ECOM40006/90013 Econometrics 3

Week 9 Extras (Part 2 Solutions)

Question 1: The Cramér-Rao lower bound

(a) Result 1. First, observe that via the Chain Rule,

$$\begin{aligned}\frac{\partial \log f(y;\theta)}{\partial \theta} &= \frac{\partial \log f(y;\theta)}{\partial f(y;\theta)} \frac{\partial f(y;\theta)}{\partial \theta} \\ &= \frac{1}{f(y;\theta)} \frac{\partial f(y;\theta)}{\partial \theta}.\end{aligned}$$

Moving the $f(y;\theta)$ to the left hand side of the equality yields

$$\frac{\partial f(y;\theta)}{\partial \theta} = \frac{\partial \log f(y;\theta)}{\partial \theta} f(y;\theta),$$

as required.

(b) Result 2. As suggested in the questions, first define

$$\begin{aligned}V &= X - \mathbb{E}(X) \\ W &= Y - \mathbb{E}(Y).\end{aligned}$$

Then consider the function

$$q(t) = \mathbb{E}\left[(V - tW)^2\right].$$

Since $(V - tW)^2$ is a squared expression, it follows that for any t, $(V - tW)^2 \geq 0$. The reason why is because the expression $(V - tW)^2$ is a random variable in its own right. If it is only ever non-negative, then its expectation should also be non-negative as well!

Expanding out the brackets, we now have

$$\begin{aligned}q(t) &= \mathbb{E}(V^2 + 2tVW + t^2W^2) \\ &= t^2\mathbb{E}(W^2) + 2t\mathbb{E}(VW) + \mathbb{E}(V^2) \\ &= at^2 + bt + c,\end{aligned}$$

where $a = \mathbb{E}(W^2)$, $b = 2\mathbb{E}(VW)$ and $c = \mathbb{E}(V^2)$, with all expressions being constants (due to the property that expectations are numbers).[1] Paired with the conclusion that $q(t) \geq 0$ and that we can write $q(t)$ as a parabola, it follows that the parabola can only

[1] Of course, conditional expectations are functions, but that's a story for another day.

ever intersect the horizontal axis (or t-axis) either once, or not at all.

This implies that the discriminant from the quadratic formula[2] must be non-positive:

$$\begin{aligned} b^2 - 4ac &\leq 0 \\ \implies 4\mathbb{E}(VW)^2 - 4\mathbb{E}(W^2)\mathbb{E}(V^2) &\leq 0 \\ \implies \mathbb{E}(VW)^2 - \mathbb{E}(W^2)\mathbb{E}(V^2) &\leq 0 \\ \implies \mathbb{E}(VW)^2 &\leq \mathbb{E}(W^2)\mathbb{E}(V^2) \\ \implies \frac{\mathbb{E}(VW)^2}{\mathbb{E}(V^2)\mathbb{E}(W^2)} &\leq 1. \end{aligned}$$

However, substituting in our expressions for V and W, we have

$$\mathbb{E}(VW) = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right] = \text{cov}(X, Y)$$

and similarly $\mathbb{E}(V^2) = \text{Var}(X)$ and $\mathbb{E}(W^2) = \text{Var}(Y)$. Making these substitutions, we have

$$\begin{aligned} \frac{\text{cov}(X,Y)^2}{\text{Var}(X)\text{Var}(Y)} &\leq 1 \\ \implies \left[\frac{\text{cov}(X,Y)}{\text{sd}(X)\text{sd}(Y)}\right]^2 &\leq 1 \\ \implies \rho^2 &\leq 1, \end{aligned}$$

as required.

(c) Consider the expression from the question

$$\mathbb{E}(\tilde{\theta}) = \int_Y \tilde{\theta}(y) f(y;\theta)\, dy.$$

How do we take this apart? A good first step would be to figure out where the source of randomness is. For this problem, the randomness comes from the dataset y, so we are integrating over y in this case. The choice of distribution on the part of the econometrician then determines the form of $f(y;\theta)$.

Finally, the realizations of y (i.e. the actual draws themselves) will determine what the estimator $\tilde{\theta}$ is. After all, in our with maximum likelihood so far, our goal is to make inferences on θ using the data y. An estimator constructed out of the data y is then a function of y (recall some of the earlier examples of maximum likelihood we've dealt with, where the MLE is the sample mean – that can be written in the form $\hat{\theta}(y)$!).[3]

[2] The quadratic formula is $t = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$, where the expression $\Delta = b^2 - 4ac$ is the discriminant. The sign of the discriminant dictates how many (real) solutions the quadratic has, with two solutions if $\Delta > 0$, one solution if $\Delta = 0$ and no solutions if $\Delta < 0$.

[3] A side question that we can also ask is what happens if the econometrician's choice of probability mass/density function is the incorrect choice – for example, choosing a normal distribution when the data is in fact uniformly distributed. This is a topic that we leave for another day, and perhaps other subjects. But if you're interested, you can look up *quasi-maximum likelihood estimation* and its close relative *pseudo-maximum likelihood estimation* to see how we deal with those issues in theory.

(d) First, observe that from above, we can write

$$\int_Y \tilde{\theta}(y) f(y;\theta)\,dy = \theta.$$

Take the partial derivative with respect to θ on both sides and exchange the order of integration and differentiation by invoking regularity conditions:

$$\begin{aligned}
\frac{\partial}{\partial\theta}\int_Y \tilde{\theta}(y) f(y;\theta)\,dy &= \frac{\partial}{\partial\theta}\theta \\
\implies \int_Y \tilde{\theta}(y)\frac{\partial f(y;\theta)}{\partial\theta}\,dy &= 1 \\
\implies \int_Y \tilde{\theta}(y)\frac{\partial \log f(y;\theta)}{\partial\theta} f(y;\theta)\,dy &= 1 \qquad \text{(Result 1)} \\
\implies \mathbb{E}\left[\tilde{\theta}(y)\frac{\partial \log f(y;\theta)}{\partial\theta}\right] &= 1,
\end{aligned}$$

as required.

(e) First, expand the expectation into integral form:

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial \log f(y;\theta)}{\partial\theta}\right] &= \int_Y \frac{\partial \log f(y;\theta)}{\partial\theta} f(y;\theta)\,dy \\
&= \int_Y \frac{\partial f(y;\theta)}{\partial\theta}\,dy \qquad \text{(Result 1 in reverse)} \\
&= \frac{\partial}{\partial\theta}\underbrace{\int_Y f(y;\theta)\,dy}_{=1} \\
&= \frac{\partial}{\partial\theta}(1) \\
&= 0,
\end{aligned}$$

where in the third last line we use regularity conditions to exchange the integral and partial derivative, also further making use of the fact that

$$\int_Y f(y;\theta) = 1,$$

since any proper density integrates to 1 over its entire support.

(f) **Deriving the CRLB.** If we define

$$u = \tilde{\theta}(y) \qquad \text{and} \qquad v = \frac{\partial \log f(y;\theta)}{\partial\theta},$$

we can recast our result from (d) as

$$\mathbb{E}(uv) = 1.$$

Furthermore, we can write the covariance between the two terms as

$$\text{cov}(u,v) = \mathbb{E}(uv) - \mathbb{E}(u)\mathbb{E}(v) = \mathbb{E}(uv),$$

since we have $\mathbb{E}(v) = 0$ from deriving that the score has zero expecation (remember, v here is the score!). Furthermore, using the corollary from Result 2, we also have

$$\begin{aligned}
\text{cov}(u, v)^2 &\leq \text{Var}(u)\text{Var}(v) \\
\implies 1 &\leq \text{Var}(u)\text{Var}(v) && (\text{cov}(u, v) = 1) \\
\implies \text{Var}(u) &\geq \frac{1}{\text{Var}(v)}.
\end{aligned}$$

Since v is the score, we have that $\text{Var}(v)$ is the variance of the score, or the *information matrix* $I(\theta)$. Replacing u and v with their definitions, we have

$$\text{Var}(\tilde{\theta}) \geq \frac{1}{I(\theta)} = I(\theta)^{-1},$$

as required.

(g) Result 3. If equality holds in Result 2, this implies there exists some number t_0 for which

$$q(t_0) = \mathbb{E}\left[(V - t_0 W)^2\right] = 0.$$

Furthermore, if we substitute our values back in, observe that

$$V - t_0 W = X - \mathbb{E}(X) - t_0[Y - \mathbb{E}(Y)]$$

and that

$$\begin{aligned}
\mathbb{E}(V - t_0 W) &= \mathbb{E}\left[X - \mathbb{E}(X) - t_0[Y - \mathbb{E}(Y)]\right] \\
&= \mathbb{E}(X) - \mathbb{E}(X) - t_0[\mathbb{E}(Y) - \mathbb{E}(Y)] \\
&= 0,
\end{aligned}$$

since $\mathbb{E}(\mathbb{E}(X)) = \mathbb{E}(X)$ and $\mathbb{E}(\mathbb{E}(Y)) = \mathbb{E}(Y)$ as expectations are treated as constant. Furthermore, at $t = t_0$ we can write

$$\begin{aligned}
\text{Var}(V - t_0 W) &= \mathbb{E}\left[(V - t_0 W)^2\right] - \mathbb{E}\left[(V - t_0 W)\right]^2 \\
&= \mathbb{E}\left[(V - t_0 W)^2\right] \\
&= q(t_0) \\
&= 0.
\end{aligned}$$

In other words, at $t = t_0$, the random variable $V - t_0 W$ has zero expectation and zero variance. That is,

$$V - t_0 W = 0 \qquad \text{with probability one.}$$

Therefore, rearrangement gives

$$\begin{aligned}
V &= t_0 W \\
\implies X - \mathbb{E}(X) &= t_0[Y - \mathbb{E}(Y)],
\end{aligned}$$

as required.[4]

[4] If you want to keep going, you can continue rearranging to get $Y = aX + b$ where $a = \frac{1}{t_0}$ and $b = \mathbb{E}(Y) - \frac{1}{t_0}\mathbb{E}(X)$.

(h) Attaining the CRLB. The critical line in our working where we bring in the inequality was the line

$$\text{cov}(u, v)^2 \leq \text{Var}(u)\text{Var}(v).$$

Result 3 tells us that if this holds with inequality, we can write

$$v - \mathbb{E}(v) = t_0[u - \mathbb{E}(u)],$$

with t_0 to be determined. Since $\mathbb{E}(v)$ is the expectation of the score, we know that $\mathbb{E}(v) = 0$. Furthermore, since $\tilde{\theta}$ is unbiased for θ, we have that $\mathbb{E}(u) = \theta$. Therefore we can write

$$v = t_0[u - \mathbb{E}(u)].$$

But what exactly is t_0? To find this out, we can go back to our earlier line and rearrange:

$$v - \mathbb{E}(v) = t_0[u - \mathbb{E}(u)] \implies u - \mathbb{E}(u) = \frac{v - \mathbb{E}(v)}{t_0}.$$

We can plug this into the covariance between u and v:

$$\begin{aligned}
\text{cov}(u, v) &= \mathbb{E}[(u - \mathbb{E}(u))(v - \mathbb{E}(v))] \\
&= \mathbb{E}\left[\frac{1}{t_0}(v - \mathbb{E}(v))^2\right] \\
&= \frac{1}{t_0}\text{Var}(v).
\end{aligned}$$

Since we also have that $\text{cov}(u, v) = 1$ from earlier, we have

$$1 = \frac{1}{t_0}\text{Var}(v) \implies \text{Var}(v) = t_0 \implies I(\theta) = t_0.$$

So to complete the derivation, a unbiased estimator $\tilde{\theta}$ of θ attains the CRLB if it can be written in the form

$$S(\theta) = I(\theta)[\tilde{\theta} - \theta],$$

as required.

(i) Bernoulli and the CRLB. The log-likelihood for the Bernoulli distribution is

$$\begin{aligned}
\log L(\theta) &= \sum_{i=1}^{n} \log p(y_i; \theta) \\
&= \sum_{i=1}^{n} \log \theta^{y_i}(1 - \theta)^{1-y_i} \\
&= \log \theta \sum_{i=1}^{n} y_i + \log(1 - \theta) \sum_{i=1}^{n}(1 - y_i).
\end{aligned}$$

The score is

$$\begin{aligned} S(\theta) &= \frac{\partial \log L(\theta)}{\partial \theta} \\ &= \frac{1}{\theta}\sum_{i=1}^{n} y_i - \frac{1}{1-\theta}\sum_{i=1}^{n}(1-y_i). \end{aligned}$$

Since we are already aware that the information matrix is $I(\theta) = \frac{n}{\theta(1-\theta)}$ we can try and manipulate the score to get it into the appropriate form. First, expanding out the $(1-y_i)$ in the summation and continuing,

$$\begin{aligned} S(\theta) &= \frac{1}{\theta}\sum_{i=1}^{n} y_i - \frac{n}{1-\theta} + \frac{1}{1-\theta}\sum_{i=1}^{n} y_i \\ &= \left(\frac{1}{\theta} + \frac{1}{1-\theta}\right)\sum_{i=1}^{n} y_i - \frac{n}{1-\theta} \\ &= \frac{1}{\theta(1-\theta)}\sum_{i=1}^{n} y_i \times \frac{n}{n} - \frac{n}{1-\theta} \times \frac{\theta}{\theta} \\ &= \frac{n}{\theta(1-\theta)}\left[\frac{1}{n}\sum_{i=1}^{n} y_i - \theta\right] \\ &= I(\theta)[\tilde{\theta} - \theta], \end{aligned}$$

where $\tilde{\theta}$ is the sample mean

$$\tilde{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

It turns out that this is also the MLE $\hat{\theta}$ for the problem! So in the case of the Bernoulli distribution, the MLE attains the Cramér-Rao lower bound.

Aside: if you want to show that $\tilde{\theta}$ is also unbiased for θ all you need to do is take the expectation using the fact that for a Bernoulli distribution, $\mathbb{E}(y_i) = \theta$:

$$\begin{aligned} \mathbb{E}(\tilde{\theta}) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(y_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} \theta \\ &= \frac{n\theta}{n} \\ &= \theta, \end{aligned}$$

so $\tilde{\theta}$ is unbiased for θ.

Question 2: MLE Asymptotics

(a) Using Jensen's inequality as directed implies that[5]

$$
\begin{aligned}
\mathbb{E}\left[\log\left(\frac{L(\theta)}{L(\theta_0)}\right)\right] &\leq \log\left[\mathbb{E}\left(\frac{L(\theta)}{L(\theta_0)}\right)\right] \\
&= \log\left[\int \frac{L(\theta)}{L(\theta_0)} f(Y|\theta_0)\, dY\right] \\
&= \log\left[\int_Y f(Y|\theta)\, dY\right] \\
&= \log 1 \\
&= 0.
\end{aligned}
$$

Note that between the second and third lines, we are using the fact that

$$L(\theta_0) = f(Y|\theta_0).$$

Namely: the likelihood function is actually just shorthand notation for a joint (conditional) probability density that depends on the true parameter vector θ_0.

Hence, this implies that

$$
\begin{aligned}
&\mathbb{E}\left[\log\left(\frac{L(\theta)}{L(\theta_0)}\right)\right] = \mathbb{E}(\log(\theta)) - \mathbb{E}(\log L(\theta_0)) \leq 0 \\
\implies\ &\mathbb{E}(\log L(\theta_0)) \geq \mathbb{E}(\log L(\theta)).
\end{aligned}
$$

Therefore, we could conclude that for any $\theta \in \Theta$, the expected log likelihood is not as large as that calculated by θ_0. So we have that θ_0 is the maximizer of the expected log-likelihood. Specifically,

$$\theta_0 = \arg\max_{\theta\in\Theta} \mathbb{E}(\log L(\theta)).$$

(b) Consider the sample log-likelihood:

$$\frac{1}{n}\sum_{i=1}^{n} \log f(y_i;\theta) \xrightarrow{p} \mathbb{E}(\log L(\theta)) \qquad \text{by the WLLN.}$$

Why does this matter? For any n, define

$$L_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log f(y_i;\theta).$$

Further, let $\hat{\theta}_n$ be the MLE associated with $L_n(\theta)$. It maximizes the value of $L_n(\theta)$ by definition. However, we have just determined that from the WLLN, $L_n(\theta) \xrightarrow{p} \mathbb{E}(\log L(\theta))$.

[5] Note that expectations in this scenario are conditional; we just omit it most of the time. Namely: what is the density of a set of data points if we knew the true value? After all, the likelihood is constructed based on the data we see. If said data is truly 'generated' from θ_0, then we can figure out what the likelihood will be, on average. It might seem a little bit odd, but that's how it works.

From part (a), we have already determined that θ_0 is the maximizer for $\mathbb{E}(\log L(\theta)$. But since $\hat{\theta}_n$ is a maximizer also, then the two are the same provided that the likelihood function has a unique maximum. This would allow us to conclude that

$$\hat{\theta}_n \xrightarrow{p} \theta_0,$$

as required.

(c) A first-order approximation is required for $G(\hat{\theta})$ around the true value θ_0:

$$G(\hat{\theta}) \approx G(\hat{\theta}_0) + G'(\theta_0)[\hat{\theta} - \theta_0].$$

At the values of $\hat{\theta}$, $G(\hat{\theta}) = 0$ by definition so

$$\begin{aligned} G(\theta_0) + H(\theta_0)[\hat{\theta} - \theta_0] &\approx 0 \\ \implies \hat{\theta} - \theta_0 &\approx -H^{-1}(\theta_0)G(\theta_0). \end{aligned}$$

We also know that

$$H(\theta) = \sum_{i=1}^{n} h_i(\theta) \qquad \text{and} \qquad G(\theta) = \sum_{i=1}^{n} g_i(\theta)$$

so

$$\begin{aligned} \hat{\theta} - \theta_0 &\approx \left(\sum_{i=1}^{n} h_i(\theta_0) \right)^{-1} \sum_{i=1}^{n} g_i(\theta_0) \\ &= \left(\frac{1}{n} \sum_{i=1}^{n} h_i(\theta_0) \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} g_i(\theta_0) \\ \implies \sqrt{n}(\hat{\theta} - \theta_0) &\approx \left(\frac{1}{n} \sum_{i=1}^{n} h_i(\theta_0) \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_i(\theta_0). \end{aligned}$$

Notice that asymptotically,

$$\frac{1}{n} \sum_{i=1}^{n} g_i(\theta_0) \xrightarrow{p} \mathbb{E}(g_i(\theta_0)) = \mathbb{E}(s_i(\theta_0)) = 0$$

where $s_i(\theta_0)$ is the score for observation i and we make use of the result that the score has zero expected value. Hence, by the CLT,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_i(\theta_0) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} g_i(\theta_0) - 0 \right) \xrightarrow{d} N(0, \mathrm{Var}(S)).$$

Now, one has

$$
\begin{aligned}
h_i(\theta) &= \frac{\partial^2 \log f(x_i;\theta)}{\partial \theta^2} \\
&= \frac{\partial}{\partial \theta}\left[\frac{\partial \log f(x_i;\theta)}{\partial \theta}\right] \\
&= \frac{\partial}{\partial \theta}\left[\frac{f'(x_i;\theta)}{f(x_i;\theta)}\right] \\
&= \frac{\partial}{\partial \theta}\left[\frac{\partial f(x_i;\theta)}{\partial \theta} f(x_i;\theta)^{-1}\right] \\
&= \frac{\partial f(x_i;\theta)}{\partial \theta}\frac{\partial}{\partial \theta} f(x_i;\theta)^{-1} + \frac{1}{f(x_i;\theta)}\frac{\partial^2 f(x_i;\theta)}{\partial \theta^2} && \text{(Product Rule)} \\
&= \underbrace{-\frac{1}{f(x_i;\theta)^2}\frac{\partial f(x_i;\theta)}{\partial \theta}\frac{\partial f(x_i;\theta)}{\partial \theta}}_{\mathbf{(1)}} + \underbrace{\frac{1}{f(x_i;\theta)}\frac{\partial^2 f(x_i;\theta)}{\partial \theta^2}}_{\mathbf{(2)}}. && \text{(Chain Rule)}
\end{aligned}
$$

Now we can deal with each of these expressions one by one. Observe that earlier results imply that

$$
\mathbf{(1)} = \left[\frac{\partial f(x_i;\theta)}{\partial \theta}\frac{1}{f(x_i;\theta)}\right]\left[\frac{\partial f(x_i;\theta)}{\partial \theta}\frac{1}{f(x_i;\theta)}\right] = g_i(\theta)^2.
$$

We also have that

$$
\begin{aligned}
\mathbb{E}[\mathbf{(2)}] &= \int \frac{\partial^2 f(x_i;\theta)}{\partial \theta^2}\frac{1}{f(x_i;\theta)} f(x_i;\theta)\, dx \\
&= \int \frac{\partial^2 f(x_i;\theta)}{\partial \theta^2}\, dx \\
&= \frac{\partial^2}{\partial \theta^2}\int f(x_i;\theta)\, dx \\
&= \frac{\partial^2}{\partial \theta^2} \times 1 \\
&= 0
\end{aligned}
$$

where we swap the orders of integration and differentiations in the third step. We thus conclude that

$$
\mathbb{E}(h(\theta)) = -\mathbb{E}[\mathbf{(1)}] + \mathbb{E}[\mathbf{(2)}] = -\mathbb{E}(g(\theta)^2).
$$

But observe that

$$
\begin{aligned}
\mathrm{Var}(S) &= \mathrm{Var}(g(\theta_0)) \\
&= \mathbb{E}(g(\theta_0)^2) - \mathbb{E}(g(\theta_0))^2 \\
&= \mathbb{E}(g(\theta_0)^2).
\end{aligned}
$$

Evaluating $h_i(\theta)$ at θ_0, we have that the expected Hessian is the negative variance of the score, as required.

Question 3: Hypothesis Testing in R

(a) The descriptive statistics, as provided by `stargazer()`'s default LaTeX output, are:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Sepal.Length	150	5.843	0.828	4.300	5.100	6.400	7.900
Sepal.Width	150	3.057	0.436	2.000	2.800	3.300	4.400
Petal.Length	150	3.758	1.765	1.000	1.600	5.100	6.900
Petal.Width	150	1.199	0.762	0.100	0.300	1.800	2.500

(b) The regression output via `stargazer()` is presented below, with some minor reformatting so that notation is consistent with that given in the question (and also fits on one page):

	Dependent variable:	
	y_i	
	(1)	(2)
$x_{1,i}$	-0.223	0.651^{***}
	(0.155)	(0.067)
$x_{2,i}$		0.709^{***}
		(0.057)
$x_{3,i}$		-0.556^{***}
		(0.128)
Intercept	6.526^{***}	1.856^{***}
	(0.479)	(0.251)
Observations	150	150
R^2	0.014	0.859
Adjusted R^2	0.007	0.856
Residual Std. Error	0.825 (df = 148)	0.315 (df = 146)
F Statistic	2.074 (df = 1; 148)	295.539^{***} (df = 3; 146)
Note:		*p<0.1; **p<0.05; ***p<0.01

(c) (i.) The variance estimator for Equation 1 is

$$\tilde{\sigma}_n^2 = 0.6808.$$

(ii.) The coefficient of determination here is $R^2 = 0.0138$.

(iii.) The coefficient estimate $\hat{\beta}_1$ on $x_{1,i}$ is -0.2234.

(d) (i.) The results from equation 2 can be seen in the `stargazer()` table in part (b).

(ii.) Refer to R for computation – very few people in general are keen on seeing the entirety of a 150-term squared residual series!

(iii.) While not asked for explicitly, you can see the auxiliary output here in a `stargazer()` table:

	Dependent variable:
	$\hat{u}_i^2$
$x_{1,i}$	-0.010
	(0.026)
$x_{2,i}$	0.005
	(0.023)
$x_{3,i}$	0.021
	(0.051)
Constant	0.083
	(0.100)
Observations	150
R^2	0.046
Adjusted R^2	0.027
Residual Std. Error	0.125 (df = 146)
F Statistic	2.368* (df = 3; 146)
Note:	*p<0.1; **p<0.05; ***p<0.01

(iv.) It can be seen from the table in part (iii) that the number of observations n is 150 and the associated coefficient of determination R^2 is 0.046. This implies the LM test statistic

$$LM = nR^2 = 6.9605 \overset{a}{\sim} \chi^2_3,$$

with an associated p-value of 0.073 when compared to the χ^2_3 distribution (which, remember, is an approximation when we have finite n). This p-value does not exceed the rejection threshold of 0.05, so we fail to reject the null hypothesis.

In any case, what even is this null hypothesis business anyway? The Breusch-Pagan test is a pretty neat test of heteroskedasticity, in the sense that intuitively, if a regression is heteroskedastic (i.e. non-constant variance) that kind of behaviour should:

- *show up in the behaviour of the residuals squared, as you use it to calculate variances*
- *also depend on the regressors – after all, that is part of what heteroskedasticity entails!*

So the idea is: if any of the ***slope*** *coefficients in the auxiliary regression are significantly non-zero, that means that the residuals squared will depend on the values of the regressors, giving evidence in favor of heteroskedasticity existing in the model. The null hypothesis here is then stated as follows:*

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$$
$$H_1 : \text{at least one restriction fails.}$$

You might notice that this only assumes a linear relationship, which is of course not the only possible type of heteroskedasticity that could occur. Nonlinearity is captured via the use of another test – the White test. Even so, at the end of the day, such tests serve as general indicators. If the test you're using doesn't fit your criteria, then perhaps it may be worth searching for a test that better suits what one may be trying to do for research!