Midterm Exam

10 April 2024

- You have 90 min.
- There are 41 points in total.
- Prepare concise answers.
- State clearly any additional assumptions, if needed.

Problem 1 (41 points)

Suppose you observe n i.i.d. observations of an outcome variable y_i and a k-dimensional vector of covariates x_i: $\{(y_i, x_i)\}_{i=1:n}$. Suppose further that you want to examine their relation using a linear regression model,

$$y_i = x_i'\beta + u_i \ , \quad u_i|x_i \sim N(0, \sigma^2) \ ,$$

or, in matrix notation,

$$Y = X\beta + U \ ,$$

where Y and U are $(n \times 1)$ and X is $(n \times k)$.

You are exclusively concerned with frequentist/classical estimation, i.e. you consider the parameters β and σ^2 to be fixed and the data $\{(y_i, x_i)\}_{i=1:n}$ to be random (i.e. your particular sample is randomly drawn from some population).

(a) (4 points) Derive the (conditional) likelihood function $\mathcal{L}(\beta, \sigma^2|Y, X) \equiv p(Y|X, \beta, \sigma^2)$, and define the Maximum Likelihood (ML) estimators $\hat{\beta}$ and $\hat{\sigma}^2$.

Solution: We have

$$y_i \mid x_i, \beta, \sigma^2 \sim N(x_i'\beta, \sigma^2) \ ,$$

i.e.

$$p(y_i|x_i, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - x_i'\beta)^2\right\} \ .$$

The likelihood function is defined as the joint pdf of all observations, which by independence is equal to the product of the marginal pdfs of the individual observations:

$$\mathcal{L}(\beta, \sigma^2|Y, X) \equiv p(Y|X, \beta, \sigma^2) = \prod_{i=1}^{n} p(y_i|x_i, \beta, \sigma^2) \ .$$

Algebraic manipulations allow us to write the likelihood out explicitly:

$$\begin{aligned}
\mathcal{L}(\beta, \sigma^2|Y, X) =& \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - x_i'\beta)^2\right\} \\
=& (2\pi\sigma^2)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i'\beta)^2\right\} \\
=& (2\pi\sigma^2)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\} \ .
\end{aligned}$$

The log-likelihood is then

$$\ell(\beta, \sigma^2|Y, X) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta) \ .$$

[3p] The ML estimator is defined as

$$(\hat{\beta}, \hat{\sigma}^2) := \arg\max_{\beta, \sigma^2} l(\beta, \sigma^2|Y, X) \ . \quad \textbf{[1p]}$$

(b) (4 points) Derive the ML estimators $\hat{\beta}$ and $\hat{\sigma}^2$.

Hint: for a matrix B and vectors a and c, it holds that $\frac{\partial (a-Bc)'(a-Bc)}{\partial c} = -2B'(a-Bc)$.

Solution:

Using the hint, we get the following First-Order Conditions (FOCs) for β and σ^2:

$$\begin{aligned}\frac{\partial}{\partial \beta} l(\beta, \sigma^2 | Y, X) &= \frac{1}{\sigma^2} X'(Y - X\beta) = 0 \;, \\ \frac{\partial}{\partial \sigma^2} l(\beta, \sigma^2 | Y, X) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(Y - X\beta)'(Y - X\beta) = 0 \;.\end{aligned}$$

[1p each] Rearrange the former to get:

$$X'Y - (X'X)\beta = 0 \quad \Rightarrow \quad (X'X)\beta = X'Y \quad \Rightarrow \quad \hat{\beta} = (X'X)^{-1}X'Y \;.$$

Define $S = (Y - X\beta)'(Y - X\beta) = \sum_{i=1}^{n} (y_i - x_i'\beta)^2$ and rearrange the latter to get:

$$-n + \frac{1}{\sigma^2} S = 0 \quad \Rightarrow \quad n = \frac{1}{\sigma^2} S \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{1}{n} S \;.$$

[1p each]

(c) (6 points) Find $\mathbb{E}[\hat{\beta}]$ and $\mathbb{V}[\hat{\beta}] = \mathbb{E}[\left(\hat{\beta} - \mathbb{E}[\hat{\beta}]\right)\left(\hat{\beta} - \mathbb{E}[\hat{\beta}]\right)']$. What else do you know about the finite sample distribution of $\hat{\beta}$?

Hint: use the Law of Iterated Expectations, computing these two quantities conditional on X first.

Solution:

First of all, note that $\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + U) = \beta + (X'X)^{-1}X'U$. The conditional expectation of $\hat{\beta}$ is

$$\begin{aligned}\mathbb{E}[\hat{\beta}|X] &= \beta + \mathbb{E}[(X'X)^{-1}X'U|X] \\ &= \beta + (X'X)^{-1}X'\mathbb{E}[U|X] \\ &= \beta \;.\end{aligned}$$

The conditional variance of $\hat{\beta}$ is

$$\begin{aligned}\mathbb{V}[\hat{\beta}|X] &= \mathbb{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'|X] = \\ &= \mathbb{E}[((X'X)^{-1}X'U)((X'X)^{-1}X'U)'|X] = \\ &= \mathbb{E}[(X'X)^{-1}X'UU'X(X'X)^{-1}|X] = \\ &= (X'X)^{-1}X'\mathbb{E}[UU'|X]X(X'X)^{-1} = \\ &= (X'X)^{-1}X'\sigma^2 I X(X'X)^{-1} = \\ &= \sigma^2(X'X)^{-1} \;.\end{aligned}$$

By LIE, we have then

$$\mathbb{E}_X[\mathbb{E}[\hat{\beta}|X]] = \mathbb{E}[\hat{\beta}] = \beta \ ,$$

as well as

$$\mathbb{V}[\hat{\beta}] = \mathbb{E}[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] = \mathbb{E}[\mathbb{E}[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'|X]] = \sigma^2\mathbb{E}[(X'X)^{-1}] \ .$$

Thus, we know that (in finite samples) $\hat{\beta}$ has mean β and variance $\sigma^2\mathbb{E}[(X'X)^{-1}]$. **[2p each]**

When conditioning on X, we know not only the mean and variance of $\hat{\beta}|X$, but we also know that $\hat{\beta}|X$ is normally distributed, as it is the (weighted) sum of Normally distributed Random Variables (RVs):

$$\hat{\beta}|X = \beta + (X'X)^{-1}X'U|X \sim N(\beta, \sigma^2(X'X)^{-1}) \ ,$$

as $U|X \sim N(0, \sigma^2 I)$ and the X-terms are just constants when we condition on X (and β is a constant under our frequentist perspective). **[2p]**

However, note that we do not know the unconditional distribution of $\hat{\beta}$ (i.e. we do not know whether it is Normally distributed or not in finite samples), we only know its mean and variance. This is because we do not know (we did not make any assumptions on) the distribution of X.

(d) (3 points) Is your estimator $\hat{\beta}$ consistent? Use the expression that you derived for it in (b) and theorems to answer this question.

Solution:

Since our sample is i.i.d., we can apply the WLLN and Slutsky's theorem to prove that $\hat{\beta}$ is consistent. We have

$$\hat{\beta} - \beta = \left(\frac{X'X}{n}\right)^{-1}\frac{X'U}{n} = \left(\frac{1}{n}\sum_i x_i x_i'\right)^{-1}\frac{1}{n}\sum_i x_i u_i \ .$$

By WLLN,

$$\frac{1}{n}\sum_i x_i x_i' \xrightarrow{p} \mathbb{E}[X'X] \equiv Q \ ,$$

and then by Slutsky, Slutsky: g( ) = ( )^-1

$$\left(\frac{X'X}{n}\right)^{-1} = \left(\frac{1}{n}\sum_i x_i x_i'\right)^{-1} \xrightarrow{p} \mathbb{E}[(X'X)^{-1}] = Q^{-1} \ .\textbf{[1p]}$$

Also by WLLN, we have

$$\frac{1}{n}X'U = \frac{1}{n}\sum_i x_i u_i \xrightarrow{p} \mathbb{E}[x_i u_i] = 0 \ . \quad \textbf{[1p]}$$

Hence, putting the two pieces together (again using Slutsky's theorem), we get

$$\hat{\beta} - \beta \xrightarrow{p} Q^{-1}0 = 0 \ . \quad \textbf{[1p]}$$

(e) (4 points) What is the asymptotic distribution of $\hat{\beta}$? Use the expression that you derived for it in (b) and theorems to answer this question.

Solution:

Since our sample is i.i.d., we can apply the CLT to prove that $\hat{\beta}$ is asymptotically Normal. We have

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n} \sum_i x_i x_i' \right)^{-1} \frac{1}{\sqrt{n}} \sum_i x_i u_i \ .$$

By CLT,

= (X'X)/n = (X'X)/sqrt(n)

$$\frac{1}{\sqrt{n}} \sum_i x_i u_i = \sqrt{n} \left(\frac{1}{n} \sum_i x_i u_i - \mathbb{E}[x_i u_i] \right) \xrightarrow{d} N(0, \mathbb{E}[(x_i u_i)(x_i u_i)']) \ ,$$

where we know that $\mathbb{E}[x_i u_i] = \mathbb{E}[\mathbb{E}[x_i u_i | x_i]] = \mathbb{E}[x_i \mathbb{E}[u_i | x_i]] = 0$ and

$$\mathbb{E}[(x_i u_i)(x_i u_i)'] = \mathbb{E}[x_i x_i' u_i^2] = \mathbb{E}[\mathbb{E}[x_i x_i' u_i^2 | x_i]] = \mathbb{E}[x_i x_i' \mathbb{E}[u_i^2 | x_i]] = \sigma^2 \mathbb{E}[x_i x_i'] = \sigma^2 Q \ .$$

[2p] From before, we know that $\left(\frac{1}{n} \sum_i x_i x_i' \right)^{-1} \xrightarrow{p} Q^{-1}$. Putting the two pieces together using Slutsky's theorem, we get

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} Q^{-1} N(0, \sigma^2 Q) Q^{-1} = N(0, \sigma^2 Q^{-1}) \ . \quad \textbf{[2p]}$$

(f) (4 points) Suppose that you cannot find the finite sample distribution of $\hat{\beta}$. Describe two approaches to approximate it.

Solution:

Approach 1: Bootstrapping. **[1p]** Take M times n random draws with replacement from the original dataset (X, Y). This gives you M different samples of size n. For each of these M newly-generated samples, find $\hat{\beta}^{(m)}$. If your original sample is truly random and representative of the true population, then the sequence of the estimated $\left\{ \hat{\beta}^{(m)} \right\}_{m=1}^{M}$ approximates the finite sample distribution of $\hat{\beta}$. **[1p]**

Approach 2: Approximate the finite sample distribution using the asymptotic distribution of $\hat{\beta}$. **[1p]** From $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$, we can argue that $\hat{\beta} \overset{approx.}{\sim} N\left(\beta, \frac{1}{n}\sigma^2 Q^{-1}\right)$. Replacing σ^2 and Q^{-1} with consistent estimators $\hat{\sigma}^2$ and $\hat{Q}^{-1} = \left(\frac{1}{n} \sum_i x_i x_i'\right)^{-1}$, we get

$$\hat{\beta} \overset{approx.}{\sim} N\left(\beta, \frac{1}{n} \hat{\sigma}^2 \hat{Q}^{-1}\right) \ .$$

This approach works well for n large. **[1p]**

(g) (6 points) Suppose that you cannot find an analytical expression for $\hat{\beta}$. Use the simplified version of the extremum estimation theory discussed in class to analyze whether $\hat{\beta}$ is consistent and to find its asymptotic distribution. For simplicity, assume you know σ^2 (i.e. you estimate only β, conditioning on σ^2).

Solution:

The extremum estimation theory shows asymptotic properties (consistency and asymptotic Normality) for an estimator defined as

$$\hat{\theta} = \arg\min_{\theta} \; Q_n(\theta|Z^n) \; ,$$

where Z^n denotes all the data. In our case, we have $\hat{\beta}_{ML} = \arg\min_{\beta} \; Q_n(\beta|Y^n, X^n)$ for

$$Q_n(\beta|Y^n, X^n) = -\frac{1}{n}\ell(\beta|X,Y) = -\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)^2 \; . \quad \textbf{[1p]}$$

Replacing $\frac{1}{n}\sum\limits_{i=1}^{n}$ by $\mathbb{E}$, we can see that $Q_n(\beta|Y^n, X^n)$ converges in probability to $Q(\beta) = \mathbb{E}[-(y_i - x_i'\beta)^2]$.[1] Moreover, $Q(\beta)$ is clearly a continuous function. Finally, $Q(\beta)$ is uniquely minimised by β_0, because $\mathbb{E}[y_i|x_i] = x_i'\beta_0$ and we know that $\mathbb{E}[(y_i - f(x_i))^2]$ is uniquely minimised by $f(x_i) = \mathbb{E}[y_i|x_i]$.[2] This proves consistency. **[2p]**

[1] The more precise calculation establishes uniform convergence in probability, i.e.

$$\mathbb{P}\left[\sup_{\beta} |Q_n(\beta|Y^n, X^n) - Q(\beta)| < \varepsilon\right] \to 1 \; , \quad \text{or} \quad \sup_{\beta} |Q_n(\beta|Y^n, X^n) - Q(\beta)| \xrightarrow{p} 0 \; .$$

To show this, first note that

$$Q(\beta) = \mathbb{E}[(y_i - x_i'\beta)^2] = \mathbb{E}[y_i^2] - 2\mathbb{E}[y_i x_i'\beta] + \mathbb{E}[(x_i'\beta)^2] \; .$$

Using this, we can see that

$$\begin{aligned} &\sup_{\beta}\left|\frac{1}{n}\sum_i y_i^2 - \frac{2}{n}\sum_i y_i x_i'\beta + \frac{1}{n}\sum_i (x_i'\beta)^2 - \mathbb{E}[y_i^2] + 2\mathbb{E}[y_i x_i'\beta] - \mathbb{E}[(x_i'\beta)^2]\right| \\ &= \sup_{\beta}\left|\frac{1}{n}\sum_i y_i^2 - \mathbb{E}[y_i^2]\right| - \sup_{\beta}\left|\frac{2}{n}\sum_i y_i x_i'\beta - 2\mathbb{E}[y_i x_i'\beta]\right| + \sup_{\beta}\left|\frac{1}{n}\sum_i (x_i'\beta)^2 - \mathbb{E}[(x_i'\beta)^2]\right| \xrightarrow{p} 0 \end{aligned}$$

by WLLN.

[2] Here's an alternative way to show that, without making use of the stated result:

$$\begin{aligned} Q(\tilde{\beta}) = \mathbb{E}[(y_i - x_i'\tilde{\beta})^2] &= \mathbb{E}[(y_i - x_i'\beta_0 - x_i'(\tilde{\beta} - \beta_0))^2] \\ &= \mathbb{E}_X[\mathbb{E}[(y_i - x_i'\beta_0 - x_i'(\tilde{\beta} - \beta_0))^2|X]] \\ &= \mathbb{E}_X[\mathbb{E}[(u_i - x_i'(\tilde{\beta} - \beta_0))^2|X]] \\ &= \mathbb{E}_X[\mathbb{E}[u_i^2|X] + \mathbb{E}[(x_i'(\tilde{\beta} - \beta_0))^2|x_i]] \\ &= \sigma^2 + \mathbb{E}_X[\mathbb{E}[(\tilde{\beta} - \beta_0)'x_i x_i'(\tilde{\beta} - \beta_0)|X]] \\ &= \sigma^2 + (\tilde{\beta} - \beta_0)'\mathbb{E}[x_i x_i'](\tilde{\beta} - \beta_0) \\ &\geq \sigma^2 = Q(\beta_0) \; , \end{aligned}$$

To calculate the asymptotic distribution, we make use of the score,

$$Q_n^{(1)}(\beta, Y^n, X^n) = \frac{\partial Q_n}{\partial \beta} = \frac{2}{n}\sum_{i=1}^n x_i'(y_i - x_i'\beta) \quad ,$$

and the Hessian,

$$Q_n^{(2)}(\beta, Y^n, X^n) = \frac{\partial^2 Q_n}{\partial \beta \partial \beta'} = \frac{\partial Q_n^{(1)}}{\partial \beta'} = \frac{2}{n}\sum_i x_i x_i' \ .$$

By CLT, the score function evaluated at β_0 and scaled by $\sqrt{n}$ converges in distribution to

$$\begin{aligned}\sqrt{n}Q_n^{(1)}(\beta_0, Y^n, X^n) &= \frac{2}{\sqrt{n}}\sum_i x_i' u_i \\ &= 2\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n x_i'(y_i - x_i'\beta) - \mathbb{E}[x_i'(y_i - x_i'\beta)]\right) \\ &\xrightarrow{d} N(0, M) \ ,\end{aligned}$$

where $M = 4\mathbb{E}[(x_i'u_i)(x_i'u_i)'] = 4\sigma^2\mathbb{E}[x_i x_i'] = 4\sigma^2 Q$, whereas – by WLLN – the Hessian converges in probability to

$$Q_n^{(2)}(\beta, Y^n, X^n) \xrightarrow{p} H = 2\mathbb{E}[x_i x_i'] = 2Q \ .$$

[1p each: score & Hessian] The extremum estimation theory tells us then that

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, H^{-1}MH^{-1}) = N(0, \sigma^2 Q^{-1}) \ . \quad \textbf{[1p]}$$

(h) (6 points) Suppose you have $k = 4$ regressors (including the intercept: $x_{i1} = 1 \ \forall \ i$), and you want to test $\mathcal{H}_0 : \{ \ \beta_2 + 3\beta_3 = 7 \ , \ \log \beta_4 = 0 \ \}$ against the alternative $\mathcal{H}_1$, specifying that at least one of the two conditions in $\mathcal{H}_0$ is not true. Describe two approaches to conduct this test and their relative advantages/disadvantages.

Solution:

[max. 3p per approach, max. 6p overall]

Approach 1: Wald test. **[1p]** We are testing $\mathcal{H}_0 : g(\beta) = 0$, where

$$g(\beta) = \begin{bmatrix} 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \beta - \begin{bmatrix} 7 \\ 1 \end{bmatrix}$$

where we used the fact that $(x_i'(\tilde{\beta} - \beta_0))^2$ is a scalar and hence can be written as $(x_i'(\tilde{\beta} - \beta_0))'(x_i'(\tilde{\beta} - \beta_0))$ as well as the fact that $\mathbb{E}[x_i x_i']$ is positive-definite, which implies that $(\tilde{\beta} - \beta_0)'\mathbb{E}[x_i x_i'](\tilde{\beta} - \beta_0) \geq 0$ and is equal to zero only at $\tilde{\beta} = \beta_0$.

(because testing $\log(\beta_4) = 0$ is equivalent to testing $\beta_4 = 1$).[3] We have the test-statistic

$$T_W = ng(\hat{\beta})[G\hat{V}G]^{-1}g(\hat{\beta})' \to \chi^2_2 \ ,$$

where

$$G = \frac{\partial g(\beta)}{\partial \beta} = \begin{bmatrix} 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \ ,$$

and

$$V = A\mathbb{V}[\hat{\beta}] = \sigma^2 Q^{-1} \ ,$$

with consistent estimator $\hat{V} = \hat{\sigma}^2\hat{Q}^{-1}$. The advantage of the Wald test is that we only need the unrestricted estimator $\hat{\beta}$, not an additional, unrestricted estimator that imposes $\mathcal{H}_0$ in the model. **[2p]**

Approach 2: Likelihood Ratio (LR) test. **[1p]** It uses the test statistic

$$T_{LR} = -2[\ell(\hat{\beta}|Y,X) - \ell(\bar{\beta}|Y,X)] \to \chi^2_2 \ ,$$

where $\ell(\beta|Y,X)$ is log-likelihood function defined above, $\hat{\beta}$ is estimator of unrestricted model,

$$y_i = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u_i \ , \quad u_i \sim N(0,\sigma^2) \ ,$$

and $\bar{\beta} = [\bar{\beta}_1, (7 - 3\bar{\beta}_3), \bar{\beta}_3, 1]$ is estimator of restricted model,

$$y_i = \beta_1 + (7 - 3\beta_3)x_2 + \beta_3 x_3 + 1x_4 + v_i \ , \quad v_i \sim N(0,\sigma^2) \ .$$

The advantage of the LR test is that it is the uniformly most powerful test. The disadvantage is that we need $\hat{\beta}$ as well as $\bar{\beta}$. **[2p]**

Approach 3: Lagrange Multiplier (LM) test. **[1p]** It uses the test-statistic

$$T_{LM} = s(\bar{\beta}|Y,X)'I(\bar{\beta})^{-1}s(\bar{\beta}|Y,X) \to \chi^2_2 \ ,$$

where

$$s(\bar{\beta}|Y,X) = \frac{\partial}{\partial \beta} l(\beta|Y,X)|_{\beta=\bar{\beta}} = \frac{1}{\sigma^2}\sum_i x_i(y_i - x_i'\bar{\beta})$$

is the score evaluated at $\bar{\beta}$, and

$$I(\bar{\beta}) = -H(\beta|Y,X) = -\frac{\partial}{\partial \beta} s(\beta|Y,X) = \frac{1}{\sigma^2}\sum_i x_i x_i' \ .$$

The advantage of this test is that we need only $\bar{\beta}$. **[2p]**

[3] Alternatively, one can also use

$$g(\beta) = \begin{bmatrix} 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\beta + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} log\, \beta - \begin{bmatrix} 7 \\ 0 \end{bmatrix} \ .$$

(i) (4 points) Construct a 95% confidence interval for $exp\{\beta_4\}$.

Solution: There are several approaches to this question. The easiest approach is to find a CI for β_4 and then turn it into a CI for $exp\{\beta_4\}$. We can find a CI for β_4 by using the finite sample distribution of $\hat{\beta}_4$. Because we do not know it, we approximate it using the asymptotic distribution (see exercise (f) above):

$$\hat{\beta}_4 \overset{approx.}{\sim} N\left(\beta_4, \hat{V}\right) \ , \quad \text{where } \hat{V} = \frac{1}{n}\hat{\sigma}^2\left(\hat{Q}^{-1}\right)_{44} \ ,$$

and $\left(\hat{Q}^{-1}\right)_{44}$ denotes the element (4,4) in the matrix $\hat{Q}^{-1}$. (Alternatively, if we condition on X, then this distribution is exact already in finite samples; see exercise (c).) **[1p]** Given this Normal distribution, we can construct the 95% CI for β_4 as usual:

$$\text{CI}_{95\%}(\hat{\beta}_4) = \left[\hat{\beta}_4 - 1.96\sqrt{\hat{V}} \ , \ \hat{\beta}_4 + 1.96\sqrt{\hat{V}}\right] \ . \quad \textbf{[1p]}$$

We can then turn it into a CI for $exp\{\beta_4\}$ by taking the exponent of both bounds of this interval:

$$\text{CI}_{95\%}(exp\{\beta_4\}) = \left[exp\left\{\hat{\beta}_4 - 1.96\sqrt{\hat{V}}\right\} \ , \ exp\left\{\hat{\beta}_4 + 1.96\sqrt{\hat{V}}\right\}\right] \ . \quad \textbf{[2p]}$$

This works because if the probability that β_4 lies in $\text{CI}_{95\%}$ is 95%, then the probability that $exp\{\beta_4\}$ lies in $\text{CI}_{95\%}(exp\{\beta_4\})$ is 95%, too, i.e. we have a valid CI for $exp\{\beta_4\}$. Note that this method does take into account that $exp\{\beta_4\} > 0$, as we get a CI that lies on $\mathbb{R}_{++}$.

A related approach (which should give the same results...) is to conclude that – because $\hat{\beta}_4$ is Normal – $exp\{\hat{\beta}_4\}$ is log-Normal, and construct directly a CI for $exp\{\hat{\beta}_4\}$ based on the log-Normal distribution.

Another approach is to use the Delta method to find the asymptotic distribution of $exp\{\hat{\beta}_4\}$ and construct an asymptotically valid CI for $exp\{\beta_4\}$. We know

$$\sqrt{n}(\hat{\beta}_4 - \beta_4) \xrightarrow{d} N(0, V) \ , \quad V = A\mathbb{V}[\hat{\beta}_4] = \sigma^2(\mathbb{E}[x_i x_i']^{-1})_{(4,4)} \ .$$

Defining $g(\hat{\beta}_4) = \exp\left\{\hat{\beta}_4\right\}$, the Delta method tells us that

$$\sqrt{n}(g(\hat{\beta}_4) - g(\beta_4)) \xrightarrow{d} N(0, GVG') = N(0, G^2V) \ ,$$

where

$$G = \frac{\partial g(\beta_4)}{\partial \beta_4} = \exp\{\beta_4\} \ .$$

Based on this, we can approximate the finite sample distribution of $g(\hat{\beta}_4)$ as

$$g(\hat{\beta}_4) \overset{\text{approx.}}{\sim} N\left(g(\beta_4) \ , \ \frac{1}{n}\hat{G}^2\hat{V}\right) = N\left(g(\beta_4) \ , \ \frac{1}{n}\exp\left\{2\hat{\beta}_4\right\}\hat{\sigma}^2\left(\left(\frac{1}{n}\sum_i x_i x_i'\right)^{-1}\right)_{(4,4)}\right) \ ,$$

[2p] and, given this Normal distribution, we can construct the 95% CI as usual:

$$\text{CI}_{95\%} = \left[\exp\left\{\hat{\beta}_4\right\} - 1.96\sqrt{n^{-1}\hat{G}^2\hat{V}} \;\; , \;\; \exp\left\{\hat{\beta}_4\right\} + 1.96\sqrt{n^{-1}\hat{G}^2\hat{V}}\right] . \quad \textbf{[2p]}$$

Note that this approach is only valid asymptotically, and it does not restrict $g(\beta_4)$ to be positive. Hence, it is very well possible that one gets a negative lower bound of the CI (in any finite sample). That's why the approaches above are clearly preferred.

[max. 4p for either of the two approaches, max. 2p for sketching the approach via log-Normal, max. 2p for discussing the relation of approaches, max. 4p overall]