Econ 203B: Single Equation Models

Solutions for Problem Set 3

Michael Powell
Department of Economics, UCLA
February 3rd, 2006

1. The regression slope $\hat{\beta}$ in a CNLR model is distributed $N\left(\beta, \sigma_{\hat{\beta}}^2\right)$ where $\sigma_{\hat{\beta}}^2 = 1$. The null hypothesis $\beta = 0$ will be tested at the 10% significance using the statistic $Z_0 = \frac{\hat{\beta}}{\sigma_{\hat{\beta}}}$. That is, the null will be rejected if and only if $|Z_0| > 1.645$.

a. Write and run a program that tabulates the power of the test at the following 9 values of the true parameter β: -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2.

Solution First, note that since $\hat{\beta} \sim N(\beta, 1)$, $\hat{\beta} - \beta \sim N(0,1)$. Recall that the power function of a test (with critical region C), as a function of the parameters (θ) of the underlying distribution is given by:

$$power\,(\theta) = \Pr\left[(x_1, \ldots, x_n) \in C|\, \theta\right]$$

For the given test, we see that $(x_1, \ldots, x_n) \in C$ if and only if $|Z_0| > 1.645$, giving us

$$\begin{aligned} power\,(\beta) &= \Pr\left[|Z_0| > 1.645|\, \beta\right] \\ &= \Pr\left[\left|\hat{\beta}\right| > 1.645\middle|\, \beta\right] \\ &= \Pr\left[\hat{\beta} > 1.645\middle|\, \beta\right] + \Pr\left[\hat{\beta} < -1.645\middle|\, \beta\right] \\ &= \Pr\left[\hat{\beta} - \beta > 1.645 - \beta\middle|\, \beta\right] + \Pr\left[\hat{\beta} - \beta < -1.645 - \beta\middle|\, \beta\right] \\ &= 1 - \Phi(1.645 - \beta) + \Phi(-1.645 - \beta) \end{aligned}$$

Where the third equality holds by the definition of absolute value and the last equality holds since $\hat{\beta} - \beta \sim N(0,1)$.

This problem then requires us simply to plug in the nine given values of β into the above equation. MATLAB gives us:

β	$power\,(\beta)$
-2	0.6388
-1.5	0.4432
-1	0.2635
-0.5	0.1421
0	0.1000
0.5	0.1421
1	0.2635
1.5	0.4432
2	0.6388

b. Redo (a) for the situation where $\sigma_{\hat{\beta}}^2 = 4$.

Solution Proceeding exactly as above, first, note that since $\hat{\beta} \sim N(\beta, 4)$, $\frac{\hat{\beta} - \beta}{2} \sim N(0,1)$.

For the given test, $(x_1, \ldots, x_n) \in C$ if and only if $|Z_0| > 1.645$, giving us

$$\begin{aligned}
power\,(\beta) &= \Pr\left[|Z_0| > 1.645|\,\beta\right] \\
&= \Pr\left[\left|\frac{\hat{\beta}}{2}\right| > 1.645\middle|\,\beta\right] \\
&= \Pr\left[\frac{\hat{\beta}}{2} > 1.645\middle|\,\beta\right] + \Pr\left[\frac{\hat{\beta}}{2} < -1.645\middle|\,\beta\right] \\
&= \Pr\left[\frac{\hat{\beta}-\beta}{2} > 1.645 - \frac{\beta}{2}\middle|\,\beta\right] + \Pr\left[\frac{\hat{\beta}-\beta}{2} < -1.645 - \frac{\beta}{2}\middle|\,\beta\right] \\
&= 1 - \Phi\left(1.645 - \frac{\beta}{2}\right) + \Phi\left(-1.645 - \frac{\beta}{2}\right)
\end{aligned}$$

Where the third equality once again holds by the definition of absolute value and the last equality holds since $\frac{\hat{\beta}-\beta}{2} \sim N(0,1)$.

Plugging in the nine given values of β into the above equation, MATLAB gives us:

β	$power\,(\beta)$
-2	0.2635
-1.5	0.1937
-1	0.1421
-0.5	0.1106
0	0.1000
0.5	0.1106
1	0.1421
1.5	0.1937
2	0.2635

c. What do your two tables tell you about the effect of $\sigma^2_{\hat{\beta}}$ on the power of the test?

Solution The tables give the intuitive result that as the standard error of the estimator increases, the power of any tests associated with it will tend to decrease.

2. The regression slopes $\hat{\beta}_1$ and $\hat{\beta}_2$ in a CNLR model are distributed as a bivariate normal:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\right)$$

where $r = 0.6$. The joint null hypothesis $\beta_1 = \beta_2 = 0$ will be tested at the 5% significance level by using the statistic

$$W_0 = \frac{\left(\hat{\beta}_1^2 + \hat{\beta}_2^2 - 2r\hat{\beta}_1\hat{\beta}_2\right)}{(1-r^2)}$$

That is, the null will be rejected if and only if $|W_0| > 5.99$.

a. Write and run a program that tabulates the power of the test at the following 9 pairs of the true parameter vector (β_1, β_2) : $(-1,1)$, $(-1,0)$, $(-1,1)$, $(0,1)$, $(0,0)$, $(0,-1)$, $(1,1)$, $(1,0)$, $(1,-1)$.

Solution We can rewrite our test statistic as:

$$W_0 = \left(\Gamma\hat{\beta} - \gamma_0\right)' \Omega_\Gamma^{-1} \left(\Gamma\hat{\beta} - \gamma_0\right)$$

Where $\Gamma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\gamma_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, and

$$\begin{aligned} \Omega_\Gamma^{-1} &= \left(\Gamma \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \Gamma' \right)^{-1} = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}^{-1} \\ &= \frac{1}{1-r^2} \begin{bmatrix} 1 & -r \\ -r & 1 \end{bmatrix} \end{aligned}$$

To verify that this is indeed equal to the given statistic:

$$\begin{aligned} W_0 &= \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}' \frac{1}{1-r^2} \begin{bmatrix} 1 & -r \\ -r & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \\ &= \frac{1}{1-r^2} \begin{bmatrix} \hat{\beta}_1 - r\hat{\beta}_2 & -r\hat{\beta}_1 + \hat{\beta}_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \\ &= \frac{1}{1-r^2} \left(\hat{\beta}_1^2 - r\hat{\beta}_1\hat{\beta}_2 - r\hat{\beta}_1\hat{\beta}_2 + \hat{\beta}_2^2 \right) \\ &= \frac{\hat{\beta}_1^2 + \hat{\beta}_2^2 - 2r\hat{\beta}_1\hat{\beta}_2}{1-r^2} \end{aligned}$$

The power function is therefore:

$$\begin{aligned} power\,(\gamma) &= \Pr\left[\left(\Gamma\hat{\beta} - \gamma_0 \right)' \Omega_\Gamma^{-1} \left(\Gamma\hat{\beta} - \gamma_0 \right) > 5.99 \middle| \Gamma\beta = \gamma \right] \\ &= \Pr\left[\left(\hat{\beta} - \gamma_0 \right)' \Omega_\Gamma^{-1} \left(\hat{\beta} - \gamma_0 \right) > 5.99 \middle| \beta = \gamma \right] \\ &= \Pr\left[\left(\hat{\beta} - \gamma + (\gamma - \gamma_0) \right)' \Omega_\Gamma^{-1} \left(\hat{\beta} - \gamma + (\gamma - \gamma_0) \right) > 5.99 \middle| \beta = \gamma \right] \\ &= 1 - \Pr\left[\left(\hat{\beta} - \gamma + (\gamma - \gamma_0) \right)' \Omega_\Gamma^{-1} \left(\hat{\beta} - \gamma + (\gamma - \gamma_0) \right) < 5.99 \middle| \beta = \gamma \right] \\ &= 1 - G_{\chi^2(2,\Delta)}\,(5.99) \end{aligned}$$

Where

$$\begin{aligned} \Delta &= (\gamma - \gamma_0)' \, \Omega_\Gamma^{-1} \, (\gamma - \gamma_0) = \frac{1}{1-r^2} \gamma' \begin{bmatrix} 1 & -r \\ -r & 1 \end{bmatrix} \gamma \\ &= 1.5625 \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix} \begin{bmatrix} 1 & -0.6 \\ -0.6 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \end{aligned}$$

Plugging the nine given values into MATLAB, we have:

(β_1, β_2)	$power\,((\beta_1, \beta_2))$
$(-1, 1)$	0.5038
$(-1, 0)$	0.1843
$(-1, -1)$	0.1553
$(0, 1)$	0.1843
$(0, 0)$	0.0500
$(0, -1)$	0.1843
$(1, 1)$	0.1553
$(1, 0)$	0.1843
$(1, -1)$	0.5038

b. Redo (a) for the situation where $r = -0.6$.

Solution The only aspect of this problem which changes from parts (a) to (b) is the noncentrality parameter:

$$\Delta = \frac{1}{1-r^2}\gamma' \begin{bmatrix} 1 & -r \\ -r & 1 \end{bmatrix} \gamma = 1.5625 \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix} \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Plugging the nine given values into MATLAB, we have:

(β_1, β_2)	$power\left((\beta_1, \beta_2)\right)$
$(-1, 1)$	0.1553
$(-1, 0)$	0.1843
$(-1, -1)$	0.5038
$(0, 1)$	0.1843
$(0, 0)$	0.0500
$(0, -1)$	0.1843
$(1, 1)$	0.5038
$(1, 0)$	0.1843
$(1, -1)$	0.1553

c. What do your two tables tell you about the effect of the correlation r on the power of the test?

Solution The power function is a multivariable function of β_1 and β_2. It appears that a change in r from 0.6 to -0.6 serves to rotate the power function around the z axis by 90 degrees.

3. Suppose that Y_i is a discrete random variable (actually a *count variable*, such as, for example, number of examples) whose conditional distribution given X_i is

$$\Pr\left(Y_i = y_i \mid X_i = x_i\right) = \frac{e^{-\beta x_i}\left(\beta x_i\right)^{y_i}}{y_i!},\ y_i = 0, 1, 2, \ldots$$

where X_i is a positive scalar random variable and $\beta > 0$. Assume that we have n independent observations $\{(Y_i, X_i)\}_{i=1}^n$.

a. Write down the (conditional) log-likelihood function of the sample and compute the maximum likelihood estimator of β, $\hat{\beta}_{MLE}$.

Solution Since $\{(Y_i, X_i)\}_{i=1}^n$ are independent, we have: (where $x = (x_i)_{i=1}^n$, and $y = (y_i)_{i=1}^n$)

$$f\left((x, y) \mid x; \beta\right) = \prod_{i=1}^n f\left(y_i \mid x_i; \beta\right) = \prod_{i=1}^n \frac{e^{-\beta x_i}\left(\beta x_i\right)^{y_i}}{y_i!}$$

Taking logs, we have:

$$\begin{aligned}
\log L\left(\beta; (x, y)\right) &= \sum_{i=1}^n \log\left[\frac{e^{-\beta x_i}\left(\beta x_i\right)^{y_i}}{y_i!}\right] \\
&= \sum_{i=1}^n \left(-\beta x_i + y_i \log\beta + y_i \log x_i - \log y_i!\right) \\
&= -\beta \sum_{i=1}^n x_i + \log\beta \sum_{i=1}^n y_i + \sum_{i=1}^n \left(y_i \log x_i - \log y_i!\right)
\end{aligned}$$

Recall that the maximum likelihood estimator is:

$$\hat{\beta}_{MLE} = \arg\min_{\beta} \log L\left(\beta; (x, y)\right)$$

Taking first order conditions, we have:

$$(\beta) : -\sum_{i=1}^{n} x_i + \frac{1}{\hat{\beta}_{MLE}} \sum_{i=1}^{n} y_i = 0$$

Which leads to

$$\hat{\beta}_{MLE} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} = \frac{\frac{1}{n}\sum_{i=1}^n y_i}{\frac{1}{n}\sum_{i=1}^n x_i} = \frac{\bar{y}}{\bar{x}}$$

b. Find the asymptotic distribution of $\hat{\beta}_{MLE}$.

Solution Recall that it is a property of maximum likelihood estimators that

$$\sqrt{n}\left(\hat{\beta}_{MLE} - \beta\right) \xrightarrow{d} N\left(0, \left(I_1\left(\hat{\beta}_{MLE}\right)\right)^{-1}\right)$$

Therefore, we need only compute $\left(I_1\left(\hat{\beta}_{MLE}\right)\right)^{-1}$. Proceeding in the usual way:

$$\log L\left(\beta; (x_i, y_i)\middle|\, x_i\right) = -\beta x_i + y_i \log\beta + y_i \log x_i - \log y_i!$$

Taking the partial derivatives with respect to β:

$$\begin{aligned} \frac{\partial L}{\partial \beta} &= -x_i + \frac{y_i}{\beta} \\ \frac{\partial^2 L}{\partial \beta^2} &= -\frac{y_i}{\beta^2} \end{aligned}$$

This yields:

$$I_1(\beta) = -E\left[\frac{\partial^2 L}{\partial \beta^2}\right] = E\left[\frac{Y_i}{\beta^2}\right] = \frac{1}{\beta^2} E\left[Y_i\right]$$

It remains to find $E[Y_i]$. Noting that $Y_i|\, x_i \sim Poisson\left(\beta x_i\right)$ and using the law of iterated expectations, we have:

$$\begin{aligned} E[Y_i] &= E\left[E\left[Y_i|\, X_i\right]\right] \\ &= E\left[\beta X_i\right] = \beta E\left[X_i\right] \end{aligned}$$

This gives us:

$$I_1(\beta) = \frac{1}{\beta^2} \beta E[x_i] = \frac{1}{\beta} E[x_i]$$

Thus, we have that:

$$\sqrt{n}\left(\hat{\beta}_{MLE} - \beta\right) \xrightarrow{d} N\left(0, \frac{\beta}{E[X_i]}\right)$$

c. Find the exact distribution of $\hat{\beta}_{MLE}$. (HINT: The sum of independent Poisson random variables with parameters λ_j is a Poisson variable with parameter $\sum_j \lambda_j$.)

Solution Here, we have:

$$\begin{aligned} \Pr\left(\hat{\beta}_{MLE} = b \,\middle|\, x\right) &= \Pr\left(\frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i} = b \,\middle|\, x\right) \\ &= \Pr\left(\sum_{i=1}^{n} Y_i = b \sum_{i=1}^{n} x_i \,\middle|\, x\right) \end{aligned}$$

Since $Y_i \sim Poisson\,(\beta x_i)$, we have that $\sum_{i=1}^n Y_i|\, x \sim Poisson\,(\beta \sum_{i=1}^n x_i)$ and therefore,

$$\Pr\left(\sum_{i=1}^n Y_i = b\sum_{i=1}^n x_i \middle|\, x\right) = \frac{\exp\left\{-\beta\sum_{i=1}^n x_i\right\}\cdot\left(\beta\sum_{i=1}^n x_i\right)^{b\sum_{i=1}^n x_i}}{\left(b\sum_{i=1}^n x_i\right)!}1_{\left\{b\sum_{i=1}^n x_i\in\mathbb{N}\right\}}$$

Thus,

$$\Pr\left(\hat\beta_{MLE} = b\middle|\, x\right) = \frac{\exp\left\{-\beta\sum_{i=1}^n x_i\right\}\cdot\left(\beta\sum_{i=1}^n x_i\right)^{b\sum_{i=1}^n x_i}}{\left(b\sum_{i=1}^n x_i\right)!}1_{\left\{b\sum_{i=1}^n x_i\in\mathbb{N}\right\}}$$

4. In the standard CNLR model,

$$Y_i = X_i\beta + \varepsilon_i$$

where $\varepsilon|\, X \sim N\left(0, \sigma^2 I_n\right)$, assume that $K = 1$ and that $\sigma^2 = \beta^2$. Obtain the maximum likelihood estimator for β and the Cramer-Rao lower bound.

Solution Under the CNLR assumptions, we know that (since $\sigma^2 = \beta^2$)

$$Y_i|\, X_i \sim N\left(X_i\beta, \beta^2\right) \text{ and } Y_i|\, X_i \amalg Y_j|\, X_j,\ i \neq j$$

Therefore, we have:

$$\begin{aligned} L(\beta; X, Y) &= \prod_{i=1}^n \frac{1}{\beta\sqrt{2\pi}}\exp\left\{-\frac{(Y_i - X_i\beta)^2}{2\beta^2}\right\} \\ &= \prod_{i=1}^n \frac{1}{\beta\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{Y_i}{\beta} - X_i\right)^2\right\} \end{aligned}$$

Taking logs:

$$\begin{aligned} \log L(\beta; X, Y) &= \sum_{i=1}^n\left(-\log\beta - \frac{1}{2}\log(2\pi) - \frac{1}{2}\left(\frac{Y_i}{\beta} - X_i\right)^2\right) \\ &= -n\log\beta - \frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^n\left(\frac{Y_i}{\beta} - X_i\right)^2 \end{aligned}$$

The first order condition with respect to β is:

$$\begin{aligned} (\beta) \quad : \quad 0 &= -\frac{n}{\hat\beta_{MLE}} - \frac{2}{2}\sum_{i=1}^n\left(\frac{Y_i}{\hat\beta_{MLE}} - X_i\right)\left(-\frac{Y_i}{\hat\beta^2_{MLE}}\right) \\ &= -\frac{n}{\hat\beta_{MLE}} + \sum_{i=1}^n\left(\frac{Y_i - \hat\beta_{MLE}X_i}{\hat\beta_{MLE}}\right)\left(\frac{Y_i}{\hat\beta^2_{MLE}}\right) \end{aligned}$$

Rearranging:

$$\frac{1}{\hat\beta^3_{MLE}}\sum_{i=1}^n\left(Y_i - \hat\beta_{MLE}X_i\right)(Y_i) - \frac{n}{\hat\beta_{MLE}} = 0$$

$$\frac{1}{\hat\beta^3_{MLE}}\sum_{i=1}^n Y_i^2 - \frac{1}{\hat\beta^2_{MLE}}\sum_{i=1}^n X_iY_i - \frac{n}{\hat\beta_{MLE}} = 0$$

$$n\hat\beta^2_{MLE} + \hat\beta_{MLE}\sum_{i=1}^n X_iY_i - \sum_{i=1}^n Y_i^2 = 0$$

Appealing to the quadratic formula:

$$\begin{aligned}\hat{\beta}_{MLE} &= \frac{-\sum_{i=1}^{n} X_iY_i \pm \sqrt{\left(\sum_{i=1}^{n} X_iY_i\right)^2 - 4n\sum_{i=1}^{n} Y_i^2}}{2n} \\ &= \frac{-X'Y \pm \sqrt{\left(X'Y\right)^2 - 4nY'Y}}{2n}\end{aligned}$$

Now, to find the Fisher information:

$$\log L\left(\beta; X, Y\right) = -n\log\beta - \frac{n}{2}\log\left(2\pi\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{Y_i}{\beta} - X_i\right)^2$$

Taking derivatives:

$$\begin{aligned}\frac{\partial \log L\left(\beta; X, Y\right)}{\partial\beta} &= -\frac{n}{\beta} + \frac{1}{\beta^3}\sum_{i=1}^{n}\left(Y_i - \beta X_i\right)\left(Y_i\right) \\ &= -\frac{n}{\beta} + \frac{1}{\beta^3}\sum_{i=1}^{n} Y_i^2 - \frac{1}{\beta^2}\sum_{i=1}^{n} X_iY_i \\ &= -\frac{n}{\beta} + \frac{1}{\beta^3}Y'Y - \frac{1}{\beta^2}X'Y\end{aligned}$$

$$\frac{\partial^2 \log L\left(\beta; X, Y\right)}{\partial\beta^2} = \frac{n}{\beta^2} - \frac{3}{\beta^4}Y'Y + \frac{2}{\beta^3}X'Y$$

This gives us:

$$\begin{aligned}I_n\left(\beta\middle| X\right) &= -E\left[\frac{\partial^2 \log L\left(\beta; X, Y\right)}{\partial\beta^2}\middle| X\right] \\ &= E\left[\frac{3}{\beta^4}Y'Y\middle| X\right] - E\left[\frac{2}{\beta^3}X'Y\middle| X\right] - \frac{n}{\beta^2} \\ &= \frac{3}{\beta^4}E\left[Y'Y\middle| X\right] - \frac{2}{\beta^3}E\left[X'Y\middle| X\right] - \frac{n}{\beta^2}\end{aligned}$$

Noticing that

$$\begin{aligned}E\left[Y'Y\middle| X\right] &= E\left[\left(X\beta+\varepsilon\right)'\left(X\beta+\varepsilon\right)\middle| X\right] \\ &= E\left[\beta'X'X\beta + \varepsilon'X\beta + \beta'X'\varepsilon + \varepsilon'\varepsilon\middle| X\right] \\ &= \beta^2X'X + 2\beta\cdot E\left[\varepsilon'X\middle| X\right] + E\left[\varepsilon'\varepsilon\middle| X\right] \\ &= \beta^2X'X + 2\beta\underbrace{E\left[\varepsilon'\middle| X\right]}_{=0}X + n\underbrace{\sigma^2}_{=\beta^2} \\ &= \beta^2X'X + n\beta^2\end{aligned}$$

And

$$E\left[X'Y\middle| X\right] = X'E\left[Y\middle| X\right] = X'X\beta$$

We have:

$$\begin{aligned}I_n\left(\beta\middle| X\right) &= \frac{3\left(\beta^2X'X + n\beta^2\right)}{\beta^4} - \frac{2X'X\beta}{\beta^3} - \frac{n}{\beta^2} \\ &= \frac{3X'X + 3n}{\beta^2} - \frac{2X'X}{\beta^2} - \frac{n}{\beta^2} = \frac{X'X + 2n}{\beta^2}\end{aligned}$$

This gives us the Cramer-Rao lower bound:

$$CRLB\left(\beta|\, X\right) = I_n\left(\beta|\, X\right)^{-1} = \frac{\beta^2}{\sum_{i=1}^n X_i^2 + 2n}$$

5. Derive the log-likelihood function, the first order conditions for maximization, and the information matrix for the model:

$$\begin{aligned} Y_i &= X_i\beta + \varepsilon_i \\ \varepsilon_i|\, X_i &\sim N\left(0, (Z_i\gamma)^2\right) \end{aligned}$$

assuming i.i.d. sampling of (Y_i, X_i) across individuals. Here Z_i is a $1 \times r$ subvector of X_i.

Solution Putting the two assumptions together, we know that

$$Y_i|\, X_i \sim N\left(X_i\beta, (Z_i\gamma)^2\right) \text{ and } Y_i|\, X_i \amalg Y_j|\, X_j,\, i \neq j$$

Since Z_i is $1 \times r$, it must be the case that γ is $r \times 1$.

The likelihood function is therefore:

$$L\left(\beta, \gamma; X, Y\right) = \prod_{i=1}^n \frac{1}{Z_i\gamma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{Y_i - X_i\beta}{Z_i\gamma}\right)^2\right\}$$

Taking logs,

$$\begin{aligned} \log L\left(\beta, \gamma; X, Y\right) &= \sum_{i=1}^n \left(-\log Z_i\gamma - \frac{1}{2}\log 2\pi - \frac{1}{2}\left(\frac{Y_i - X_i\beta}{Z_i\gamma}\right)^2\right) \\ &= -\frac{n}{2}\log 2\pi - \sum_{i=1}^n \log Z_i\gamma - \frac{1}{2}\sum_{i=1}^n \frac{\left(Y_i - X_i\beta\right)^2}{\left(Z_i\gamma\right)^2} \end{aligned}$$

Since (β, γ) is a $(k + r) \times 1$ vector, we will have $k + r$ first order conditions. I will group them together by using the tools of matrix calculus. Throughout this problem, I will make use of the fact that $Y_i - X_i\beta$ is a scalar. The first order conditions are:

$$\begin{aligned} (\beta) \quad &: \quad 0 = \frac{\partial \log L}{\partial \beta} = -\frac{2}{2}\sum_{i=1}^n \frac{\left(Y_i - X_i\hat{\beta}_{MLE}\right)}{\left(Z_i\hat{\gamma}_{MLE}\right)^2}\left(-X_i'\right) \\ &= \sum_{i=1}^n \frac{X_i'Y_i}{\left(Z_i\hat{\gamma}_{MLE}\right)^2} - \sum_{i=1}^n \frac{X_i'X_i\hat{\beta}_{MLE}}{\left(Z_i\hat{\gamma}_{MLE}\right)^2} \end{aligned}$$

$$\begin{aligned} (\gamma) \quad &: \quad 0 = \frac{\partial \log L}{\partial \gamma} = -\sum_{i=1}^n \frac{1}{Z_i\hat{\gamma}_{MLE}} Z_i' + \frac{2}{2}\sum_{i=1}^n \frac{\left(Y_i - X_i\hat{\beta}_{MLE}\right)^2}{\left(Z_i\hat{\gamma}_{MLE}\right)^3} Z_i' \\ &= -\sum_{i=1}^n \frac{Z_i'}{Z_i\hat{\gamma}_{MLE}} + \sum_{i=1}^n \frac{Z_i'\left(Y_i - X_i\hat{\beta}_{MLE}\right)^2}{\left(Z_i\hat{\gamma}_{MLE}\right)^3} \end{aligned}$$

Where I used the chain rule and the result that if x is a $1 \times n$ vector and y is an $n \times 1$ vector, then

$$\frac{\partial xy}{\partial y} = x'$$

Now, moving on to calculate the Fisher information, the first partial derivatives are (from above):

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{n} \frac{X_i' Y_i}{(Z_i\gamma)^2} - \sum_{i=1}^{n} \frac{X_i' X_i \beta}{(Z_i\gamma)^2}$$

$$\frac{\partial \log L}{\partial \gamma} = -\sum_{i=1}^{n} \frac{Z_i'}{Z_i\gamma} + \sum_{i=1}^{n} \frac{Z_i' (Y_i - X_i\beta)^2}{(Z_i\gamma)^3}$$

The second order partial derivatives are:

$$\frac{\partial^2 \log L}{\partial \beta \partial \beta'} = -\sum_{i=1}^{n} \frac{X_i' X_i}{(Z_i\gamma)^2}$$

$$\begin{aligned} \frac{\partial^2 \log L}{\partial \beta \partial \gamma'} &= -2\sum_{i=1}^{n} \frac{X_i' Y_i}{(Z_i\gamma)^3} Z_i - 2\sum \frac{X_i' X_i \beta}{(Z_i\gamma)^3} Z_i \\ &= -2\sum_{i=1}^{n} \frac{X_i' Z_i (Y_i - X_i\beta)}{(Z_i\gamma)^3} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \log L}{\partial \gamma \partial \beta'} &= -2\sum_{i=1}^{n} \frac{Z_i' (Y_i - X_i\beta)}{(Z_i\gamma)^3} (X_i) \\ &= -2\sum_{i=1}^{n} \frac{Z_i' X_i (Y_i - X_i\beta)}{(Z_i\gamma)^3} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \log L}{\partial \gamma \partial \gamma'} &= \sum_{i=1}^{n} \frac{Z_i'}{(Z_i\gamma)^2} Z_i - 3\sum_{i=1}^{n} \frac{Z_i' (Y_i - X_i\beta)^2}{(Z_i\gamma)^4} Z_i \\ &= \sum_{i=1}^{n} \frac{Z_i' Z_i}{(Z_i\gamma)^2} - 3\sum_{i=1}^{n} \frac{Z_i' Z_i (Y_i - X_i\beta)^2}{(Z_i\gamma)^4} \end{aligned}$$

Where I used the results that if x is a $1 \times n$ vector, y is an $n \times 1$ vector, then

$$\frac{\partial xy}{\partial y'} = x$$

Before putting this into the information matrix, I will simplify by computing the expected values, conditional on X.

$$-E\left[\left. \frac{\partial^2 \log L}{\partial \beta \partial \beta'} \right| X \right] = E\left[\left. \sum_{i=1}^{n} \frac{X_i' X_i}{(Z_i\gamma)^2} \right| X \right] = \sum_{i=1}^{n} \frac{X_i' X_i}{(Z_i\gamma)^2}$$

$$\begin{aligned} -E\left[\left. \frac{\partial^2 \log L}{\partial \beta \partial \gamma'} \right| X \right] &= E\left[\left. 2\sum_{i=1}^{n} \frac{X_i' Z_i (Y_i - X_i\beta)}{(Z_i\gamma)^3} \right| X \right] \\ &= 2\sum_{i=1}^{n} \frac{X_i' Z_i (E[Y_i | X] - X_i\beta)}{(Z_i\gamma)^3} \\ &= 2\sum_{i=1}^{n} \frac{X_i' Z_i (X_i\beta - X_i\beta)}{(Z_i\gamma)^3} = 0 \end{aligned}$$

$$
\begin{aligned}
-E\left[\left.\frac{\partial^2 \log L}{\partial\gamma\partial\beta'}\right| X\right] &= E\left[\left. 2\sum_{i=1}^{n}\frac{Z_i' X_i\left(Y_i - X_i\beta\right)}{\left(Z_i\gamma\right)^3}\right| X\right] \\
&= 2\sum_{i=1}^{n}\frac{Z_i' X_i\left(E\left[\left. Y_i\right| X\right] - X_i\beta\right)}{\left(Z_i\gamma\right)^3} \\
&= 2\sum_{i=1}^{n}\frac{Z_i X_i\left(X_i\beta - X_i\beta\right)}{\left(Z_i\gamma\right)^3} = 0
\end{aligned}
$$

$$
\begin{aligned}
-E\left[\left.\frac{\partial^2 \log L}{\partial\gamma\partial\gamma'}\right| X\right] &= E\left[\left. 3\sum_{i=1}^{n}\frac{Z_i' Z_i\left(Y_i - X_i\beta\right)^2}{\left(Z_i\gamma\right)^4} - \sum_{i=1}^{n}\frac{Z_i' Z_i}{\left(Z_i\gamma\right)^2}\right| X\right] \\
&= 3\sum_{i=1}^{n}\frac{Z_i' Z_i E\left[\left.\left(Y_i - X_i\beta\right)^2\right| X\right]}{\left(Z_i\gamma\right)^4} - \sum_{i=1}^{n}\frac{Z_i' Z_i}{\left(Z_i\gamma\right)^2} \\
&= 3\sum_{i=1}^{n}\frac{Z_i' Z_i Var\left(\left. Y_i\right| X\right)}{\left(Z_i\gamma\right)^4} - \sum_{i=1}^{n}\frac{Z_i' Z_i}{\left(Z_i\gamma\right)^2} \\
&= 3\sum_{i=1}^{n}\frac{Z_i' Z_i\left(Z_i\gamma\right)^2}{\left(Z_i\gamma\right)^4} - \sum_{i=1}^{n}\frac{Z_i' Z_i}{\left(Z_i\gamma\right)^2} = 2\sum_{i=1}^{n}\frac{Z_i' Z_i}{\left(Z_i\gamma\right)^2}
\end{aligned}
$$

Putting this together, we have:

$$
\begin{aligned}
I\left(\beta,\gamma\right| X) &= \begin{bmatrix} -E\left[\left.\frac{\partial^2 \log L}{\partial\beta\partial\beta'}\right| X\right] & -E\left[\left.\frac{\partial^2 \log L}{\partial\beta\partial\gamma'}\right| X\right] \\ -E\left[\left.\frac{\partial^2 \log L}{\partial\gamma\partial\beta'}\right| X\right] & -E\left[\left.\frac{\partial^2 \log L}{\partial\gamma\partial\gamma'}\right| X\right] \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^{n}\frac{X_i' X_i}{(Z_i\gamma)^2} & 0 \\ 0 & 2\sum_{i=1}^{n}\frac{Z_i' Z_i}{(Z_i\gamma)^2} \end{bmatrix}
\end{aligned}
$$

Using the law of iterated expectations to find the unconditional information matrix:

$$
I\left(\beta,\gamma\right) = E\left[I\left(\beta,\gamma\right| X)\right] = \begin{bmatrix} \sum_{i=1}^{n} E\left[\frac{X_i' X_i}{(Z_i\gamma)^2}\right] & 0 \\ 0 & 2\sum_{i=1}^{n} E\left[\frac{Z_i' Z_i}{(Z_i\gamma)^2}\right] \end{bmatrix}
$$

6. Five sample observations are

X	4	1	5	8	2
Y	6	3	12	15	4

Assume a linear model, $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$, with heteroskedasticity of the form $Var\left(Y_i\right) = Var\left(\varepsilon_i\right) = \sigma_i^2 = \sigma^2 X_i^2$ where σ^2 is a positive constant. Calculate the OLS and GLS estimates of β_1 and β_2 and the corresponding standard errors.

Solution The relevant matrices for this problem are:

$$
X = \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 5 \\ 1 & 8 \\ 1 & 2 \end{bmatrix},\ Y = \begin{bmatrix} 6 \\ 3 \\ 12 \\ 15 \\ 4 \end{bmatrix},\ \Omega = \sigma^2 \begin{bmatrix} 16 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 25 & 0 & 0 \\ 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}
$$

Then, calculating the estimators:

$$
\hat\beta_{OLS} = \begin{bmatrix} \hat\beta_{1,OLS} \\ \hat\beta_{2,OLS} \end{bmatrix} = \left(X'X\right)^{-1} X'Y
$$

$$= \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 1 & 5 & 8 & 2 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 5 \\ 1 & 8 \\ 1 & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 1 & 5 & 8 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 3 \\ 12 \\ 15 \\ 4 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 20 \\ 20 & 110 \end{bmatrix}^{-1} \begin{bmatrix} 40 \\ 215 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{11}{6} \end{bmatrix}$$

And

$$\hat{\beta}_{GLS} = \begin{bmatrix} \hat{\beta}_{1,GLS} \\ \hat{\beta}_{2,GLS} \end{bmatrix} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y$$

$$= \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 1 & 5 & 8 & 2 \end{bmatrix} \begin{bmatrix} 16 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 25 & 0 & 0 \\ 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 5 \\ 1 & 8 \\ 1 & 2 \end{bmatrix} \right)^{-1} \cdot$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 1 & 5 & 8 & 2 \end{bmatrix} \begin{bmatrix} 16 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 25 & 0 & 0 \\ 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ 3 \\ 12 \\ 15 \\ 4 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1000}{507} & -\frac{415}{507} \\ -\frac{415}{507} & \frac{2189}{4056} \end{bmatrix} \begin{bmatrix} \frac{8143}{1600} \\ \frac{431}{40} \end{bmatrix} = \begin{bmatrix} \frac{2471.}{2028} \\ \frac{26759.}{16224} \end{bmatrix} \approx \begin{bmatrix} 1.2184 \\ 1.6493 \end{bmatrix}$$

Next, proceeding to compute the variance-covariance matrices for $\hat{\beta}_{OLS}$ and $\hat{\beta}_{GLS}$, from which we will be able to get the respective standard errors, recall that

$$\begin{aligned} Var\left(\hat{\beta}_{OLS}\middle| X\right) &= Var\left((X'X)^{-1} X'Y\middle| X\right) \\ &= (X'X)^{-1} X'Var\left(Y\middle| X\right) X \left(X'X\right)^{-1} \\ &= \sigma^2 \left(X'X\right)^{-1} X'\Omega X \left(X'X\right)^{-1} \end{aligned}$$

And

$$\begin{aligned} Var\left(\hat{\beta}_{GLS}\middle| X\right) &= Var\left(\left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y\middle| X\right) \\ &= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Var\left(Y\middle| X\right) \left(\Omega^{-1}\right)' X \left(X'\Omega^{-1}X\right)^{-1} \\ &= \sigma^2 \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}\Omega \left(\Omega'\right)^{-1} X \left(X'\Omega^{-1}X\right)^{-1} \\ &= \sigma^2 \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}X \left(X'\Omega^{-1}X\right)^{-1} \\ &= \sigma^2 \left(X'\Omega^{-1}X\right)^{-1} \end{aligned}$$

Where $\Omega = \Omega'$ since Ω is a variance-covariance matrix and is hence symmetric.

Plugging in the values, we have:

$$Var\left(\hat{\beta}_{OLS}\middle| X\right) = \sigma^2 \left(X'X\right)^{-1} X'\Omega X \left(X'X\right)^{-1}$$

$$= \quad \sigma^2 \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 1 & 5 & 8 & 2 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 5 \\ 1 & 8 \\ 1 & 2 \end{bmatrix} \right)^{-1} \cdot$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 1 & 5 & 8 & 2 \end{bmatrix} \begin{bmatrix} 16 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 25 & 0 & 0 \\ 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 5 \\ 1 & 8 \\ 1 & 2 \end{bmatrix} \cdot$$

$$\left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 1 & 5 & 8 & 2 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 5 \\ 1 & 8 \\ 1 & 2 \end{bmatrix} \right)^{-1}$$

$$= \sigma^2 \begin{bmatrix} \frac{11}{15} & -\frac{2}{15} \\ -\frac{2}{15} & \frac{1}{30} \end{bmatrix} \begin{bmatrix} 110 & 710 \\ 710 & 4994 \end{bmatrix} \begin{bmatrix} \frac{11}{15} & -\frac{2}{15} \\ -\frac{2}{15} & \frac{1}{30} \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} \frac{682}{75} & -\frac{223}{75} \\ -\frac{223}{75} & \frac{179}{150} \end{bmatrix} \approx \sigma^2 \begin{bmatrix} 9.0933 & -2.9733 \\ -2.9733 & 1.1933 \end{bmatrix}$$

And,

$$Var\left(\hat{\beta}_{GLS}\middle| X\right) = \sigma^2 \left(X'\Omega^{-1}X\right)^{-1}$$

$$= \sigma^2 \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 1 & 5 & 8 & 2 \end{bmatrix} \begin{bmatrix} 16 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 25 & 0 & 0 \\ 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 5 \\ 1 & 8 \\ 1 & 2 \end{bmatrix} \right)^{-1}$$

$$= \sigma^2 \begin{bmatrix} \frac{1000}{507} & -\frac{415}{507} \\ -\frac{415}{507} & \frac{2189}{4056} \end{bmatrix} = \sigma^2 \begin{bmatrix} 1.9724 & -0.8185 \\ -0.8185 & 0.5397 \end{bmatrix}$$

Finally, what is σ^2? Using the two respective estimates, we have:

$$\begin{aligned} \hat{\sigma}^2_{OLS} &= \frac{1}{5-2}\left(Y - X\hat{\beta}_{OLS}\right)'\left(Y - X\hat{\beta}_{OLS}\right) \\ &= \frac{1}{3}\left(\begin{bmatrix} 6 \\ 3 \\ 12 \\ 15 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 5 \\ 1 & 8 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \frac{2}{3} \\ \frac{11}{6} \end{bmatrix} \right)' \left(\begin{bmatrix} 6 \\ 3 \\ 12 \\ 15 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 5 \\ 1 & 8 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \frac{2}{3} \\ \frac{11}{6} \end{bmatrix} \right) \\ &= \frac{55}{18} \end{aligned}$$

And,

$$\hat{\sigma}^2_{GLS} = \frac{1}{5-2}\left(\Omega^{-\frac{1}{2}}Y - \Omega^{-\frac{1}{2}}X\hat{\beta}_{GLS}\right)'\left(\Omega^{-\frac{1}{2}}Y - \Omega^{-\frac{1}{2}}X\hat{\beta}_{GLS}\right)$$

$$= \frac{1}{3}\left(\begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{5} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{8} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 6 \\ 3 \\ 12 \\ 15 \\ 4 \end{bmatrix} - \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{5} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{8} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 5 \\ 1 & 8 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1.2184 \\ 1.6493 \end{bmatrix}\right)' \cdot$$

$$\left(\begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{5} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{8} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 6 \\ 3 \\ 12 \\ 15 \\ 4 \end{bmatrix} - \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{5} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{8} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 5 \\ 1 & 8 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1.2184 \\ 1.6493 \end{bmatrix}\right)$$

$$= \frac{1}{3}\begin{bmatrix} -0.4539 \\ 0.1323 \\ 0.50702 \\ 0.0734 \\ -0.2585 \end{bmatrix}'\begin{bmatrix} -0.4539 \\ 0.1323 \\ 0.50702 \\ 0.0734 \\ -0.2585 \end{bmatrix} = 0.1843$$

Using OLS, we have, then:

$$\hat{Y}_i = \underset{(5.2711)}{0.6667} + \underset{(1.9095)}{1.8333}X_i$$

And using GLS, we have:

$$\hat{Y}_i = \underset{(0.6029)}{1.2184} + \underset{(0.3153)}{1.1649}X_i$$

The GLS estimator produces coefficients that are statistically significant. Clearly, in the presence of heteroskedasticity, the standard errors of the GLS estimator will be significantly less than those of the OLS estimator.

7. Determine whether the following statement is true or false: Suppose that the CLR model applies to

$$E(Y|X) = X\beta$$

that T is a nonstochastic nonsingular matrix and that $Y^* = TY$ and $X^* = TX$; then the GLS regression of Y^* on X^* gives the same coefficient estimates as OLS of Y on X.

Solution Intuitively, this statement appears to be true. Pre-multiplying everything by a nonsingular matrix T would have the effect of re-scaling both the dependent and independent variables. This could potentially be problematic, since a linear transformation of a random variable would also transform the variance.

However, this is exactly what the GLS estimator is designed to deal with. The GLS estimator acts by pre-multiplying the entire model by a matrix that corrects for a change in the variance of the dependent variable. It will be shown that this will exactly cancel out the effect of pre-multiplying by T.

First, note that, since the CLR assumptions apply to this model, we have

$$Var(TY|X) = TVar(Y|X)T' = \sigma^2 TT' \equiv \Omega$$

Therefore, our GLS estimator (under the T transformation) is

$$\begin{aligned} \hat{\beta}^*_{GLS} &= \left((TX)'\Omega^{-1}TX\right)^{-1}(TX)'\Omega^{-1}TY \\ &= \left(X'T'\left(\sigma^2TT'\right)^{-1}TX\right)^{-1}X'T'\left(\sigma^2TT'\right)^{-1}TY \\ &= \left(\frac{1}{\sigma^2}X'T'(T')^{-1}T^{-1}TX\right)^{-1}\frac{1}{\sigma^2}X'T'(T')^{-1}T^{-1}TY \\ &= \frac{\sigma^2}{\sigma^2}(X'X)^{-1}X'Y = (X'X)^{-1}X'Y = \hat{\beta}_{OLS} \end{aligned}$$

That is, the GLS estimator of the $Y^* = X^*\beta^* + \varepsilon^*$ is the same as the OLS estimator of the $Y = X\beta + \varepsilon$.