P218 Econometrics I

TA Session 2

Gabriel Simões Gaspar

London Business School

Fall 2022

Roadmap

1. Question 1
2. Question 2
3. Question 3
4. Question 4
5. Question 5

Question 1

Part (a)

We can start by noting that:

$$f_{x_T,\cdots,x_1|x_0} = \frac{f_{x_T,\cdots,x_0}}{f_{x_0}}$$

But we have that:

$$\begin{aligned}
f_{x_T,\cdots,x_0} &= f_{x_T|x_{T-1},\cdots,x_0} f_{x_{T-1},\cdots,x_0} \\
f_{x_{T-1},\cdots,x_0} &= f_{x_{T-1}|x_{T-2},\cdots,x_0} f_{x_{T-2},\cdots,x_0} \\
f_{x_{T-2},\cdots,x_0} &= f_{x_{T-2}|x_{T-3},\cdots,x_0} f_{x_{T-3},\cdots,x_0} \\
&\cdots \\
f_{x_1,x_0} &= f_{x_1|x_0} f_{x_0}
\end{aligned}$$

Part (a)

We can plug this into the first equation to get:

$$\begin{aligned} f_{x_T,\cdots,x_1|x_0} &= \frac{f_{x_T,\cdots,x_0} f_{x_{T-1},\cdots,x_0} \cdots f_{x_0}}{f_{x_0}} \\ &= f_{x_T,\cdots,x_0} f_{x_{T-1},\cdots,x_0} \\ &= \prod_{t=1}^{T} f_{x_t|x_{t-1}\cdots x_0} \end{aligned}$$

Part (b)

The likelihood function will be:

$$\mathcal{L} = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_t - x_{t-1} - \delta)^2}{2\sigma^2}\right\}$$

So log-likelihood becomes:

$$L = \log(\mathcal{L}) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2}\sum_{t=1}^{T}\frac{(x_t - x_{t-1} - \delta)^2}{\sigma^2}$$

Part (b)

We get the following FOC for δ:

$$\frac{\partial L}{\partial \delta} = \sum_{t=1}^{T}(x_t - x_{t-1} - \hat{\delta}_{MLE}) = 0$$

$$\begin{aligned}
T\hat{\delta}_{MLE} &= \sum_{t=1}^{T} x_t - x_{t-1} \\
T\hat{\delta}_{MLE} &= x_1 - x_0 + x_2 - x_1 + \cdots x_T - x_{T-1} \\
\hat{\delta}_{MLE} &= \frac{x_T}{T}
\end{aligned}$$

Part (b)

And for σ^2:

$$\frac{\partial L}{\partial \sigma^2} = \sum_{t=1}^{T} \frac{(x_t - x_{t-1} - \hat{\delta}_{MLE})^2}{2\hat{\sigma}^4_{MLE}} - \frac{T}{2\hat{\sigma}^2_{MLE}} = 0$$

$$\sum_{t=1}^{T} (x_t - x_{t-1} - \frac{x_T}{T})^2 = T\hat{\sigma}^2_{MLE}$$

$$\hat{\sigma}^2_{MLE} = \frac{1}{T}\sum_{t=1}^{T} (x_t - x_{t-1} - \frac{x_T}{T})^2$$

Part (c)

We can check the lowest variance of δ by finding the inverse of the Fisher information matrix. In this case, that is simply:

$$
\begin{aligned}
(\mathcal{I}(\delta))^{-1} &= \left(-\operatorname{E}\left[\frac{\partial^2 L}{\partial \delta^2}\right]\right)^{-1} \\
&= (T)^{-1} \\
&= \frac{1}{T}
\end{aligned}
$$

Part (d)

Check code for solution. You should get:

$$\hat{\delta}_{MLE} \approx 0.0055$$
$$\hat{\sigma}_{MLE} \approx 0.0422$$

Part (d)

Now we are told that

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

and all ε_i are i.i.d. random variables distributed as $\mathcal{N}(0, \gamma^2)$ and independent from x_i. Does this model satisfy the GM assumptions?

- GM1 : matrix X will have rank 2.
- GM2: $\mathbb{E}[\varepsilon \mid X] = 0$ since ε is independent from X.
- GM3: $\mathrm{Var}(\varepsilon \mid X) = \gamma^2 I_T$ by assumption.

But (y_t, x_t) are not i.i.d. across $t = 1, \cdots, T$. Do we need to assume random sample for GM? How would you test the hypothesis that $\beta = 1$? if $T = 3$? Does GM4 hold?

Question 2

Part (a)

This generates a long output since there will be a dummy for each category of the education and experience variables (check R code for results). R automatically drops the first dummies of education and experience. Why is that? Should we drop any other dummies?

Part (b)

We want to test the null hypothesis that

$$H_0 : R\alpha = q$$

where α is the vector of coefficients being estimated such that α would lead to the linear specification

$$\mathbb{E}[lwage \mid educ, exper] = \beta_0 + \beta_1 \times educ + \beta_2 \times exper + \beta_3 \times exper^2$$

We want to model the CEF of log-wage for the population of people having a given education and experience level.

Part (b)

Imagine we want to look at a population with education levels ranging from 0 to p and with experience levels ranging from 0 to g. We will then have an unrestricted and a restricted model.
Our unrestricted model can be found using the regression we used in part (a) as

$$\begin{aligned} \textit{lwage} = \kappa + \delta_1 \times \textit{educ}_1 + \delta_2 \times \textit{educ}_2 + \cdots + \delta_p \times \textit{educ}_p \\ + \gamma_1 \times \textit{exper}_1 + \gamma_2 \times \textit{exper}_2 + \cdots + \gamma_g \times \textit{exper}_g + \varepsilon \end{aligned}$$

where $\textit{educ}_i$ and $\textit{exper}_j$ are dummies for different levels of education and experience.

Part (b)

Similarly, we will have that our restricted model is

$$lwage = \beta_0 + \beta_1 \times educ + \beta_2 \times exper + \beta_3 \times exper^2 + \varepsilon$$

We would then expect that for an individual with education level i and experience level j the predicted wage from both specifications must equal

$$\kappa + \delta_i + \gamma_j = \beta_0 + \beta_1 \times i + \beta_2 \times j + \beta_3 \times j^2$$

This should hold for all $i = 0, \cdots, p$ and all $j = 0, \cdots, g$. What happened to δ_0 and γ_0?

Part (b)

From the previous relation we can pin down the coefficients of the restricted model. Note that when $i = j = 0$ we will have that

$$\kappa = \beta_0$$

Similarly, for $i \neq 0$ and $j = 0$

$$\delta_i = \beta_1 \times i$$

And if $i = 0$ and $j \neq 0$

$$\gamma_j = \beta_2 \times j + \beta_3 \times j^2$$

Part (b)

In particular, $\gamma_1 = \beta_2 + \beta_3$ and $\gamma_2 = 2\beta_2 + 4\beta_3$. This means that

$$\beta_2 = \frac{4\gamma_1 - \gamma_2}{2}$$
$$\beta_3 = \frac{\gamma_2 - 2\gamma_1}{2}$$

Note that

$$\alpha = \begin{bmatrix} \kappa & \delta_1 & \delta_2 & \cdots & \delta_p & \gamma_1 & \gamma_2 & \cdots & \gamma_g \end{bmatrix}'$$

which is a $(1 + p + g) \times 1$ column vector.

Part (b)

We can summarize the relations between the coefficients in α and the β's from the restricted model in $R\alpha = q$. This leads us to the following:

$$R = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}$$

where $\mathbf{A}$ and $\mathbf{B}$ are, respectively, $(p-3) \times (p+1)$ and $g \times g$ matrices. How do we find those? What is q in this case?

Part (b)

A simple alternative is to set

$$\mathbf{A} = \begin{bmatrix} 0 & -2 & 1 & 0 & \cdots & 0 \\ 0 & -3 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -p & 0 & 0 & \cdots & 1 \end{bmatrix}$$

And

$$\mathbf{B} = \begin{bmatrix} 3 & -3 & 1 & 0 & \cdots & 0 \\ 8 & -6 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -2g + g^2 & \frac{g-g^2}{2} & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Why does this work? Can we compute the F-statistic?

Part (b)

What we can do is model the CEF of log-wage for the population of individuals with only those education and experience levels that we observe. If we assume that α is now a vector where we have dropped δ_2, δ_5, δ_6 and δ_{20} (why?) and γ_{21}, γ_{23} and γ_{24} (why?), we can test the hypothesis.

If we had data for each level of education and experience, we would be testing $20 + 25 - 3 = 42$ restrictions. Now, we have $(20 - 4) + (25 - 3) - 3 = 35$ restrictions on the specification of the CEF. The F-test is

$$F = \frac{(RSS_R - RSS_U)/35}{RSS_U/(1500 - 39)} = \frac{(987.21 - 958.3)/35}{958.3/1461} \approx 1.26 \sim F(35, 1461)$$

Since the 0.95 critical value distribution is 1.4311 we cannot reject the null.

Part (c)

Now we want to test the assumption that

$$\mathbb{E}[wage \mid educ, exper] = \beta_0 + \beta_1 \times educ + \beta_2 \times exper + \beta_3 \times exper^2$$

To do that proceed as follows:

1. Regress wage on the dummies for education and experience and get RSS_U
2. Regress wage on constant, education, experience and experience square and get RSS_R
3. Calculate the F-test

If you did everything correctly you should have

$$F = \frac{(RSS_R - RSS_U)/35}{RSS_U/(1500-39)} = \frac{(2.7406 - 2.6050)/35}{2.6050/1461} \approx 2.17 \sim F(35, 1461)$$

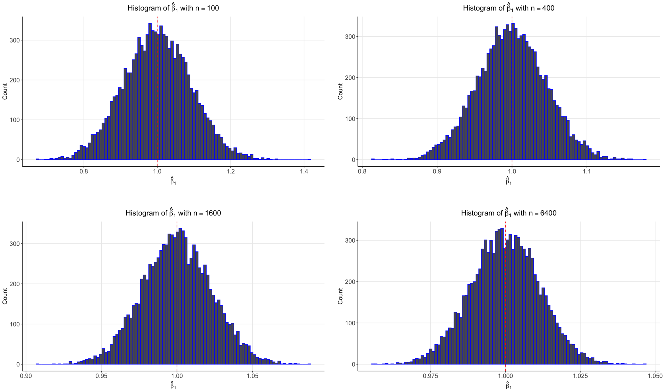
This is larger than the 95% quantile, so the restriction can be rejected. What does this mean?

Part (d)

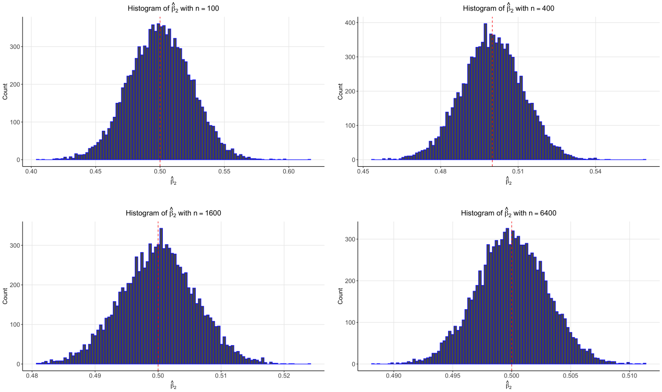
Is this methodology a good way of testing the assumption of a linear conditional expectation function?

Question 3

Simulation for $\hat{\beta}_1$



Simulation for $\hat{\beta}_2$



Question 4

Part (a)

For a regression with no intercept, we can find the formula for the OLS estimator by minimizing the sum of square errors

$$\hat{\beta}_{OLS} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \beta x_i)^2 \implies \hat{\beta}_{OLS} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Plug in the true model to get

$$\hat{\beta}_{OLS} = \beta + \frac{\sum_{i=1}^{n} x_i \varepsilon_i}{\sum_{i=1}^{n} x_i^2}$$

Part (a)

Consider the case where we have an even n observations. This means that

$$\sum_{i=1}^{n} x_i^2 = \sum_{j=1}^{n/2} (2j)^2 = \frac{n(n/2+1)(n+1)}{3}$$

$$\sum_{i=1}^{n} x_i \varepsilon_i = \sum_{j=1}^{n/2} 2j\varepsilon_{2j} \sim \mathcal{N}\left(0, \frac{n(n/2+1)(n+1)}{3}\right)$$

where the last equality is true since this is just a sum of i.i.d. ε_{2j}, which is $\mathcal{N}(0,1)$.

Part (a)

This means that our OLS estimator is consistent since

$$\hat{\beta}_{OLS} \sim \mathcal{N}\left(\beta, \frac{3}{n(n/2+1)(n+1)}\right)$$

And

$$\frac{3}{n(n/2+1)(n+1)} \to 0$$

This means that $\Pr(|\hat{\beta}_{OLS} - \beta|) > \delta, \forall \delta > 0$ converges to zero by Chebyshev's inequality.

Part (b)

Now you are told that $x_i = \lambda^i$, for $\mid \lambda \mid < 1$. Now, we will have that

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} \lambda^{2i} = \frac{\lambda^2 - \lambda^{2(n+1)}}{1-\lambda^2}$$

This means that

$$\begin{aligned}\frac{\sum_{i=1}^{n} x_i \varepsilon_i}{\sum_{i=1}^{n} x_i^2} &\sim \mathcal{N}\left(0, \left[\sum_{i=1}^{n} x_i^2\right]^{-1}\right) \\ &\sim \mathcal{N}\left(0, \frac{1-\lambda^2}{\lambda^2 - \lambda^{2(n+1)}}\right)\end{aligned}$$

This variance does not converge to zero as $n \to \infty$. Does that mean that our OLS estimator is not consistent?

Part (c)

Do the GM assumptions hold in this case?

- $\text{rank}(X) = m$?
- $\mathbb{E}[Y \mid X] = X\beta \iff \mathbb{E}[\varepsilon \mid X] = 0$?
- $\text{Var}(Y \mid X) = \text{Var}(\varepsilon \mid X) = \sigma^2 I_n$?

What about the fact that the estimator is not consistent in (b)?

Part (d)

From part (b) we can get that as $n \to \infty$ we will have that

$$\frac{1-\lambda^2}{\lambda^2 - \lambda^{2(n+1)}} \to \frac{1}{\lambda^2} - 1$$

And so

$$\hat{\beta}_{OLS} \xrightarrow{d} \mathcal{N}\left(\beta, \frac{1}{\lambda^2} - 1\right)$$

Can you give a more rigorous proof?

Question 5

Part (a)

The question tells us that a random sample is taken from the exponential distribution with density

$$f_\theta(y) = \frac{1}{\sqrt{\theta}} \exp(-y_i/\sqrt{\theta}), \qquad \text{for } y > 0$$

We know that the Fisher's Information can be found by

$$\mathcal{I}(\theta) = \text{Var}\left(\frac{d}{d\theta}\sum_{i=1}^{n} \log(f_\theta(y))\right) = -\mathbb{E}\Bigg[H(\theta)\Bigg]$$

where $H(\theta)$ is the Hessian matrix of second derivatives of the log of $f_\theta(y)$

Part (a)

We have all the ingredients, so let's compute this!

$$
\begin{aligned}
\mathcal{L}(\theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{\theta}} \exp(-y_i/\sqrt{\theta}) \\
&= \frac{1}{\theta^{n/2}} \exp\left(-\sum_{i=1}^{n} y_i/\sqrt{\theta} \right)
\end{aligned}
$$

Defining $L(\theta) = \log(\mathcal{L}(\theta))$ we then have that

$$
L(\theta) = -\frac{n}{2}\log(\theta) - \frac{\sum_{i=1}^{n} y}{\sqrt{\theta}} \implies \frac{d^2}{d\theta^2} L(\theta) = \frac{n}{2\theta^2} - \frac{3}{4}\frac{\sum_{i=1}^{n} y}{\theta^{5/2}}
$$

Part (a)

Finally, we have

$$
\begin{aligned}
\mathcal{I}(\theta) &= -\mathbb{E}\left[\frac{d^2}{d\theta^2} L(\theta)\right] \\
&= -\frac{n}{2\theta^2} + \frac{3}{4\theta^{5/2}} \mathbb{E}\left[\sum_{i=1}^{n} y_i\right] \\
&= -\frac{n}{2\theta^2} + \frac{3}{4\theta^{5/2}} n\sqrt{\theta} \\
&= \frac{n}{4\theta^2}
\end{aligned}
$$

Part (b)

To find the MLE estimator we can simply solve

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta)$$

Taking the first-order conditions of the log-likelihood function:

$$\frac{d}{d\theta} L(\theta) = -\frac{n}{2\hat{\theta}_{ML}} + \frac{\sum_{i=1}^{n} y_i}{2\hat{\theta}_{ML}^{3/2}} = 0 \iff \hat{\theta}_{MLE} = \left(\frac{\sum_{i=1}^{n} y_i}{n} \right)^2 = \bar{y}^2$$

Part (b)

To find the bias, we can take the expectation of our estimator

$$\mathbb{E}[\hat{\theta}_{MLE}] = \frac{1}{n^2}\mathbb{E}\left[\left(\sum_{i=1}^{n} y_i\right)^2\right]$$

But note that

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{i=1}^{n} y_i\right)^2\right] &= \text{Var}\left(\sum_{i=1}^{n} y_i\right) + \mathbb{E}\left[\sum_{i=1}^{n} y_i\right]^2 \\ &= n\theta + (n\sqrt{\theta})^2\end{aligned}$$

Given the information the question gives us

Part (b)

So we have

$$\mathbb{E}[\hat{\theta}_{MLE}] = \frac{n\theta + n^2\theta}{n^2} = \left(\frac{1+n}{n}\right)\theta$$

The bias is then

$$\mathbb{E}[\hat{\theta}_{MLE}] - \theta = \frac{\theta}{n}$$

Part (c)

Uh oh! Our model is misspecified and the sample actually comes from a $\chi^2(1)$ with density

$$g(y) = \frac{1}{\sqrt{2\pi y}} \exp(-y/2), \qquad \text{for } y > 0$$

We can try to approximate the expected log-likelihood given that the true density is $g(\cdot)$ using our sample analogue as:

$$\mathbb{E}_{\hat{g}}[\log f_\theta(y)] \approx \frac{1}{n} \sum_{i=1}^{n} \log f_\theta(y) = \frac{1}{n} L(\theta)$$

Part (c)

But given that y_i is i.i.d. and its first moment is finite, Khinchine's LLN gives that our sample average approaches the expected value in probability:

$$\frac{1}{n}L(\theta) \xrightarrow{p} \mathbb{E}_g[\log f_\theta(y)]$$

Given (regularity conditions) this becomes:

$$\hat{\theta}_{ML} = \arg\max_\theta \frac{1}{n}L(\theta) \xrightarrow{p} \arg\max_\theta \mathbb{E}_g[\log f_\theta(y)]$$

What is the value? By Kinchine's LLN $\bar{y}$ converges in probability to the mean of $\chi^2(1)$, which is 1. So by the continuous mapping theorem $\bar{y}^2$ converges to $1^2 = 1$.

Part (d)

We are interested in finding

$$\tilde{\theta} = \arg\min_{\theta} \int_0^{\infty} g(y) \log \frac{g(y)}{f_\theta(y)} dy = \arg\max_{\theta} \int_0^{\infty} g(y) \log(f_\theta(y)) dy$$

We can plug in the values for $g(y)$ and $\log(f_\theta)(y)$:

$$\begin{aligned}\int_0^{\infty} g(y) \log(f_\theta(y)) dy &= \int_0^{\infty} \frac{1}{\sqrt{2\pi y}} \exp(-y/2)(-\log\sqrt{\theta} - y/\sqrt{\theta}) dy \\ &= -\log\sqrt{\theta} - \frac{1}{\sqrt{\theta}}\end{aligned}$$

Part (d)

Take the first-order condition of the previous equation and we find that $\tilde{\theta} = 1$, just as in (c)! Why?

Part (e)

We will have the following:

$$
\begin{aligned}
-\mathbb{E}\left[\frac{d^2}{d\theta^2}\log f_\theta(y_i)\right] &= -\mathbb{E}\left[\frac{1}{2\theta^2} - \frac{3y_i}{4\theta^{5/2}}\right] \\
&= -\frac{1}{2\theta^2} + \frac{3}{4\theta^{5/2}}\mathbb{E}_g[y_i] \\
&= -\frac{1}{2\theta^2} + \frac{3}{4\theta^{5/2}}
\end{aligned}
$$

where above we use the fact that the mean of a $\chi^2(1) = 1$.

Part (e)

For the variance, we can compute:

$$\begin{aligned}
\mathrm{Var}\left(\frac{d}{d\theta}\log f_\theta(y_i)\right) &= \mathrm{Var}\left(-\frac{1}{2\theta}+\frac{y_i}{2\theta^{3/2}}\right)\\
&= \frac{1}{4\theta^3}\,\mathrm{Var}_g(y_i)\\
&= \frac{1}{2\theta^3}
\end{aligned}$$

where above we use the fact that the variance of a $\chi^2(1) = 2$.

Part (e)

This shows us that:

$$-\mathbb{E}\left[\frac{d^2}{d\theta^2}\log f_\theta(y_i)\right] < \mathrm{Var}\left(\frac{d}{d\theta}\log f_\theta(y_i)\right), \forall\theta > 0$$

What would happen if f was the true distribution?