Geneva Graduate Institute (IHEID)

Econometrics I (EI035), Fall 2024

Marko Mlikota

Final Exam Solutions

Tuesday, 10 December

- You have 1h 30min.
- There are 42 points in total.
- Prepare concise answers.
- Write your answers on separate sheets, not on the exam copy.
- State clearly any additional assumptions, if needed.
- For full credit, you need to explain your answers.

Problem 1 (42 points)

Suppose you observe a sample of n unemployed invidiuals. Let y_i denote the time (in weeks) that individual i spent in unemployment, and let $\bar{y} \equiv \frac{1}{n}\sum_{i=1}^{n} y_i$ be the average unemployment span in your sample. You can assume that your observations are independent.

One can model y_i using an expontential distribution:

$$p(y_i|\ \lambda) = \lambda exp\{-\lambda y_i\} \quad , \quad \lambda > 0 \ .$$

The parameter λ is the job-finding rate. It tells you how many (acceptable) job offers per week an individual gets. For example, $\lambda = 3$ would tell you that (on average) an individual receives three offers every week, while $\lambda = 1/2$ would tell you that (on average), an individual receives an offer every two weeks. For now, we assume that this λ is the same for all individuals in the sample.

(a) (4 points) The mean and variance of y_i are given by

$$\mathbb{E}[y_i] = \frac{1}{\lambda} \quad \text{and} \quad \mathbb{V}[y_i] = \frac{1}{\lambda^2} \ .$$

Interpret these expressions, relying on the two examples $\lambda = 3$ and $\lambda = 1/2$.

Solution: The expected time spent in unemployment is the inverse of the rate at which (acceptable) job offers arrive. For example, if $\lambda = 3$, the individual receives 3 job offers per week, which means that they are expected to be unemployed for 1/3 of a week. If $\lambda = 1/2$, they receive one job offer every five weeks, which means that they are expected to be unemployed for 2 weeks. **[2p.]**

There is more uncertainty around this mean estimate (i.e. the variance of the actually spent time in unemployment is higher) if λ is high. For $\lambda = 3$, we get a variance of 1/9, which means that the number of weeks an individual spends in unemployment is rather concentrated around the mean value of 3. In contrast, for $\lambda = 1/2$, the variance is 4, which means that the time an individual spends in unemployment can be considerably shorter or longer than the mean estimate of 2 weeks. **[2p.]**

(b) (3 points) Derive the log-likelihood $\ell(\lambda|Y)$ and find the Maximum Likelihood (ML) estimator

$$\hat{\lambda} \equiv \arg\max_{\lambda} \ell(\lambda|Y) \ .$$

Solution: Since the sample is i.i.d., we can set up the likelihood function:

$$p(Y|\lambda) = \prod_{i=1}^{n} p(y_i|\lambda) = \prod_{i=1}^{n} \lambda \exp\{-\lambda y_i\} = \lambda^n \exp\left\{-\lambda \sum_{i=1}^{n} y_i\right\} \ . \quad [\mathbf{1p}.]$$

The log-likelihood is then:

$$l(\lambda|Y) = \log p(Y|\lambda) = n\log(\lambda) - \lambda \sum_{i=1}^{n} y_i \; . \quad [\mathbf{1p}.]$$

Take the FOC to find the maximum w.r.t. λ:

$$\frac{\partial l(\lambda|Y)}{\partial \lambda} = n\frac{1}{\lambda} - \sum_{i=1}^{n} y_i = 0 \quad \rightarrow \quad \hat{\lambda} = \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^{-1} = \bar{Y}^{-1} \; . \quad [\mathbf{1p}.]$$

(c) (3 points) What is the probability limit of the average unemployment span in your sample, $\bar{y}$? Based on that result, is $\hat{\lambda}$ consistent?

Solution: Given i.i.d.-ness, we can invoke the WLLN:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i \xrightarrow{p} \mathbb{E}[y_i] = \frac{1}{\lambda} \; . \quad [\mathbf{1p}.]$$

Since the function $g(x) = x^{-1}$ is continuous, by Slutsky's theorem we have that:

$$\hat{\lambda} = \bar{Y}^{-1} \xrightarrow{p} \left(\frac{1}{\lambda}\right)^{-1} = \lambda \; . \quad [\mathbf{1p}.]$$

Thus, $\hat{\lambda}$ is consistent. **[1p.]**

(d) (4 points) What is the approximate distribution of $\bar{y}$ for large n? Based on that result, what is the approximate distribution of $\hat{\lambda}$ for large n?

Solution: By i.i.d-ness, we can invoke the CLT:

$$\sqrt{n}\left(\bar{Y} - \mathbb{E}[y_i]\right) \xrightarrow{d} N(0, \mathbb{V}[y_i]) \; .$$

That is, for n large, $\bar{Y}$ follows the approximate distribution:

$$\bar{Y} \overset{approx.}{\sim} N\left(\mathbb{E}[y_i], \frac{1}{n}\mathbb{V}[y_i]\right) \; ,$$

where $\mathbb{E}[y_i] = \frac{1}{\lambda}$ and $\mathbb{V}[y_i] = \frac{1}{\lambda^2}$. **[3p.]**

Thus, $\hat{\lambda}$ is approximately distributed as the inverse of a Normal with mean $\frac{1}{\lambda}$ and variance $\frac{1}{n\lambda^2}$. **[1p.]** (Note that $\hat{\lambda}$ cannot be Normally distributed, as it has to be positive, whereas the Normal distribution goes from $-\infty$ to ∞.)

(e) (3 points) Describe another, numerical approach to approximate the distribution of $\hat{\lambda}$.

Solution: The finite sample distribution of $\hat{\lambda}$ can be numerically approximated by bootstrapping. **[1p.]** This consists in randomly drawing with replacement M samples of size n from our original sample and using them to compute M estimates $\{\hat{\lambda}_m\}_{m=1}^M$, which approximate the distribution of $\hat{\lambda}$ in repeated sampling. **[2p.]**

More formally, given our i.i.d. sample $\{y_i\}_{i=1}^n$, for $m = 1 : M$ (and M large),

- draw n observations with replacement from the original sample, yielding a sample $\{y_i^m\}_{i=1}^n$
- compute $\hat{\lambda}_m = (\bar{Y}_m)^{-1}$ based on this sample

Then, the set $\{\hat{\lambda}_m\}_{m=1}^M$ approximates the distribution of $\hat{\lambda}$.

Now let's make the job-finding rate heterogeneous. Specifically, suppose x_i denotes the number of applications per week that individual i sent out, and let

$$\lambda_i = \exp\{\alpha + \beta x_i\} \ .$$

Note that this implies that

$$\mathbb{E}[y_i|x_i] = \frac{1}{\lambda_i} = \exp\{-(\alpha + \beta x_i)\} \quad \text{and} \quad \mathbb{V}[y_i|x_i] = \frac{1}{\lambda_i^2} = \exp\{-2(\alpha + \beta x_i)\} \ . \tag{1}$$

For simplicity, let's suppose you know α (so it's just a constant) and you only need to estimate β. The ML estimator can be defined as

$$\hat{\beta} = \arg\min_{\beta} Q_n(\beta) \ , \quad Q_n(\beta) = \frac{1}{n}\sum_{i=1}^n -(\alpha + \beta x_i) + y_i \exp\{\alpha + \beta x_i\} \ .$$

(f) (5 points) How do you interpret the parameters α and β?

Solution: Parameter α: Assume that $x_i = 0$, then $\lambda_i = \exp\{\alpha\}$. That is, α is the log of the job-finding rate for individuals who do not send out any applications. Higher (lower) values of α mean that such individuals have it easier (harder) to find a job. **[2p.]**

Parameter β: Tells you how much the job-finding rate changes as individuals send out more or less applications. **[1p.]** Concretely, we have

$$\frac{\partial \mathbb{E}[y_i|x_i]}{\partial x_i} = -\beta \exp\{-(\alpha + \beta x_i)\} = -\beta \mathbb{E}[y_i|x_i] \ ,$$

which implies

$$\frac{\Delta \mathbb{E}[y_i|x_i]}{\Delta x_i} \approx -\beta \mathbb{E}[y_i|x_i] \quad \Leftrightarrow \quad \frac{\Delta \mathbb{E}[y_i|x_i]}{\mathbb{E}[y_i|x_i]} \approx -\beta \Delta x_i$$

for small Δx_i. Therefore β tells us the approximate percentage reduction of the expected weeks spent in unemployment when the individual sends out one more application per week. **[2p.]**

(Clearly, the exact calculation is:

$$\mathbb{E}[y_i|x_i+1] - \mathbb{E}[y_i|x_i] = \exp\{-(\alpha+\beta x_i+1)\} - \exp\{-(\alpha+\beta x_i)\}$$

but it is not very helpful for intuition).

(g) (4 points) How can you find $\hat{\beta}$? Derive the first-order condition associated with the above optimization problem.

Solution: Take the FOC:

$$\begin{aligned} S_n(\beta) = Q_n^{(1)}(\beta) = \frac{\partial \mathcal{Q}(\beta)}{\partial \beta} &= \frac{1}{n}\sum_{i=1}^{n} -x_i + y_i \exp\{\alpha+\beta x_i\} x_i \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i(y_i \exp\{\alpha+\beta x_i\} - 1) = 0 \ . \quad [\mathbf{1p}.] \end{aligned}$$

We cannot solve this for β. **[1p.]** Hence, we need to use a numerical algorithm to obtain the $\hat{\beta}$ that minimizes $Q_n(\beta)$. **[2p.]** (**1p for saying if we solved FOC, we'd get** $\hat{\beta}$)

(h) (6 points) Is $\hat{\beta}$ a consistent estimator for β_0, the true value for β?
Hint: Note that $\mathbb{E}[y_i|x_i] = \exp\{-(\alpha+\beta_0 x_i)\}$, *and remember the Law of Iterated Expectations (LIE). Also, note that* $x_i \geq 0$ *and a function like* $\mathbb{E}[x_i exp\{x_i\beta\}]$ *is strictly increasing in* β.[1]

Solution: As $\hat{\beta}$ is not analytically available, we can verify consistency using extremum estimation theory (its simplified version seen in class). To verify consistency, i.e. $\hat{\beta} \xrightarrow{p} \beta_0$, we need to check that the following four conditions are satisfied: **[1p.]**

1) Let $\beta \in \mathcal{B} = [-c, c]$, for some large c. Then $\mathcal{B}$, the parameter space we consider, is compact. **[1p.]** **(even just mentioning compactness gives 1p)**

2) The function $Q_n(\beta)$ converges in probability to the true $Q(\beta)$:

$$\begin{aligned} Q_n(\beta) &= \frac{1}{n}\sum_{i=1}^{n} -\log(\lambda_i) + y_i\lambda_i \\ &\xrightarrow{p} \mathbb{E}[-log(\lambda_i) + y_i\lambda_i] \\ &= \mathbb{E}[-(\alpha+\beta x_i) + y_i \exp\{\alpha+\beta x_i\}] = Q(\beta) \quad [\mathbf{1p}.] \end{aligned}$$

(**0.5p for saying the above without calculations**)

[1]Strictly speaking, this holds provided that $\mathbb{E}[x_i] > 0$.

3) $Q(\beta)$ is a continuous function of β. **[1p.]**

4) To show that $Q(\beta)$ is uniquely minimised at β_0 **[1p.]**, some work is required. By the LIE, we have that:

$$\begin{aligned} Q(\beta) &= \mathbb{E}[-(\alpha + \beta x_i) + y_i \exp\{\alpha + \beta x_i\}] \\ &= \mathbb{E}\left[\mathbb{E}[-(\alpha + \beta x_i) + y_i \exp\{\alpha + \beta x_i\} \,|x_i]\right] \\ &= \mathbb{E}\left[-\alpha - \beta x_i + \mathbb{E}[y_i|x_i] \exp\{\alpha + \beta x_i\}\right] \\ &= \mathbb{E}[-\alpha - \beta x_i + \exp\{(\beta - \beta_0)x_i\}] \\ &= -\alpha - \beta\mathbb{E}[x_i] + \mathbb{E}[\exp\{x_i(\beta - \beta_0)\}] \ . \end{aligned}$$

Therefore, the FOC is:

$$\frac{\partial Q(\beta)}{\partial \beta} = -\mathbb{E}[x_i] + \mathbb{E}[x_i \exp\{x_i(\beta - \beta_0)\}] = 0 \ .$$

Note that this FOC is satisfied at $\beta = \beta_0$. This solution is unique (i.e. $Q(\beta)$ attains a global minimum at β_0) because:

- For β very low (as $\beta \to -\infty$), the second term goes to zero, meaning that the overall expression is $-\mathbb{E}[x_i]$, which is smaller than zero.
- For β very high (as $\beta \to +\infty$), the second term goes to $+\infty$, meaning that the overall expression is larger than zero,
- In-between, the function is strictly increasing in β.

Hence, there can only be a single point at which $\dfrac{\partial Q(\beta)}{\partial \beta} = 0$, which is $\beta = \beta_0$. **[1p.]**

(i) (6 points) Show that the asymptotic distribution of $\hat{\beta}$ is given by

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N\left(0, \mathbb{E}[x_i^2]^{-1}\right) \ .$$

Hint: To simplify notation, write λ_{i0} for exp$\{\alpha + \beta_0 x_i\}$. Also, your formula for the asymptotic variance simplifies thanks to the LIE and the results in Eq. (1)*.*

Solution: Again, as $\hat{\beta}$ is not analytically available, we can use extremum estimation theory to find the asymptotic distribution of $\hat{\beta}$. For this, we need to verify three more conditions:**[1p.]**

1) We assume $\beta_0 \in int(\mathcal{B})$ (which is given, provided that β_0 is some finite value). **[1p.]**

2) Using the result from the exercise (g),

$$\begin{aligned}\sqrt{n}Q_n^{(1)}(\beta) &= \sqrt{n}\frac{1}{n}\sum_{i=1}^n x_i(y_i \exp \alpha + \beta x_i - 1) \\ &= \sqrt{n}\frac{1}{n}\sum_{i=1}^n x_i(y_i\lambda_0 - 1) \\ &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n x_i(y_i\lambda_0 - 1) - \mathbb{E}[x_i(y_i\lambda_{i,0} - 1)]\right) .\end{aligned}$$

The second term in brackets can be added because $\mathbb{E}[x_i(y_i\lambda_{i,0} - 1)] = 0$ by LIE, owing to the fact that $\mathbb{E}[y_i|x_i] = \lambda_{i,0}^{-1}$.

By CLT, then:

$$\sqrt{n}Q_n^{(1)} \xrightarrow{d} N(0, M) , \quad M = \mathbb{V}[x_i(y_i\lambda_{i,0} - 1)] . \quad [\mathbf{1p}.]$$

Thereby, we can simplify

$$\begin{aligned}M = \mathbb{V}[x_i(y_i\lambda_{i,0} - 1)] &= \mathbb{E}[x_i^2(y_i\lambda_{i,0} - 1)^2] \\ &= \mathbb{E}[x_i^2(y_i^2\lambda_{i,0}^2 + 1 - 2y_i\lambda_{i,0})] \\ &= \mathbb{E}[x_i^2] + \mathbb{E}[x_i^2y_i^2\lambda_{i,0}^2] - 2\mathbb{E}[x_i^2y_i\lambda_{i,0}] \\ &= \mathbb{E}[x_i^2] + 2\mathbb{E}[x_i^2] - 2\mathbb{E}[x_i^2] \\ &= \mathbb{E}[x_i^2] ,\end{aligned}$$

whereby we apply the LIE to the terms $\mathbb{E}[x_i^2y_i^2\lambda_{i,0}^2]$ and $2\mathbb{E}[x_i^2y_i\lambda_{i,0}]$ and use the facts that

$$\begin{aligned}&\mathbb{E}[y_i|x_i] = \lambda_{i,0}^{-1} , \\ &\mathbb{E}[y_i|x_i] = \mathbb{V}[y_i|x_i] + \mathbb{E}[y_i|x_i]^2 = \lambda_{i,0}^{-2} + (\lambda_{i,0}^2)^{-1} = 2\lambda_{i,0}^{-2} . \quad [\mathbf{1p}.]\end{aligned}$$

3)

$$\begin{aligned}Q_n^{(2)}(\beta) = \frac{1}{n}\sum_{i=1}^n x_i^2 y_i \exp\{\alpha + \beta x_i\} &= \frac{1}{n}\sum_{i=1}^n x_i^2 y_i \lambda_i \\ &\xrightarrow{p} \mathbb{E}[x_i^2 y_i \lambda_{i,0}] = \mathbb{E}[x_i^2] \equiv H\end{aligned}$$

[1p.]

Combining the three together, we obtain that:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, H^{-1}MH^{-1}) ,$$

where $H^{-1}MH^{-1} = \mathbb{E}[x_i^2]^{-1}\mathbb{E}[x_i^2]\mathbb{E}[x_i^2]^{-1} = \mathbb{E}[x_i^2]$. **[1p.]** (**1p for saying that $H = M$ because of Information Matrix Equality**)

(j) (4 points) Construct a hypothesis test with size $\alpha = 0.05$ for testing

$$\mathscr{H}_0 : \mathbb{E}[y_i|x_i = 10] = \frac{1}{2}\mathbb{E}[y_i|x_i = 5] \quad \text{vs.} \quad \mathscr{H}_0 : \mathbb{E}[y_i|x_i = 10] \neq \frac{1}{2}\mathbb{E}[y_i|x_i = 5] \ ,$$

i.e. testing whether an individual submitting 10 applications per week spends (in expectation) exactly half as long in unemployment than a person sending out only 5 applications per week. More concretely, defining the test as $\varphi = \mathbf{1}\{T(X) < c_\alpha\}$, define the test-statistic $T(X)$ and find the critical value c_α.

Solution: One can apply several tests here: t-test, Likelihood Ratio (LR) test, Wald test. The following solution is based on the t-test, which requires rewriting our hypothesis:

$$\begin{aligned}
\mathbb{E}[y_i|x_i = 10] &= \frac{1}{2}\mathbb{E}[y_i|x_i = 5] \\
\Leftrightarrow \quad \exp\{-(\alpha + 10\beta)\} &= \frac{1}{2}\exp -(\alpha + 5\beta) \\
\Leftrightarrow \quad -\alpha - 10\beta &= \log\left(\frac{1}{2}\right) - \alpha - 5\beta \\
\Leftrightarrow \quad -5\beta &= -\log(2) \\
\Leftrightarrow \quad \beta &= \frac{\log(2)}{5} \quad [\mathbf{1p}.]
\end{aligned}$$

Under $\mathscr{H}_0 : \beta = \beta_0 = \log(2)/5$, we have that

$$\hat{\beta} \overset{approx.}{\sim} N\left(\beta_0, \frac{1}{n}\hat{\mathbb{E}[x_i^2]}^{-1}\right) \ ,$$

where $\hat{\mathbb{E}[x_i^2]} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$. **[1p.]** Therefore,

$$\frac{\hat{\beta} - \beta_0}{(n\hat{\mathbb{E}[x_i^2]})^{-1/2}} \overset{approx.}{\sim} N(0,1) \ ,$$

and we can set up a t-test $\varphi_i = \{T(X) < c_\alpha\}$, where:

$$T(X) = \left| \frac{\hat{\beta} - \beta_0}{(\widehat{n\mathbb{E}[x_i^2]})^{-1/2}} \right| = \left| \frac{\hat{\beta} - (\log(2)/5)}{(\widehat{n\mathbb{E}[x_i^2]})^{-1/2}} \right| \ ,$$

and where c_α is the 97.5-th quantile of the standard Normal distribution. **[2p.]**

Similarly, one would get [**1p**] for mentioning Wald test/LR test, [**3p**] for sketching it correctly.

Thereby, the LR test would compare the likelihood of the unrestricted model (where β is estimated) to the likelihood of the restricted model (where $\beta = \log(2)/5$ is imposed). In contrast, the Wald test does not (necessarily) require simplifying $\mathscr{H}_0$ to arrive at $\beta = \log(2)/5$. Instead, it suffices to write $\mathscr{H}_0$ as

$$\mathscr{H}_0 : g(\beta) = 0 \; , \quad g(\beta) = \exp\left\{-(\alpha + 10\beta)\right\} - \frac{1}{2}\exp -(\alpha + 5\beta) \; .$$

(This is equivalent to $g(\beta) = \beta - \log(2)/5$.)