Econ 203B Econometrics

Raphael Lam

1 Problem Set 3 Solutions[1]

Problem 1:

The regression slope $\hat{\beta}$ in a CNLR model is distributed $N\left(\beta, \sigma^2_{\hat{\beta}}\right)$ where $\sigma^2_{\hat{\beta}} = 1$. The null hypothesis $\beta = 0$ will be tested at the 10% significance level by using the statistic $Z_0 = \hat{\beta}/\sigma_{\hat{\beta}}$. 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.

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

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

Since the current model is CNLR and the variance is know($\sigma^2_{\hat{\beta}} =$ 1 or 4 in parts a) and b) repsectively), we can write down the Z-statistics

$$z = \frac{\Gamma\widehat{\beta} - \gamma_0}{(\sigma^2 \Gamma (X'X)^{-1} \Gamma')^{1/2}}$$

It follows a standard normal distribution under the null hypothesis.

$H_0 : \beta = 0 (= \gamma_0)$ with $\alpha = 0.1$

$H_a : \beta \neq 0$

The power function is

[1] I have included in part the solutions of D. Miyakawa on questions 1 and 2. Some common mistakes in this homework are that some of you are confused about convergence in distribution and asymptotic variance, especially under the ML case with the information matrix.

$$
\begin{aligned}
\Pr(|z_0| \geq c^* \mid \Gamma\beta = \gamma) &= \Pr\left(\left|\frac{\Gamma\widehat{\beta} - \gamma_0}{(\sigma^2\Gamma(X'X)^{-1}\Gamma')^{1/2}}\right| \geq c^* \mid \Gamma\beta = \gamma\right) \\
&= \Pr\left(\frac{\Gamma\widehat{\beta} - \gamma_0}{(\Omega)^{1/2}} \geq c^* \mid \Gamma\beta = \gamma\right) + \Pr\left(\frac{\Gamma\widehat{\beta} - \gamma_0}{(\Omega)^{1/2}} \leq -c^* \mid \Gamma\beta = \gamma\right) \\
with \;\; \Omega &\equiv \sigma^2\Gamma(X'X)^{-1}\Gamma' \\
&= \Pr\left(\frac{\Gamma\widehat{\beta} - \gamma}{(\Omega)^{1/2}} + \frac{\gamma - \gamma_0}{(\Omega)^{1/2}} \geq c^* \mid \Gamma\beta = \gamma\right) + \Pr\left(\frac{\Gamma\widehat{\beta} - \gamma}{(\Omega)^{1/2}} + \frac{\gamma - \gamma_0}{(\Omega)^{1/2}} \leq -c^* \mid \Gamma\beta = \gamma\right) \\
&= \Pr\left(\frac{\Gamma\widehat{\beta} - \gamma}{(\Omega)^{1/2}} + \frac{\gamma - \gamma_0}{(\Omega)^{1/2}} \geq c^* \mid \Gamma\beta = \gamma\right) + \Pr\left(\frac{\Gamma\widehat{\beta} - \gamma}{(\Omega)^{1/2}} + \frac{\gamma - \gamma_0}{(\Omega)^{1/2}} \leq -c^* \mid \Gamma\beta = \gamma\right) \\
&= \Pr\left(z \geq c^* - \frac{\gamma - \gamma_0}{(\Omega)^{1/2}} \mid \Gamma\beta = \gamma\right) + \Pr\left(z \leq -c^* - \frac{\gamma - \gamma_0}{(\Omega)^{1/2}} \mid \Gamma\beta = \gamma\right) \\
with \;\; z &= \frac{\Gamma\widehat{\beta} - \gamma}{(\Omega)^{1/2}} \\
&= \left[1 - \Phi(c^* - \frac{\gamma - \gamma_0}{(\Omega)^{1/2}})\right] + \Phi(-c^* - \frac{\gamma - \gamma_0}{(\Omega)^{1/2}}) \quad \text{with } \Phi \text{ is the CDF of std normal} \\
&= \left[1 - \Phi(1.645 - \frac{\gamma - 0}{(1)^{1/2}})\right] + \Phi(-1.645 - \frac{\gamma - 0}{(1)^{1/2}})
\end{aligned}
$$

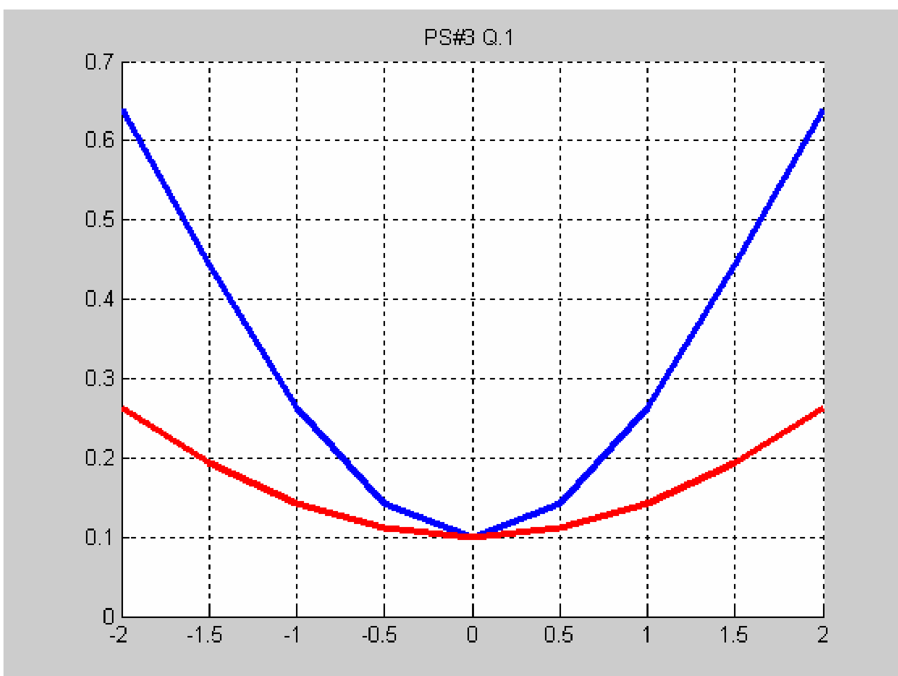
```
clear all;
clc
%Data
Betanull=0;
Sigmahat1=1;
Sigmahat2=4;
Truebeta=[-2 -1.5 -1 -0.5 0 0.5 1 1.5 2];

%Calculate the power for each true beta
Power1=ones(1,9)-normcdf(1.645*ones(1,9)-Truebeta./(Sigmahat1.^(1/2)))...
 +normcdf(-1.645*ones(1,9)-Truebeta./(Sigmahat1.^(1/2)));
Power2=ones(1,9)-normcdf(1.645*ones(1,9)-Truebeta./(Sigmahat2.^(1/2)))...
 +normcdf(-1.645*ones(1,9)-Truebeta./(Sigmahat2.^(1/2)));
Power=[Power1; Power2];

%Display output
TruebetaPower12=[Truebeta' Power1' Power2']
%Plot the two data
figure(1)
```

```
hold on
plot(Truebeta,Power1,'b','LineWidth',3)
plot(Truebeta,Power2,'r','LineWidth',3)
title('PS#3 Q.1')
grid on
```

Output Result:

True β	Power1 ($\sigma^2_{\hat{\beta}} = 1$)	Power2 ($\sigma^2_{\hat{\beta}} = 1$)
-2.0000	0.6388	0.2635
-1.5000	0.4432	0.1937
-1.0000	0.2635	0.1421
-0.5000	0.1421	0.1106
0	0.1000	0.1000
0.5000	0.1421	0.1106
1.0000	0.2635	0.1421
1.5000	0.4432	0.1937
2.0000	0.6388	0.2635

The blue and red curves corresponds to Power1 and 2 respectively



(c) It can be seen that, given the true paratmer, power decreases with increasing variance $\sigma^2_{\hat{\beta}}$ (implying the estimated parameter will be estimated with less precision). At the true parameter $\beta_0 = 0$, when the null $\beta = 0$, the power is equal to 10% assigned initially. Otherwise, when the sampling distribution of the estimator is wide (less precise), the observations from the data can be consistent with many alternative hypotheses and therefore the power of test decreases over all the range of the power function, except one point.

Problem 2:

The regression slopes $\hat{\beta}_1$ and $\hat{\beta}_2$ in a CNLR model are distributed as 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**

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

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) : $(-1,1)$, $(-1,0)$, $(-1,-1)$, $(0,1)$, $(0,0)$, $(0,-1)$, $(1,1)$, $(1,0)$, $(1,-1)$.

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

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

We first construct the Wald statistic $\quad W_0 = \left[\Gamma\widehat{\beta} - \gamma_0\right]' \left[\Gamma \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \Gamma'\right]^{-1} \left[\Gamma\widehat{\beta} - \gamma_0\right]$

The hypothesis is

$H_0 : \Gamma\beta = \gamma_0$
$H_a : \Gamma\beta \neq \gamma_0$ $\quad$ where $\Gamma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \gamma_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

Given the regression slopes, $\quad \widehat{\beta} \sim N\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \right)$

$$\Gamma\widehat{\beta} - \gamma_0 \sim N\left((\Gamma\beta) - \gamma_0, \Gamma \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \Gamma' \right)$$

with $\Gamma \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \Gamma' \equiv \Sigma$

Generally, we use the χ^2 distribution on the Wald-statistics. Note that the above is not a sum of the squares of the standard normal distribution with mean equal to zero. Here the normally distributed quantities have a mean other than zero (given the true parameters). the sum of squares yields a noncentral χ^2 distribution that contains two parameters including the degrees of freedom and the noncentrality parameter. The latter is the sum of the squared means of the normally distributed quantities. In this particular example, the mean of the estimated parameters is $(\Gamma\beta) - \gamma_0$, while it is valid under the null only. Given the true parameter vector, the mean should be $(\Gamma\beta_{TRUE}) - \gamma_0$. The non-centrality parameter is in turn equal to $[(\Gamma\beta) - \gamma_0]' \begin{pmatrix} 1\, r \\ r\, 1 \end{pmatrix}^{-1} [(\Gamma\beta) - \gamma_0]$. The basic idea is that we have to correct for the 'non-central' to turn it into a standard normal distribution and apply the χ^2 distribution properties. We know that a non-central χ^2 distribution with p degree of freedom and non-centrality parameter $\triangle$, denoted as $\chi^2(p, \triangle)$ is defined as "taking sum of squares of a multivariate normal with non-zero mean and unit variance. In the current case,

$$
\begin{aligned}
W_0 &= (\Gamma\widehat{\beta} - \gamma_0)'\Sigma^{-1}(\Gamma\widehat{\beta} - \gamma_0) \sim \chi^2(p, \triangle) \\
&= \frac{1}{1-r^2}(\beta_1^2 + \beta_2^2 - 2r\beta_1 + \beta_2) \sim \chi^2(p, \triangle) \\
where\ \triangle &= [(\Gamma\beta) - \gamma_0]' \begin{pmatrix} 1\, r \\ r\, 1 \end{pmatrix}^{-1} [(\Gamma\beta) - \gamma_0]
\end{aligned}
$$

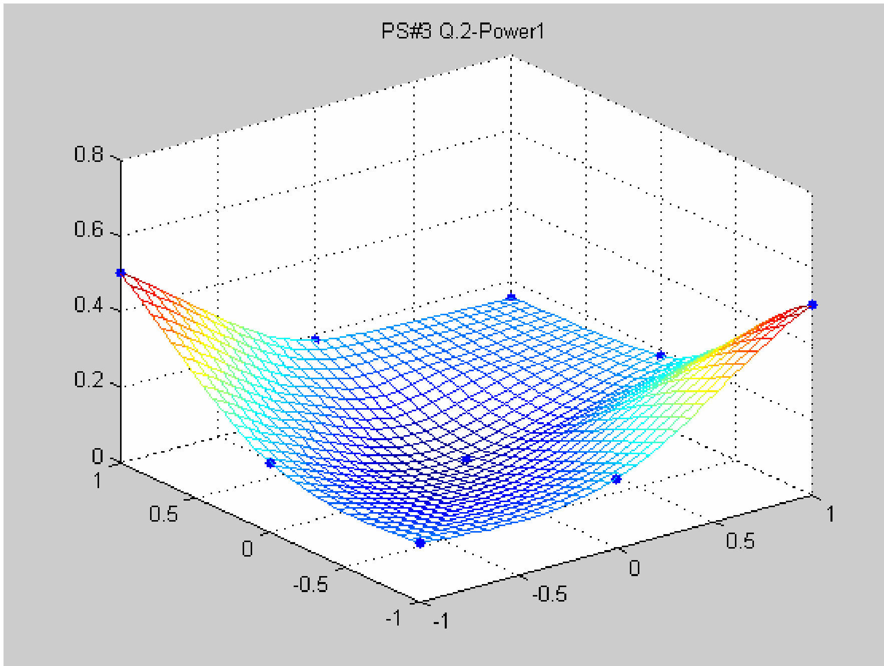
Matlab Code:

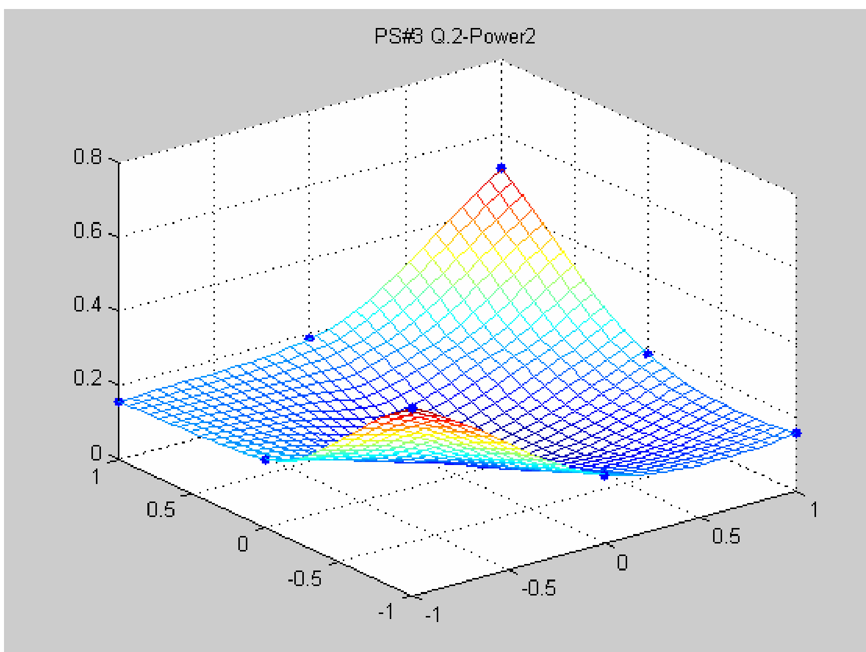
```
clear all
clc
%Data
Gammanull=[0 0];
Betanull=[0 0];
R1=[1 0.6;0.6 1];
R2=[1 -0.6;-0.6 1];
G=[1 1];
Truebeta=[-1 1;-1 0;-1 -1;0 1;0 0;0 -1;1 1;1 0;1 -1];

%Calculate Wnull-statistic
for i=1:9
 Wnull1(i,1)=(Truebeta(i,:)*inv(R1)*Truebeta(i,:)');
 Wnull2(i,1)=(Truebeta(i,:)*inv(R2)*Truebeta(i,:)');
end
```

```
%Calculate Power
Power1=(1-ncx2cdf(5.99,2,Wnull1))';
Power2=(1-ncx2cdf(5.99,2,Wnull2))';

%Display output
TrueBetaPower12=[Truebeta Power1' Power2']
%Plot the first power
figure(1)
Xlin=linspace(min(Truebeta(:,1)),max(Truebeta(:,1)),30);
Ylin=linspace(min(Truebeta(:,2)),max(Truebeta(:,2)),30);
[X,Y]=meshgrid(Xlin,Ylin);
Z1=griddata(Truebeta(:,1),Truebeta(:,2),Power1,X,Y,'cubic');
mesh(X,Y,Z1);
hold on
plot3(Truebeta(:,1),Truebeta(:,2),Power1,'.','Markersize',15)
title('PS#3 Q.2-Power1')
grid on

%Plot the second power
figure(2)
Xlin=linspace(min(Truebeta(:,1)),max(Truebeta(:,1)),30);
Ylin=linspace(min(Truebeta(:,2)),max(Truebeta(:,2)),30);
[X,Y]=meshgrid(Xlin,Ylin);
Z2=griddata(Truebeta(:,1),Truebeta(:,2),Power2,X,Y,'cubic');
mesh(X,Y,Z2);
hold on
plot3(Truebeta(:,1),Truebeta(:,2),Power2,'.','Markersize',15)
title('PS#3 Q.2-Power2')
grid on
```

And the result is

True beta1	True beta2	Power1 ($r = 0.6$)	Power2 ($r = -0.6$)
-1.0000	1.0000	0.5038	0.1553
-1.0000	0	0.1843	0.1843
-1.0000	-1.0000	0.1553	0.5038
0	1.0000	0.1843	0.1843
0	0	0.0500	0.0500
0	-1.0000	0.1843	0.1843
1.0000	1.0000	0.1553	0.5038
1.0000	0	0.1843	0.1843
1.0000	-1.0000	0.5038	0.1553





(c) When the sign of two betas are different, the positive (negative) correlated case provides high (low) power. When the sign of two betas are same, the positive (negative) correlated case provides low (high) power. When one of the betas is 0, the pattern of correlation does not affect the power. And, when truebeta 1 & 2 are 0, power is 0.05 (=significant level). At the true values of parameter vector, the power gives the significant level 0.05 as constructed, but strictly speaking the power function at that particular point is not defined.

Problem 3:

Remark 1 *This question is the same as one of the exam questions in the past.*

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

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

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, $\hat{\beta}_{MLE}$.

Given the observations are independent $\{(Y_i, X_i)\}_{i=1}^n$. We can write down the sample conditional likelihood function as following:

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

Note: We do not know the joint likelihood as we are not given the marginal distribution of x_i. The marginal distribution of x_i is not affected by β.

The conditional log-likelihood function of the sample

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

Find the MLE estimator, we derive the first order conditions with respect to β:

$$\frac{\partial \mathcal{L}\left(y_i|x_i;\beta\right)}{\partial \beta}|_{\beta=\hat{\beta}_{MLE}} = 0$$
$$\sum_{i=1}^{n}\left(\frac{y_i}{\hat{\beta}_{MLE}} - x_i\right) = 0$$
$$\hat{\beta}_{MLE}\Sigma_{i=1}^{n}x_i = \Sigma_{i=1}^{n}y_i$$
$$\hat{\beta}_{MLE}\Sigma_{i=1}^{n}x_i = \Sigma_{i=1}^{n}y_i$$
$$\hat{\beta}_{MLE} = \frac{\Sigma_{i=1}^{n}y_i}{\Sigma_{i=1}^{n}x_i} = \frac{\bar{y}}{\bar{x}}$$

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

Remark 2 *We know from lecture note that $\hat{\beta}_{MLE}$ is consistent and asymptotic normal under some general conditions (I believe that you only need to know those conditions, but not the proof of consistency of MLE estimator. I will soon post a note for your reference (probably not for exam, but i cannot guarantee) for a proof. Once we take for granted on the consistency of MLE estimator, the asymptotic distribution is relatively easy.*

Remark 3 *The distribution of y_i conditional on x_i has a Poisson distribution with parameter $x_i\beta$. Therefore, the conditional mean $E(y_i|x_i) = x_i\beta$*

Therefore, the conditional expected value of $\hat{\beta}_{MLE}$ is

$$E(\hat{\beta}_{MLE}|X) = E\left[\frac{\Sigma_{i=1}^{n}y_i}{\Sigma_{i=1}^{n}x_i}|X\right] = \frac{\Sigma_{i=1}^{n}x_i\beta}{\Sigma_{i=1}^{n}x_i} = \beta$$

So, $\hat{\beta}_{MLE}$ is unbiased and the variance of the estimator attains the the Cramer-Rao lower bound asymptotically.

Provided the MLE estimator is consistent. The asymptotic distribution of the MLE estimator $\hat{\beta}_{MLE}$

$$\sqrt{n}\left(\hat{\beta}_{MLE} - \beta\right) \xrightarrow{d} N\left(0, \left(\lim_{n\to\infty}\frac{1}{n}I\left(\beta\right)\right)^{-1}\right)$$
$$\hat{\beta}_{MLE} \overset{A}{\sim} N\left(\beta, \hat{I}\left(\beta\right)^{-1}\right)$$

The variance is computed as the inverse of the information matrix.

$$\begin{aligned} I(\beta|X) &= -E\left[\frac{\partial^2\mathcal{L}(\beta)}{\partial\beta^2}|X\right] = -E\left[-\sum_{i=1}^{n}\frac{y_i}{\beta^2}|X\right] \\ &= \frac{1}{\beta^2}E\left[\sum_{i=1}^{n}y_i|X\right] = \frac{1}{\beta^2}\left(\sum_{i=1}^{n}E\left[y_i|X\right]\right) \\ &= \frac{1}{\beta^2}\left(\sum_{i=1}^{n}E\left[y_i|X\right]\right) = \sum_{i=1}^{n}\frac{x_i}{\beta} = \frac{n\bar{x}}{\beta} \end{aligned} \tag{1}$$

using the condition given above that observation y_i follows a Poisson distribution with a conditional mean βx_i. Substituting into the conditional asymptotic distribution

$$\sqrt{n}\left(\hat{\beta}_{MLE}-\beta\right) \xrightarrow{d} N\left(0,\left(\lim_{n\to\infty}\frac{1}{n\beta}\sum_{i=1}^{n}x_i\right)^{-1}\right)$$

$$\hat{\beta}_{MLE} \overset{A}{\sim} N\left(\beta,\frac{\beta}{\sum_{i=1}^{n}x_i}\right)$$

So, β is consistent and has an asymptotics variance equal to $\beta/\sum_{i=1}^{n}x_i$. Since the true parameter β is unknown, we may estimate the variance term by using the $\hat{\beta}_{MLE} = \frac{\Sigma_{i=1}^{n}y_i}{\Sigma_{i=1}^{n}x_i}$ such that the variance term $\widehat{AV(\hat{\beta}_{MLE})} = \frac{\Sigma_{i=1}^{n}y_i}{\left(\Sigma_{i=1}^{n}x_i\right)^2}$.

(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 $\sum_j\lambda_j$.)

Given the hint, we define $\lambda = \sum y_i$. Find the expectation of λ,i.e.

$$E(\lambda|X) = E(\sum_{i=1}^{n}y_i) = \sum_{i=1}^{n}\beta x_i \tag{2}$$

To find the exact distribution of $\hat{\beta}_{MLE}$:

$$\begin{aligned} \Pr\left(\hat{\beta}_{MLE}=b\right) &= \Pr\left(\frac{\Sigma_{i=1}^{n}y_i}{\Sigma_{i=1}^{n}x_i}=b|X_i=x_i\right) \\ &= \frac{e^{-\left(\beta\Sigma_{i=1}^{n}x_i\right)}\left(\beta\Sigma_{i=1}^{n}x_i\right)^{b\Sigma_{i=1}^{n}x_i}}{(b\Sigma_{i=1}^{n}x_i)!} \end{aligned}$$

where b is a discrete set of numbers such that $b\Sigma_{i=1}^{n}x_i = 0, 1, 2, ...$

Problem 4:
In the standard CNLR model,

$$y_i = X_i\beta + \varepsilon_i$$

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

Under CNLR model, $y_i = X_i\beta + \varepsilon_i$, we assume for simplicity i.i.d. observations on $\{y_i, x_i\}$

$$\begin{aligned} L\left(\beta, \sigma^2; Y|X\right) &= L(\beta) \qquad \text{with } \beta^2 = \sigma^2 \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(y_i - x_i\beta)^2}{2\beta^2}\right) \\ &= (2\pi)^{-n/2}\left(b^2\right)^{-n/2} \exp\left(-\frac{1}{2b^2}(y - Xb)'(y - Xb)\right) \end{aligned}$$

The log-likelihood function is then

$$\begin{aligned} \mathcal{L}(\beta) &= \ln(L(\beta) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\beta^2\right) - \frac{1}{2\beta^2}(y - X\beta)'(y - X\beta) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\beta^2\right) - \frac{y'y}{2\beta^2} + \frac{y'X}{\beta} - \frac{1}{2}X'X \end{aligned}$$

Taking first order conditions we have that

$$\begin{aligned} \frac{\partial\mathcal{L}(\beta|X)}{\partial\beta}|_{\beta=\hat{\beta}_{MLE}} &= 0 \\ -\frac{n}{\hat{\beta}_{MLE}} - \frac{y'X}{\hat{\beta}_{MLE}^2} + \frac{y'y}{\hat{\beta}_{MLE}^3} &= 0 \\ -n\hat{\beta}_{MLE}^2 - y'X\hat{\beta}_{MLE} + y'y &= 0 \end{aligned}$$

$$\hat{\beta}_{MLE} = \frac{-y'X \pm \sqrt{(y'X)^2 + 4n(y'y)}}{2n} \qquad (3)$$

We check the second order conditions for a maximum. The maximum depends on the sign of $y'X$. In particular, if $y'X > 0$, $\frac{-y'X+\sqrt{(y'X)^2+4n(y'y)}}{2n}$

is a maximum and $\frac{-y'X-\sqrt{(y'X)^2+4n(y'y)}}{2n}$ is a minimum. Also , if $y'X < 0$, $\frac{-y'X+\sqrt{(y'X)^2+4n(y'y)}}{2n}$ is a minimum and $\frac{-y'X-\sqrt{(y'X)^2+4n(y'y)}}{2n}$ is a maximum. The quadratic nature of $\sigma^2 = \beta^2$ makes it satisfies the FOC at two points, but only one of those is a maximum.

Second order condition requires at the maximum:

$$\frac{\partial^2 \mathcal{L}(\beta|X)}{\partial \beta^2}|_{\beta=\hat{\beta}_{MLE}} = \frac{n}{\hat{\beta}^2} - 3\frac{y'y}{\hat{\beta}^4} + 2\frac{y'X}{\hat{\beta}^3}$$

We may assume $y'X > 0$ such that the aboe SOC holds only for a particular root

$$\hat{\beta}_{MLE} = \frac{-y'X + \sqrt{(y'X)^2 + 4ny'y}}{2n} \tag{4}$$

The information matrix gives the CramerRao Lower Bound with the true parameter β

$$\begin{aligned}
I(\beta|X) &= -E\left[\frac{\partial^2 \mathcal{L}(b|X)}{\partial b^2}|_{b=\beta}|X\right] = -E\left[\frac{n}{\beta^2} + \frac{2y'X}{\beta^3} - \frac{3y'y}{\beta^4}|X\right] \\
&= -E\left[\frac{n}{\beta^2}|X\right] - E\left[\frac{2y'X}{\beta^3}|X\right] + E\left[\frac{3y'y}{\beta^4}|X\right] \\
&= -\frac{n}{\beta^2} - \frac{2}{\beta^3}E\left[y'X|X\right] + \frac{3}{\beta^4}E\left[y'y|X\right] \\
&= -\frac{n}{\beta^2} - \frac{2}{\beta^3}E\left[(X\beta+\varepsilon)' X|X\right] + \frac{3}{\beta^4}E\left[(X\beta+\varepsilon)'(X\beta+\varepsilon)|X\right] \\
&= -\frac{n}{\beta^2} - \frac{2}{\beta^3}E\left[X'X\beta + X'\varepsilon|X\right] + \frac{3}{\beta^4}E\left[\beta^2X'X + 2\beta X'\varepsilon + \varepsilon'\varepsilon|X\right] \\
&= -\frac{n}{\beta^2} - \frac{2(X'X)\beta}{\beta^3} + \frac{3\left(\beta^2X'X + n\beta^2\right)}{\beta^4} \\
&= \frac{X'X + 2n}{\beta^2}
\end{aligned} \tag{5}$$

Conditional on X, the Cramer-Rao lower bound is then

$$I(\beta|X)^{-1} = \frac{\beta^2}{X'X + 2n}$$

Problem 5:

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

$$Y_i = X_i\beta + \varepsilon_i$$
$$\varepsilon_i|X_i \sim N\left(0, (Z_i\gamma)^2\right)$$

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

The conditional likelihood and log-likelihood are:

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

The first order conditions are the following:

$$\frac{\partial \ln L(\beta, \gamma; Y_i, X_i)}{\partial \beta}|_{\beta=\hat{\beta}_{MLE}} = \sum_{i=1}^{n} \frac{X_i'(Y_i - X_i\beta)}{(Z_i\gamma)^2} = 0$$
$$\frac{\partial \ln L(\beta, \gamma; Y_i, X_i)}{\partial \gamma}|_{\gamma=\hat{\gamma}_{MLE}} = -\sum_{i=1}^{n} \frac{Z_i'}{Z_i\gamma} + \sum_{i=1}^{n} \frac{Z_i'(Y_i - X_i\beta)^2}{(Z_i\gamma)^3} = 0$$
$$\frac{\partial^2 \ln L(\beta, \gamma; Y_i, X_i)}{\partial \beta^2} = -\sum_{i=1}^{n} \frac{X_i'X_i}{(Z_i\gamma)^2}$$
$$\frac{\partial^2 \ln L(\beta, \gamma; Y_i, X_i)}{\partial \gamma \partial \gamma'} = \sum_{i=1}^{n} \frac{Z_i'Z_i}{(Z_i\gamma)^2} - 3\sum_{i=1}^{n} \frac{Z_i'Z_i(Y_i - X_i\beta)^2}{(Z_i\gamma)^4}$$
$$\frac{\partial^2 \ln L(\beta, \gamma; Y_i, X_i)}{\partial \gamma \partial \beta} = -\sum_{i=1}^{n} \frac{2Z_i'X_i(Y_i - X_i\beta)}{(Z_i\gamma)^3}$$
$$\frac{\partial^2 \ln L(\beta, \gamma; Y_i, X_i)}{\partial \beta \partial \gamma} = -\sum_{i=1}^{n} \frac{2X_i'Z_i(Y_i - X_i\beta)}{(Z_i\gamma)^3}$$

The information matrix for the sample conditional on X

$$I(\beta, \gamma|X) = -E\left[\begin{array}{cc} -\sum_{i=1}^{n} \frac{X_i'X_i}{(Z_i\gamma)^2} & -\sum_{i=1}^{n} \frac{2X_i'Z_i(Y_i - X_i\beta)}{(Z_i\gamma)^3} \\ -\sum_{i=1}^{n} \frac{2Z_i'X_i(Y_i - X_i\beta)}{(Z_i\gamma)^3} & \sum_{i=1}^{n} \frac{Z_i'Z_i}{(Z_i\gamma)^2} - 3\sum_{i=1}^{n} \frac{Z_i'Z_i(Y_i - X_i\beta)^2}{(Z_i\gamma)^4} \end{array}\right| X\right]$$

$$I(\beta,\gamma|X)=E\left\{\begin{array}{cc} \sum_{i=1}^{n}\frac{X_i'X_i}{(Z_i\gamma)^2} & \sum_{i=1}^{n}\frac{2X_i'Z_i(Y_i-X_i\beta)}{(Z_i\gamma)^3} \\ \sum_{i=1}^{n}\frac{2Z_i'X_i(Y_i-X_i\beta)}{(Z_i\gamma)^3} & -\sum_{i=1}^{n}\frac{Z_i'Z_i}{(Z_i\gamma)^2}+3\sum_{i=1}^{n}\frac{Z_i'Z_i(Y_i-X_i\beta)^2}{(Z_i\gamma)^4} \end{array}\right| X\right\}$$

Using the law of interated expectation to find the information matrix

$$I(\beta,\gamma)=E\begin{bmatrix} \sum_{i=1}^{n}\frac{X_i'X_i}{(Z_i\gamma)^2}{}_{K\times K=1\times 1} & \mathbf{0}_{1\times r} \\ \mathbf{0_{r\times 1}} & 2\sum_{i=1}^{n}\frac{Z_i'Z_i}{(Z_i\gamma)^2}{}_{r\times r} \end{bmatrix}$$

given $E\left[(Y_i-X_i\beta)/X\right]=E\left[(Y_i-X_i\beta)/X_i\right]=E\left[\varepsilon_i/X_i\right]=0$
and $E\left[(Y_i-X_i\beta)^2/X\right]=E\left[(Y_i-X_i\beta)^2/X_i\right]=E\left[\varepsilon_i^2/X_i\right]=(Z_i\gamma)^2$

Problem 6:
Five sample observations are

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

Assume a linear model, $Y_i=\beta_1+\beta_2X_i+\varepsilon_i$, with heteroskedasticity of the form $Var\left(Y_i\right)\equiv Var\left(\varepsilon_i\right)\equiv\sigma_i^2=\sigma^2X_i^2$ where σ^2 is a positive constant. Calculate the OLS and GLS estimates of β_1 and β_2 and the corresponding standard errors.

Let $X=\begin{pmatrix}1&4\\1&1\\1&5\\1&8\\1&2\end{pmatrix}$, $y=\begin{pmatrix}6\\3\\12\\15\\4\end{pmatrix}$ *and* $Var\left(y\right)=Var\left(\varepsilon\right)=\Sigma=\sigma^2\Omega=$
$\sigma^2\begin{pmatrix}16&0&0&0&0\\0&1&0&0&0\\0&0&25&0&0\\0&0&0&64&0\\0&0&0&0&4\end{pmatrix}$
Then

$$\widehat{\beta}_{OLS}=\left(X'X\right)^{-1}\left(X'y\right)$$
$$\widehat{\beta}_{GLS}=\left(X'\Omega^{-1}X\right)^{-1}\left(X'\Omega^{-1}y\right)$$

and

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

The standard error are the square root of the elements in the diagonal of the corresponding variance-covariance matrices and they can be left expressed as a function of σ

You may put into matlab to find the values of corresponding matrices.

Problem 7:

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

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

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.

If the CLR model applies, the following assumptions hold:

1) Linearity: $Y = X\beta$

2) Strict Exogeneity: $E(\varepsilon_i|X) = 0; (i = 1, 2, ..., n)$

3) No multicollinearity: This means that the matrix X is a full rank one.

4) Spherical error variance:

Homoskedasticity: $E(\varepsilon_i^2|X) = \sigma^2 > 0; \quad (i = 1, 2, ..., n)$

No correlation between observations: $E(\varepsilon_i \varepsilon_j|X) = 0; (i = 1, ., n; i \neq j)$

We can write down $V(\varepsilon|X) = \sigma^2 I_n$ with $Cov(\varepsilon_i \varepsilon_j) = 0$ and $V(Y|X) = \sigma^2 I_n$

Consider a nonstochastic nonsingular matrix T such that $Y^* = TY$ and $X^* = TX$

Then $V(Y^*) = V(TY)$
$= TV(Y)T'$
$= \sigma^2 TT'$

In general, when $V(Y) = \sigma^2\Omega$, being Ω a known matrix, the GLS estimator of β would be given by $\widehat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$

In our case $\Omega = TT'$. Then, the GLS regression of Y^* on X^* , knowing $V(Y^*) = \sigma^2 TT'$, implies the following estimated coefficients:

$$\widehat{\beta}_{GLS} = (X^{*\prime}(TT')^{-1}X^*)^{-1}X^{*\prime}(TT')^{-1}Y^*$$

and because $Y^* = TY$ and $X^* = TX$

$$\widehat{\beta}_{GLS} = [(TX)'(TT')^{-1}(TX)]^{-1}(TX)'(TT')^{-1}(TY)$$

$$\widehat{\beta}_{GLS} = [X'T'(TT')^{-1}TX]^{-1}X'T'(TT')^{-1}TY$$

with $T'(TT')^{-1}T = I_n$. We may postmultiply by the identity given by $(TT')^{-1}TT' = I_n$.. Then $TT' = TT'(TT')^{-1}TT'$, which necessarily implies that $T'(TT')^{-1}T = I_n$)

$$\widehat{\beta}_{GLS} = (X'X)^{-1}X'Y$$

that is equal to the OLS regression of Y on X, given by $\widehat{\beta}_{OLS} = (X'X)^{-1}X'Y$.

$$\widehat{\beta}_{GLS,(Y^* on\ X^*)} = (X'X)^{-1}X'Y = \widehat{\beta}_{OLS,(Y\ on\ X)}$$

They are the same.