} \right] \tag{3}$$

Question 4 Categorical variables $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$, and $\mathbf{1}_n$ each is a vector of size $n \times 1$, and that $\boldsymbol{d}_2 = \mathbf{1}_n - \boldsymbol{d}_1$ with $n = n_1 + n_2$ (n_1: number of men and n_2: number of women) such that:

$$d_{1,i} = \begin{cases} 1 & \text{if man} \\ 0 & \text{if woman} \end{cases}$$

suppose:

$$\boldsymbol{y} = \boldsymbol{d}_1 \widehat{\gamma}_1 + \boldsymbol{d}_2 \widehat{\gamma}_2 + \widehat{\boldsymbol{u}}$$

(4.1) Show that $(\widehat{\gamma}_1, \ \widehat{\gamma}_2)' = (\overline{y}_1, \ \overline{y}_2)'$.

(4.2) Compare $\widehat{\gamma}_1$ and $\widetilde{\gamma}_1$ from two OLS regressions:

$$\widehat{\boldsymbol{y}} \quad = \quad \boldsymbol{d}_1 \widehat{\gamma}_1 + \boldsymbol{d}_2 \widehat{\gamma}_2 \tag{4}$$

$$\widehat{\boldsymbol{y}} \quad = \quad \boldsymbol{d}_1 \widetilde{\gamma}_1 \tag{5}$$

Question 5 Consider the regression model ($\boldsymbol{y}$ and $\boldsymbol{u}$ each is $N \times 1$, $\boldsymbol{X}$ is $N \times k$ and $\boldsymbol{\beta}$ is $k \times 1$):

$$\boldsymbol{y} \quad = \quad \boldsymbol{X\beta} + \boldsymbol{u}$$

and that we additionally wish to examine $\boldsymbol{R\beta} = \boldsymbol{r}$ where $\boldsymbol{R}$ is $q \times k$ and $\boldsymbol{r}$ is $q \times 1$. Let RSS_U and RSS_R denote the unrestricted and restricted sum of squared residuals, respectively.

(5.1) Write a formal expression for the null and alternative hypotheses.

(5.2) Write the problem in terms of a constrained problem (Lagrange problem).

(5.3) Derive the first order conditions and solve.

(5.4) What the value of Lagrange multiplier? Interpret the Lagrange multiplier. What is the sign?

(5.5) What are the equations for RSS_U and RSS_R?

(5.6) Derive an expression in terms of regression residuals for, $RSS_R - RSS_U$.

(5.7) Interpret the term $RSS_R - RSS_U$. What is the sign and why?

Question 6 Consider the model $y_i = \alpha + \exp(x_i \beta) + u_i$. Derive the NLS estimators for α and β.

Answers

Problem 1

(a) Given the discrete random variables X and Y, the marginal densities are (probability mass functions):

	0	1	2
marginal density (X)	0.4	0.2	0.4

	-5	2	8
marginal density (Y)	0.4	0.35	0.25

then the (unconditional) expectations are,

$$\begin{aligned}\mathbb{E}[X] &= \sum_i p_{x,i}x_i \;=\; 0.4\times 0 + 0.2\times 1 + 0.4\times 2 \;=\; 1.0\\ \mathbb{E}[Y] &= \sum_i p_{y,i}y_i \;=\; 0.4\times(-5) + 0.35\times 2 + 0.35\times 8 \;=\; 0.7\end{aligned}$$

(b) Conditioning on a given X (assuming that a certain value has occurred) then conditional probabilities are,

	given $X = 0$	given $X = 1$	given $X = 2$
$\mathbb{P}[Y = -5 \mid X = x_i]$	$\frac{0.1}{0.1+0.2+0.1} = 0.25$	$\frac{0.2}{0+0.2+0} = 1$	$\frac{0.1}{0.1+0.15+0.15} = 0.25$
$\mathbb{P}[Y = 2 \mid X = x_i]$	$\frac{0.2}{0.1+0.2+0.1} = 0.50$	$\frac{0}{0+0.2+0} = 0$	$\frac{0.15}{0.1+0.15+0.15} = 0.375$
$\mathbb{P}[Y = 8 \mid X = x_i]$	$\frac{0.1}{0.1+0.2+0.1} = 0.25$	$\frac{0}{0+0.2+0} = 0$	$\frac{0.15}{0.1+0.15+0.15} = 0.375$

Conditioning on a given Y,

	$\mathbb{P}[X = 0 \mid Y = y_i]$	$\mathbb{P}[X = 1 \mid Y = y_i]$	$\mathbb{P}[X = 2 \mid Y = y_i]$
given $Y = -5$	$\frac{0.1}{0.1+0.2+0.1} = 0.25$	$\frac{0.2}{0.1+0.2+0.1} = 0.5$	$\frac{0.1}{0.1+0.2+0.1} = 0.25$
given $Y = 2$	$\frac{0.2}{0.2+0+0.15} \approx 0.57$	$\frac{0}{0.2+0+0.15} = 0$	$\frac{0.15}{0.2+0+0.15} \approx 0.43$
given $Y = 8$	$\frac{0.1}{0.1+0+0.15} = 0.40$	$\frac{0}{0.1+0+0.15} = 0$	$\frac{0.15}{0.1+0+0.15} = 0.60$

(c) Conditional expectation of Y given different values for X,

$$\begin{aligned}\mathbb{E}[Y|X=0] &= \frac{0.1}{0.4}\times(-5) + \frac{0.2}{0.4}\times 2 + \frac{0.1}{0.4}\times 8 \;=\; 1.75\\ \mathbb{E}[Y|X=1] &= \frac{0.2}{0.2}\times(-5) \;=\; -5\\ \mathbb{E}[Y|X=2] &= \frac{0.1}{0.4}\times(-5) + \frac{0.15}{0.4}\times 2 + \frac{0.15}{0.4}\times 8 \;=\; 2.5\end{aligned}$$

(d) Using law of iterated expectations, where $\mathbb{P}[x_i]$ is given by the marginal density of X,

$$\begin{aligned}\mathbb{E}[\mathbb{E}[Y|X]] &= \sum_i \mathbb{P}[x_i]\times\mathbb{E}[Y|x_i]\\ &= 0.4\times 1.75 + 0.2\times(-5) + 0.4\times 2.5 \;=\; 0.7\end{aligned}$$

Question 2

(2.1) 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.

(2.2) No, in general, the projection property implies idempotency however the reverse is not always true as the projection property 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.

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

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

(2.5) 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!

Problem 3

(3.1) The regressor x_i is a dummy variable (0 or 1 values) then $\overline{x} = \frac{N_x}{N}$ because:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N}\left(\sum_{i\in\text{Financial Sector}}^{N_x} x_i + \sum_{i\notin\text{Financial Sector}}^{N-N_x} x_i\right) = \frac{1}{N}\sum_{i\in\text{Financial Sector}}^{N_x} 1$$

(3.2) Re-write the regression in matrix form $\boldsymbol{y} = \begin{bmatrix}\mathbf{1} & \boldsymbol{x}\end{bmatrix}$ where $\mathbf{1}$ and $\boldsymbol{x}$ each is $N \times 1$ let $\boldsymbol{B} = (\beta_0, \beta_1)'$, then:

$$\boldsymbol{B} = \left(\begin{bmatrix}\mathbf{1}' \\ \boldsymbol{x}'\end{bmatrix}' \begin{bmatrix}\mathbf{1} & \boldsymbol{x}\end{bmatrix}\right)^{-1} \begin{bmatrix}\mathbf{1}' \\ \boldsymbol{x}'\end{bmatrix} \boldsymbol{y} = \begin{pmatrix} N & N_x \\ N_x & N_x \end{pmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i\in\text{Fin-Sector}} y_i \end{bmatrix}$$

the inverse exists[1] since $N_x(N-N_x) \neq 0$ and is equal to $\frac{1}{N_x(N-N_x)}\begin{pmatrix} N_x & -N_x \\ -N_x & N \end{pmatrix}$ then,

$$\begin{aligned}
\widehat{\beta}_1^{\text{ols}} &= \frac{1}{N_x(N-N_x)}\left[-N_x\sum_{i}^{N} y_i + N\sum_{i\in\text{Fin-Sector}} y_i\right] \\
&= \frac{1}{N_x(N-N_x)}\left[-N_x\left(\sum_{i\in\text{Fin-Sector}} y_i + \sum_{i\notin\text{Fin-Sector}} y_i\right) + N\sum_{i\in\text{Fin-Sector}} y_i\right] \\
&= \frac{1}{N_x(N-N_x)}\left[(N-N_x)\sum_{i\in\text{Fin-Sector}} y_i - N_x\sum_{i\notin\text{Fin-Sector}} y_i\right] \\
&= \frac{1}{N_x(N-N_x)} \times (N-N_x)\sum_{i\in\text{Fin-Sector}} y_i - \frac{1}{N_x(N-N_x)} \times N_x\sum_{i\notin\text{Fin-Sector}} y_i
\end{aligned}$$

Question 4

[1]Unless $N = N_x$ indicating that $x_i = 1$, $\forall_i$ (x is constant), which is not true as long as at least one firm is unlisted.

(4.1) In order to show that $(\widehat{\gamma}_1,\ \widehat{\gamma}_2)' = (\overline{y}_1,\ \overline{y}_2)'$, I re-write the regression model in matrix form and compute the inverse of the quadratic form for explanatory variables. Next, note that because $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ are binary variables and are perfectly orthogonal, then their inner product is zero, the inverse form simplifies to a inverse of diagonal elements inside the matrix:

$$
\begin{aligned}
\begin{pmatrix} \widehat{\gamma}_1 \\ \widehat{\gamma}_2 \end{pmatrix} &= \begin{pmatrix} \boldsymbol{d}_1'\boldsymbol{d}_1 & \boldsymbol{d}_1'\boldsymbol{d}_2 \\ \boldsymbol{d}_2'\boldsymbol{d}_1 & \boldsymbol{d}_2'\boldsymbol{d}_2 \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{d}_1' \\ \boldsymbol{d}_2' \end{pmatrix} \boldsymbol{y} \\
&= \begin{pmatrix} \sum_{i\in n_1} d_{1,i}^2 & 0 \\ 0 & \sum_{i\in n_1} d_{2,i}^2 \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{d}_1' \\ \boldsymbol{d}_2' \end{pmatrix} \boldsymbol{y} \\
&= \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{d}_1'\boldsymbol{y} \\ \boldsymbol{d}_2'\boldsymbol{y} \end{pmatrix} = \begin{pmatrix} \frac{1}{n_1} & 0 \\ 0 & \frac{1}{n_2} \end{pmatrix} \begin{pmatrix} \sum_{i\in n_1} y_i \\ \sum_{i\in n_2} y_i \end{pmatrix} = \begin{pmatrix} \frac{1}{n_1}\sum_{i\in n_1} y_i \\ \frac{1}{n_2}\sum_{i\in n_2} y_i \end{pmatrix} = \begin{pmatrix} \overline{y}_1 \\ \overline{y}_2 \end{pmatrix}
\end{aligned}
$$

(4.2) Under the definition of the model, we have $\widehat{\boldsymbol{\gamma}}_1 = \widetilde{\boldsymbol{\gamma}}_1$ since $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ are perfectly orthogonal, indicating that we can split the regression for each sample and run individual regressions. Alternatively, we note that because of orthogonality, each original coefficient $\widehat{\gamma}_i$ only measures the impact from its respective variable. Algebraically, running the regression without either of the variables, amounts to no estimation bias due to omitted variable.

Question 5

(5.1) The test setup is:

$$
\begin{aligned}
H_0 &: \boldsymbol{R\beta} - \boldsymbol{r} = \boldsymbol{0} \\
H_A &: \boldsymbol{R\beta} - \boldsymbol{r} \neq \boldsymbol{0}
\end{aligned}
$$

at a given confidence level and using theoretical F-distribution critical values.

(5.2) The Lagrangian to this constrained problem is:

$$
\mathcal{L} = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) + \boldsymbol{\lambda}'(\boldsymbol{R\beta} - \boldsymbol{r})
$$

where $\boldsymbol{\beta}$ is the parameter vector of restricted least squares or constrained least squares (CLS).

(5.3) The first order condition(s) with respect to $\boldsymbol{\beta}$ is:

$$
\boldsymbol{X}'\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\beta}} = \boldsymbol{R}'\boldsymbol{\lambda} \tag{6}
$$

or just $\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{R}'\boldsymbol{\lambda}$. The first order condition(s) with respect to $\boldsymbol{\lambda}$ is:

(5.4) The solution including the Lagrange multiplier (vector) $\boldsymbol{\lambda}$ is:

$$
\boldsymbol{0} = \boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r} \tag{7}
$$

Now we'll need to solve the two first order conditions for the values of $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\lambda}$ that satisfy the Lagrangian problem. First pre-multiply by $\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}$:

$$
\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\left[\boldsymbol{X}'\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\beta}}\right] = \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\left[\boldsymbol{R}'\boldsymbol{\lambda}\right]
$$

and simplify to get:

$$
\begin{aligned}
\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\beta}} &= \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\boldsymbol{\lambda} \\
\boldsymbol{R}\widetilde{\boldsymbol{\beta}} - \boldsymbol{R}\widehat{\boldsymbol{\beta}} &= \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\boldsymbol{\lambda}
\end{aligned}
$$

where $\tilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ and $(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X}) = \boldsymbol{I}$. Note that under the null hypothesis $\boldsymbol{R}\widehat{\boldsymbol{\beta}} = \boldsymbol{r}$:

$$\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r} \quad = \quad \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\boldsymbol{\lambda}$$

Solving for $\boldsymbol{\lambda}$:

$$\boldsymbol{\lambda} \quad = \quad \left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r}\right)$$

The expression above does not include $\widehat{\boldsymbol{\beta}}$ and hence is the solution for $\boldsymbol{\lambda}$. Substituting this into equation (6) gives:

$$\begin{aligned}
\boldsymbol{X}'\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\beta}} &= \boldsymbol{R}'\boldsymbol{\lambda} \\
\boldsymbol{X}'\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\beta}} &= \boldsymbol{R}'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r}\right) \\
&\Leftrightarrow \\
(\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\beta}} &= \boldsymbol{X}'\boldsymbol{y} - \boldsymbol{R}'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r}\right) \\
&\Leftrightarrow \\
\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r}\right) \\
\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r}\right) \\
&= \tilde{\boldsymbol{\beta}} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1}\left(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r}\right)
\end{aligned}$$

The solution to $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\lambda}$ is:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \tilde{\boldsymbol{\beta}} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r}) \\
\boldsymbol{\lambda} &= [\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}]^{-1}(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r})
\end{aligned}$$

The Lagrange multiplier is a vector with non-negative elements which measures whether constraint is binding or slack. The interpretation of the Lagrange multiplier depends on economic meaning of each variable. In general, however, this links the constraint to the sum of squared residuals.

(5.5) Let $\widehat{\boldsymbol{\epsilon}} \equiv \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ be residuals from the restricted regression:

$$\begin{aligned}
\widehat{\boldsymbol{\epsilon}} &\equiv \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \\
&= \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{X}\tilde{\boldsymbol{\beta}} - \boldsymbol{X}\tilde{\boldsymbol{\beta}} \\
&= (\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}) + \boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}})
\end{aligned}$$

Hence $SSR_R \equiv \widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}}$ is:

$$\begin{aligned}
\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}} &= \left[(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}) + \boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}})\right]'\left[(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}) + \boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}})\right] \\
&= (\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}) + (\tilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}})
\end{aligned}$$

Noting, firstly that $\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = 0$, and secondly, $SSR_U = (\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})$, then:

$$SSR_R - SSR_U \quad = \quad (\tilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}})$$

Using the results $\widehat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r})$, to simplify the RHS:

$$\begin{aligned}
. \quad &= (\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r}) \\
&= (\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r}) \\
&= (\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r})
\end{aligned}$$

Using $\boldsymbol{\lambda} = [\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}]^{-1}(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r})$, then:

$$\begin{aligned} SSR_R - SSR_U &= (\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r}) \\ &= (\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r})'\boldsymbol{\lambda} \\ &= (\boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r})'\left\{[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']\right\}\boldsymbol{\lambda} \\ &= \boldsymbol{\lambda}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}]\boldsymbol{\lambda} \end{aligned}$$

Use first order conditions, $\boldsymbol{X}'\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\beta}} = \boldsymbol{R}'\boldsymbol{\lambda}$, then $\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{R}'\boldsymbol{\lambda}$, and then $\boldsymbol{X}'\widehat{\epsilon} = \boldsymbol{R}'\boldsymbol{\lambda}$. The interpretation is that, given the constraint, then the inner product $\boldsymbol{X}'\widehat{\epsilon}$ is not zero and indeed equal to a function of $\boldsymbol{\lambda}$:

$$\begin{aligned} SSR_R - SSR_U &= \boldsymbol{\lambda}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']\boldsymbol{\lambda} \\ &= \widehat{\boldsymbol{\epsilon}}'\boldsymbol{P}\widehat{\boldsymbol{\epsilon}} \end{aligned}$$

where $\boldsymbol{P}$ is the projection matrix.

Question 6 Let SSR be defined by, $SSR = \sum_{i=1}^{n}(y_i - \alpha - \exp(\beta x_i))^2$. Then the first order conditions are:

$$\begin{aligned} \text{w.r.t. } \alpha\text{: } 0 &= -2\sum_{i=1}^{n}(y_i - \widehat{\alpha} - \exp(\widehat{\beta}x_i)) \\ \text{w.r.t. } \theta\text{: } 0 &= -2\sum_{i=1}^{n}(y_i - \widehat{\alpha} - \exp(\widehat{\beta}x_i))x_i\exp(\widehat{\beta}x_i) \end{aligned}$$

We can re-arrange to construct a matrix $\boldsymbol{w}_i = \begin{bmatrix} 1 & x_i\exp(\theta x_i) & z_i \end{bmatrix}$ which gives the moment conditions $\boldsymbol{W'}(\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta})) = \boldsymbol{0}$, but note that the moment condition driven by the system of FOCs may not necessarily have a unique solution due to nonlinearity of the RSS function in parameters. We can simplify the system but analytical solutions are only found in special cases (not in the present example). In this case, the first equations gives one parameter in terms of $\widehat{\beta}$:

$$\begin{aligned} 0 &= -2\sum_{i=1}^{n}y_i - 2\sum_{i=1}^{n}\widehat{\alpha} - -2\sum_{i=1}^{n}\exp(\widehat{\beta}x_i) \\ \widehat{\alpha} &= \overline{Y} - \frac{1}{n}\sum_{i=1}^{n}\exp(\widehat{\beta}x_i) \end{aligned}$$

substituting this in the second equation (FOC w.r.t. $\widehat{\beta}$):

$$\begin{aligned} 0 &= -2\sum_{i=1}^{n}y_ix_i\exp(\widehat{\beta}x_i) - 2\widehat{\alpha}\sum_{i=1}^{n}x_i\exp(\widehat{\beta}x_i) - 2\sum_{i=1}^{n}\exp(\widehat{\beta}x_i)x_i\exp(\widehat{\beta}x_i) \\ &= -\sum_{i=1}^{n}y_ix_i\exp(\widehat{\beta}x_i) - \left[\overline{Y} - \frac{1}{n}\sum_{i=1}^{n}\exp(\widehat{\beta}x_i)\right]\sum_{i=1}^{n}x_i\exp(\widehat{\beta}x_i) - \sum_{i=1}^{n}x_i\exp(2\widehat{\beta}x_i) \end{aligned}$$

depending on the values $\{x_i\}_{i=1}^{n}$, this expression may yield multiple maxima, but under regularity condition, a unique maximizer to the RSS function is guaranteed. Although the final equation has only one unknown $\widehat{\beta}$, an analytical solution does not exist and it should be solved numerically.