MONASH University

2021 Semester Two Assignment One
(August 2021)

STUDENT ID: 28723740 _ _ _ _ _ _ _ _ _ _ _ _ _

SURNAME: Li _ _ _ _ _ _ _ _ _ _ _ _ _

GIVEN NAME: Weihao _ _ _ _ _ _ _ _ _ _ _ _

UNIT CODE: BEX6510

UNIT TITLE: **Foundations of Econometrics**

TASK DURATION: **Wednesday September 1ˢᵗ 16:00 hrs – Monday September 6ᵗʰ 16:00 hrs**

Upon completion of this task please upload to Moodle as a Word or PDF document through Moodle submission.

This submission must occur by 4:00pm on Monday 6ᵗʰ September 2021. (Australian Eastern Standard Time).

Please attach a copy of this page as a cover page to your submission with your personal details completed as asked for above.

Question 1

(a)

Let $\boldsymbol{D}$ be a length n column vector where the i'th element is one, $1 \leq i \leq n$, but other elements are zeros. Suppose the OLS regression equation is given as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{D}d + \boldsymbol{r},$$

where $\boldsymbol{X}$ is an $n \times k$ matrix, $\boldsymbol{y}$, $\boldsymbol{b}$ and $\boldsymbol{r}$ are length n column vectors, and d is a scalar parameter.

By the Frisch-Waugh-Lovell Theorem, the OLS regression coefficient d is

$$d = (\boldsymbol{D}'\boldsymbol{R}_X\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{R}_X\boldsymbol{y},$$

where $\boldsymbol{R}_X = \boldsymbol{I}_T - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}$.

It can then be expressed as

$$
\begin{aligned}
d &= (\boldsymbol{D}'(\boldsymbol{I}_T - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X})\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{R}_X\boldsymbol{y} \\
&= (\boldsymbol{D}'\boldsymbol{D} - \boldsymbol{D}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{R}_X\boldsymbol{y} \\
&= (\boldsymbol{D}'\boldsymbol{D} - \boldsymbol{D}'\boldsymbol{H}_X\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{R}_X\boldsymbol{y} \\
&= (1 - h_{ii})^{-1}\boldsymbol{D}'\boldsymbol{R}_X\boldsymbol{y} \\
&= (1 - h_{ii})^{-1}\boldsymbol{D}'(\boldsymbol{I}_T - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X})\boldsymbol{y} \\
&= (1 - h_{ii})^{-1}\boldsymbol{D}'(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{y}) \\
&= (1 - h_{ii})^{-1}\boldsymbol{D}'\boldsymbol{e} \\
&= (1 - h_{ii})^{-1}e_i,
\end{aligned}
$$

where h_{ii} is the i'th diagonal element of $\boldsymbol{H}_X = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}$, and e_i is the i'th residual from the OLS regression of $\boldsymbol{y}$ on $\boldsymbol{X}$.

(b)

Let $\boldsymbol{e}$ be the residual from the OLS regression of $\boldsymbol{y}$ on $\boldsymbol{X}$, and $\boldsymbol{r}$ be the residual from the OLS regression of $\boldsymbol{y}$ on $\boldsymbol{X}$ and $\boldsymbol{D}$.

By the Frisch-Waugh-Lovell Theorem,

$$\boldsymbol{r} = \boldsymbol{R}_X(\boldsymbol{y} - \boldsymbol{D}d),$$

where $\boldsymbol{R}_X = \boldsymbol{I}_T - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}$.

Then the residual sum of square for the second regression is

$$
\begin{aligned}
\boldsymbol{r'r} &= (\boldsymbol{R}_X(\boldsymbol{y} - \boldsymbol{D}d))'(\boldsymbol{R}_X(\boldsymbol{y} - \boldsymbol{D}d)) \\
&= (\boldsymbol{R}_X\boldsymbol{y} - \boldsymbol{R}_X\boldsymbol{D}d)'(\boldsymbol{R}_X\boldsymbol{y} - \boldsymbol{R}_X\boldsymbol{D}d) \\
&= (\boldsymbol{e} - \boldsymbol{R}_X\boldsymbol{D}d)'(\boldsymbol{e} - \boldsymbol{R}_X\boldsymbol{D}d) \\
&= \boldsymbol{e'e} - 2\boldsymbol{e'}\boldsymbol{R}_X\boldsymbol{D}d + (\boldsymbol{R}_X\boldsymbol{D}d)'(\boldsymbol{R}_X\boldsymbol{D}d) \\
&= \boldsymbol{e'e} - 2\boldsymbol{e'}(\boldsymbol{I}_T - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X})\boldsymbol{D}d + (\boldsymbol{R}_X\boldsymbol{D}d)'(\boldsymbol{R}_X\boldsymbol{D}d) \\
&= \boldsymbol{e'e} - 2(\boldsymbol{e'} - \boldsymbol{0})\boldsymbol{D}d + (\boldsymbol{R}_X\boldsymbol{D}d)'(\boldsymbol{R}_X\boldsymbol{D}d) \\
&= \boldsymbol{e'e} - 2\boldsymbol{e'}\boldsymbol{D}d + (\boldsymbol{R}_X\boldsymbol{D}d)'(\boldsymbol{R}_X\boldsymbol{D}d) \\
&= \boldsymbol{e'e} - 2e_i d + (\boldsymbol{R}_X\boldsymbol{D}d)'(\boldsymbol{R}_X\boldsymbol{D}d) \\
&= \boldsymbol{e'e} - 2e_i d + d^2 \boldsymbol{D'}\boldsymbol{R}_X'\boldsymbol{R}_X\boldsymbol{D} \\
&= \boldsymbol{e'e} - 2e_i d + d^2 \boldsymbol{D'}\boldsymbol{R}_X\boldsymbol{D} \\
&= \boldsymbol{e'e} - 2e_i d + d^2 \boldsymbol{D'}(\boldsymbol{I}_T - \boldsymbol{H}_X)\boldsymbol{D} \\
&= \boldsymbol{e'e} - 2e_i d + d^2 (1 - h_{ii}) \\
&= \boldsymbol{e'e} - 2e_i \frac{e_i}{1 - h_{ii}} + \frac{e_i^2}{(1 - h_{ii})^2}(1 - h_{ii}) \\
&= \boldsymbol{e'e} - 2\frac{e_i^2}{1 - h_{ii}} + \frac{e_i^2}{(1 - h_{ii})} \\
&= \boldsymbol{e'e} - \frac{e_i^2}{1 - h_{ii}}.
\end{aligned}
$$

Hence,

$$
\boldsymbol{e'e} = \boldsymbol{r'r} + \frac{e_i^2}{1 - h_{ii}}.
$$

(c)

By the Frisch-Waugh-Lovell Theorem, the regression coefficient $\boldsymbol{b}$ from the OLS regression of $\boldsymbol{y}$ on $\boldsymbol{X}$ and $\boldsymbol{D}$ is

$$
\boldsymbol{b} = (\boldsymbol{X'}\boldsymbol{R}_D\boldsymbol{X})^{-1}\boldsymbol{X'}\boldsymbol{R}_D\boldsymbol{y},
$$

where $\boldsymbol{R}_D = \boldsymbol{I}_T - \boldsymbol{D}(\boldsymbol{D'D})^{-1}\boldsymbol{D'}$.

Given $\boldsymbol{R}_D = \boldsymbol{R}_D' = \boldsymbol{R}_D'\boldsymbol{R}_D = \boldsymbol{R}_D\boldsymbol{R}_D'$. Then $\boldsymbol{b}$ can be expressed as

$$
\boldsymbol{b} = (\boldsymbol{X'}\boldsymbol{R}_D'\boldsymbol{R}_D\boldsymbol{X})^{-1}\boldsymbol{X'}\boldsymbol{R}_D'\boldsymbol{R}_D\boldsymbol{y}.
$$

Using the fact that $\boldsymbol{R}_D = \boldsymbol{I}_T - \boldsymbol{D}(\boldsymbol{D'D})^{-1}\boldsymbol{D'} = \boldsymbol{I}_T - \boldsymbol{DD'}$, which equals a matrix with ones on the leading diagonal, apart from the i'th location which is zero, and is zero everywhere else.

Thus, $\boldsymbol{R}_D\boldsymbol{X} = \boldsymbol{X}_{(i)}$, which is $\boldsymbol{X}$ with data in i'th row replaced by zeros. Similarly, $\boldsymbol{R}_D\boldsymbol{y} = \boldsymbol{y}_{(i)}$, which is $\boldsymbol{y}$ with i'th value replaced by zero.

It follows that $\boldsymbol{b}$ is the regression coefficient obtained from the regression of $\boldsymbol{y}_{(i)}$ on $\boldsymbol{X}_{(i)}$, because

$$
\boldsymbol{b} = (\boldsymbol{X'}\boldsymbol{R}_D'\boldsymbol{R}_D\boldsymbol{X})^{-1}\boldsymbol{X'}\boldsymbol{R}_D'\boldsymbol{R}_D\boldsymbol{y} = (\boldsymbol{X}_{(i)}'\boldsymbol{X}_{(i)})^{-1}\boldsymbol{X}_{(i)}'\boldsymbol{y}_{(i)}.
$$

Again, by the Frisch-Waugh-Lovell Theorem, we have

$$
\boldsymbol{r} = \boldsymbol{R}_D(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) = \boldsymbol{y}_{(i)} - \boldsymbol{X}_{(i)}\boldsymbol{b}.
$$

Thus, $\boldsymbol{r}$ is the residual obtained from the regression of $\boldsymbol{y}_{(i)}$ on $\boldsymbol{X}_{(i)}$. And hence, using the result from (2), $\boldsymbol{r'r} = \boldsymbol{e'e} - e_i^2(1 - h_{ii})^{-1}$ is the residual sum of square of this regression.

(d)

Cook and Weisberg (1982) show that the externally studentized residual can be defined by

$$t_i = \frac{e_i}{\hat{\sigma}_{(i)}\sqrt{(1 - h_{ii})}},$$

where e_i is the residual from the OLS regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$, $\hat{\sigma}_{(i)}$ is the estimated standard deviation of the regression computed without i'th case, and h_{ii} is the leverage of the i'th observation.

Using the result from (b) and (c), $\hat{\sigma}_{(i)}$ can be expressed as

$$\hat{\sigma}_{(i)} = \sqrt{\hat{\sigma}_{(i)}^2} = \sqrt{\boldsymbol{r'r}/(n - k - 1)} = \sqrt{[\boldsymbol{e'e} - e_i^2/(1 - h_{ii})]/(n - k - 1)}.$$

Under normality, the distribution of t_i is Student's t with $n - k - 1$ degree of freedom. Hence, we can construct a test based on the externally studentized residual. We will reject the null hypothesis that the errors are independent and normally distributed with expected value 0 and variance σ^2 when

$$|t_i| > t_{1-\alpha/2}(n - k - 1),$$

where $t_{1-\alpha/2}(n - k - 1)$ is the $1 - \alpha/2$ quantile point of the Student's t distribution with $n - k - 1$ degree of freedom. If we reject the null hypothesis for the i'th observation, then we may consider it as an "outlier."

Question 2

(a)

The MLE of β subject to the J linear constraints $\boldsymbol{L\beta} = \boldsymbol{r}$ can be obtained by maximizing the Lagrangian

$$\mathcal{L} = logL(\boldsymbol{\theta}) - \boldsymbol{\lambda}'h(\boldsymbol{\theta}) = -\frac{n}{2}log(2\pi\sigma^2) - \frac{(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})}{2\sigma^2} - \boldsymbol{\lambda}'(\boldsymbol{L\beta} - \boldsymbol{r}).$$

By solving the first order condition of the Lagrangian,

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} = \frac{\boldsymbol{X'y} - \boldsymbol{X'X\beta}}{\sigma^2} - \boldsymbol{L'\lambda} = \boldsymbol{0},$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})}{2\sigma^4} = \boldsymbol{0}, \text{and}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}} = \boldsymbol{L\beta} - \boldsymbol{r} = \boldsymbol{0},$$

we have a system of equations:

$$\begin{cases} \hat{\boldsymbol{b}} = (\boldsymbol{X'X})^{-1}(\boldsymbol{X'y} - \hat{\sigma}^2\boldsymbol{L'\lambda}) & (1) \\ \hat{\sigma}^2 = (\boldsymbol{y} - \boldsymbol{X\hat{b}})'(\boldsymbol{y} - \boldsymbol{X\hat{b}})/n & (2) \\ \boldsymbol{L\hat{b}} = \boldsymbol{r} & (3) \end{cases}$$

.

Substitute (3) with (1), we then have

$$\hat{\boldsymbol{\lambda}} = (\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{r})\hat{\sigma}^{-2}. \tag{4}$$

Substitute (1) with (4), we obtain the closed form solution of $\hat{\boldsymbol{b}}$,

$$
\begin{aligned}
\hat{\boldsymbol{b}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{y} - \hat{\sigma}^2 \boldsymbol{L}'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{r})\hat{\sigma}^{-2}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{L}'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{r})) \\
&= \boldsymbol{b} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{r}) \\
&= \boldsymbol{b} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{L}\boldsymbol{b} - \boldsymbol{r}) \\
&= \boldsymbol{b} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{r} - \boldsymbol{L}\boldsymbol{b}),
\end{aligned}
$$

where $\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$.

(b)

By definition

$$
\begin{aligned}
\hat{\boldsymbol{e}} &= \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}} \\
&= \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}} - \boldsymbol{X}(\hat{\boldsymbol{b}} - \boldsymbol{b}) \\
&= \boldsymbol{e} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{r} - \boldsymbol{L}\boldsymbol{b}).
\end{aligned}
$$

Thus

$$
\begin{aligned}
\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}} &= \boldsymbol{e}'\boldsymbol{e} + (\boldsymbol{r} - \boldsymbol{L}\boldsymbol{b})'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{r} - \boldsymbol{L}\boldsymbol{b}) \\
&= \boldsymbol{e}'\boldsymbol{e} + (\boldsymbol{r} - \boldsymbol{L}\boldsymbol{b})'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{r} - \boldsymbol{L}\boldsymbol{b}) \\
&= \boldsymbol{e}'\boldsymbol{e} + (\boldsymbol{r} - \boldsymbol{L}\boldsymbol{b})'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{r} - \boldsymbol{L}\boldsymbol{b}),
\end{aligned}
$$

because $\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$.

(c)

Given the unrestricted estimator for $\boldsymbol{\beta}$ is the same in MLE and OLS, which will be denoted as $\boldsymbol{b}$, but the unrestricted estimator for σ^2 in MLE is $\hat{\sigma}^2 = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})n^{-1}$. The Likelihood Ratio test of the null hypothesis $H_0 : \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{r}$ is

$$TLR_n = 2[logL(\boldsymbol{\theta}_U) - logL(\boldsymbol{\theta}_R)]$$

$$= 2[-\frac{n}{2}log(2\pi\hat{\sigma}_U^2) - \frac{(\boldsymbol{y} - \boldsymbol{Xb})'(\boldsymbol{y} - \boldsymbol{Xb})}{2\hat{\sigma}_U^2} + \frac{n}{2}log(2\pi\hat{\sigma}_R^2) + \frac{(\boldsymbol{y} - \boldsymbol{X\hat{b}})'(\boldsymbol{y} - \boldsymbol{X\hat{b}})}{2\hat{\sigma}_R^2}]$$

$$= -nlog(\hat{\sigma}_U^2) - \frac{(\boldsymbol{y} - \boldsymbol{Xb})'(\boldsymbol{y} - \boldsymbol{Xb})}{\hat{\sigma}_U^2} + nlog(\hat{\sigma}_R^2) + \frac{(\boldsymbol{y} - \boldsymbol{X\hat{b}})'(\boldsymbol{y} - \boldsymbol{X\hat{b}})}{\hat{\sigma}_R^2}$$

$$= -nlog(\hat{\sigma}_U^2) - n + nlog(\hat{\sigma}_R^2) + n$$

$$= nlog(\hat{\sigma}_R^2/\hat{\sigma}_U^2)$$

$$= nlog(\frac{\boldsymbol{e}_R'\boldsymbol{e}_R}{\boldsymbol{e}_U'\boldsymbol{e}_U})$$

$$= nlog(\frac{\boldsymbol{e}_U'\boldsymbol{e}_U + (\boldsymbol{r} - \boldsymbol{Lb})'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{r} - \boldsymbol{Lb})}{\boldsymbol{e}_U'\boldsymbol{e}_U})$$

$$= nlog(1 + \frac{(\boldsymbol{r} - \boldsymbol{Lb})'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{r} - \boldsymbol{Lb})}{\boldsymbol{e}_U'\boldsymbol{e}_U}).$$

Given the F test statistic is

$$TF_n = \frac{(n-k)(\boldsymbol{r} - \boldsymbol{Lb})'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{r} - \boldsymbol{Lb})}{\boldsymbol{e}_U'\boldsymbol{e}_U J},$$

where J is the number of restriction.

Then, the LR test statistic is a monotonic transformation of the F test statistic

$$TLR_n = nlog(1 + \frac{(\boldsymbol{r} - \boldsymbol{Lb})'(\boldsymbol{L}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{r} - \boldsymbol{Lb})}{\boldsymbol{e}_U'\boldsymbol{e}_U}) = nlog(1 + \frac{J}{n-k}TF_n).$$

Hence, the LR test and the F test are equivalent.

Question 3

The RESET test aims to test if there exists non linearity in the linear model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

by testing $H_0 : \boldsymbol{\gamma} = \boldsymbol{0}$ in a proposed model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{\varepsilon}$$

using a standard F test, where $\boldsymbol{Z}$ is a collection of higher order terms of the predicted values.

In this question, we includes the squares and cubes of the predicted values to test two models. The first one is a regular regression model with the house price being the response and 4 other variables being the explanatory variables. The second one uses the same setting except log transformation is applied on all variables. We then obtain the test statistic for these two models, which are 4.67 and 2.56. Given the number of observations is 88 and the number of additional variables in the unrestricted model is 2, the degrees of freedom of the null distribution are 2 and $88 - 5 - 2 = 81$. By using the cdf calculator in R, we obtain the corresponding p-value, which are 0.0120365 and 0.0835469.

With 5% significance level, we reject the null hypothesis for the first one but do not reject for the second one. This means the first model could potentially omits some higher order terms or interaction terms. The indication of this is the assumptions of the regression model may no longer hold, and we may observe endogeneity. The second model tries to use the log-log form to capture the non linearity, and the test does not suggest any potential issue. Thus, we will choose the log-log form as our preferred model.

Question 4

(1)

Given a_1, a_2, a_3 and a_4 are constants, $E[N_i] = n(\theta^{i-1} - \theta^i \mathbb{1}(i \neq 4))$ for $i = 1, .., 4$, and $E[T] = \theta$. We have

$$E[T] = E[a_1N1 + a_2N_2 + a_3N_3 + a_4N_4] = n[a_1(1-\theta) + a_2(\theta - \theta^2) + a_3(\theta^2 - \theta^3) + a_4\theta^3] = \theta.$$

Since T is a statistic, a_1, a_2, a_3 and a_4 do not depend on θ. Hence, $a_1 = 0$, $a_2 = a_3 = a_4 = N^{-1}$.

(2)

Because $N = N_1 + N_2 + N3 + N_4$, the estimator T can be expressed by

$$T = a_1N1 + a_2N_2 + a_3N_3 + a_4N_4 = (N_2 + N_3 + N_4)N^{-1} = (N - N_1)N^{-1} = 1 - N_1/N.$$

Given $V[N_i] = Np_i(1 - p_i)$ and $p_i = \theta^{i-1} - \theta^i \mathbb{1}(i \neq 4)$ for $i = 1, .., 4$. The variance of T is

$$V[T] = V[1 - N_1/N] = N^{-2}V[N_1] = N^{-2}Np_1(1 - p_1) = \theta(1 - \theta)N^{-1}.$$

(3)

The pmf of X is given by

$$f(x|\theta) = (1 - \theta)^{\mathbb{1}(x=1)}(\theta - \theta^2)^{\mathbb{1}(x=2)}(\theta^2 - \theta^3)^{\mathbb{1}(x=3)}(\theta^3)^{\mathbb{1}(x=4)},$$

for $x \in \{1, 2, 3, 4\}$ and $0 \leq \theta \leq 0$.

The support function is

$$logL(\theta|x_1, ...x_N) = logZ(x_1, ..x_N) + logf(x_1, ..x_N|\theta)$$

$$= logZ(x_1, .., x_N) + \sum_{i=1}^{N} \mathbb{1}(x = 1)log(1 - \theta) + \mathbb{1}(x = 2)log(\theta - \theta^2) + \mathbb{1}(x = 3)log(\theta^2 - \theta^3) + \mathbb{1}(x = 4)log(\theta^3)$$

$$= logZ(x_1, .., x_N) + N_1log(1 - \theta) + N_2log(\theta - \theta^2) + N_3log(\theta^2 - \theta^3) + N_4log(\theta^3).$$

The score function is

$$\frac{dlogL(\theta|x_1, .., x_N)}{d\theta} = \frac{-N_1}{(1 - \theta)} + \frac{(1 - 2\theta)N_2}{(\theta - \theta^2)} + \frac{(2\theta - 3\theta^2)N_3}{(\theta^2 - \theta^3)} + \frac{(3\theta^2)N_4}{\theta^3}.$$

By setting the score function equals to zero, we have

$$\frac{-\hat{\theta}^2 N_1 + \hat{\theta}(1 - 2\hat{\theta})N_2 + (2\hat{\theta} - 3\hat{\theta}^2)N_3 + 3(\hat{\theta} - \hat{\theta}^2)N_4}{(\hat{\theta}^2 - \hat{\theta}^3)} = 0.$$

Thus, the MLE estimator is

$$\hat{\theta} = \frac{N_2 + 2N_3 + 3N_4}{N_1 + 2N_2 + 3N_3 + 3N_4}.$$

The second order derivative of the support function is

$$\frac{d^2 logL(\theta|x_1,..,x_N)}{d\theta^2} = \frac{-N_1}{(1-\theta)^2} - \frac{(1-2\theta)^2 N_2}{(\theta-\theta^2)^2} - \frac{(2\theta-3\theta^2)^2 N_3}{(\theta^2-\theta^3)^2} - \frac{(3\theta^2)^2 N_4}{(\theta^3)^2} < 0,$$

because N_1, N_2, N_3 and N_4 are non-negative. Hence, it is a maximum.

(4)

The Fisher's information per observation is

$$
\begin{aligned}
i(\theta_0) &= -E\left[\frac{dlogL(\theta|x)}{d\theta^2}\right]\Bigg|_{\theta=\theta_0} \\
&= -E\left[\frac{-\mathbb{1}(x=1)}{(1-\theta)^2} - \frac{(1-2\theta)^2\mathbb{1}(x=2)}{(\theta-\theta^2)^2} - \frac{(2\theta-3\theta^2)^2\mathbb{1}(x=3)}{(\theta^2-\theta^3)^2} - \frac{(3\theta^2)^2\mathbb{1}(x=4)}{(\theta^3)^2}\right]\Bigg|_{\theta=\theta_0} \\
&= -\left[\frac{-p_1}{(1-\theta)^2} - \frac{(1-2\theta)^2 p_2}{(\theta-\theta^2)^2} - \frac{(2\theta-3\theta^2)^2 p_3}{(\theta^2-\theta^3)^2} - \frac{(3\theta^2)^2 p_4}{(\theta^3)^2}\right]\Bigg|_{\theta=\theta_0} \\
&= -\left[\frac{-1}{(1-\theta)} - \frac{(1-2\theta)^2}{(\theta-\theta^2)} - \frac{(2\theta-3\theta^2)^2}{(\theta^2-\theta^3)} - \frac{(3\theta^2)^2}{(\theta^3)}\right]\Bigg|_{\theta=\theta_0} \\
&= -\left[\frac{-1}{(1-\theta)} - \frac{(1-2\theta)^2}{\theta(1-\theta)} - \frac{(2-3\theta)^2}{(1-\theta)} - \frac{(3\theta^2)^2}{(\theta^3)}\right]\Bigg|_{\theta=\theta_0} \\
&= -\left[\frac{-\theta}{\theta(1-\theta)} - \frac{(1-2\theta)^2}{\theta(1-\theta)} - \frac{\theta(2-3\theta)^2}{\theta(1-\theta)} - \frac{9\theta^2(1-\theta)}{\theta(1-\theta)}\right]\Bigg|_{\theta=\theta_0} \\
&= -\left[\frac{-\theta-1-\theta^2}{\theta(1-\theta)}\right]\Bigg|_{\theta=\theta_0} \\
&= \frac{\theta_0+1+\theta_0^2}{\theta_0(1-\theta_0)}.
\end{aligned}
$$

Thus, the CRLB for an ubiased estimator T is

$$V[T] \geq \frac{(\phi'(\theta_0))^2}{Ni(\theta_0)} = \frac{1}{Ni(\theta_0)} = \frac{\theta_0(1-\theta_0)}{N(\theta_0+1+\theta_0^2)}.$$

Since the MLE will achieve the CRLB asymptotically, we know for large n, the approximate variance of MLE is $V[\hat{\theta}] = \theta_0(1-\theta_0)[n(\theta_0+1+\theta_0^2)]^{-1}$. Hence, the asymptotic efficiency of T relative to $\hat{\theta}$ is

$$\frac{e(T)}{e(\hat{\theta})} = \frac{\frac{[\phi'(\theta_0)]^2}{Ni(\theta_0)}\frac{1}{V[T]}}{\frac{[\phi'(\theta_0)]^2}{Ni(\theta_0)}\frac{1}{V[\hat{\theta}]}} = \frac{V[\hat{\theta}]}{V[T]} = \frac{\frac{\theta_0(1-\theta_0)}{N(\theta_0+1+\theta_0^2)}}{\frac{\theta_0(1-\theta_0)}{N}} = (\theta_0+1+\theta_0^2)^{-1}.$$

Reference

Cook, R Dennis, and Sanford Weisberg. 1982. *Residuals and Influence in Regression.* New York: Chapman; Hall.