1. From the data set "`Midterm_q1_no.dta`":

Estimate the following models

$$y_{1t} = \beta_{10} + \gamma_{12} y_{2t} + \beta_{13} x_{3t} + u_{1t} \quad (1)$$

$$y_{2t} = \beta_{20} + \gamma_{21} y_{1t} + \beta_{21} x_{1t} + \beta_{22} x_{2t} + u_{2t} \quad (2)$$

a. Estimate model (1) and (2) using Ordinary Least Squares (OLS) and state consequences of using OLS in this case (5 Points)

Ans. The estimate will be biased, inconsistent, and inefficient.

```
. reg y1 y2 x3

      Source |       SS           df       MS      Number of obs   =        50
-------------+----------------------------------   F(2, 47)        =    106.18
       Model |  53133.7009         2  26566.8504   Prob > F        =    0.0000
    Residual |  11759.2791        47  250.197428   R-squared       =    0.8188
-------------+----------------------------------   Adj R-squared   =    0.8111
       Total |    64892.98        49  1324.34653   Root MSE        =    15.818

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          y2 |   .0009441   .0001926     4.90   0.000     .0005566    .0013315
          x3 |   .0074941   .0010981     6.82   0.000     .0052851    .0097031
       _cons |     42.099   16.83374     2.50   0.016     8.233888    75.96411
------------------------------------------------------------------------------

. reg y2 y1 x1 x2

      Source |       SS           df       MS      Number of obs   =        50
-------------+----------------------------------   F(3, 46)        =     30.49
       Model |  8.2348e+09         3  2.7449e+09   Prob > F        =    0.0000
    Residual |  4.1417e+09        46  90036236.9   R-squared       =    0.6654
-------------+----------------------------------   Adj R-squared   =    0.6435
       Total |  1.2376e+10        49   252580510   Root MSE        =    9488.7

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          y1 |   303.4508   46.84582     6.48   0.000     209.1551    397.7466
          x1 |   237.4617   134.2505     1.77   0.084    -32.77052    507.6939
          x2 |  -.0004671   .0003288    -1.42   0.162    -.0011289    .0001947
       _cons |  -36320.06   9029.637    -4.02   0.000    -54495.77   -18144.34
------------------------------------------------------------------------------
```

b. Estimate model (1) and (2) using Two Stage Least Squares (2SLS) and state reduced form and structural form of these simultaneous equation models. Specify endogenous variables and exogenous variables. Then, estimate reduced form models using OLS and structural form models using IV technique from the predicted endogenous variables from reduced form estimated results. (7 points)

Ans. Endogenous variable: y1 y2. Exogenous variable: x1 x2 x3

$$y_{1t} = \beta_{10} + \gamma_{12} y_{2t} + \beta_{13} x_{3t} + u_{1t}$$

Structural form: $$y_{2t} = \beta_{20} + \gamma_{21} y_{1t} + \beta_{21} x_{1t} + \beta_{22} x_{2t} + u_{2t}$$

β π

Reduced form: $y_{1t} = \pi_0 + \pi_1 * x_{1t} + \pi_2 * x_{2t} + \pi_3 * x_{3t} + w_{1t}$

$Y_{2t} = \pi_4 + \pi_5 * x_{1t} + \pi_6 * x_{2t} + \pi_7 * x_{3t} + w_{2t}$

```
. reg3 (y1 y2 x3) (y2 y1 x1 x2), 2sls nodfk inst(x1 x2 x3)
```

Two-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	F-Stat	P
y1	50	2	16.36851	0.8059	96.88	0.0000
y2	50	3	9492.366	0.6651	26.66	0.0000

	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
y1						
y2	.0012956	.0005021	2.58	0.011	.0002986	.0022925
x3	.0061423	.0020951	2.93	0.004	.0019819	.0103028
_cons	51.43261	20.89625	2.46	0.016	9.936801	92.92842
y2						
y1	294.67	57.5455	5.12	0.000	180.396	408.9439
x1	252.4501	142.6724	1.77	0.080	-30.86907	535.7692
x2	-.0004777	.0003185	-1.50	0.137	-.0011102	.0001547
_cons	-35214.46	9774.202	-3.60	0.001	-54624.09	-15804.83

Endogenous variables: y1 y2
Exogenous variables: x1 x2 x3

```
. reg y1 x1 x2 x3
```

Source	SS	df	MS
Model	48898.4924	3	16299.4975
Residual	15994.4876	46	347.706253
Total	64892.98	49	1324.34653

Number of obs	=	50
F(3, 46)	=	46.88
Prob > F	=	0.0000
R-squared	=	0.7535
Adj R-squared	=	0.7375
Root MSE	=	18.647

y1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x1	.5660038	.2509797	2.26	0.029	.060808	1.0712
x2	-5.53e-07	6.41e-07	-0.86	0.393	-1.84e-06	7.37e-07
x3	.0097182	.0011453	8.48	0.000	.0074128	.0120237
_cons	9.11207	19.23536	0.47	0.638	-29.60671	47.83085

```
8 . predict y1hat
   (option xb assumed; fitted values)

9 . reg y2 x1 x2 x3

      Source |       SS           df       MS      Number of obs   =        50
-------------+----------------------------------   F(3, 46)        =     17.69
       Model |  6.6305e+09         3  2.2102e+09   Prob > F        =    0.0000
    Residual |  5.7460e+09        46   124912002   R-squared       =    0.5357
-------------+----------------------------------   Adj R-squared   =    0.5055
       Total |  1.2376e+10        49   252580510   Root MSE        =     11176

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   419.2344     150.43     2.79   0.008     116.4346    722.0342
          x2 |  -.0006408   .0003843    -1.67   0.102    -.0014143    .0001327
          x3 |   2.863673   .6864876     4.17   0.000     1.481846    4.245501
       _cons |  -32529.41   11529.12    -2.82   0.007    -55736.33   -9322.491
------------------------------------------------------------------------------

10 . predict y2hat
   (option xb assumed; fitted values)

11 . reg y1 y2hat x3

      Source |       SS           df       MS      Number of obs   =        50
-------------+----------------------------------   F(2, 47)        =     71.26
       Model |  48799.1291         2  24399.5646   Prob > F        =    0.0000
    Residual |  16093.8509        47  342.422359   R-squared       =    0.7520
-------------+----------------------------------   Adj R-squared   =    0.7414
       Total |    64892.98        49  1324.34653   Root MSE        =    18.505

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       y2hat |   .0012956   .0005854     2.21   0.032     .0001179    .0024732
          x3 |   .0061423   .0024429     2.51   0.015     .0012278    .0110569
       _cons |   51.43261   24.36555     2.11   0.040     2.415447    100.4498
------------------------------------------------------------------------------

12 . reg y2 y1hat x1 x2

      Source |       SS           df       MS      Number of obs   =        50
-------------+----------------------------------   F(3, 46)        =     17.69
       Model |  6.6305e+09         3  2.2102e+09   Prob > F        =    0.0000
    Residual |  5.7460e+09        46   124912001   R-squared       =    0.5357
-------------+----------------------------------   Adj R-squared   =    0.5055
       Total |  1.2376e+10        49   252580510   Root MSE        =     11176

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       y1hat |     294.67   70.63909     4.17   0.000     152.4809    436.8591
          x1 |   252.4501   175.1353     1.44   0.156     -100.079    604.9792
          x2 |  -.0004777    .000391    -1.22   0.228    -.0012647    .0003092
       _cons |  -35214.46   11998.17    -2.93   0.005    -59365.53    -11063.4
------------------------------------------------------------------------------
```

```
13 . reg (y1 y2hat x3) (y2 y1hat x1 x2), ols
  option ols not allowed
  r(198);

14 . reg3 (y1 y2hat x3) (y2 y1hat x1 x2), ols

  Multivariate regression
  ----------------------------------------------------------------------
  Equation             Obs   Parms        RMSE    "R-sq"       F-Stat        P
  ----------------------------------------------------------------------
  y1                    50       2    18.50466    0.7520        71.26   0.0000
  y2                    50       3     11176.4    0.5357        17.69   0.0000
  ----------------------------------------------------------------------

  ------------------------------------------------------------------------------
               |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
  y1           |
         y2hat |   .0012956   .0005854     2.21   0.029     .0001331    .0024581
            x3 |   .0061423   .0024429     2.51   0.014     .0012912    .0109935
         _cons |   51.43261   24.36555     2.11   0.037     3.047457    99.81776
  -------------+----------------------------------------------------------------
  y2           |
         y1hat |     294.67   70.63909     4.17   0.000     154.3947    434.9452
            x1 |   252.4501   175.1353     1.44   0.153    -95.33393    600.2341
            x2 |  -.0004777    .000391    -1.22   0.225    -.0012541    .0002986
         _cons |  -35214.46   11998.17    -2.93   0.004    -59040.45   -11388.48
  ------------------------------------------------------------------------------
```

c). 2sls estimation uses ols method to estimate the reduced form and predict endogenous variable, then use the predicted value, estimating using IV technique.

3sls then use 2sls and fgls to estimate.

The advantage of using 2sls instead of ols when we have simultaneous bias is that it will lead ro a consistent and asymptotically efficient, but still bias.

2sls and ols are single equation estimation methods, unlike 3sls as it is a system equation estimation method. When we think there is a correlation between error terms across equation, system equation method should be used as it can lead to more asymptotic efficiency. However, if there is specification error in one equation, the problem will spread through all the system. Then, it will be inconsistent.

```
15 . reg3 (y1 y2 x3) (y2 y1 x1 x2), 3sls inst( x1 x2 x3 )

Three-stage least-squares regression
----------------------------------------------------------------------
Equation             Obs   Parms        RMSE    "R-sq"       chi2        P
----------------------------------------------------------------------
y1                    50       2    15.86986    0.8059     193.76   0.0000
y2                    50       3    9127.985    0.6634      80.39   0.0000
----------------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y1           |
          y2 |   .0012956   .0005021     2.58   0.010     .0003116    .0022796
          x3 |   .0061423   .0020951     2.93   0.003      .002036    .0102486
       _cons |   51.43261   20.89625     2.46   0.014     10.47671     92.3885
-------------+----------------------------------------------------------------
y2           |
          y1 |   289.0278    56.8401     5.08   0.000     177.6233    400.4324
          x1 |   264.9876   141.2692     1.88   0.061      -11.895    541.8702
          x2 |  -.0003677    .000266    -1.38   0.167     -.000889    .0001536
       _cons |  -35235.29   9774.146    -3.60   0.000    -54392.26   -16078.32
------------------------------------------------------------------------------
Endogenous variables:  y1 y2
Exogenous variables:   x1 x2 x3
------------------------------------------------------------------------------
```

2).

a. Estimate the model (4) using NLS estimation method using initial values of $\ln\lambda=1$, $\theta=0.5$, $\beta=0.5$, $\alpha=-0.5$. Determine the estimated value of efficiency parameter (λ), distribution parameter (θ), parameter (β), and substitution parameter (α), and elasticity of substitution (σ). Perform F-test to test whether $\theta=0$, $\alpha=0$, and $\beta=0$. (6 points)

```
5 . nl (lnC = {ln(lamda)}-({beta}/{alpha})*ln({theta}*R^(-{alpha})+(1-{theta})*W^(-{alpha}))), init(ln(lamda
> ) 1 theta 0.5 beta 0.5 alpha -0.5)
(obs = 250)

Iteration 0:  residual SS =  315.7594
Iteration 1:  residual SS =  298.0254
Iteration 2:  residual SS =  297.5135
Iteration 3:  residual SS =  297.5127
Iteration 4:  residual SS =  297.5127
Iteration 5:  residual SS =  297.5127
Iteration 6:  residual SS =  297.5127
Iteration 7:  residual SS =  297.5127


      Source |      SS            df       MS
-------------+----------------------------------    Number of obs =        250
       Model |  44.833222          3  14.9444072    R-squared     =     0.1310
    Residual |  297.51272        246  1.20940131    Adj R-squared =     0.1204
-------------+----------------------------------    Root MSE      =   1.099728
       Total |  342.34594        249  1.37488331    Res. dev.     =   752.9683
```

```
Sunday March 21 10:30:49 2021   Page 2

------------------------------------------------------------------------------
         lnC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   ln(lamda) |   2.150055   .7760475     2.77   0.006     .6215102    3.678601
       /beta |   .8120362   .1479719     5.49   0.000     .5205828     1.10349
      /alpha |  -1.070749   .7145536    -1.50   0.135    -2.478172    .3366746
      /theta |    .225971   .2422622     0.93   0.352    -.2512017    .7031437
------------------------------------------------------------------------------
  Parameter ln(lamda) taken as constant term in model & ANOVA table
```

```
6 . est store nls1

7 . sca sigma=1/(1-(_b[/theta])
  too few ')' or ']'
  r(132);

8 . sca sigma=1/(1-(_b[/theta]))

9 . test (_b[/theta]=0) (_b[/alpha]=0) (_b[/beta]=0)

   ( 1)  [theta]_cons = 0
   ( 2)  [alpha]_cons = 0
   ( 3)  [beta]_cons = 0

         F(  3,   246) =   37.78
              Prob > F =    0.0000
```

b. What will happy if we change initial values to $\ln\lambda$=0.5, θ=0.1, β=0.1, α= –0.1? Will the estimated results be the same as (a)? What are the differences between the previous result in (a) and the new result? Give explanation why? (6 points)

Ans. The result remains the same. Although we change the initial values of ln(lamda), theta, beta, and alpha, it does not affect the estimate as we already transform the model into a log form.

```
10 . nl (lnC = {ln(lamda)}-({beta}/{alpha})*ln({theta}*R^(-{alpha})+(1-{theta})*W^(-{alpha}))), init(ln(lamda
   > ) 0.5 theta 0.1 beta 0.1 alpha -0.1)
   (obs = 250)

   Iteration 0:  residual SS =  4522.524
   Iteration 1:  residual SS =  2782.494
   Iteration 2:  residual SS =  499.7826
   Iteration 3:  residual SS =  492.9607
   Iteration 4:  residual SS =  333.3436
   Iteration 5:  residual SS =  299.3888
   Iteration 6:  residual SS =  297.9689
   Iteration 7:  residual SS =  297.5269
   Iteration 8:  residual SS =  297.5129
   Iteration 9:  residual SS =  297.5127
   Iteration 10:  residual SS =  297.5127
   Iteration 11:  residual SS =  297.5127
   Iteration 12:  residual SS =  297.5127
   Iteration 13:  residual SS =  297.5127
```

Source	SS	df	MS
Model	44.833222	3	14.9444072
Residual	297.51272	246	1.20940131
Total	342.34594	249	1.37488331

Number of obs	=	250
R-squared	=	0.1310
Adj R-squared	=	0.1204
Root MSE	=	1.099728
Res. dev.	=	752.9683

lnC	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln(lamda)	2.150055	.7760475	2.77	0.006	.6215102	3.678601
/beta	.8120362	.1479719	5.49	0.000	.5205828	1.10349
/alpha	-1.070749	.7145523	-1.50	0.135	-2.47817	.3366723
/theta	.2259711	.2422623	0.93	0.352	-.2512019	.7031441

Parameter ln(lamda) taken as constant term in model & ANOVA table

```
11 . est store nls2

12 . est table nls1 nls2, star(.1 .05 .01) stat(N rss r2 r2_a)
```

Variable	nls1	nls2
ln(lamda)		
_cons	2.1500554***	2.1500554***
beta		
_cons	.81203618***	.81203616***
alpha		
_cons	-1.0707489	-1.0707487
theta		
_cons	.22597098	.22597108
Statistics		
N	250	250
rss	297.51272	297.51272
r2	.13095882	.13095882
r2_a	.12036076	.12036076

legend: * p<.1; ** p<.05; *** p<.01

c. From (b), if we change convergence value from default of 0.00001 or (1e-5) to (i) 0.1 or (1e-1) and (ii) (1e-15) with maximum iteration of 40, what will happen to the estimated result? Interpret the estimated result and why do we get this kind of result? (Make comparison between previous result in (b) and the new result) (6 points)

Ans. The result nearly the same as we use the transformed model to estimate. Because it only has to iterate to 0.1, no need to go further under this command, and as we increase the convergence value form 1e-5 to 0.1, the value of residual SS iteration will be a bit higher than 1e-5.

Under convergence value of 1e-15, the final result of iteration yields exactly the same with that of 1e-5. Although the convergence value of 1e-5 is higher but the iteration no need to go further as we can see that the iteration result remains the same for about 20 times.

```
14 . est store nls3

15 . nl (lnC = {ln(lamda)}-({beta}/{alpha})*ln({theta}*R^(-{alpha})+(1-{theta})*W^(-{alpha}))), init(ln(lamda
> ) 0.5 theta 0.1 beta 0.1 alpha -0.1) eps(1e-15) iter(40)
  (obs = 250)

  Iteration 0:  residual SS =  4522.524
  Iteration 1:  residual SS =  2782.494
  Iteration 2:  residual SS =  499.7826
  Iteration 3:  residual SS =  492.9607
  Iteration 4:  residual SS =  333.3436
  Iteration 5:  residual SS =  299.3888
  Iteration 6:  residual SS =  297.9689
  Iteration 7:  residual SS =  297.5269
  Iteration 8:  residual SS =  297.5129
  Iteration 9:  residual SS =  297.5127
  Iteration 10:  residual SS =  297.5127
  Iteration 11:  residual SS =  297.5127
  Iteration 12:  residual SS =  297.5127
  Iteration 13:  residual SS =  297.5127
  Iteration 14:  residual SS =  297.5127
  Iteration 15:  residual SS =  297.5127
  Iteration 16:  residual SS =  297.5127
  Iteration 17:  residual SS =  297.5127
  Iteration 18:  residual SS =  297.5127
  Iteration 19:  residual SS =  297.5127
  Iteration 20:  residual SS =  297.5127
  Iteration 21:  residual SS =  297.5127
  Iteration 22:  residual SS =  297.5127
  Iteration 23:  residual SS =  297.5127
  Iteration 24:  residual SS =  297.5127
  Iteration 25:  residual SS =  297.5127
  Iteration 26:  residual SS =  297.5127
  Iteration 27:  residual SS =  297.5127
  Iteration 28:  residual SS =  297.5127
  Iteration 29:  residual SS =  297.5127
  Iteration 30:  residual SS =  297.5127
  Iteration 31:  residual SS =  297.5127
  Iteration 32:  residual SS =  297.5127
  Iteration 33:  residual SS =  297.5127

        Source |      SS            df       MS
  -------------+----------------------------------    Number of obs =        250
         Model |  11566.902          4  2891.72555    R-squared     =     0.9749
      Residual |  297.51272        246  1.20940131    Adj R-squared =     0.9745
  -------------+----------------------------------    Root MSE      =   1.099728
         Total |  11864.415        250  47.4576596    Res. dev.     =   752.9683

  ------------------------------------------------------------------------------
           lnC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
     ln(lamda) |   2.150055   .7760475     2.77   0.006     .6215102      3.6786
         /beta |   .8120362   .1479719     5.49   0.000     .5205828     1.10349
        /alpha |  -1.070749   .7145544    -1.50   0.135    -2.478174     .3366759
        /theta |   .2259709   .2422621     0.93   0.352    -.2512016     .7031435
  ------------------------------------------------------------------------------

13 . nl (lnC = {ln(lamda)}-({beta}/{alpha})*ln({theta}*R^(-{alpha})+(1-{theta})*W^(-{alpha}))), init(ln(lamda
> ) 0.5 theta 0.1 beta 0.1 alpha -0.1) eps(1e-1)
  (obs = 250)

  Iteration 0:  residual SS =  4522.524
  Iteration 1:  residual SS =  2782.494
  Iteration 2:  residual SS =  499.7826
  Iteration 3:  residual SS =  492.9607
  Iteration 4:  residual SS =  333.3436
  Iteration 5:  residual SS =  299.3888
  Iteration 6:  residual SS =  297.9689
  Iteration 7:  residual SS =  297.5269
  Iteration 8:  residual SS =  297.5129

        Source |      SS            df       MS
  -------------+----------------------------------    Number of obs =        250
         Model |  44.833086          3  14.9443621    R-squared     =     0.1310
      Residual |  297.51286        246  1.20940186    Adj R-squared =     0.1204
  -------------+----------------------------------    Root MSE      =   1.099728
         Total |  342.34594        249  1.37488331    Res. dev.     =   752.9684

  ------------------------------------------------------------------------------
           lnC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
     ln(lamda) |   2.150928   .7764215     2.77   0.006     .6216464     3.68021
         /beta |   .8115734   .1478932     5.49   0.000     .5202749     1.102872
        /alpha |  -1.063781   .6803851    -1.56   0.119    -2.403904     .2763427
        /theta |   .2281466   .2427406     0.94   0.348    -.2499685     .7062617
  ------------------------------------------------------------------------------
    Parameter beta taken as constant term in model & ANOVA table
```

d. Why do we prefer to estimate nonlinear regression model in log-form instead of its original functional form? (2 points)

Ans. Because as we use log form it makes the model more robust to change in initial value. Convergence value, and times iteration. The result remains nearly the same.

a. Estimate the above models using MLE with (i) Newton-Ralphson algorithm; (ii) Berndt-Hall-Hall-Hausman algorithm; and (iii) Broyden-Fletcher-Goldfarb-Shanno algorithm, make comparison of the estimated results using different algorithm, and give explanation why are they different? (5 points)

Ans. Because the algorithm is different where they use different methods to compute delta. (iteration: $\theta hat_{t+1} = \theta hat_t - delta_t$). Newton(the default) uses inverse of hessian and gradient matrix to compute. Bhhh does not use hessian because hessian inverse sometimes is difficult to calculate, it uses only gradient.

```
4 . program ml_logit
  1.    args lnf theta
  2.    quietly replace `lnf'=ln(1/(1+exp(-`theta'))) if $ML_y1==1
  3.    quietly replace `lnf'=ln(1-(1/(1+exp(-`theta')))) if $ML_y1==0
  4. end

5 .
end of do-file

6 . ml model lf ml_logit (y=x1 x2)

7 . ml max

initial:       log likelihood = -277.25887
alternative:   log likelihood = -266.63079
rescale:       log likelihood = -266.63079
Iteration 0:   log likelihood = -266.63079
Iteration 1:   log likelihood = -220.43542
Iteration 2:   log likelihood = -219.32057
Iteration 3:   log likelihood = -219.31866
Iteration 4:   log likelihood = -219.31866

                                                Number of obs     =        400
                                                Wald chi2(2)      =      67.00
Log likelihood = -219.31866                     Prob > chi2       =     0.0000
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
x1	.2717348	.1165076	2.33	0.020	.043384	.5000855
x2	-.5636146	.0705617	-7.99	0.000	-.701913	-.4253162
_cons	.2278471	.1946263	1.17	0.242	-.1536134	.6093075

```
8 . est store ur

9 . ml model lf ml_logit (y=x1 x2), tech(bhhh)

10 . ml max

initial:       log likelihood = -277.25887
alternative:   log likelihood = -266.63079
rescale:       log likelihood = -266.63079
Iteration 0:   log likelihood = -266.63079
Iteration 1:   log likelihood = -220.09581
Iteration 2:   log likelihood = -219.32406
Iteration 3:   log likelihood = -219.31868
Iteration 4:   log likelihood = -219.31866

                                          Number of obs     =        400
                                          Wald chi2(2)      =      65.32
Log likelihood = -219.31866               Prob > chi2       =     0.0000
```

y	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf. Interval]	
x1	.2717805	.1176237	2.31	0.021	.0412423	.5023188
x2	-.5636101	.0719945	-7.83	0.000	-.7047167	-.4225035
_cons	.2277883	.1939254	1.17	0.240	-.1522984	.607875

```
11 . est store bhhh

12 . help ml

13 . ml model lf ml_logit (y=x1 x2), technique(bfgs)

14 . ml max

initial:       log likelihood = -277.25887
alternative:   log likelihood = -266.63079
rescale:       log likelihood = -266.63079
Iteration 0:   log likelihood = -266.63079
Iteration 1:   log likelihood = -258.84807  (backed up)
Iteration 2:   log likelihood = -256.80725  (backed up)
Iteration 3:   log likelihood =  -251.8522
Iteration 4:   log likelihood =  -221.9644
Iteration 5:   log likelihood = -219.41513
Iteration 6:   log likelihood = -219.33249
Iteration 7:   log likelihood = -219.31868
Iteration 8:   log likelihood = -219.31866

                                          Number of obs     =        400
                                          Wald chi2(2)      =      66.98
Log likelihood = -219.31866               Prob > chi2       =     0.0000
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
x1	.2716236	.1165017	2.33	0.020	.0432844	.4999628
x2	-.5634727	.0705532	-7.99	0.000	-.7017544	-.4251909
_cons	.2275914	.1946156	1.17	0.242	-.1538483	.609031

```
15 . est store bf
```

b. Perform hypothesis testing whether $\beta_1 = \beta_2 = 0$ using LR-test and Wald test. Make comparison between the two tests. Which test is preferable? Why? (5 points)

Ans. LR test is preferable because it uses both unrestricted and restricted model to test. unlike wald or LM test, it uses only unrestricted or restricted to test.

```
16 . ml model lf ml_logit (y=)

17 . ml max

  initial:       log likelihood = -277.25887
  alternative:   log likelihood = -266.63079
  rescale:       log likelihood = -266.63079
  Iteration 0:   log likelihood = -266.63079
  Iteration 1:   log likelihood = -266.58356
  Iteration 2:   log likelihood = -266.58356

                                                Number of obs     =        400
                                                Wald chi2(0)      =          .
  Log likelihood = -266.58356                   Prob > chi2       =          .

------------------------------------------------------------------------------
             y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+--------------------------------------------------------------
         _cons |  -.4683789   .1027548    -4.56   0.000    -.6697746   -.2669832
------------------------------------------------------------------------------

18 . est store res

19 . lrtest ur res

  Likelihood-ratio test                               LR chi2(2)  =     94.53
  (Assumption: res nested in ur)                      Prob > chi2 =    0.0000

21 . ml model lf ml_logit (y=x1 x2)

22 . ml max

  initial:       log likelihood = -277.25887
  alternative:   log likelihood = -266.63079
  rescale:       log likelihood = -266.63079
  Iteration 0:   log likelihood = -266.63079
  Iteration 1:   log likelihood = -220.43542
  Iteration 2:   log likelihood = -219.32057
  Iteration 3:   log likelihood = -219.31866
  Iteration 4:   log likelihood = -219.31866

                                                Number of obs     =        400
                                                Wald chi2(2)      =      67.00
  Log likelihood = -219.31866                   Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+--------------------------------------------------------------
            x1 |   .2717348   .1165076     2.33   0.020      .043384    .5000855
            x2 |  -.5636146   .0705617    -7.99   0.000     -.701913   -.4253162
         _cons |   .2278471   .1946263     1.17   0.242    -.1536134    .6093075
------------------------------------------------------------------------------

23 . test x1 x2

 ( 1)  [eq1]x1 = 0
 ( 2)  [eq1]x2 = 0

           chi2(  2) =   67.00
         Prob > chi2 =    0.0000
```

c. Why overall test in MLE is Chi-square test instead of F-test? Can we still employ F-test as overall test? Why or why not? (2 points)

Ans. Due to the computation of wald test which is unrestricted/overall test, the result will not be negative (chi^2). No, we cannot use F-test because in this case we compare the value of loglikelihood instead of R-squared.

d. Estimate the models with heteroskedasticity using MLE with Newton-Ralphson algorithm. Perform LR-test whether there exists significant heteroskedasticity.

```
4 . program ml_probit
  1.     args lnf theta
  2.     tempvar z
  3.     quietly g double `z'=`theta'
  4.     quietly replace `lnf'=ln(normal(`z')) if $ML_y1==1
  5.     quietly replace `lnf'=ln(1-normal(`z')) if $ML_y1==0
  6. end

5 .
end of do-file

6 . ml model lf ml_probit (y=x1 x2)

7 . ml max

initial:       log likelihood = -277.25887
alternative:   log likelihood = -271.85123
rescale:       log likelihood = -266.80559
Iteration 0:   log likelihood = -266.80559
Iteration 1:   log likelihood = -219.83398
Iteration 2:   log likelihood = -219.35724
Iteration 3:   log likelihood = -219.35696
Iteration 4:   log likelihood = -219.35696

                                                Number of obs     =        400
                                                Wald chi2(2)      =      76.45
Log likelihood = -219.35696                     Prob > chi2       =     0.0000
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	.1585088	.069309	2.29	0.022	.0226656	.294352
x2	-.3350693	.039372	-8.51	0.000	-.412237	-.2579016
_cons	.1398805	.1162269	1.20	0.229	-.08792	.3676809

```
8 . est store probit

9 . do "C:\Users\a\AppData\Local\Temp\STD4024_000000.tmp"

10 . program ml_probit_het
  1.     args lnf theta sigma
  2.     tempvar z s
  3.     quietly g double `s'=exp(`sigma')
  4.     quietly g double `z'=`theta'/`s'
  5.     quietly replace `lnf'=ln(normal(`z')) if $ML_y1==1
  6.     quietly replace `lnf'=ln(1-normal(`z')) if $ML_y1==0
  7. end

11 .
end of do-file

12 . ml model lf ml_probit_het (y=x1 x2) (x3, noconstant)

13 . ml max

initial:       log likelihood = -277.25887
alternative:   log likelihood =   -271.834
rescale:       log likelihood =  -266.8011
rescale eq:    log likelihood = -266.61459
Iteration 0:   log likelihood = -266.61459
Iteration 1:   log likelihood = -241.98701
Iteration 2:   log likelihood = -233.69245
Iteration 3:   log likelihood =  -219.0346
Iteration 4:   log likelihood = -217.26136
Iteration 5:   log likelihood = -217.02236
Iteration 6:   log likelihood = -217.02113
Iteration 7:   log likelihood = -217.02113

                                                Number of obs     =        400
                                                Wald chi2(2)      =      74.28
Log likelihood = -217.02113                     Prob > chi2       =     0.0000
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
eq1						
x1	.1744698	.0664826	2.62	0.009	.0441662	.3047733
x2	-.3334951	.0399312	-8.35	0.000	-.4117589	-.2552314
_cons	.0952407	.1122231	0.85	0.396	-.1247127	.315194
eq2						
x3	-.2116171	.0998641	-2.12	0.034	-.407347	-.0158872

```
14 . est store hetprob

15 . lrtest hetprob probit

   Likelihood-ratio test                                 LR chi2(1)  =      4.67
   (Assumption: probit nested in hetprob)                Prob > chi2 =    0.0307

16 . est table probit hetprob, star(.1 .05 .01) stat( N ll chi2 p )
```

Variable	probit	hetprob
eq1		
x1	.15850879**	.17446976***
x2	-.33506928***	-.33349513***
_cons	.13988047	.09524065
eq2		
x3		-.2116171**
Statistics		
N	400	400
ll	-219.35696	-217.02113
chi2	76.4453	74.277847
p	2.513e-17	7.426e-17

legend: * p<.1; ** p<.05; *** p<.01

Ans. From Prob > chi2 = 0.037, the p-value is less than 0.05 as we reject the null hypothesis, meaning that there is significant heteroskedasticicty.

In this case, can we perform Wald-test or LM-test to test heteroskedasticity? Why? or why not? (5 points)

Ans. No, we cannot. Because LM and Wald use only one of the two model so we cannot test the difference of the two model under LM and Wald tests.

e. Why individual test in MLE is z-test instead of t-test? Why don't we have R-square in MLE? If we would like to make comparison between the two estimated results, which statistical index should be employed? (3 points)

Ans. We use z-test because we assume asymptotic properties to get unbiased estimate of variance. Because we treat x and y as given and we maximize the value of loglikelihood to get the estimate of the theta. Unlike least square, we treat x and y as variables and we minimize the SSR. Also, from that, we can get the meaning of that how much independent variable can explain dependent variable.

If to compare models in MLE. We should look at the value of loglikelihood where the higher is the better.

a. Estimate Unrestricted model using method of moment, Merton model using GMM, and Vasicek using GMM. Make evaluation of the estimated result of Merton model. (5 points).

Ans. From Hansen's J chi2(2) = 7.36003 (p = 0.0252), merton reject the null hypothesis, indicating that some of the moment conditions under Merton model are not true.

```
3 . tsset time
        time variable:  time, 1 to 1326
                delta:  1 unit

4 . g dr=f.r-r
  (1 missing value generated)

5 . gmm (dr-{alpha}-{beta}*r) ((dr-{alpha}-{beta}*r)*r) ((dr-{alpha}-{beta}*r)^2- {sigma2}*r^(2*{ga
  > ma})) (((dr-{alpha}-{beta}*r)^2-{sigma2}*r^(2*{gamma}))*r) winitial(identity)
  note: 1 missing value returned for equation 1 at initial values
  note: 1 missing value returned for equation 2 at initial values
  note: 1 missing value returned for equation 3 at initial values
  note: 1 missing value returned for equation 4 at initial values

  Step 1
  numerical derivatives are approximate
  flat or discontinuous region encountered
  Iteration 0:   GMM criterion Q(b) =  .00001191
  Iteration 1:   GMM criterion Q(b) =  8.702e-06  (backed up)
  Iteration 2:   GMM criterion Q(b) =  6.127e-06  (not concave)
  Iteration 3:   GMM criterion Q(b) =  5.698e-06  (backed up)
  Iteration 4:   GMM criterion Q(b) =  5.645e-06  (backed up)

  Step 2
  Iteration 0:   GMM criterion Q(b) =  .00791639
  Iteration 1:   GMM criterion Q(b) =  .00781166  (backed up)
  Iteration 2:   GMM criterion Q(b) =  .00702483
  Iteration 3:   GMM criterion Q(b) =  .00133279
  Iteration 4:   GMM criterion Q(b) =   .0001404
  Iteration 5:   GMM criterion Q(b) =  1.292e-06
  Iteration 6:   GMM criterion Q(b) =  5.361e-11
  Iteration 7:   GMM criterion Q(b) =  4.436e-20

  note: model is exactly identified
```

```
GMM estimation

Number of parameters =   4
Number of moments    =   4
Initial weight matrix: Identity                 Number of obs   =      1,325
GMM weight matrix:     Robust
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/alpha	-.0023916	.0011632	-2.06	0.040	-.0046713	-.0001118
/beta	.0004321	.0002872	1.50	0.132	-.0001308	.000995
/sigma2	.0005129	.000328	1.56	0.118	-.0001301	.0011558
/gamma	.0944296	.1808315	0.52	0.602	-.2599936	.4488528

```
Instruments for equation 1: _cons
Instruments for equation 2: _cons
Instruments for equation 3: _cons
Instruments for equation 4: _cons
```

```
6 . gmm (dr-{alpha}) ((dr-{alpha})*r) ((dr-{alpha})^2-{sigma2}) (((dr-{alpha})^2- {sigma2})*r) win
  > ial(identity)
  note: 1 missing value returned for equation 1 at initial values
  note: 1 missing value returned for equation 2 at initial values
  note: 1 missing value returned for equation 3 at initial values
  note: 1 missing value returned for equation 4 at initial values

  Step 1
  Iteration 0:   GMM criterion Q(b) =  .00001191
  Iteration 1:   GMM criterion Q(b) =  4.145e-08
  Iteration 2:   GMM criterion Q(b) =  4.144e-08

  Step 2
  Iteration 0:   GMM criterion Q(b) =  .00795661
  Iteration 1:   GMM criterion Q(b) =  .00555474
  Iteration 2:   GMM criterion Q(b) =  .00555474

  GMM estimation

  Number of parameters =   2
  Number of moments    =   4
  Initial weight matrix: Identity                 Number of obs   =      1,325
  GMM weight matrix:     Robust
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/alpha	-.0008231	.0006876	-1.20	0.231	-.0021708	.0005245
/sigma2	.000435	.0002932	1.48	0.138	-.0001396	.0010095

```
  Instruments for equation 1: _cons
  Instruments for equation 2: _cons
  Instruments for equation 3: _cons
  Instruments for equation 4: _cons

7 . estat overid

    Test of overidentifying restriction:

    Hansen's J chi2(2) = 7.36003 (p = 0.0252)

8 . est store Merton

9 . gmm (dr-{alpha}-{beta}*r) ((dr-{alpha}-{beta}*r)*r) ((dr-{alpha}-{beta}*r)^2- {sigma2}) (((dr-
  > lpha}-{beta}*r)^2-{sigma2})*r) winitial(identity)
  note: 1 missing value returned for equation 1 at initial values
  note: 1 missing value returned for equation 2 at initial values
  note: 1 missing value returned for equation 3 at initial values
  note: 1 missing value returned for equation 4 at initial values

  Step 1
  Iteration 0:   GMM criterion Q(b) =  .00001191
  Iteration 1:   GMM criterion Q(b) =  3.158e-10
  Iteration 2:   GMM criterion Q(b) =  3.092e-10

  Step 2
  Iteration 0:   GMM criterion Q(b) =  .00059096
  Iteration 1:   GMM criterion Q(b) =  .00019358
  Iteration 2:   GMM criterion Q(b) =  .00019357

  GMM estimation

  Number of parameters =   3
  Number of moments    =   4
  Initial weight matrix: Identity                 Number of obs   =      1,325
  GMM weight matrix:     Robust
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/alpha	-.0027106	.0009773	-2.77	0.006	-.004626	-.0007952
/beta	.0005365	.0001996	2.69	0.007	.0001452	.0009278
/sigma2	.0005957	.0003005	1.98	0.047	6.68e-06	.0011847

```
  Instruments for equation 1: _cons
  Instruments for equation 2: _cons
  Instruments for equation 3: _cons
  Instruments for equation 4: _cons

10 . estat overid

    Test of overidentifying restriction:

    Hansen's J chi2(1) = .256486 (p = 0.6125)

11 . est store Vasicek
```

```
12 . gmm (dr-{alpha}-{beta}*r) ((dr-{alpha}-{beta}*r)*r) ((dr-{alpha}-{beta}*r)^2- {sigma2}*r^(2*{ga
> ma})) (((dr-{alpha}-{beta}*r)^2-{sigma2}*r^(2*{gamma}))*r) winitial(identity)
note: 1 missing value returned for equation 1 at initial values
note: 1 missing value returned for equation 2 at initial values
note: 1 missing value returned for equation 3 at initial values
note: 1 missing value returned for equation 4 at initial values

Step 1
numerical derivatives are approximate
flat or discontinuous region encountered
Iteration 0:   GMM criterion Q(b) =  .00001191
Iteration 1:   GMM criterion Q(b) =  8.702e-06  (backed up)
Iteration 2:   GMM criterion Q(b) =  6.127e-06  (not concave)
Iteration 3:   GMM criterion Q(b) =  5.698e-06  (backed up)
Iteration 4:   GMM criterion Q(b) =  5.645e-06  (backed up)

Step 2
Iteration 0:   GMM criterion Q(b) =  .00791639
Iteration 1:   GMM criterion Q(b) =  .00781166  (backed up)
Iteration 2:   GMM criterion Q(b) =  .00702483
Iteration 3:   GMM criterion Q(b) =  .00133279
Iteration 4:   GMM criterion Q(b) =   .0001404
Iteration 5:   GMM criterion Q(b) =  1.292e-06
Iteration 6:   GMM criterion Q(b) =  5.361e-11
Iteration 7:   GMM criterion Q(b) =  4.436e-20

note: model is exactly identified

GMM estimation

Number of parameters =   4
Number of moments    =   4
Initial weight matrix: Identity                 Number of obs   =      1,325
GMM weight matrix:     Robust
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/alpha	-.0023916	.0011632	-2.06	0.040	-.0046713	-.0001118
/beta	.0004321	.0002872	1.50	0.132	-.0001308	.000995
/sigma2	.0005129	.000328	1.56	0.118	-.0001301	.0011558
/gamma	.0944296	.1808315	0.52	0.602	-.2599936	.4488528

```
Instruments for equation 1: _cons
Instruments for equation 2: _cons
Instruments for equation 3: _cons
Instruments for equation 4: _cons

13 . est store ur

14 . est table ur Merton Vasicek, star(0.1 0.05 0.01) stat( N J )
```

Variable	ur	Merton	Vasicek
alpha			
_cons	-.00239157**	-.00082314	-.00271059***
beta			
_cons	.0004321		.0005365***
sigma2			
_cons	.00051289	.00043498	.00059569**

gamma			
_cons	.09442959		
Statistics			
N	1325	1325	1325
J	5.878e-17	7.360027	.25648614

legend: * p<.1; ** p<.05; *** p<.01

```
15 . test (_b[/beta]=0) (_b[/gamma]=0)

 ( 1)  [beta]_cons = 0
 ( 2)  [gamma]_cons = 0

           chi2(  2) =    7.90
         Prob > chi2 =    0.0193

16 . test (_b[/gamma]=0)

 ( 1)  [gamma]_cons = 0

           chi2(  1) =    0.27
         Prob > chi2 =    0.6015
```

b. Determine which model is the most appropriated model. Provide and give explanation of your selected criteria. (5 points)

Ans. Vasicek. Because from

```
16 . test (_b[/gamma]=0)

 ( 1)  [gamma]_cons = 0

           chi2(  1) =    0.27
         Prob > chi2 =    0.6015
```

it means we accpect null hypothesis and Vasicek is the most appropriated.

c. What will happen to the estimated results if we estimate this model (10) using OLS? (2 points)

Ans. The estimate under ols method will be biased, inconsistent, and inefficient.

If $E(z_{1i}u_i)=0,\ E(z_{2i}u_i)=0,\ E(z_{3i}u_i)=0$, and z_{1i}, z_{2i}, and z_{3i} are highly correlated with x_{1i} and x_{2i}

d. Estimate the model (10) using GMM and 2SLS by employing z_{1i}, z_{2i}, and z_{3i} as instrumental variables for x_{1i} and x_{2i}. Perform the test to check whether GMM is appropriated. (5 points).

D). Ans. Gmm is appropriated as p-value > 0.05 and we accept the null hypothesis which means all conditions are true.

```
3 . ivregress gmm y (x1 x2 = z1 z2 z3)

Instrumental variables (GMM) regression          Number of obs   =        500
                                                 Wald chi2(2)    =      76.89
                                                 Prob > chi2     =     0.0000
                                                 R-squared       =     0.5443
GMM weight matrix: Robust                        Root MSE        =     21.684
```

y	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	3.540811	1.063452	3.33	0.001	1.456483	5.625139
x2	1.813119	.5669784	3.20	0.001	.701862	2.924376
_cons	9.737418	8.191478	1.19	0.235	-6.317583	25.79242

```
Instrumented:  x1 x2
Instruments:   z1 z2 z3

4 . estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(1) = .000449 (p = 0.9831)
```

```
5 . ivregress 2sls y (x1 x2 = z1 z2 z3)

Instrumental variables (2SLS) regression          Number of obs   =        500
                                                  Wald chi2(2)    =      70.06
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.5442
                                                  Root MSE        =     21.685
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
x1	3.540565	1.070329	3.31	0.001	1.442759	5.638371
x2	1.813013	.5711847	3.17	0.002	.6935111	2.932514
_cons	9.742898	8.637345	1.13	0.259	-7.185987	26.67178

```
Instrumented:  x1 x2
Instruments:   z1 z2 z3
```

e. According to the estimated results of (d), give explanation of the differences between 2SLS and GMM estimated results in this case. (3 points)

Ans. In 2sls, we state the reduced form first where x_i that correlates with the error term is dependent variable and the variables that highly correlated with x_i is the dependent variables. Then run the reduced form using ols method and obtain x_i hat, and use those instead of actual one to run the structural form, thus, get the result.

For gmm, we come o with moments conditions where we use the variables that uncorr with error term but correlate with x_i instead of x_i; plug into the conditions. Then, we test whether all conditions are true.

5).

(a) Estimate the model assuming that the probability function is cumulative normal distribution function and logistic probability distribution. Can we compare the estimated coefficients of the two estimated functional forms? Why? or why not? Also, make interpret the estimated result of the Logit model (Overall test, individual test, pseudo R^2, counted R^2). (8 points)

Ans. No, because we assume different distribution so we cannot compare the coefficients, but we can compare the marginal effect of both models.

McFadden 0.088 ; the higher is better when compare model; no meaning

McFadden(adjusted) 0.055

Count 0.735 ; predictability

```
. logit y x1 x2 x3

Iteration 0:   log likelihood = -124.34324
Iteration 1:   log likelihood = -114.02587
Iteration 2:   log likelihood = -113.45555
Iteration 3:   log likelihood = -113.45347
Iteration 4:   log likelihood = -113.45347

Logistic regression                             Number of obs     =        215
                                                LR chi2(3)        =      21.78
                                                Prob > chi2       =     0.0001
Log likelihood = -113.45347                     Pseudo R2         =     0.0876
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	.0078462	.0026161	3.00	0.003	.0027187	.0129736
x2	.0204767	.0155317	1.32	0.187	-.0099648	.0509182
x3	.2007257	.080111	2.51	0.012	.0437111	.3577404
_cons	.4287884	.1939767	2.21	0.027	.048601	.8089758

```
15 . fitstat
```

	logit
Log-likelihood	
Model	-113.453
Intercept-only	-124.343
Chi-square	
Deviance(df=211)	226.907
LR(df=3)	21.780
p-value	0.000
R2	
McFadden	0.088
McFadden(adjusted)	0.055
McKelvey & Zavoina	0.192
Cox-Snell/ML	0.096
Cragg-Uhler/Nagelkerke	0.141
Efron	0.089
Tjur's D	0.090
Count	0.735
Count(adjusted)	0.000
IC	
AIC	234.907
AIC divided by N	1.093
BIC(df=4)	248.389
Variance of	
e	3.290
y-star	4.071

```
16 . sum x1 x2 x3

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          x1 |       215    60.87276      79.649  -75.47762   332.6589
          x2 |       215   -1.596953    13.82847  -54.81482    40.2438
          x3 |       215    1.625622    2.873515  -3.865531   13.81453

17 . mfx

Marginal effects after logit
      y  = Pr(y) (predict)
         =  .76852643
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .0013958      .00044    3.18   0.001   .000534  .002257   60.8728
      x2 |   .0036427       .0027    1.35   0.178  -.001658  .008943  -1.59695
      x3 |   .0357078      .01354    2.64   0.008   .009172  .062244   1.62562
------------------------------------------------------------------------------

18 . mfx, at(median)

Marginal effects after logit
      y  = Pr(y) (predict)
         =  .61785883
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .0018525      .00066    2.82   0.005   .000565   .00314   .780464
      x2 |   .0048347      .00368    1.31   0.189  -.002384  .012054   .549624
      x3 |   .0473932      .01947    2.43   0.015    .00924  .085546   .170897
------------------------------------------------------------------------------

19 . logit y x1 x2 x3

Iteration 0:   log likelihood = -124.34324
Iteration 1:   log likelihood = -114.02587
Iteration 2:   log likelihood = -113.45555
Iteration 3:   log likelihood = -113.45347
Iteration 4:   log likelihood = -113.45347

Logistic regression                               Number of obs   =        215
                                                  LR chi2(3)      =      21.78
                                                  Prob > chi2     =     0.0001
Log likelihood = -113.45347                       Pseudo R2       =     0.0876

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0078462   .0026161     3.00   0.003     .0027187    .0129736
          x2 |   .0204767   .0155317     1.32   0.187    -.0099648    .0509182
          x3 |   .2007257    .080111     2.51   0.012     .0437111    .3577404
       _cons |   .4287884   .1939767     2.21   0.027      .048601    .8089758
------------------------------------------------------------------------------
```

```
20 . est store ur

21 . logit y

Iteration 0:   log likelihood = -124.34324
Iteration 1:   log likelihood = -124.34324

Logistic regression                             Number of obs     =        215
                                                LR chi2(0)        =       0.00
                                                Prob > chi2       =          .
Log likelihood = -124.34324                     Pseudo R2         =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   1.019544   .1545088     6.60   0.000     .7167121    1.322375
------------------------------------------------------------------------------

22 . est store res

23 . lrtest ur res

Likelihood-ratio test                                 LR chi2(3)  =     21.78
(Assumption: res nested in ur)                        Prob > chi2 =    0.0001

24 . help lr

25 . help lrtest

26 . test x1=x2
x1 not found
r(111);

27 . logit y x1 x2 x3

Iteration 0:   log likelihood = -124.34324
Iteration 1:   log likelihood = -114.02587
Iteration 2:   log likelihood = -113.45555
Iteration 3:   log likelihood = -113.45347
Iteration 4:   log likelihood = -113.45347

Logistic regression                             Number of obs     =        215
                                                LR chi2(3)        =      21.78
                                                Prob > chi2       =     0.0001
Log likelihood = -113.45347                     Pseudo R2         =     0.0876

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0078462   .0026161     3.00   0.003     .0027187    .0129736
          x2 |   .0204767   .0155317     1.32   0.187    -.0099648    .0509182
          x3 |   .2007257    .080111     2.51   0.012     .0437111    .3577404
       _cons |   .4287884   .1939767     2.21   0.027      .048601    .8089758
------------------------------------------------------------------------------

28 . test x1=x2

 ( 1)  [y]x1 - [y]x2 = 0

           chi2(  1) =    0.71
         Prob > chi2 =    0.3981

29 . estat clas

Logistic model for y

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       158            57  |        215
     -     |         0             0  |          0
-----------+--------------------------+-----------
   Total   |       158            57  |        215

Classified + if predicted Pr(D) >= .5
True D defined as y != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)  100.00%
Specificity                     Pr( -|~D)    0.00%
Positive predictive value       Pr( D| +)   73.49%
Negative predictive value       Pr(~D| -)       .%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)  100.00%
False - rate for true D         Pr( -| D)    0.00%
False + rate for classified +   Pr(~D| +)   26.51%
False - rate for classified -   Pr( D| -)       .%
--------------------------------------------------
Correctly classified                        73.49%
--------------------------------------------------
```

(b) Using logistic probability distribution, compute marginal effect at mean and at median. Make interpretation of marginal effects at mean of x_{1i}. (5 points)

Ans. If x1 changes by 1 unit (from 60.8728 to 60.8728+1), probability(hat) of Y = 1 will increase by .0013958, holding x2 x3 at their means -1.59695 and 1.62562.

(c) Perform hypothesis testing whether $\beta_1 = \beta_2 = \beta_3 = 0$ using LR-test and Wald test. Perform hypothesis testing whether $\beta_1 = \beta_2$ using LR-test. Make conclusion of the tests. (5 points)

Ans.

```
LR chi2(3)      =     21.78
Prob > chi2     =    0.0001
```

not reject null; overall test is significant

```
28 . test x1=x2

 ( 1)  [y]x1 - [y]x2 = 0

           chi2(  1) =    0.71
         Prob > chi2 =    0.3981
```

reject null; it is not equal

```
19 . logit y x1 x2 x3

Iteration 0:   log likelihood = -124.34324
Iteration 1:   log likelihood = -114.02587
Iteration 2:   log likelihood = -113.45555
Iteration 3:   log likelihood = -113.45347
Iteration 4:   log likelihood = -113.45347

Logistic regression                             Number of obs     =        215
                                                LR chi2(3)        =      21.78
                                                Prob > chi2       =     0.0001
Log likelihood = -113.45347                     Pseudo R2         =     0.0876
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	.0078462	.0026161	3.00	0.003	.0027187	.0129736
x2	.0204767	.0155317	1.32	0.187	-.0099648	.0509182
x3	.2007257	.080111	2.51	0.012	.0437111	.3577404
_cons	.4287884	.1939767	2.21	0.027	.048601	.8089758

(d) Why is threshold in computing Counted R^2 important? If we change the threshold, can the value of counted R^2 change? Why? (2 points)

Ans. YES, THE VALUE can change when we change the threshold. As can be seen that the sensitivity is 100 percent, indicating that it is very sensitive to the threshold. If the threshold for y=1 changes from >.5 to >.7, it can be implied that we weight the other more.

```
. estat clas

Logistic model for y

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       158            57  |        215
     -     |         0             0  |          0
-----------+--------------------------+-----------
   Total   |       158            57  |        215

Classified + if predicted Pr(D) >= .5
True D defined as y != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)  100.00%
Specificity                     Pr( -|~D)    0.00%
Positive predictive value       Pr( D| +)   73.49%
Negative predictive value       Pr(~D| -)       .%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)  100.00%
False - rate for true D         Pr( -| D)    0.00%
False + rate for classified +   Pr(~D| +)   26.51%
False - rate for classified -   Pr( D| -)       .%
--------------------------------------------------
Correctly classified                        73.49%
--------------------------------------------------
```

6).

a. Estimate model (13) using Panel Least Squares estimation method and PGLS assuming Heteroskedasticity, and test whether there exists Heteroskedasticity problem. What will happen if Heteroscedasticity occurs in the model (5 points)

Ans. No, there is no hetero problem as we do not reject the null hypothesis; p-value > 0.05. If there is hetero problem, the least square method still leads to an unbiased estimate but it does not have the lowest variance in which we call it is the best when it has lowest variance.

```
3 . xtset id t
        panel variable:  id (strongly balanced)
         time variable:  t, 1 to 12
                 delta:  1 unit

4 . xtgls y x1 x2 x3 x4 x5 x6 x7, igls panels(heteroskedastic)
  variable x4 not found
  r(111);

5 . xtgls y x1 x2 x3, igls panels(heteroskedastic)
  Iteration 1: tolerance = .00590371
  Iteration 2: tolerance = .00177471
  Iteration 3: tolerance = .00055434
  Iteration 4: tolerance = .000177
  Iteration 5: tolerance = .00005721
  Iteration 6: tolerance = .00001862
  Iteration 7: tolerance = 6.085e-06
  Iteration 8: tolerance = 1.993e-06
  Iteration 9: tolerance = 6.534e-07
  Iteration 10: tolerance = 2.144e-07
  Iteration 11: tolerance = 7.040e-08


  Cross-sectional time-series FGLS regression

  Coefficients:  generalized least squares
  Panels:        heteroskedastic
  Correlation:   no autocorrelation

  Estimated covariances      =       100          Number of obs     =      1,200
  Estimated autocorrelations =         0          Number of groups  =        100
  Estimated coefficients     =         4          Time periods      =         12
                                                  Wald chi2(3)      =   33298.40
  Log likelihood             = -6165.009          Prob > chi2       =     0.0000
```

```
          y        Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

         x1     .3168738   .0100633    31.49   0.000     .2971501    .3365975
         x2     1.311977   .0139878    93.79   0.000     1.284562    1.339393
         x3    -.4630102    .011065   -41.84   0.000    -.4846972   -.4413232
      _cons    -132.2058   6.140582   -21.53   0.000    -144.2411   -120.1705

  6 . est store het

  7 . xtgls y x1 x2 x3

    Cross-sectional time-series FGLS regression

    Coefficients:  generalized least squares
    Panels:        homoskedastic
    Correlation:   no autocorrelation

    Estimated covariances      =          1        Number of obs     =      1,200
    Estimated autocorrelations =          0        Number of groups  =        100
    Estimated coefficients     =          4        Time periods      =         12
                                                   Wald chi2(3)      =   29882.41
    Log likelihood             =   -6211.29        Prob > chi2       =     0.0000

          y        Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

         x1     .3183087   .0109146    29.16   0.000     .2969164    .3397011
         x2     1.320714   .0149495    88.34   0.000     1.291413    1.350014
         x3    -.4642183    .011884   -39.06   0.000    -.4875104   -.4409262
      _cons    -136.2145   6.645908   -20.50   0.000    -149.2402   -123.1887

  8 . est store pgls

  9 . local df=e(N_g)-1

 10 . lrtest het, df(`df')

    Likelihood-ratio test                              LR chi2(99) =      92.56
    (Assumption: pgls nested in het)                   Prob > chi2 =     0.6629
```

b. Estimate the above three models including Panel Least Squares model (13), Fixed effects model (14), and Random-effects model (15). Perform fixed effects tests and random effects test, also state null hypothesis of the tests. Then, determine the most appropriated model. Also, give explanation of the choosing criterion (perform the tests), and make interpretation of the estimated models. (10 points)

Ans. The null hypothesis of fe test: $\alpha_1 = \ldots = \alpha_i = 0$

If we reject it means fixed effects exist. In this case we reject.

The criteria is that if the unobserved variable or the fe is highly correlated with independent variable in the model, then it is better to use fe model. However, when there exists fe, we must choose between fe model and re model, using hausman test to choose. If we reject, we should use fe model. In this case, we reject and use fe model.

```
11 . xtgls y x1 x2 x3

   Cross-sectional time-series FGLS regression

   Coefficients:  generalized least squares
   Panels:        homoskedastic
   Correlation:   no autocorrelation

   Estimated covariances      =         1          Number of obs     =      1,200
   Estimated autocorrelations =         0          Number of groups  =        100
   Estimated coefficients     =         4          Time periods      =         12
                                                   Wald chi2(3)      =   29882.41
   Log likelihood             =  -6211.29          Prob > chi2       =     0.0000

   ------------------------------------------------------------------------------
              y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
   -------------+----------------------------------------------------------------
             x1 |   .3183087   .0109146    29.16   0.000     .2969164    .3397011
             x2 |   1.320714   .0149495    88.34   0.000     1.291413    1.350014
             x3 |  -.4642183    .011884   -39.06   0.000    -.4875104   -.4409262
          _cons |  -136.2145   6.645908   -20.50   0.000    -149.2402   -123.1887
   ------------------------------------------------------------------------------

12 . xtreg y x1 x2 x3, fe

   Fixed-effects (within) regression               Number of obs     =      1,200
   Group variable: id                              Number of groups  =        100

   R-sq:                                           Obs per group:
        within  = 0.9502                                         min =         12
        between = 0.9848                                         avg =       12.0
        overall = 0.8846                                         max =         12

                                                   F(3,1097)         =    6980.72
   corr(u_i, Xb)  = 0.8057                         Prob > F          =     0.0000

   ------------------------------------------------------------------------------
              y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
   -------------+----------------------------------------------------------------
             x1 |   .1014426   .0044866    22.61   0.000     .0926393    .1102459
             x2 |   .6951028   .0085422    81.37   0.000      .678342    .7118637
             x3 |  -.4953168   .0041895  -118.23   0.000    -.5035372   -.4870965
          _cons |    208.659   4.373144    47.71   0.000     200.0784    217.2397
   -------------+----------------------------------------------------------------
        sigma_u |  123.46108
        sigma_e |  14.376737
            rho |  .98662139   (fraction of variance due to u_i)
   ------------------------------------------------------------------------------
   F test that all u_i=0: F(99, 1097) = 96.47                   Prob > F = 0.0000

13 . est store fixed
```

```
14 . xtreg y x1 x2 x3, re

Random-effects GLS regression                   Number of obs     =      1,200
Group variable: id                              Number of groups  =        100

R-sq:                                           Obs per group:
     within  = 0.8956                                         min =         12
     between = 0.9953                                         avg =       12.0
     overall = 0.9543                                         max =         12

                                                Wald chi2(3)      =    8707.50
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .2162513   .0090869    23.80   0.000     .1984414    .2340613
          x2 |   1.028607   .0152844    67.30   0.000     .9986503    1.058564
          x3 |  -.4779339   .0091412   -52.28   0.000    -.4958503   -.4600176
       _cons |   25.24251   8.000206     3.16   0.002     9.562392    40.92262
-------------+----------------------------------------------------------------
     sigma_u |  12.351151
     sigma_e |  14.376737
         rho |  .42464732   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
17 . hausman re fixed

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      re          fixed        Difference          S.E.
-------------+----------------------------------------------------------------
          x1 |    .2162513     .1014426        .1148087         .007902
          x2 |    1.028607     .6951028        .3335042        .0126745
          x3 |   -.4779339    -.4953168        .0173829        .0081246
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =     1451.21
                Prob>chi2 =      0.0000
```

c. What are the differences between Fixed effects estimation method and First difference estimation method? What are the differences between Fixed effects model and Random effects model? (3 points)

Ans. Fe and first diff: Both will have the same result if T=2. But not for T>2. First diff can only used for balanced panel but fe can do both unbalanced and balanced panel.

Fe and re : normally we assume independent variable correlate with unobserved so we use fe model. Nut if it not that highly correlate, we can use re model. Cited from the equation of random effect, lamda can be seen as correlation. If it is one. Which is very high, we use fe model.

d. Give explanation of Within R-squares, Overall R-squares, and Between R-squares of the estimated results of the Fixed-effects model. (2 points)

Ans. The yhat and ybar are not the same under three different R2; overall R2 is more prefer because its computation is the same with normal R2 computation(actual yhat and ybar)