PhD Econometrics 1: Study Questions Class 4
Imperial College London
Hormoz Ramian

Solutions

Question 1 Let SSR be defined by, $SSR = \sum_{i=1}^{n}(y_i - \alpha - \exp{(\beta x_i)})^2$. Then the first order conditions are:

$$\begin{aligned}\text{w.r.t. } \alpha\text{: } 0 &= -2\sum_{i=1}^{n}(y_i - \widehat{\alpha} - \exp{(\widehat{\beta}x_i)}) \\ \text{w.r.t. } \theta\text{: } 0 &= -2\sum_{i=1}^{n}(y_i - \widehat{\alpha} - \exp{(\widehat{\beta}x_i)})x_i \exp{(\widehat{\beta}x_i)}\end{aligned}$$

We can re-arrange to construct a matrix $\boldsymbol{w}_i = \begin{bmatrix}1 & x_i \exp{(\theta x_i)} & z_i\end{bmatrix}$ which gives the moment conditions $\boldsymbol{W'}(\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta})) = \boldsymbol{0}$, but note that the moment condition driven by the system of FOCs may not necessarily have a unique solution due to nonlinearity of the RSS function in parameters. We can simplify the system but analytical solutions are only found in special cases (not in the present example). In this case, the first equations gives one parameter in terms of $\widehat{\beta}$:

$$\begin{aligned}0 &= -2\sum_{i=1}^{n} y_i - 2\sum_{i=1}^{n}\widehat{\alpha} - -2\sum_{i=1}^{n}\exp{(\widehat{\beta}x_i)} \\ \widehat{\alpha} &= \overline{Y} - \frac{1}{n}\sum_{i=1}^{n}\exp{(\widehat{\beta}x_i)}\end{aligned}$$

substituting this in the second equation (FOC w.r.t. $\widehat{\beta}$):

$$\begin{aligned}0 &= -2\sum_{i=1}^{n} y_i x_i \exp{(\widehat{\beta}x_i)} - 2\widehat{\alpha}\sum_{i=1}^{n} x_i \exp{(\widehat{\beta}x_i)} - 2\sum_{i=1}^{n}\exp{(\widehat{\beta}x_i)}x_i \exp{(\widehat{\beta}x_i)} \\ &= -\sum_{i=1}^{n} y_i x_i \exp{(\widehat{\beta}x_i)} - \left[\overline{Y} - \frac{1}{n}\sum_{i=1}^{n}\exp{(\widehat{\beta}x_i)}\right]\sum_{i=1}^{n} x_i \exp{(\widehat{\beta}x_i)} - \sum_{i=1}^{n} x_i \exp{(2\widehat{\beta}x_i)}\end{aligned}$$

depending on the values $\{x_i\}_{i=1}^{n}$, this expression may yield multiple maxima, but under regularity condition, a unique maximizer to the RSS function is guaranteed. Although the final equation has only one unknown $\widehat{\beta}$, an analytical solution does not exist and it should be solved numerically.

Question 2 Start by the definition:

$$1 = \int_{x\in\mathcal{X}} f(x_i|\boldsymbol{\theta}_0)dx_i \tag{1}$$

where $\mathcal{X}$ is the support for continuous variate x over which the density function is defined. Continuity property implies that the likelihood function is also a continuously differentiable function of $\boldsymbol{\theta} \in \Theta$ (hence support $\mathcal{X}$ does not depend on $\boldsymbol{\theta}$) is essential as it allows interchanging differentiation and integration. Differentiating this equation and noting that under Leibnitz theorem[1], integration

[1]The theorem is very general and is applicable to most density functions satisfying the boundry condition: $\lim_{x_i\to\overline{\boldsymbol{\theta}}} f(x_i|\boldsymbol{\theta}_0) = \lim_{x_i\to\underline{\boldsymbol{\theta}}} f(x_i|\boldsymbol{\theta}_0) = 0$ where $\overline{\boldsymbol{\theta}}$ and $\underline{\boldsymbol{\theta}}$ are the upper and lower bounds of the parameter space.

and differentiation can be re-arranged as:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\theta}_0} 1 &= \frac{\partial}{\partial \boldsymbol{\theta}_0} \int_{x\in\mathcal{X}} f(x_i|\boldsymbol{\theta}_0)dx_i \\ \mathbf{0}_k &= \int_{x\in\mathcal{X}} \frac{\partial}{\partial \boldsymbol{\theta}_0} f(x_i|\boldsymbol{\theta}_0)dx_i\end{aligned}$$

where the left hand side is the derivative of the likelihood function with respect to k-parameters in vector $\boldsymbol{\theta}$ therefore there are k partial derivatives on the right hand side equal to a zeros vector $\mathbf{0}_k$. Note that the derivative of a scalar function $\frac{\partial}{\partial \boldsymbol{\theta}_0} f(x_i|\boldsymbol{\theta}_0)\frac{f(x_i|\boldsymbol{\theta}_0)}{f(x_i|\boldsymbol{\theta}_0)}$ with respect to a column vector is a row (alternatively[2] column) vector:

$$\mathbf{0}_k = \int_{x\in\mathcal{X}} \frac{\partial}{\partial \boldsymbol{\theta}_0} f(x_i|\boldsymbol{\theta}_0)\frac{f(x_i|\boldsymbol{\theta}_0)}{f(x_i|\boldsymbol{\theta}_0)}dx_i \tag{2}$$

$$\mathbf{0}_k = \int_{x\in\mathcal{X}} \frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0} f(x_i|\boldsymbol{\theta}_0)dx_i \tag{3}$$

$$\mathbf{0}_k = \mathbb{E}\left[\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\right] \tag{4}$$

In order to show that the quadratic form of the score related to the hessian, we need to differentiate equation (3) again with respect to $\boldsymbol{\theta}'$. In order to find the derivative, we first note that there are k-scalar-equations inside (3), where each of these single equations should be differentiated w.r.t. the row vector $\boldsymbol{\theta}'$, overall resulting in a $k \times k$ sytem:

$$\begin{aligned}\mathbf{0}_{k\times k} &= \frac{\partial}{\partial \boldsymbol{\theta}_0'} \int_{x\in\mathcal{X}} \frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0} f(x_i|\boldsymbol{\theta}_0)dx_i \\ &= \int_{x\in\mathcal{X}} \left[\frac{\partial^2 \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'} f(x_i|\boldsymbol{\theta}_0)dx_i + \frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\frac{\partial f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0'}\right] dx_i \end{aligned}\tag{5}$$

Note that:

$$\frac{\partial f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0'} = f(x_i|\boldsymbol{\theta}_0)\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0'}$$

then the first partial in the second term of right-hand-side of equation (5) can be substitute for:

$$\begin{aligned}\mathbf{0}_{k\times k} &= \int_{x\in\mathcal{X}} \left[\frac{\partial^2 \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'} f(x_i|\boldsymbol{\theta}_0) + \frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\frac{\partial f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0'}\right] dx_i \\ &= \int_{x\in\mathcal{X}} \left[\frac{\partial^2 \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'} f(x_i|\boldsymbol{\theta}_0) + \frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0'} f(x_i|\boldsymbol{\theta}_0)\right] dx_i\end{aligned}$$

Re-arrange:

$$-\int_{x\in\mathcal{X}} \frac{\partial^2 \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'} f(x_i|\boldsymbol{\theta}_0)dx_i = \int_{x\in\mathcal{X}} \frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0'} f(x_i|\boldsymbol{\theta}_0)dx_i \tag{6}$$

The LHS of equation (6) can be written as:

$$-\int_{x\in\mathcal{X}} \frac{\partial^2 \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'} f(x_i|\boldsymbol{\theta}_0)dx_i = -\mathbb{E}\left[\frac{\partial^2 \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'}\right] \tag{7}$$

Before writing the RHS, let's define the covariance of the score:

$$\operatorname{cov}\left[\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\right] = \mathbb{E}\left\{\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0} - \mathbb{E}\left[\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\right]\right\}\left\{\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0'} - \mathbb{E}\left[\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0'}\right]\right\}$$

[2]Derivative of a scalar function w.r.t. a col-vector is a row-vector which is the denominator layout, also this can be written as derivative of a scalar function w.r.t. to a col-vector as a col-vector that is referred to as the numerator layout. Both conventions are acceptable as long as one remains consistent.

But we know that from equation (4), $\mathbb{E}\left[\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\right] = 0$, hence:

$$\text{cov}\left[\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\right] = \mathbb{E}\left[\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0'}\right]$$

Subsequently, the RHS of equation (6) can be written as:

$$\begin{aligned}\int_{x\in\mathcal{X}} \frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0'} f(x_i|\boldsymbol{\theta}_0)dx_i &= \mathbb{E}\left[\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0'}\right] \\ &= \text{cov}\left[\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\right] \qquad (8)\end{aligned}$$

Therefore, from equations (7) and (7), we have:

$$-\mathbb{E}\left[\frac{\partial^2 \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'}\right] = \text{cov}\left[\frac{\partial \ln f(x_i|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\right] \qquad (9)$$

Question 3 The likelihood function $\mathcal{L}(\lambda; x)$, under the independence assumption, is:

$$\begin{aligned}\mathcal{L}(\lambda; x) &= f(x_1, x_2, \ldots, x_T; \lambda) \\ &= f(x_1;\lambda) \times f(x_1;\lambda)\times, \ldots, \times f(x_T;\lambda) = \prod_{i=1}^{T} f(x_i;\lambda)\end{aligned}$$

Taking logs gives the log-likelihood function $\ell(\lambda; x)$:

$$\begin{aligned}\ell(\lambda; x) &= \log \prod_{i=1}^{T} f(x_i;\lambda) = \sum_{i=1}^{T} \log f(x_i;\lambda) \\ &= \sum_{i=1}^{T} \log \lambda e^{-\lambda x} = \sum_{i=1}^{T} \log \lambda - \sum_{i=1}^{T} \lambda x_i \\ &= T \log \lambda - \lambda \sum_{i=1}^{T} x_i\end{aligned}$$

Taking derivative with respect to the parameter give:

$$0 = \left.\frac{\partial \ell(\lambda; x)}{\partial \lambda}\right|_{\lambda=\widehat{\lambda}_{ML}} = \frac{T}{\widehat{\lambda}_{ML}} - \sum_{i=i}^{T} x_i$$

giving the ML estimator,

$$\widehat{\lambda}_{ML} = \left(\frac{1}{T}\sum_{i=1}^{T} x_i\right)^{-1} \qquad (10)$$

To derive the asymptotic variance, we form the second derivative of the log-likelihood function with respect to λ to obtain the hessian. Then the Fisher's information, its large sample counterpart and asymptotic variance of the ML estimator are:

$$\begin{aligned}\mathcal{H} &= -\frac{T}{\lambda^2} \\ I &= -\mathbb{E}\mathcal{H} = \frac{T}{\lambda^2} \\ \mathcal{I} &= \frac{1}{T}\text{plim}\frac{T}{\lambda^2} = \frac{1}{\lambda^2} \\ \text{AsyVar}(\widehat{\lambda}) &= \frac{1}{\mathcal{I}} = \lambda^2\end{aligned}$$

Question 4

(4.1) We wish to maximize the joint density of random variables given the parameters, which under independence, simplifies to the following product:

$$\mathcal{L} \quad = \quad \max_{\boldsymbol{\theta}\in\Theta} \prod_{i=1}^{n} f(y_i;\boldsymbol{\theta}) \tag{11}$$

The ML estimator $\widehat{\boldsymbol{\theta}}_{ML}$ or for short $\widehat{\boldsymbol{\theta}}$ is:

$$\widehat{\boldsymbol{\theta}} \quad = \quad \arg\max_{\boldsymbol{\theta}\in\Theta} \mathcal{L}(\boldsymbol{\theta}) \tag{12}$$

where $\Theta \subset \mathbb{R}^2$ is the two-dimensional parameter space, where each parameter belongs to the real line. Moreover, since logarithmic transformation is a monotonic transformation, which retains the parameter estimates unchanged, we can re-write the previous equation as:

$$\begin{aligned}\widehat{\boldsymbol{\theta}} \quad &= \quad \arg\max_{\boldsymbol{\theta}\in\Theta} \mathcal{L}(\boldsymbol{\theta}) \\ &= \quad \arg\max_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}) \end{aligned}\tag{13}$$

(4.2) The probability density function for Normal distribution is:

$$\widehat{\boldsymbol{\theta}} \quad = \quad \arg\max_{\boldsymbol{\theta}\in\Theta} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(y_i-\alpha-\beta x_i)^2}{\sigma^2}\right\} \tag{14}$$

$$= \quad \arg\max_{\boldsymbol{\theta}\in\Theta} \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} \prod_{i=1}^{n} \exp\left\{-\frac{1}{2}\frac{(y_i-\alpha-\beta x_i)^2}{\sigma^2}\right\} \tag{15}$$

$$= \quad \arg\max_{\boldsymbol{\theta}\in\Theta} -\frac{n}{2}\log(2\pi\sigma^2) + \sum_{i=i}^{n} \log\left(\exp\left\{-\frac{1}{2}\frac{(y_i-\alpha-\beta x_i)^2}{\sigma^2}\right\}\right) \tag{16}$$

Because $\frac{n}{2}\log(2\pi)$ in the first term is a constant and can be separated from the rest of the objective function, we can drop it from the optimization. In other words, $\frac{n}{2}\log(2\pi)$ is only a constant shift and in inconsequential to the value of $\widehat{\boldsymbol{\theta}}$:

$$\widehat{\boldsymbol{\theta}} \quad = \quad \arg\max_{\boldsymbol{\theta}\in\Theta} -\cancel{\frac{n}{2}\log(2\pi)} - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2 \tag{17}$$

Furthermore, since the question has give the true value of σ^2, we do not estimate it and as a result the objective function reduces to:

$$\widehat{\boldsymbol{\theta}} \quad = \quad \arg\max_{\boldsymbol{\theta}\in\Theta} -\cancel{\frac{n}{2}\log(\sigma^2)} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2 \tag{18}$$

$$= \quad \arg\max_{\boldsymbol{\theta}\in\Theta} -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2 \tag{19}$$

Differentiating with respect to each parameter gives the first order conditions, or the score functions:

$$\begin{aligned}\left.\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \quad &= \quad -\frac{1}{2\sigma^2}\frac{\partial}{\partial \alpha}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2 \\ \left.\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \quad &= \quad -\frac{1}{2\sigma^2}\frac{\partial}{\partial \beta}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2 \end{aligned}$$

Setting the equations equal to zero:

$$0 = \sum_{i=1}^{n}(y_i - \widehat{\alpha} - \widehat{\beta}x_i) \tag{20}$$

$$0 = \sum_{i=1}^{n} x_i(y_i - \widehat{\alpha} - \widehat{\beta}x_i) \tag{21}$$

Solving the system above gives:

$$\widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x} \tag{22}$$

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{23}$$

The estimates are identical to those from the OLS.

(4.3) We start by differentiating the objective function twice, or differentiating the first order conditions once, with respect to the parameter to construct the hessian matrix $\mathcal{H}(\boldsymbol{\theta}) = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$:

$$\begin{aligned}\mathcal{H}(\boldsymbol{\theta}) &= \begin{bmatrix} \partial^2\ell(\boldsymbol{\theta})/\partial\alpha^2 & \partial^2\ell(\boldsymbol{\theta})/\partial\alpha\partial\beta \\ \partial^2\ell(\boldsymbol{\theta})/\partial\beta\partial\alpha & \partial^2\ell(\boldsymbol{\theta})/\partial\beta^2 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i \\ -\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i & -\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2 \end{bmatrix} \end{aligned} \tag{24}$$

The information matrix $I(\boldsymbol{\theta})$ is defined as $I(\boldsymbol{\theta}) = -\mathbb{E}\left(\mathcal{H}\right)$ which can be simplified to $I(\boldsymbol{\theta}) = -\mathcal{H}$ under non-stochastic regressors assumption. Therefore,

$$\begin{aligned} I(\boldsymbol{\theta}) &= -\mathbb{E}(\mathcal{H}) \\ &= \begin{bmatrix} \frac{n}{\sigma^2} & \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i \\ \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i & \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2 \end{bmatrix} \\ &= \frac{n}{\sigma^2}\begin{bmatrix} 1 & \frac{1}{n}\sum_{i=1}^{n} x_i \\ \frac{1}{n}\sum_{i=1}^{n} x_i & \frac{1}{n}\sum_{i=1}^{n} x_i^2 \end{bmatrix} \end{aligned}$$

then:

$$\begin{aligned} \mathcal{I}(\boldsymbol{\theta}) &= \text{plim}_{n\to\infty}\frac{1}{n}I(\boldsymbol{\theta}) \\ &= \frac{1}{\sigma^2}\begin{bmatrix} 1 & \frac{1}{n}\sum_{i=1}^{n} x_i \\ \frac{1}{n}\sum_{i=1}^{n} x_i & \frac{1}{n}\sum_{i=1}^{n} x_i^2 \end{bmatrix} \\ &= \frac{1}{\sigma^2}\frac{\boldsymbol{X}'\boldsymbol{X}}{n} = \frac{1}{\sigma^2}Q_{\boldsymbol{X}} \end{aligned}$$

where $Q_{\boldsymbol{X}} = \frac{\boldsymbol{X}'\boldsymbol{X}}{n}$ and we have $\mathcal{I}^{-1}(\boldsymbol{\theta}) = \sigma^2 Q_{\boldsymbol{X}}^{-1}$. We conclude that the asymptotic covariance matrix of the MLE coincides with the asymptotic covariance matrix of the OLS estimators (The correlation between parameters is only equal to zero iff $\sum_{i=1}^{n} x_i = 0$). The ML estimators obtained in equations (22) and (23), $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\beta})'$ are identical to those of the OLS estimation and therefore unbiased $\mathbb{E}\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}$ also note that expressing the estimator as an average:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \beta + \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})u_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{25}$$

noting that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} x_i u_i &\xrightarrow{p} 0 \\
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i &\xrightarrow{d} \mathcal{N}(0, \sigma^2 Q_{xx})
\end{aligned}$$

where $Q_{xx} = \text{plim}\frac{1}{n}\sum_{i=1}^{n} x_i^2$ because, under i.i.d. assumption:

$$\begin{aligned}
\text{var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(x_i - \overline{x})u_i\right] &= \frac{1}{n}\,\text{var}\left[\sum_{i=1}^{n}(x_i - \overline{x})u_i\right] \overset{\text{i.i.d.}}{=} \frac{1}{n}\sum_{i=1}^{n}\text{var}((x_i - \overline{x})u_i) \\
&= \sigma^2\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right) \qquad (26)
\end{aligned}$$

re-arranging equation (25) gives:

$$\sqrt{n}(\widehat{\beta} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(x_i - \overline{x})u_i \qquad (27)$$

now taking limit of the random variable $\sqrt{n}(\widehat{\beta} - \beta)$, as the sample size goes to infinity $\lim_{n\to\infty}\sqrt{n}(\widehat{\beta} - \beta)$ yields the following aymptotic results:

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}^{-1}) \qquad (28)$$

first, we use normal distribution by assuming the Central Limit Theorem as the estimator can be constructed as an average (under some regularity conditions, mainly finite variance, the averages converge in distribution to a normal distribution), second, using the unbiasedness property, the limiting distribution is centered at zero, and third the asymptotic variance is determined by the results in equation (26). repeating for $\widehat{\alpha}$, and collecting both parameter in $\boldsymbol{\theta} = \begin{bmatrix}\alpha & \beta\end{bmatrix}'$ then,

$$\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta} \qquad (29)$$

and then,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathcal{I}^{-1}) \qquad (30)$$

shows asymptotic normality of the ML estimator.