Exo7 version 2.0 Attention il peut y avoir des errata.

Q1 Estimer β par MLE.

We need to write the density of y. This transformation gives us $f(y) = f(\varepsilon)\left|\frac{\partial \varepsilon}{\partial y'}\right|$ where $\left|\frac{\partial \varepsilon}{\partial y'}\right|$ is the absolute value of the determinant of the matrix of partial derivatives (Jacobian) such that:

$$\frac{\partial \varepsilon}{\partial y'} = \begin{bmatrix} \frac{\partial \varepsilon_1}{\partial y_1} & \frac{\partial \varepsilon_1}{\partial y_2} & \cdots & \frac{\partial \varepsilon_1}{\partial y_n} \\ \frac{\partial \varepsilon_2}{\partial y_1} & \frac{\partial \varepsilon_2}{\partial y_2} & \cdots & \frac{\partial \varepsilon_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \varepsilon_n}{\partial y_1} & \frac{\partial \varepsilon_n}{\partial y_2} & \cdots & \frac{\partial \varepsilon_n}{\partial y_n} \end{bmatrix}$$ This is called the Jacobian.

And in the our case we have that the determinant of the Jacobian $\left|\frac{\partial \varepsilon}{\partial y'}\right| = \left|I_n\right| = 1$ since $\varepsilon_i = y_i - \beta$

So we get that $f(y) = f(\varepsilon)\left|\frac{\partial \varepsilon}{\partial y'}\right| = f(\varepsilon)$

Since we have $\varepsilon \sim N(0, \sigma^2 I)$, the density of $\varepsilon_i \sim N(0, \sigma^2)$ is given by the PDF of the normal distribution:

$$f(\varepsilon_i) = \frac{1}{\left(2\pi\sigma^2\right)^{1/2}} \exp\left(\frac{-1}{2}\frac{(y_i - \beta)^2}{\sigma^2}\right)$$

In that case because the ε_i are independent (that we know for sure), we can write $f(\varepsilon) = \prod_{i=1}^{n} f(\varepsilon_i)$

In matrix notation we thus have: $f(\varepsilon) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(\frac{-1}{2}\frac{\varepsilon'\varepsilon}{\sigma^2}\right)$

Hence we can write the likelihood function for y as:

$$L(\theta \mid y) \equiv f(y) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(\frac{-1}{2}\frac{(y - \iota\beta)'(y - \iota\beta)}{\sigma^2}\right) \text{ where } \theta = \begin{bmatrix} \beta \\ \sigma^2 \end{bmatrix}$$

Knowing that it is equivalent to maximize the likelihood or the log-likelihood. We can write the maximization problem as:

$$\max_{\theta\in\Theta} \ln L(\theta \mid y) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y - \iota\beta)'(y - \iota\beta)$$

The FOC are given by:

$$\frac{\partial \ln L(\theta \mid y)}{\partial \beta} = -\frac{1}{2\hat{\sigma}^2_{ML}}(-2y'\iota + 2(\iota'\iota)\hat{\beta}_{ML}) = -\frac{1}{\hat{\sigma}^2_{ML}}(-y'\iota + (\iota'\iota)\hat{\beta}_{ML}) = 0$$ which gives us

$$\hat{\beta}_{ML} = (\iota'\iota)^{-1}\iota'y = \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y}$$

$$\frac{\partial \ln L(\theta \mid y, X)}{\partial \sigma^2} = -\frac{n}{2\hat{\sigma}^2_{ML}} + \frac{1}{2\hat{\sigma}^4_{ML}}(y - \iota\hat{\beta}_{ML})'(y - \iota\hat{\beta}_{ML}) = 0$$

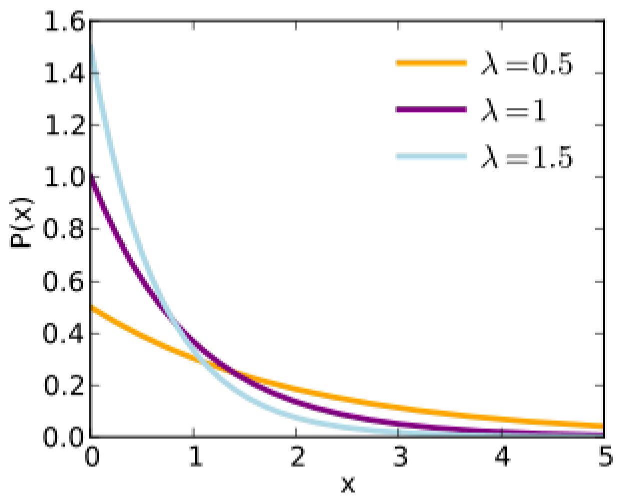
which delivers $\hat{\sigma}^2_{ML} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} = \frac{(y - \iota\hat{\beta}_{ML})'(y - \iota\hat{\beta}_{ML})}{n}$

To sum up:

$$\hat{\beta}_{ML} = \frac{\sum_{i=1}^{n} y_i}{n} = \overline{y}$$

$$\hat{\sigma}^2_{ML} = \frac{\sum_{i=1}^{n}(y_i - \hat{\beta}_{ML})^2}{n}$$

Q2 exponential distribution
http://en.wikipedia.org/wiki/Exponential_distribution



Probability density function
The probability density function (pdf) of an exponential distribution is

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0, \\ 0, & x < 0. \end{cases}$$

Here $\lambda > 0$ is the parameter of the distribution, often called the rate parameter. The distribution is supported on the interval [0, ∞). If a random variable X has this distribution, we write X ~ Exp(λ).

Maximum likelihood

The likelihood function for λ, given an independent and identically distributed sample x = (x_1, ..., x_n) drawn from the variable, is

$$L(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda x_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right) = \lambda^n \exp\left(-\lambda n \overline{x}\right),$$

where

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

is the sample mean.

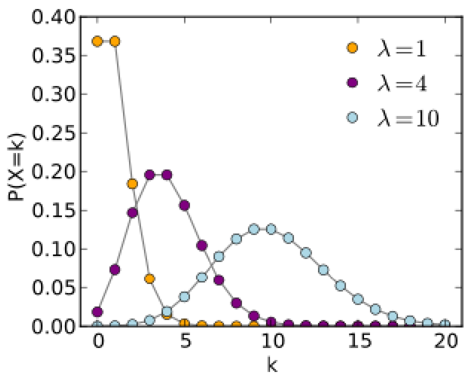
The derivative of the likelihood function's logarithm is

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \ln L(\lambda) = \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(n \ln(\lambda) - \lambda n \overline{x}\right) = \frac{n}{\hat{\lambda}_{ML}} - n\overline{x} = 0$$

Consequently the maximum likelihood estimate for the rate parameter is

$$\hat{\lambda}_{ML} = \frac{1}{\overline{x}}.$$

Q3 poison distribution   http://en.wikipedia.org/wiki/Poisson_distribution



Maximum likelihood

Given a sample of n measured values x_i we wish to estimate the value of the parameter λ of the Poisson population from which the sample was drawn. To calculate the maximum likelihood value, we form the log-likelihood function

$$\begin{aligned}\ell(\lambda) = \ln L(\lambda) &= \ln \prod_{i=1}^{n} f(x_i \mid \lambda) \\ &= \sum_{i=1}^{n} \ln\left(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right) \\ &= -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln(\lambda) - \sum_{i=1}^{n}\ln(x_i!).\end{aligned}$$

Take the derivative of $\ell(\lambda)$ with respect to λ and equate it to zero:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\ell(\lambda) = 0 \Leftrightarrow -n + \left(\sum_{i=1}^{n} x_i\right)\frac{1}{\hat{\lambda}_{ML}} = 0.$$

Solving for λ yields a stationary point, which if the second derivative is negative is the maximum-likelihood estimate of λ:

$$\hat{\lambda}_{\mathrm{ML}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

Checking the second derivative, it is found that it is negative for all λ and x_i greater than zero, therefore this stationary point is indeed a maximum of the initial likelihood function:

$$\frac{\partial^2 L}{\partial \lambda^2} = \sum_{i=1}^{n} -\lambda^{-2} x_i$$

Since each observation has expectation λ so does this sample mean. Therefore it is an unbiased estimator of λ. It is also an efficient estimator, i.e. its estimation variance achieves the Cramér-Rao lower bound (CRLB). Hence it is MVUE (minimum variance unbiased estimator). Also it can be proved that the sample mean is complete and sufficient statistic for λ.

Q4 Solutions

a) La solution est d'estimer par OLS le modèle transformé log-log. Ainsi en prenant le logarithme naturel des deux côtés on obtient le modèle suivant: $\ln \breve{y}_i = \underbrace{\ln \beta_1}_{\alpha_1} + \beta_2 \ln(\breve{x}_{i2}) + \beta_3 \ln(\breve{x}_{i3}) + \varepsilon_i \underbrace{\ln(e)}_{1}$

En définissant le modèle comme

$$y_i = \ln \breve{y}_i$$
$$x_{i2} = \ln \breve{x}_{i2}$$
$$x_{i3} = \ln \breve{x}_{i3}$$

Avec $X = \begin{bmatrix} \iota & x_2 & x_3 \end{bmatrix}$

Et $\beta = \begin{bmatrix} \alpha_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$

Pour la dérivation voir les notes du cours.

L'estimateur correspondant des MCO suivant $\hat{\beta}_{LS} = (X'X)^{-1} X'y$ est non biaisé (Si A1 et A2 et A3 tiennent avec le modèle transformé) et à variance minimale $\text{var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ sous les assomptions A1, A2, A3 et A4, ainsi il est BLUE comme démontré par le théoreme de Gauss-Markov.

Les tests (t et F) en petit échantillons ne seront pas valides par contre en l'absence de la normalité, il faudra utiliser les tests asymptotiques (la loi normale avec le test Z et les test Chi2) parce que l'on a pas la normalité des aléas.

4b) L'estimateur des MCO est non-biaisé.
Montrez cela en prenant l'espérance. Voir les notes du cours **section 5.1**.

L'estimateur des MCO n'est plus à variance minimale par contre et donc il n'est pas BLUE. Voir les notes du cours **section 5.1**.

Les tests (t et F) en petit échantillons ne seront pas valides par contre en l'absence de la normalité, il faudra utiliser les tests asymptotiques (la loi normale avec le test Z et les test Chi2) parce que l'on a pas la normalité des aléas. **En plus il faut utiliser une autre matrice de variance que celle de l'estimateur des MCO.**

On sait que l'estimateur GLS est BLUE, et que l'estimateur ML est BLUE pour β dans ce cas s'ils sont bien définis. Le problème avec l'estimateur de ML est que l'on a pas la forme de la densité des aléas.

Ainsi le seul estimateur possible est l'estimateur GLS à moins d'imposer une distribution

L'estimateur GLS (MCG) est donné par $\hat{\beta}_{GLs} = \underset{\beta \epsilon \mathbb{R}^K}{\arg\min}\, s(\beta) = (y - X\beta)'\Omega^{-1}(y - X\beta)$

Le critère se réécrit comme

$$s(\beta) = (y - X\beta)'\Omega^{-1}(y - X\beta) = (y - X\beta)'(\Omega^{-1}y - \Omega^{-1}X\beta)$$
$$= y'\Omega^{-1}y - \beta'X'\Omega^{-1}y - y'\Omega^{-1}X\beta + \beta'X'\Omega^{-1}X\beta$$
$$= y'\Omega^{-1}y - 2y'\Omega^{-1}X\beta + \beta'X'\Omega^{-1}X\beta$$

The FOC are given by:

$\dfrac{\partial s(\beta)}{\partial \beta} = (-2X'\Omega^{-1}y + 2X'\Omega^{-1}X\hat{\beta}_{GLs}) = 0$ which gives us

$$\hat{\beta}_{GLs} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = (X'\tilde{P}'\underbrace{\tilde{P}X}_{\tilde{X}})^{-1}X'\tilde{P}'\underbrace{\tilde{P}y}_{\tilde{y}} = \underbrace{(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}}_{\substack{\text{MCO avec le modèle} \\ \text{transformé par } \tilde{P}}}$$
$$= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y = (X'P'\underbrace{PX}_{X^*})^{-1}X'P'\underbrace{Py}_{y^*} = \underbrace{(X^{*}{}'X^*)^{-1}X^{*}{}'y^*}_{\substack{\text{MCO avec le modèle} \\ \text{transformé par } P}}$$

La variance est données par

$$\text{var}(\hat{\beta}_{GLs}) = \sigma^2(X'\Omega^{-1}X)^{-1} = \sigma^2(X'\tilde{P}'\tilde{P}X)^{-1} = \sigma^2(\tilde{X}'\tilde{X})^{-1} = (X'\Sigma^{-1}X)^{-1} == (X'P'PX)^{-1} = (X^{*}{}'X^*)^{-1}$$

Ainsi on estimera le modèle transformé par MCO avec la matrice P où $P'P = \Sigma^{-1} = (\sigma^2\Omega)^{-1}$

$$Py = PX\beta + P\varepsilon,$$

Que l'on réécrit comme

$$y^* = X^*\beta + \varepsilon^*.$$

ou bien avec la matrice $\tilde{P}$ où $\tilde{P}'\tilde{P} = \Omega^{-1}$

$$\tilde{P}y = \tilde{P}X\beta + \tilde{P}\varepsilon,$$

Que l'on réécrit comme

$$\tilde{y} = \tilde{X}\beta + \tilde{\varepsilon}.$$

4c) We need to write the density of y. This transformation gives us $f(y) = f(\varepsilon)\left|\dfrac{\partial \varepsilon}{\partial y'}\right|$ where $\left|\dfrac{\partial \varepsilon}{\partial y'}\right|$ is the absolute value of the determinant of the matrix of partial derivatives (Jacobian) such that:

$$\frac{\partial \varepsilon}{\partial y\ '} = \begin{bmatrix} \dfrac{\partial \varepsilon_1}{\partial y_1} & \dfrac{\partial \varepsilon_1}{\partial y_2} & \cdots & \dfrac{\partial \varepsilon_1}{\partial y_n} \\ \dfrac{\partial \varepsilon_2}{\partial y_1} & \dfrac{\partial \varepsilon_2}{\partial y_2} & \cdots & \dfrac{\partial \varepsilon_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial \varepsilon_n}{\partial y_1} & \dfrac{\partial \varepsilon_n}{\partial y_2} & \cdots & \dfrac{\partial \varepsilon_n}{\partial y_n} \end{bmatrix}$$ This is called the Jacobian.

And in the the case of the CLRM we have that the determinant of the Jacobian $\left|\dfrac{\partial \varepsilon}{\partial y'}\right| = \left|I_n\right| = 1$ since

$\varepsilon_i = y_i - x_{(i)}'\beta$

So we get that $f(y\ ) = f(\varepsilon)\left|\dfrac{\partial \varepsilon}{\partial y\ '}\right| = f(\varepsilon)$

If we truly had $\varepsilon \sim N(0, \sigma^2 I)$, the density of $\varepsilon_i \sim N(0, \sigma^2)$ would be given by the PDF of the normal distribution:

$$f(\varepsilon_i) = \frac{1}{\left(2\pi\sigma^2\right)^{1/2}} \exp\left(\frac{-1}{2}\frac{(y_i - x_{(i)}'\beta)^2}{\sigma^2}\right)$$

We are wreckless childs and we impose it no matter what. It's like driving a car with closed eyes… WHOA!!

In that case because the ε_i are independent (that we know for sure), we can write $f(\varepsilon) = \prod_{i=1}^{n} f(\varepsilon_i)$

We will have the pseudo-ML or Quas-ML

In matrix notation we thus have: $f(\varepsilon) = \dfrac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(\dfrac{-1}{2}\dfrac{\varepsilon'\varepsilon}{\sigma^2}\right)$

Hence we can write the likelihood function for y as:

$$L(\theta \mid y\ , X\ ) \equiv f(y\ ) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(\frac{-1}{2}\frac{(y\ - X\ \beta)'(y\ - X\ \beta)}{\sigma^2}\right)$$ where $\theta = \begin{bmatrix} \beta \\ \sigma^2 \end{bmatrix}$

Knowing that it is equivalent to maximize the likelihood or the log-likelihood. We can write the maximization problem as:

$$\max_{\theta\in\Theta} \ln L(\theta \mid y, X) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

The FOC are given by:

$$\frac{\partial \ln L(\theta \mid y, X)}{\partial \beta} = -\frac{1}{2\hat{\sigma}^2_{ML}}(-2X'y + 2X'X\hat{\beta}_{ML}) = \frac{1}{\hat{\sigma}^2_{ML}}(X'y + X'X\hat{\beta}_{ML}) = 0$$ which gives us

$$\hat{\beta}_{ML} = (X'X)^{-1}X'y$$

$$\frac{\partial \ln L(\theta \mid y, X)}{\partial \sigma^2} = -\frac{n}{2\hat{\sigma}^2_{ML}} + \frac{1}{2\hat{\sigma}^4_{ML}}(y - X\hat{\beta}_{ML})'(y - X\hat{\beta}_{ML}) = 0$$

which delivers $\hat{\sigma}^2_{ML} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} = \frac{(y - X\hat{\beta}_{ML})'(y - X\hat{\beta}_{ML})}{n}$

Note that the MLE of β is the same as the OLS estimator, so in the case of the CLRM it is BLUE.

But unlike the OLS estimator of the variance of ε, the MLE of σ^2 is biased downward, but it has a lower variance too.

Small sample test won`t be valid because in reality we don't have the normality assumption, we need to rely on asymptotic tests since normality is not present.

Les tests (t et F) en petit échantillons ne seront pas valides par contre en l'absence de la normalité, il faudra utiliser les tests asymptotiques (la loi normale avec le test Z et les test Chi2) parce que l'on a pas la normalité des aléas.

The Cramer-Rao theorem
If there exist an estimator for which the variance is equal to the inferior boundery of the variance, it is given by the ML estimator.
The Cramer Rao theorem basically states that bourdery. It represents the inferior limit of the variance for all unbiased estimator (linear or non-linear). Saddly we cannot always attain that bourdery.

The Formal Cramer-Rao theorem:
The matrix $Var(\hat{\theta}\ ) - I^{-1}(\theta)$ is a semi-definite matrix where
$I(\theta) = -E\left(\frac{\partial^2 \ln L}{\partial\theta\partial\theta'}\right) = E\left(\frac{\partial \ln L}{\partial\theta}\frac{\partial \ln L}{\partial\theta'}\right)$ is called the Fisher information matrix and that matrix is positive definite.

$$I(\theta) = -E\begin{bmatrix} \frac{\partial^2 \ln L}{\partial\theta_1^2} & \frac{\partial^2 \ln L}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2 \ln L}{\partial\theta_1\partial\theta_K} \\ \frac{\partial^2 \ln L}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 \ln L}{\partial\theta_2^2} & \cdots & \frac{\partial^2 \ln L}{\partial\theta_2\partial\theta_K} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln L}{\partial\theta_K\partial\theta_1} & \frac{\partial^2 \ln L}{\partial\theta_K\partial\theta_1} & \cdots & \frac{\partial^2 \ln L}{\partial\theta_K^2} \end{bmatrix}$$

and for the Maximum Likelihood estimator, $Var(\hat{\theta}_{ML}) = I^{-1}(\theta)$, so the MLE is for sure a minimal variance estimator MVE. If it is also unbiased it is MVUE.

In the Linear model $y = X\beta + \varepsilon$

The second order derivatives would give us:

$$\frac{\partial \ln L(\theta \mid y, X)}{\partial\beta\partial\beta'} = -\frac{1}{2\sigma^2}(2X'X) = -\frac{(X'X)}{\sigma^2}$$

$$\frac{\partial \ln L(\theta \mid y, X)}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{(y - X\beta)'(y - X\beta)}{\sigma^6}$$

$$\frac{\partial \ln L(\theta \mid y, X)}{\partial\beta\partial(\sigma^2)} = \frac{1}{\sigma^4}(X'y - X'X\beta) = \frac{X'\varepsilon}{\sigma^4}$$

Taking the Expectation of each terms.

$$E\frac{\partial \ln L(\theta \mid y, X)}{\partial\beta\partial\beta'} = -\frac{(X'X)}{\sigma^2}$$

$$E\frac{\partial \ln L(\theta \mid y, X)}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} - n\frac{\sigma^2}{\sigma^6} = \frac{n}{2\sigma^4} - \frac{n2}{2\sigma^4} = -\frac{n}{2\sigma^4}$$

$$E\frac{\partial \ln L(\theta \mid y, X)}{\partial\beta\partial(\sigma^2)} = -\frac{1}{\sigma^4}(X'y - X'X\beta) = -\frac{E(X'\varepsilon)}{\sigma^4} = 0$$

Thus the ML variance is given by $\mathrm{var}(\hat{\theta}_{ML}) = I(\theta)^{-1} = \begin{bmatrix} \frac{(X'X)}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}^{-1} = \begin{bmatrix} \sigma^2 (X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$

Which is also the CLRM Cramer-Rao bound.

In practice, For more complex non-linear Likelihood estimator, to get an actual estimator of the variance of the parameters of the MLE we have the following equality

$\widehat{var}(\hat{\theta}) = I(\hat{\theta})^{-1}$ where $I(\hat{\theta})^{-1} = \left[-E\left(\frac{\partial^2 \ln}{\partial\theta\partial\theta'} \right)_{\theta=\hat{\theta}} \right]^{-1}$ evaluated at the found estimate $\theta = \hat{\theta}$

But since $I(\hat{\theta})^{-1}$ is often very complicated to obtain. More simpy we can calculate an estimate via the gradients

$I(\hat{\theta})^{-1} = \left[\sum_{i=1}^{n} \hat{g}_i \, \hat{g}_i \, ' \right]^{-1}$ where $\hat{g}_i = \left[\frac{\partial \ln L}{\partial\theta} \right]_{\theta=\hat{\theta}}$

In small samples, **no biased estimator** can attain the Cramer Rao lower bound.

Nevertheless in the CLRM asymtotically.

$$\begin{bmatrix} \hat{\beta}_{ML} \\ \hat{\sigma}^2_{ML} \end{bmatrix} \overset{a}{\sim} N\left[\begin{bmatrix} \beta \\ \sigma^2 \end{bmatrix}, \begin{bmatrix} \sigma^2 (X'X)^{-1} & 0 \\ 0 & 2\sigma^4 / n \end{bmatrix} \right]$$

Q4d) We need to write the density of y. This transformation gives us $f(y) = f(\varepsilon)\left|\frac{\partial \varepsilon}{\partial y'}\right|$ where $\left|\frac{\partial \varepsilon}{\partial y'}\right|$ is the absolute value of the determinant of the matrix of partial derivatives such that:

$$\frac{\partial \varepsilon}{\partial y'} = \begin{bmatrix} \frac{\partial \varepsilon_1}{\partial y_1} & \frac{\partial \varepsilon_1}{\partial y_2} & \cdots & \frac{\partial \varepsilon_1}{\partial y_n} \\ \frac{\partial \varepsilon_2}{\partial y_1} & \frac{\partial \varepsilon_2}{\partial y_2} & \cdots & \frac{\partial \varepsilon_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \varepsilon_n}{\partial y_1} & \frac{\partial \varepsilon_n}{\partial y_2} & \cdots & \frac{\partial \varepsilon_n}{\partial y_n} \end{bmatrix}$$ This is called the Jacobian.

And in the our case we have that $\left|\frac{\partial \varepsilon}{\partial y'}\right| = \left|I_n\right| = 1$ since $\varepsilon_i = y_i - x_{(i)}'\beta$

So we get that $f(y) = f(\varepsilon)\left|\frac{\partial \varepsilon}{\partial y'}\right| = f(\varepsilon)$

If we had $\varepsilon \sim N(0, \sigma^2\Omega)$, the density of ε would be given by the PDF of the normal distribution:

In matrix notation we thus have: $f(\varepsilon) = \frac{1}{(2\pi)^{n/2}}\left|\Sigma\right|^{-1/2}\exp\left(\frac{-1}{2}\frac{\varepsilon'\Sigma^{-1}\varepsilon}{1}\right) = \frac{1}{(2\pi)^{n/2}}\left|\sigma^2\Omega\right|^{-1/2}\exp\left(\frac{-1}{2}\frac{\varepsilon'\Omega^{-1}\varepsilon}{\sigma^2}\right)$

Playing with danger, here we can write the likelihood function for y as:

$$L(\theta \mid y, X) \equiv f(y) = \frac{1}{(2\pi\sigma^2)^{n/2}}\left|\Omega\right|^{-1/2}\exp\left(\frac{-1}{2}\frac{(y - X\beta)'\Omega^{-1}(y - X\beta)}{\sigma^2}\right) \text{ where } \theta = \begin{bmatrix} \beta \\ \sigma^2 \end{bmatrix}$$

Knowing that it is equivalent to maximize the likelihood or the log-likelihood. We can write the maximization problem as:

$$\max_{\theta\in\Theta} \ln L(\theta \mid y, X) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\sigma^2\Omega\right| - \frac{1}{2}(y - X\beta)'(\sigma^2\Omega)^{-1}(y - X\beta)$$

Or

$$\max_{\theta\in\Theta} \ln L(\theta \mid y, X) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left|\sigma^2\right| - \frac{1}{2}\ln\left|\Omega\right| - \frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta)$$

Since $\left|\sigma^2\Omega\right| = (\sigma^2)^n\left|\Omega\right|$

Note that the term in the brackets in the last term is

$$(y - X\beta)'\Omega^{-1}(y - X\beta) = (y - X\beta)'(\Omega^{-1}y - \Omega^{-1}X\beta)$$
$$= y'\Omega^{-1}y - \beta'X'\Omega^{-1}y - y'\Omega^{-1}X\beta + \beta'X'\Omega^{-1}X\beta$$
$$= y'\Omega^{-1}y - 2y'\Omega^{-1}X\beta + \beta'X'\Omega^{-1}X\beta$$

The FOC are given by:

$$\frac{\partial \ln L(\theta \mid y, X\ )}{\partial \beta} = -\frac{1}{2\hat{\sigma}^2_{ML}}(-2X\ '\Omega^{-1}y\ + 2X\ '\Omega^{-1}X\ \hat{\beta}_{ML}) = 0 = \frac{1}{\hat{\sigma}^2_{ML}}(X\ '\Omega^{-1}y\ + X\ '\Omega^{-1}X\ \hat{\beta}_{ML}) = 0$$

which gives us

$$\hat{\beta}_{ML} = (X\ '\Omega^{-1}X\ )^{-1}X\ '\Omega^{-1}y\ = (X\ '\tilde{P}'\underbrace{\tilde{P}X}_{\tilde{X}}\ )^{-1}X\ '\tilde{P}'\underbrace{\tilde{P}y}_{\tilde{y}}\ = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$$
$$= (X\ '\Sigma^{-1}X\ )^{-1}X\ '\Sigma^{-1}y\ = (X\ 'P'\underbrace{PX}_{X^*}\ )^{-1}X\ 'P'\underbrace{Py}_{y^*}\ = (X^*{}'X^*)^{-1}X^*{}'y^*$$

$$\frac{\partial \ln L(\theta \mid y, X)}{\partial \sigma^2} = -\frac{n}{2\hat{\sigma}^2_{ML}} + \frac{1}{2\hat{\sigma}^4_{ML}}(y\ - X\ \hat{\beta})'\Omega^{-1}(y\ - X\ \hat{\beta}_{ML}) = 0$$ which delivers $\hat{\sigma}^2_{ML} = \dfrac{\hat{\varepsilon}'\Omega^{-1}\hat{\varepsilon}}{n}$

Note that the MLE of β is the same as the GLS estimator, so in the case where we regress the properly transformed model (same as in the GLS case) that transformed model corresponds to the CLRM and it is also BLUE.

But unlike the GLS estimator of the variance of ε, the MLE of σ^2 is biased downward.

The variance of the estimator is defined by the information matrix, hence:
$$\text{var}(\hat{\beta}) = \sigma^2(X'\Omega^{-1}X)^{-1} = \sigma^2(X'\tilde{P}'\tilde{P}X)^{-1} = \sigma^2(\tilde{X}'\tilde{X})^{-1} = (X'\Sigma^{-1}X)^{-1} == (X'P'PX)^{-1} = (X^*{}'X^*)^{-1}$$

In our case since we don’t actually have normality in small samples,
Typical small sample tests won`t be valid, we will need to rely on asymptotic tests since normality is not present.

Nevertheless asymptotically
$$\begin{bmatrix}\hat{\beta}_{ML}\\ \hat{\sigma}^2_{ML}\end{bmatrix} \overset{a}{\sim} N\left[\begin{bmatrix}\beta\\ \sigma^2\end{bmatrix},\begin{bmatrix}\sigma^2(X'\Omega^{-1}X)^{-1} & 0\\ 0 & 2\sigma^4/n\end{bmatrix}\right]$$

The other problem that arise in practice is that Ω is unknown, but it is a symmetric matrix which contains $\dfrac{n(n+1)}{2}$ different elements. Saddly, we only have n observations to estimate $\dfrac{n(n+1)}{2}$ elements.
So we would have to impose some restrictions to proceed to actual estimation when in practice Ω is unknown.