1. From the data set “Midterm_q1_7.dta”: Estimate the following models

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

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

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

```
. reg3 (y1 y2 x3) (y2 y1 x1 x2), ols


Multivariate regression
--------------------------------------------------------------------------
Equation             Obs   Parms        RMSE    "R-sq"     F-Stat        P
--------------------------------------------------------------------------
y1                    50       2    15.69495    0.8241     110.13   0.0000
y2                    50       3    9461.687    0.6690      30.99   0.0000
--------------------------------------------------------------------------


------------------------------------------------------------------------------
             |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y1           |
          y2 |   .0009024   .0001951     4.62   0.000     .0005149      .00129
          x3 |   .0076394    .001097     6.96   0.000      .005461    .0098178
       _cons |   42.12549   16.46368     2.56   0.012     9.431886     74.8191
-------------+----------------------------------------------------------------
y2           |
          y1 |   301.3366   46.72295     6.45   0.000      208.554    394.1191
          x1 |   237.1728   131.9271     1.80   0.075    -24.80819    499.1538
          x2 |  -.0004591   .0003309    -1.39   0.169    -.0011162    .0001981
       _cons |  -35835.83    8913.26    -4.02   0.000     -53535.8   -18135.86
------------------------------------------------------------------------------
```

From the equation, we can see clearly that these are simultaneous equation causing our ols estimates to be bias (endogeneity bias)

$E(u_{1t},y_{2t}) \neq 0$ and $E(u_{2t},y_{1t}) \neq 0$

b. Estimate model (1) and (2) using Two Stage Least Squares (2SLS)

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


Two-stage least-squares regression

----------------------------------------------------------------------

Equation            Obs   Parms        RMSE    "R-sq"     F-Stat        P

----------------------------------------------------------------------

y1                   50       2    16.13454    0.8141      96.60   0.0000

y2                   50       3    9462.845    0.6689      25.50   0.0000

----------------------------------------------------------------------



------------------------------------------------------------------------------

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

-------------+----------------------------------------------------------------

y1           |

          y2 |   .0012213   .0005449     2.24   0.027     .0001393    .0023033

          x3 |   .0063975   .0022727     2.81   0.006     .0018844    .0109107

       _cons |   50.76203   21.78882     2.33   0.022     7.493746    94.03031

-------------+----------------------------------------------------------------

y2           |

          y1 |   296.3783   59.13098     5.01   0.000     178.9559    413.8007

          x1 |   245.6199   145.6696     1.69   0.095    -43.65105    534.8909

          x2 |  -.0004649   .0003337    -1.39   0.167    -.0011275    .0001978

       _cons |  -35207.64   10026.98    -3.51   0.001    -55119.22   -15296.05

------------------------------------------------------------------------------

Endogenous variables:  y1 y2

Exogenous variables:   x1 x2 x3

------------------------------------------------------------------------------
```

and state reduced form and structural form of these simultaneous equation models. Specify endogenous variables and exogenous variables.

Reduced Form:

$Y_{2t} = \pi_{10} + \pi_{11}x_1 + \pi_{12}x_2 + \pi_{13}x_3 + \omega_{1t}$

$Y_{1t} = \pi_{20} + \pi_{21}x_1 + \pi_{22}x_2 + \pi_{23}x_3 + \omega_{2t}$

Structural Form:

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

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

Where y_{1t} and y_{2t} are endogenous variables and x_1, x_2, x_3 are exogenous variables.

Then, estimate reduced form models using OLS

Estimate $Y_{2t} = \pi_{10} + \pi_{11}x_1 + \pi_{12}x_2 + \pi_{13}x_3 + \omega_{1t}$

```
reg y2 x1 x2 x3

      Source |       SS           df       MS      Number of obs   =        50
-------------+----------------------------------   F(3, 46)        =     18.78
       Model |  6.8500e+09         3  2.2833e+09   Prob > F        =    0.0000
    Residual |  5.5922e+09        46   121570103   R-squared       =    0.5505
-------------+----------------------------------   Adj R-squared   =    0.5212
       Total |  1.2442e+10        49   253923367   Root MSE        =     11026

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   394.2228   147.9601     2.66   0.011     96.39459     692.051
          x2 |  -.0005688   .0003845    -1.48   0.146    -.0013428    .0002051
          x3 |   2.910982   .6767057     4.30   0.000     1.548844     4.27312
       _cons |  -31729.77   11092.16    -2.86   0.006    -54057.12   -9402.414
------------------------------------------------------------------------------

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

Estimate $Y_{1t} = \pi_{20} + \pi_{21}x_1 + \pi_{22}x_2 + \pi_{23}x_3 + \omega_{2t}$

```
. reg y1 x1 x2 x3

      Source |       SS           df       MS      Number of obs   =        50
-------------+----------------------------------   F(3, 46)        =     50.22
       Model |  50434.3503         3  16811.4501   Prob > F        =    0.0000
    Residual |  15398.5297        46  334.750647   R-squared       =    0.7661
-------------+----------------------------------   Adj R-squared   =    0.7508
       Total |    65832.88        49  1343.52816   Root MSE        =    18.296

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
          x1 |   .5013959   .2455228     2.04   0.047      .007184    .9956077
          x2 |  -3.51e-07   6.38e-07    -0.55   0.585    -1.64e-06    9.34e-07
          x3 |   .0098218   .0011229     8.75   0.000     .0075615    .0120822
       _cons |   11.73457   18.40616     0.64   0.527    -25.31511    48.78426
------------------------------------------------------------------------------

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

and structural form models using IV technique from the predicted endogenous variables from reduced form estimated results. (7 points)

```
. reg y1 y2hat x3

      Source |       SS           df       MS      Number of obs   =        50
-------------+----------------------------------   F(2, 47)        =     76.07
       Model |  50294.6657         2  25147.3328   Prob > F        =    0.0000
    Residual |  15538.2143        47  330.600305   R-squared       =    0.7640
-------------+----------------------------------   Adj R-squared   =    0.7539
       Total |    65832.88        49  1343.52816   Root MSE        =    18.182

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       y2hat |   .0012213    .000614     1.99   0.053     -.000014    .0024565
          x3 |   .0063975   .0025612     2.50   0.016     .0012451    .0115499
       _cons |   50.76202   24.55437     2.07   0.044        1.365     100.159
------------------------------------------------------------------------------

. reg y2 y1hat x1 x2

      Source |       SS           df       MS      Number of obs   =        50
-------------+----------------------------------   F(3, 46)        =     18.78
       Model |  6.8500e+09         3  2.2833e+09   Prob > F        =    0.0000
    Residual |  5.5922e+09        46   121570107   R-squared       =    0.5505
-------------+----------------------------------   Adj R-squared   =    0.5212
```

```
      Total |  1.2442e+10        49   253923367   Root MSE        =     11026

------------------------------------------------------------------------------
         y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      y1hat |   296.3783   68.89802     4.30   0.000     157.6938    435.0628
         x1 |   245.6199   169.7308     1.45   0.155    -96.03046    587.2703
         x2 |  -.0004649   .0003888    -1.20   0.238    -.0012475    .0003177
      _cons |  -35207.64   11683.19    -3.01   0.004    -58724.69   -11690.59
------------------------------------------------------------------------------
```

c. Estimate model (1) and (2) using Three Stage Least Squares (3SLS) and give the explanation of the differences among the three estimation methods (OLS, 2SLS, and 3SLS) (conceptually). Point out the advantage and disadvantage in term of properties of estimated results from each method (single equation estimation vs system equations estimation methods). (8 points)

```
. 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    15.64302    0.8141     205.53   0.0000
y2                    50       3    9103.966    0.6669      83.67   0.0000
----------------------------------------------------------------------


------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y1           |
          y2 |   .0012213   .0005283     2.31   0.021     .0001859    .0022567
          x3 |   .0063975   .0022035     2.90   0.004     .0020788    .0107163
       _cons |   50.76203   21.12505     2.40   0.016     9.357698    92.16636
-------------+----------------------------------------------------------------
y2           |
          y1 |   291.0428   56.27507     5.17   0.000     180.7457    401.3399
```

```
          x1 |    256.281    139.007     1.84   0.065    -16.16773    528.7298

          x2 |   -.000329   .0002647    -1.24   0.214    -.0008478    .0001899

       _cons |  -35255.72   9617.326    -3.67   0.000    -54105.33    -16406.1
------------------------------------------------------------------------------

Endogenous variables:  y1 y2

Exogenous variables:   x1 x2 x3
------------------------------------------------------------------------------
```

For the simultaneous equations, OLS estimation method will be biased and inconsistent. Instead, if 2SLS is being used, then the estimates will be consistent. Next, 3SLS is a system equation estimation methods is better than 2SLS in term of being more asymptotically efficient. However, if there exists the specification error in just only one equation in the system, the problem will spread through all equations in the system. Meanwhile, the single equation estimation methods (OLS, 2SLS) the problem of specification error will not spread through because they are not the system, but in return, the estimates will be less asymptotically efficient compared with 3SLS.

2. From the data set "`Final_q2_no.dta`":

The study of cost function assumes the model follows CES cost function:

$$c = \lambda \left[\theta R^{-\alpha} + (1-\theta) W^{-\alpha} \right]^{-\frac{\beta}{\alpha}} + \varepsilon \qquad (3)$$

where:

c = Total cost
R = Capital cost
W = Labor cost
λ = Efficiency Parameter
θ = Distribution Parameter
β = Parameter
α = Substitution Parameter
ε = Disturbance Term

Elasticity of Substitution can be computed from: $\sigma = \dfrac{1}{1-\theta}$

The model can then be transformed as:

$$\ln c = \ln \lambda - \frac{\beta}{\alpha} \ln \left[\theta R^{-\alpha} + (1-\theta) W^{-\alpha} \right] + \varepsilon \qquad (4)$$

a. Estimate the model (4) using NLS estimation method using initial values of $\ln\lambda$=1, θ=0.5, β=0.5, α= –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 θ=0, α=0, and β=0. (6 points)

```
 . nl (lnC = ({lnramda})-({beta}/{alpha})*ln(({theta}*R^(-{alpha}))+(1-{theta})*W^(-{alpha}))),
init(lnramda 1 theta 0.5 beta 0.5 alpha -0.5) nolog

(obs = 252)



      Source |       SS              df       MS

-------------+----------------------------------    Number of obs =        252

       Model |  45.117021           3   15.039007    R-squared     =     0.1382

    Residual |  281.28663         248  1.13422026    Adj R-squared =     0.1278

-------------+----------------------------------    Root MSE      =   1.064998

       Total |  326.40365         251  1.30041293    Res. dev.     =   742.8512


------------------------------------------------------------------------------

         lnC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]

-------------+----------------------------------------------------------------
```

```
    /lnramda |   1.996386   .7811621     2.56   0.011     .4578285    3.534944
       /beta |   .8545114    .147163     5.81   0.000     .5646628     1.14436
      /alpha |   -.931979   .5400816    -1.73   0.086    -1.995711    .1317527
      /theta |   .2102829   .1833049     1.15   0.252    -.1507501    .5713159
------------------------------------------------------------------------------
  Parameter lnramda taken as constant term in model & ANOVA table

. est store lnnls

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

. sca list sigmaλ
     sigma =  1.2662762

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

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

       F(  3,   248) =   36.86
            Prob > F =    0.0000

. sca ramda = exp((_b[/lnramda]))

. sca list ramda
     ramda =  7.3624037
```

According to the estimated result provided above, the estimated value of λ=7.36, θ=0.21, β=0.85, α=-0.93, and σ=1.27. From the F-test, the null hypothesis that θ=α=β=0 is rejected.

b. What will happy if we change initial values to $\ln\lambda=0.5$, $\theta=0.1$, $\beta=0.1$, $\alpha=-0.1$? 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)

```
. nl (lnC = ({lnramda})-({beta}/{alpha})*ln(({theta}*R^(-{alpha}))+(1-{theta})*W^(-{alpha}))),
init(lnramda 0.5 theta 0.1 beta 0.1 alpha -0.1) nolog

(obs = 252)


      Source |      SS            df       MS

-------------+----------------------------------    Number of obs =        252

       Model |  45.117021          3   15.039007    R-squared     =     0.1382

    Residual |  281.28663        248  1.13422026    Adj R-squared =     0.1278

-------------+----------------------------------    Root MSE      =   1.064998

       Total |  326.40365        251  1.30041293    Res. dev.     =   742.8512


------------------------------------------------------------------------------

         lnC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]

-------------+----------------------------------------------------------------

    /lnramda |   1.996386   .7811621     2.56   0.011     .4578286    3.534944

       /beta |   .8545114    .147163     5.81   0.000     .5646628     1.14436

      /alpha |   -.931979   .5400817    -1.73   0.086    -1.995711    .1317529

      /theta |   .2102829   .1833049     1.15   0.252      -.15075    .5713158

------------------------------------------------------------------------------

  Parameter lnramda taken as constant term in model & ANOVA table


. est store lnnls2


. est table lnnls lnnls2, star(.1 .05 .01) stat(N rss r2 r2_a)


----------------------------------------------

    Variable |     lnnls          lnnls2

-------------+--------------------------------

lnramda      |

       _cons |  1.9963865**     1.9963865**

-------------+--------------------------------

beta         |

       _cons |  .85451141***    .85451141***
```

```
-------------+--------------------------------
alpha        |
       _cons |   -.931979*     -.93197896*
-------------+--------------------------------
theta        |
       _cons |   .2102829       .21028292
-------------+--------------------------------
Statistics   |
           N |        252             252
         rss |  281.28663       281.28663
          r2 |  .13822462       .13822462
        r2_a |  .12779992       .12779992
----------------------------------------------
           legend: * p<.1; ** p<.05; *** p<.01
```

According to the above estimated result compared with the result in (a) shown in the table, the estimated coefficient, r2, r2_a are indifferent, because our models are in the log-form. The log-form gives the robust estimated results.

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)

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

(obs = 252)


Iteration 0:  residual SS =  4418.289

Iteration 1:  residual SS =  3876.752

. . .

Iteration 11:  residual SS =  281.2873



      Source |      SS            df       MS

-------------+----------------------------------    Number of obs =        252

       Model |  45.116337          3   15.038779    R-squared     =     0.1382
```

```
    Residual |  281.28731        248  1.13422302    Adj R-squared =     0.1278

-------------+----------------------------------    Root MSE      =   1.064999

       Total |  326.40365        251  1.30041293    Res. dev.     =   742.8518


------------------------------------------------------------------------------

         lnC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]

-------------+----------------------------------------------------------------

    /lnramda |   1.996213   .7820293     2.55   0.011     .4559475    3.536479

       /beta |   .8550992   .1471821     5.81   0.000     .5652128    1.144986

      /alpha |  -.9333344   .5161881    -1.81   0.072    -1.950006    .0833371

      /theta |    .208476   .1838451     1.13   0.258    -.1536208    .5705728

------------------------------------------------------------------------------

  Parameter beta taken as constant term in model & ANOVA table


. est store lnnls3


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

(obs = 252)


Iteration 0:  residual SS =  4418.289

Iteration 1:  residual SS =  3876.752

. . .

Iteration 39:  residual SS =  281.2866



      Source |      SS            df       MS

-------------+----------------------------------    Number of obs =        252

       Model |  11625.479          4  2906.36982    R-squared     =     0.9764

    Residual |  281.28663        248  1.13422026    Adj R-squared =     0.9760

-------------+----------------------------------    Root MSE      =   1.064998

       Total |  11906.766        252  47.2490711    Res. dev.     =   742.8512


------------------------------------------------------------------------------

         lnC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]

-------------+----------------------------------------------------------------

    /lnramda |   1.996386   .7811622     2.56   0.011     .4578281    3.534944
```

```
       /beta |   .8545116    .147163     5.81   0.000      .564663     1.14436
      /alpha |  -.9319802   .5400797    -1.73   0.086    -1.995708    .1317477
      /theta |   .2102824   .1833053     1.15   0.252    -.1507512    .5713161
------------------------------------------------------------------------------
convergence not achieved
r(430);


. est table lnnls2 lnnls3, star(.1 .05 .01) stat(N rss r2 r2_a)
----------------------------------------------
    Variable |    lnnls2          lnnls3
-------------+--------------------------------
lnramda      |
       _cons |  1.9963865**     1.9962133**
-------------+--------------------------------
beta         |
       _cons |  .85451141***    .85509918***
-------------+--------------------------------
alpha        |
       _cons | -.93197896*     -.93333442*
-------------+--------------------------------
theta        |
       _cons |  .21028292       .20847598
-------------+--------------------------------
Statistics   |
           N |       252             252
         rss |  281.28663       281.28731
          r2 |  .13822462       .13822253
        r2_a |  .12779992        .1277978
----------------------------------------------
           legend: * p<.1; ** p<.05; *** p<.01
```

According to the estimated results above, the result when the convergence value is changed to 0.1 compared with the result in (b) which has the convergence value equal to 0.00001 are slightly different. This is because the convergence value has changed (to be more precise, the convergence value now is too high). The smaller convergence value should be better; however, the result can be in the case like changing the convergence value to 1e-15 with maximum iteration equal to 40 times where the convergence value is not achieved (it might require more iterations to reach the given convergence value.

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

Because the log-form gives the robust estimated result. From (a),(b), and(c), even the initial values are changed, the estimated results are indifferent. Or changing the convergence value the result might be only slightly different.

3. From the data set "`Midterm_q3_no.dta`": (using do-file from assignment 4)

Index function for the decision model can be stated as.

$$I_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \tag{5}$$

The log-likelihood function of this model is as follows:

$$\ln L = \begin{cases} \ln \Phi(z_i) \; if \;\; y_i = 1 \\ \ln \Phi(-z_i) \; if \;\; y_i = 0 \end{cases} \tag{6}$$

where: y_i = 1 for yes and 0 for no.

x_{1i} = independent variable 1 of individual *i*

x_{2i} = independent variable 2 of individual *i*

Let $\Phi(.)$ = Logistic probability distribution function. $\Phi(z_i) = \dfrac{1}{1+e^{-z_i}}$

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)

```
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


.

end of do-file


. ml model lf ml_logit (y=x1 x2)


. ml maximize


initial:       log likelihood = -277.25887

alternative:   log likelihood = -279.13079

rescale:       log likelihood = -275.12577

Iteration 0:   log likelihood = -275.12577

Iteration 1:   log likelihood = -227.96269

Iteration 2:   log likelihood = -227.67192

Iteration 3:   log likelihood = -227.67158
```

```
Iteration 4:   log likelihood = -227.67158

                                                Number of obs     =        400
                                                Wald chi2(2)      =      66.43
Log likelihood = -227.67158                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .2508771   .1098901     2.28   0.022     .0354965    .4662577
          x2 |  -.5626563   .0706508    -7.96   0.000    -.7011294   -.4241832
       _cons |   .6249761   .1975087     3.16   0.002     .2378661    1.012086
------------------------------------------------------------------------------

. est store newton

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

. ml maximize

initial:       log likelihood = -277.25887
alternative:   log likelihood = -279.13079
rescale:       log likelihood = -275.12577
Iteration 0:   log likelihood = -275.12577
Iteration 1:   log likelihood = -227.90673
Iteration 2:   log likelihood = -227.67573
Iteration 3:   log likelihood = -227.67164
Iteration 4:   log likelihood = -227.67159
Iteration 5:   log likelihood = -227.67158

                                                Number of obs     =        400
                                                Wald chi2(2)      =      81.09
Log likelihood = -227.67158                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
```

```
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .2508881   .1109659     2.26   0.024      .033399    .4683772
          x2 |  -.5626591   .0646113    -8.71   0.000    -.6892949   -.4360233
       _cons |   .6249569   .2014597     3.10   0.002     .2301032    1.019811
------------------------------------------------------------------------------

. est store bhhh

. help tech

. help ml

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

. ml maximize

initial:       log likelihood = -277.25887
alternative:   log likelihood = -279.13079
rescale:       log likelihood = -275.12577
Iteration 0:   log likelihood = -275.12577
Iteration 1:   log likelihood = -252.09628  (backed up)
Iteration 2:   log likelihood = -233.07538  (backed up)
Iteration 3:   log likelihood = -229.55938
Iteration 4:   log likelihood = -227.71884
Iteration 5:   log likelihood = -227.67316
Iteration 6:   log likelihood =  -227.6716
Iteration 7:   log likelihood = -227.67158

                                                  Number of obs     =        400
                                                  Wald chi2(2)      =      66.43
Log likelihood = -227.67158                       Prob > chi2       =     0.0000

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

```
          x1 |   .2508827   .1098903     2.28   0.022     .0355017    .4662636
          x2 |  -.5626583    .070651    -7.96   0.000    -.7011317   -.4241849
       _cons |   .6249707   .1975088     3.16   0.002     .2378605    1.012081
------------------------------------------------------------------------------
```

```
. est store bfgs
```

```
. est table newton bhhh bfgs, star(0.1 0.05 0.01) stat(N ll chi2 p)
```

```
--------------------------------------------------------------
    Variable |    newton           bhhh            bfgs
-------------+------------------------------------------------
          x1 |  .25087707**     .25088813**     .25088266**
          x2 | -.56265628***   -.56265912***   -.56265833***
       _cons |  .62497609***    .62495691***    .62497069***
-------------+------------------------------------------------
           N |        400             400             400
          ll | -227.67158      -227.67158      -227.67158
        chi2 |  66.430535        81.09243       66.430831
           p |  3.757e-15       2.460e-18       3.756e-15
--------------------------------------------------------------
                           legend: * p<.1; ** p<.05; *** p<.01
```

According to the estimated result using different algorithm provided in the table, the estimated results are slightly different because the term to be converged (Δ_t) varies with different technique.

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)**

```
. est restore newton
(results newton are active now)
```

Wald Test

```
. test x1 x2
```

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

           chi2(  2) =   66.43
         Prob > chi2 =    0.0000
```

<u>LR-test</u>

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

. ml maximize

initial:       log likelihood = -277.25887
alternative:   log likelihood = -279.13079
rescale:       log likelihood = -275.12577
Iteration 0:   log likelihood = -275.12577
Iteration 1:   log likelihood =  -275.0498
Iteration 2:   log likelihood =  -275.0498

                                                Number of obs     =        400
                                                Wald chi2(0)      =          .
Log likelihood =  -275.0498                     Prob > chi2       =          .

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -.2107769   .1005559    -2.10   0.036    -.4078627   -.0136911
------------------------------------------------------------------------------

. est store res

. lrtest newton res

Likelihood-ratio test                                 LR chi2(2)  =     94.76
(Assumption: res nested in newton)                    Prob > chi2 =    0.0000
```

According to the Wald test, the null hypothesis that $\beta_1 = \beta_2 = 0$ is rejected. Also, the LR-test, LR-statistic = 2(-227.67-(-275.05)) = 94.76 which also rejects the null hypothesis as well.

From my point of view, LR-test should be employed because it uses both unrestricted model and restricted model, while Wald test uses only unrestricted model.

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)

F-test as overall test cannot be used because MLE does not provide the sum of square terms (SSE, SST, SSR). It maximizes the log-likelihood function instead of minimizing sum of square residuals. So, we use LR test; LR = 2($\log L_{ur}$ – $\log L_r$). It is Chi-square because we use the asymptotic property in MLE.

Let $\Phi(.)$ = Cumulation standard normal probability distribution function and

$$z_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \tag{7}$$

Assume that there exists heteroskedasticity in the model as: $\sigma_i^2 = \exp(\gamma x_{3i})^2$, then,

$\Phi(.)$ = Cumulation standard normal probability distribution function $\Phi \;\; z_i / \exp(\gamma x_{3i})$

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

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

```
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


end of do-file


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


. ml maximize


initial:       log likelihood = -277.25887
```

```
alternative:   log likelihood =  -291.8231

rescale:       log likelihood = -275.03884

rescale eq:    log likelihood = -271.46187

Iteration 0:   log likelihood = -271.46187  (not concave)

Iteration 1:   log likelihood = -259.02903  (not concave)

Iteration 2:   log likelihood = -254.52045

Iteration 3:   log likelihood = -234.56833

Iteration 4:   log likelihood = -229.78224

Iteration 5:   log likelihood = -228.69789

Iteration 6:   log likelihood = -225.67317

Iteration 7:   log likelihood = -225.58267

Iteration 8:   log likelihood = -225.58236

Iteration 9:   log likelihood = -225.58236


                                                Number of obs     =        400

                                                Wald chi2(2)      =      64.18

Log likelihood = -225.58236                     Prob > chi2       =     0.0000


------------------------------------------------------------------------------

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

-------------+----------------------------------------------------------------

eq1          |

          x1 |   .1580213   .0626383     2.52   0.012     .0352526    .2807901

          x2 |  -.3231145   .0414919    -7.79   0.000    -.4044373   -.2417918

       _cons |   .3665324   .1132645     3.24   0.001     .1445381    .5885268

-------------+----------------------------------------------------------------

eq2          |

          x3 |   .3778106   .1707856     2.21   0.027      .043077    .7125442

------------------------------------------------------------------------------


. est store hetprob


. do "C:\Users\Chaiyapong M\Desktop\BE\EE426\Midterm\Midterm_q3 - ml_probit.do"


. 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

end of do-file

. ml model lf ml_probit (y=x1 x2)

. ml maximize

initial:       log likelihood = -277.25887
alternative:   log likelihood = -292.02536
rescale:       log likelihood = -275.05597
Iteration 0:   log likelihood = -275.05597
Iteration 1:   log likelihood = -228.81014
Iteration 2:   log likelihood = -228.59561
Iteration 3:   log likelihood = -228.59552
Iteration 4:   log likelihood = -228.59552

                                                Number of obs     =        400
                                                Wald chi2(2)      =      77.26
Log likelihood = -228.59552                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .1428731   .0655991     2.18   0.029     .0143013     .271445
          x2 |  -.3256139   .0381417    -8.54   0.000    -.4003703   -.2508575
       _cons |   .3642278   .1170545     3.11   0.002     .1348053    .5936503
------------------------------------------------------------------------------

. est store probit

. lrtest hetprob probit
```

```
Likelihood-ratio test                                 LR chi2(1)  =      6.03
(Assumption: probit nested in hetprob)                Prob > chi2 =    0.0141


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


----------------------------------------------
    Variable |    probit          hetprob
-------------+--------------------------------
eq1          |
          x1 |  .14287314**     .15802132**
          x2 | -.32561388***   -.32311454***
       _cons |  .36422782***    .36653245***
-------------+--------------------------------
eq2          |
          x3 |                  .37781057**
-------------+--------------------------------
Statistics   |
           N |       400             400
          ll | -228.59552      -225.58236
        chi2 |  77.255828       64.182443
           p |  1.675e-17       1.156e-14
----------------------------------------------
           legend: * p<.1; ** p<.05; *** p<.01
```

Since p-value of the LR-test < 0.05, the null hypothesis of no heteroscedasticity is rejected. So, there exists significant heteroscedasticity problem. In this case, we cannot use neither Wald-test nor LM-test to test heteroscedasticity because we require to use 2 models (having heteroscedasticity and not having heteroscedasticity) which are unrestricted model and restricted model. So, we have to use LR-test.

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)

Because in MLE, distribution of the estimated parameters depends on the distribution model (not necessary to be t-distributed) (distribution of the parameters in MLE is flexible). So, the individual test can be performed by using Z-test acquiring asymptotic normal property. We do not have R-squared because MLE is derived from maximizing the likelihood function, not by minimizing sum of square residuals like the OLS. Then, if we want to compare the estimated results, we look at the log-likelihood value instead. But, this number is to compare only (more is better). It cannot give the interpretation like R-squared.

4. From the data set "`Midterm_q4-1_no.dta`":

The model of interest rate structure, continuous time model can be specified as:

$$r_{t+\Delta t} - r_t = (\alpha + \beta r_t)\Delta t + \varepsilon_{t+\Delta t} \qquad (8)$$

where: $E[\varepsilon_{t+\Delta t}] = 0$ and $E[\varepsilon^2_{t+\Delta t}] = \sigma^2 r_t^{2\gamma} \Delta t$

Then, the model can be transformed to be discrete time model by setting $\Delta t = 1$. The discrete time model can be stated as:

$$r_{t+1} - r_t = \Delta r_t = \alpha + \beta r_t + \varepsilon_{t+1} \qquad (9)$$

where: $E[\varepsilon_{t+1}] = 0$ and $E[\varepsilon^2_{t+1}] = \sigma^2 r_t^{2\gamma}$

The model consists of four parameters, including $\alpha, \beta, \sigma^2, \gamma$. Four moment condition equations of the model can be stated as:

(1) Zero mean condition: $E(\varepsilon_{t+1}) = 0$

(2) Orthogonality condition: $E(\varepsilon_{t+1} r_t) = 0$

(3) Variance condition: $E(\varepsilon^2_{t+1} - \sigma^2 r_t^{2\gamma}) = 0$

(4) Zero covariance condition: $E\left(\left(\varepsilon^2_{t+1} - \sigma^2 r_t^{2\gamma}\right) r_t\right) = 0$

The above model can be claimed as unrestricted model for other two interest rate structure models which can be indicated as follows:

Model	α	β	σ^2	γ
(1) Unrestricted				
(2) Merton		0		0
(3) Vasicek				0

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).

```
tsset time

        time variable:  time, 1 to 1326

                delta:  1 unit

. gen dr=f.r-r

(1 missing value generated)
```

Unrestricted Model

```
. gmm (dr-{alpha}-{beta}*r) ((dr-{alpha}-{beta}*r)*r) ((dr-{alpha}-{beta}*r)^2-
{sigma2}*r^(2*{gamma})) (((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) =  8.722e-06  (backed up)
Iteration 2:   GMM criterion Q(b) =  6.129e-06  (not concave)
Iteration 3:   GMM criterion Q(b) =  5.573e-06  (backed up)
Iteration 4:   GMM criterion Q(b) =  5.489e-06  (backed up)

Step 2
Iteration 0:   GMM criterion Q(b) =   .0076731
Iteration 1:   GMM criterion Q(b) =   .0072924
Iteration 2:   GMM criterion Q(b) =  .00453733
Iteration 3:   GMM criterion Q(b) =  .00151095
Iteration 4:   GMM criterion Q(b) =   .0006721
Iteration 5:   GMM criterion Q(b) =  6.065e-07
Iteration 6:   GMM criterion Q(b) =  2.058e-13
Iteration 7:   GMM criterion Q(b) =  1.852e-24

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

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /alpha |  -.0023917   .0011635    -2.06   0.040    -.0046722   -.0001112
```

```
       /beta |   .0004321   .0002872     1.50   0.132    -.0001309    .0009951
     /sigma2 |   .0005133   .0003281     1.56   0.118    -.0001299    .0011564
      /gamma |   .0942137   .1807945     0.52   0.602    -.2601371    .4485645
------------------------------------------------------------------------------
Instruments for equation 1: _cons
Instruments for equation 2: _cons
Instruments for equation 3: _cons
Instruments for equation 4: _cons

. estimates store unres
```

<u>Merton</u>

```
. gmm (dr-{alpha}) ((dr-{alpha})*r) ((dr-{alpha})^2-{sigma2}) (((dr-{alpha})^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) =  4.144e-08
Iteration 2:   GMM criterion Q(b) =  4.144e-08

Step 2
Iteration 0:   GMM criterion Q(b) =  .00793798
Iteration 1:   GMM criterion Q(b) =  .00554765
Iteration 2:   GMM criterion Q(b) =  .00554765

GMM estimation

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

------------------------------------------------------------------------------
```

```
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /alpha |  -.0008227   .0006877    -1.20   0.232    -.0021706    .0005252
     /sigma2 |   .0004354   .0002932     1.49   0.137    -.0001392      .00101
------------------------------------------------------------------------------
Instruments for equation 1: _cons
Instruments for equation 2: _cons
Instruments for equation 3: _cons
Instruments for equation 4: _cons
. estat overid


  Test of overidentifying restriction:


  Hansen's J chi2(2) = 7.35063 (p = 0.0253)


. estimates store merton
```

Vasicek

```
. gmm (dr-{alpha}-{beta}*r) ((dr-{alpha}-{beta}*r)*r) ((dr-{alpha}-{beta}*r)^2-{sigma2}) (((dr-
{alpha}-{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) =  3.145e-10
Iteration 2:   GMM criterion Q(b) =  3.079e-10


Step 2
Iteration 0:   GMM criterion Q(b) =   .0005879
Iteration 1:   GMM criterion Q(b) =  .00019279
Iteration 2:   GMM criterion Q(b) =  .00019279


GMM estimation
```

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

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /alpha |  -.0027101   .0009776    -2.77   0.006    -.0046262    -.000794
       /beta |   .0005363   .0001997     2.69   0.007     .0001449    .0009277
     /sigma2 |   .0005959   .0003005     1.98   0.047     6.90e-06     .001185
------------------------------------------------------------------------------
Instruments for equation 1: _cons
Instruments for equation 2: _cons
Instruments for equation 3: _cons
Instruments for equation 4: _cons

. estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(1) = .255441 (p = 0.6133)

. estimates store vas


. est table unres merton vas, star(.1 .05 .01) stat(N J)

--------------------------------------------------------------
    Variable |     unres          merton           vas
-------------+------------------------------------------------
alpha        |
       _cons | -.00239168**     -.0008227      -.00271012***
-------------+------------------------------------------------
```

```
beta         |
       _cons |  .00043212                        .00053631***
-------------+---------------------------------------------
sigma2       |
       _cons |  .00051327       .00043542       .00059595**
-------------+---------------------------------------------
gamma        |
       _cons |   .0942137
-------------+---------------------------------------------
Statistics   |
           N |       1325            1325            1325
           J |  2.454e-21       7.3506315       .25544087
-----------------------------------------------------------
                           legend: * p<.1; ** p<.05; *** p<.01
```

```
The Merton model should not be employed because, from the overidentifying test, it
reveals that some of the moment conditions are not true (rejecting the null
hypothesis).
```

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

```
. est restore unres
(results unres are active now)
```

Test Unrestricted VS Merton

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

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

           chi2(  2) =    7.89
         Prob > chi2 =    0.0194
```

Test Unrestricted VS Vasicek

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

 ( 1)  [gamma]_cons = 0

           chi2(  1) =    0.27
         Prob > chi2 =    0.6023
```

According to the Wald-test, Vasicek is the most appropriated one because the p-value is very large. It indicates that the null hypothesis of γ = 0 is failed to be rejected.

From the data set "`Midterm_q4-2_no.dta`":

The model: $$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i \qquad (10)$$

where: $u_i \sim N(0,\sigma^2)$, $E(x_{3i}u_i)=0$ and $E(x_{4i}u_i)=0$ but $E(x_{1i}u_i)\neq 0$ and $E(x_{2i}u_i)\neq 0$

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

If we run the model by using OLS, the estimated results will be bias and inconsistent, since the independent variables (x_1 and x_2) are correlated with the error term.

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).

From the given information, z_1, z_2, z_3 should be used as the instrumental variables.

```
. ivregress gmm y x3 x4 (x1 x2 = z1 z2 z3)


Instrumental variables (GMM) regression           Number of obs   =        500
                                                  Wald chi2(4)    =     175.43
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.6174
```

```
GMM weight matrix: Robust                         Root MSE        =      19.85

------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   3.461043   .9124621     3.79   0.000      1.67265    5.249436
          x2 |   2.007654   .4974433     4.04   0.000     1.032684    2.982625
          x3 |   1.037275   .3047988     3.40   0.001     .4398801     1.63467
          x4 |   1.336469   .1726284     7.74   0.000     .9981233    1.674814
       _cons |    .564682   7.290525     0.08   0.938    -13.72448    14.85385
------------------------------------------------------------------------------
Instrumented:  x1 x2
Instruments:   x3 x4 z1 z2 z3

. estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(1) = .018463 (p = 0.8919)

. estat endogenous x1 x2

  Test of endogeneity (orthogonality conditions)
  Ho: variables are exogenous

  GMM C statistic chi2(2) =  174.972  (p = 0.0000)

. est store gmm

. ivregress 2sls y x3 x4 (x1 x2 = z1 z2 z3)

Instrumental variables (2SLS) regression          Number of obs   =        500
                                                  Wald chi2(4)    =     161.74
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.6173
```

```
                                                   Root MSE        =     19.851

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   3.462966   .9026394     3.84   0.000     1.693826    5.232107
          x2 |   2.005854   .5008566     4.00   0.000     1.024193    2.987515
          x3 |   1.035421   .3052038     3.39   0.001     .4372322    1.633609
          x4 |   1.337074    .176902     7.56   0.000     .9903522    1.683795
       _cons |   .5920357   7.907207     0.07   0.940    -14.90581    16.08988
------------------------------------------------------------------------------
Instrumented:  x1 x2
Instruments:   x3 x4 z1 z2 z3

. est store twostage
```

From the overidentifying test, it suggests that all moment conditions are true (fail to reject the null hypothesis). Thus, GMM is appropriated.

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

```
. est table gmm twostage, star(.1 .05 .01) stat(N rss chi2 r2 r2_a J)

----------------------------------------------
    Variable |      gmm          twostage
-------------+--------------------------------
          x1 |   3.461043***    3.4629664***
          x2 |  2.0076544***    2.0058537***
          x3 |  1.0372749***    1.0354206***
          x4 |  1.3364688***    1.3370738***
       _cons |  .56468196       .59203565
-------------+--------------------------------
           N |        500             500
         rss |  197002.88       197037.95
        chi2 |  175.42866       161.73731
          r2 |  .61739875       .61733066
        r2_a |  .61430703       .61423838
           J |  .01846258
```

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

According to the results above, it is clearly that the estimated results are slightly different. This is because the 2SLS technique has only 5 moment conditions, while GMM has 6 moment conditions in this case.

Generalized Method of Moment

→ ivregress $\boxed{\text{gmm}}$ $y\ x_3\ x_4\ (x_1\ x_2 = z_1\ z_2\ z_3)$

Endogeneity Problem.

Moment Conditions

1. $\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 \underline{x_{1i}} - \beta_2 \underline{x_{2i}} - \beta_3 x_{3i} - \beta_4 x_{4i}) = 0$
2. $\frac{1}{n}\sum_{i=1}^{n}x_{1i}(y_i - \beta_0 - \beta_1 \underline{x_{1i}} - \beta_2 \underline{x_{2i}} - \beta_3 x_{3i} - \beta_4 x_{4i}) = 0$
3. $\frac{1}{n}\sum_{i=1}^{n}x_{2i}(y_i - \beta_0 - \beta_1 \underline{x_{1i}} - \beta_2 \underline{x_{2i}} - \beta_3 x_{3i} - \beta_4 x_{4i}) = 0$
4. $\frac{1}{n}\sum_{i=1}^{n}z_{1i}(y_i - \beta_0 - \beta_1 \underline{x_{1i}} - \beta_2 \underline{x_{2i}} - \beta_3 x_{3i} - \beta_4 x_{4i}) = 0$
5. $\frac{1}{n}\sum_{i=1}^{n}z_{2i}(y_i - \beta_0 - \beta_1 \underline{x_{1i}} - \beta_2 \underline{x_{2i}} - \beta_3 x_{3i} - \beta_4 x_{4i}) = 0$
6. $\frac{1}{n}\sum_{i=1}^{n}z_{3i}(y_i - \beta_0 - \beta_1 \underline{x_{1i}} - \beta_2 \underline{x_{2i}} - \beta_3 x_{3i} - \beta_4 x_{4i}) = 0$

2-stage least square

→ ivregress $\boxed{\text{2sls}}$ $y\ x_3\ x_4\ (x_1\ x_2 = z_1\ z_2\ z_3)$

1st Reduced Form

$$\left.\begin{aligned} x_1 &= \alpha_{10} + \alpha_{11} z_1 + \alpha_{12} z_2 + \alpha_{13} z_3 + w_{1i} \\ x_2 &= \alpha_{20} + \alpha_{21} z_1 + \alpha_{22} z_2 + \alpha_{23} z_3 + w_{2i} \end{aligned}\right\} \hat{x}_1,\ \hat{x}_2$$

2nd Structural Form

$$y = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2 + \beta_3 x_3 + \beta_4 x_3 + u_i$$

1. $E(u_i) = 0$: Zero-Mean Condition

   $\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 \hat{x}_{1i} - \beta_2 \hat{x}_{2i} - \beta_3 x_{3i} - \beta_4 x_{4i}) = 0$

2. $E(\hat{x}_{1i} u_i) = 0$: Exogeneity Condition

   $\frac{1}{n}\sum_{i=1}^{n}\hat{x}_{1i}(y_i - \beta_0 - \beta_1 \hat{x}_{1i} - \beta_2 \hat{x}_{2i} - \beta_3 x_{3i} - \beta_4 x_{4i}) = 0$

3. $E(\hat{x}_{2i} u_i) = 0$
4. $E(x_{3i} u_i) = 0$
5. $E(x_{4i} u_1) = 0$

5. From the data set "`Midterm_q5_no.dta`":
In the study of decision to pay dividend of listed company in the stock market, the study states the following probability function:

$$I_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} \tag{11}$$

The log-likelihood function of this model is as follows:

$$\ln L = \begin{cases} \ln \Lambda(I_i) \; if \;\; y_i = 1 \\ \ln \Lambda(-I_i) \; if \;\; y_i = 0 \end{cases} \tag{12}$$

where: y_i = 1 for dividend paying firm and 0 otherwise.

x_{1i} = Liquidity ratio of firm *i*

x_{2i} = log of firm size firm *i*

x_{3i} = Profitability index of firm *i*

(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)

```
. probit y x1 x2 x3

Iteration 0:   log likelihood = -124.34324
Iteration 1:   log likelihood = -113.56596
. . .
Iteration 4:   log likelihood = -113.39588

Probit regression                               Number of obs     =        215
                                                LR chi2(3)        =      21.89
                                                Prob > chi2       =     0.0001
Log likelihood = -113.39588                     Pseudo R2         =     0.0880

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0045192   .0014459     3.13   0.002     .0016853    .0073531
          x2 |   .0103081    .008238     1.25   0.211    -.0058381    .0264544
          x3 |   .1124807   .0428669     2.62   0.009     .0284631    .1964983
       _cons |    .270818   .1195302     2.27   0.023     .0365431     .505093
```

```
------------------------------------------------------------------------------

. est store probit

. logit y x1 x2 x3

Iteration 0:   log likelihood = -124.34324
Iteration 1:   log likelihood = -114.01368
Iteration 2:   log likelihood = -113.44921
Iteration 3:   log likelihood = -113.44715
Iteration 4:   log likelihood = -113.44715

Logistic regression                               Number of obs     =        215
                                                  LR chi2(3)        =      21.79
                                                  Prob > chi2       =     0.0001
Log likelihood = -113.44715                       Pseudo R2         =     0.0876

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0077872   .0026035     2.99   0.003     .0026845      .01289
          x2 |   .0196843   .0155294     1.27   0.205    -.0107528    .0501213
          x3 |   .2020492   .0803094     2.52   0.012     .0446457    .3594527
       _cons |    .428184   .1940002     2.21   0.027     .0479506    .8084174
------------------------------------------------------------------------------

. fitstat

Measures of Fit for logit of y

Log-Lik Intercept Only:     -124.343     Log-Lik Full Model:         -113.447
D(211):                      226.894     LR(3):                        21.792
                                         Prob > LR:                     0.000
McFadden's R2:                 0.088     McFadden's Adj R2:             0.055
Maximum Likelihood R2:         0.096     Cragg & Uhler's R2:            0.141
McKelvey and Zavoina's R2:     0.191     Efron's R2:                    0.089
```

```
Variance of y*:                 4.068     Variance of error:              3.290

Count R2:                       0.735     Adj Count R2:                   0.000

AIC:                            1.093     AIC*n:                        234.894

BIC:                         -906.310     BIC':                          -5.680


. est store logit
```

We cannot compare the estimated results of Probit and Logit models because they assume different probability distribution. About the Logit model, p-value of the overall test is 0.0001 < 0.005 meaning that all regressors are statistically jointly significant. About the individual test using Z-test, it suggests that x_2 is insignificant while others are significant. Pseudo-R^2 = 0.088 (quite low), and Counted R^2 = 0.735

(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)

```
. mfx

Marginal effects after logit

      y  = Pr(y) (predict)

         =  .76842297

------------------------------------------------------------------------------

variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X

---------+--------------------------------------------------------------------

      x1 |   .0013857      .00044    3.16   0.002    .000527  .002244   60.8578

      x2 |   .0035028      .00271    1.29   0.196  -.001807  .008813   -1.6447

      x3 |   .0359545      .01357    2.65   0.008    .009359   .06255   1.63181


. sum x*


    Variable |        Obs        Mean    Std. Dev.       Min        Max

-------------+---------------------------------------------------------

          x1 |        215    60.85779    79.66036  -75.47762   332.6589

          x2 |        215   -1.644697     13.8192  -54.81482    40.2438

          x3 |        215    1.631809     2.87174  -3.865531   13.81453
```

The predicted probability of being a dividend paying firm (Phat(y=1)|x) will increase by 0.0014 (0.14%) if x_1 changes by 1 unit from its mean (from 60.8575 to 61.8578), holding other independent variables at their means.

(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)

```
. test (x1=0) (x2=0) (x3=0)


 ( 1)  [y]x1 = 0

 ( 2)  [y]x2 = 0

 ( 3)  [y]x3 = 0


           chi2(  3) =   16.37

         Prob > chi2 =    0.0010
```

According to the Wald-test, the null hypothesis of $\beta_1 = \beta_2 = \beta_3 = 0$ is rejected (0.05>0.01).

```
. qui logit y

. est store res

. lrtest logit res


Likelihood-ratio test                                 LR chi2(3)  =     21.79

(Assumption: res nested in logit)                     Prob > chi2 =    0.0001
```

According to the LR-test, the null hypothesis of $\beta_1 = \beta_2 = \beta_3 = 0$ is rejected (0.05>0.01).

(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)

The threshold in computing Counted R^2 is important because if we set threshold to be 0.5 then any Phat that greater than 0.5 will be predicted to be Yhat = 1. Therefore, if we change the threshold, the counted R^2 might change as well. For example,

```
est restore logit

(results logit are active now)


. predict pr

(option pr assumed; Pr(y))
```

```
. gen yhat1 = 1 if pr>0.5


. replace yhat1 = 0 if pr<=0.5

(0 real changes made)


. tab y yhat1


           |   yhat1

         Y |         1 |     Total

-----------+-----------+----------

         0 |        57 |        57

         1 |       158 |       158

-----------+-----------+----------

     Total |       215 |       215
```

Counted R^2 = (158+0)/215 = 0.7349

In this case, it is clearly that we fail to predict when Y=0 (we made 0% corrected predicted result) which is unsatisfying at all. So, I try to change the threshold to be 0.7 instead of 0.5

```
. gen yhat2 = 1 if pr>0.7

(117 missing values generated)

. replace yhat2 = 0 if pr<=0.7

(117 real changes made)


. tab y yhat2


           |         yhat2

         Y |         0          1 |     Total

-----------+----------------------+----------

         0 |        44         13 |        57

         1 |        73         85 |       158

-----------+----------------------+----------

     Total |       117         98 |       215
```

Counted R^2 = (85+44)/215 = 0.6

Now, even the Counted R^2 decreases to 0.6 but

Y=0 Counted R^2 = 44/57 =77.19%

Y=1 Counted R^2 = 85/158 =53.8%

6. From the data set "`Midterm_q6_no.dta`":

<u>Panel Data Model</u>

$$y_{it} = \alpha + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it} \qquad (13)$$

where: y_{it} is Dependent variable.
x_{kit} is Independent variable *k*.
u_{it} is Stochastic disturbance term.

<u>Fixed Effects Model</u>

$$y_{it} = \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it} \qquad (14)$$

where: α_i is Cross-sectional fixed effects

<u>Random Effects Model</u>

$$y_{it} = \alpha + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it} \qquad (15)$$

and $u_{it} = v_i + \varepsilon_{it}$

where: v_i = Cross-section random effects
ε_{it} = residual terms

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)

```
. xtset id t

       panel variable:  id (strongly balanced)

        time variable:  t, 1 to 12

                delta:  1 unit


. 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
------------------------------------------------------------------------------

. est store hetero

. 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
------------------------------------------------------------------------------

. est store pgls

. local df = e(N_g) - 1

. lrtest hetero . , df(`df')

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

According to the LR-test, the null hypothesis of no heteroscedasticity cannot be rejected. Thus, there is insignificant heteroscedasticity.

If there exists the heteroscedasticity, then our estimated will no longer be Best Linear Unbiased Estimator but it is still unbiased.

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)

```
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
------------------------------------------------------------------------------

. 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

. est store fixed

. 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)
------------------------------------------------------------------------------
```

```
. est store random

hausman fixed random


                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed        random       Difference          S.E.
-------------+----------------------------------------------------------------
          x1 |    .1014426     .2162513       -.1148087               .
          x2 |    .6951028     1.028607       -.3335042               .
          x3 |   -.4953168    -.4779339       -.0173829               .
------------------------------------------------------------------------------
                           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    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test
. hausman random fixed


                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     random       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
```

According to the fixed effect test, the null hypothesis of $\alpha_1 = \alpha_2 = \ldots = \alpha_n$ is rejected. It means that there exists significant fixed effect.

Then, I perform the Hausman test to see whether it should be the fixed effect or random effect. Hausman test, H_0: $\beta_{fixed} = \beta_{random}$ is rejected. Therefore, Fixed effect model is the most appropriated model.

From the fixed effect model, the x_1 and x_2 have positive effect on y while x_3 has a negative effect on y. Both Overall test and individual test suggest that all independent variables are significant. The result can be interpreted as follows;

If x_1 increases by 1 unit, y will increase by 0.10 unit, ceteris paribus.

If x_2 increases by 1 unit, y will increase by 0.70 unit, ceteris paribus.

If x_3 increases by 1 unit, y will decrease by 0.50 unit, ceteris paribus.

If all x are 0, y will be predicted to be 208.66

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)

First different method and Fixed effect estimation method are used to difference-out the unobserved fixed effects, but the different is that the First difference estimation method take the difference between period t and period t-1 to be the new independent variables

$$y_{it} = \beta_1 + \beta_2 x_{2it} + \ldots + \beta_k x_{kit} + \alpha_i + \varepsilon_{it} \qquad (1)$$

$$y_{it-1} = \beta_1 + \beta_2 x_{2it-1} + \ldots + \beta_k x_{kit-1} + \alpha_i + \varepsilon_{it-1} \qquad (2)$$

$$\Delta y_{it} = \beta_2 \Delta x_{2it} + \ldots + \Delta_k x_{kit} + \Delta\varepsilon_{it} \qquad (1)-(2)$$

However, this method works only when the case of balance panel.

For the Fixed effect estimation method, instead of using the different between period t and t-1, it uses the deviation from the average cross sectional group.

$$y_{it}-y_i\text{bar} = \beta_1 + \beta_2(x_{2it}-x_{2i}\text{bar}) + \ldots + \beta_k(x_{kit}-x_{ki}\text{bar}) + (\alpha_i-\alpha\text{bar}) + (\varepsilon_{it} - \varepsilon_i\text{bar})$$

$$y_{it}-y_i\text{bar} = \beta_1 + \beta_2(x_{2it}-x_{2i}\text{bar}) + \ldots + \beta_k(x_{kit}-x_{ki}\text{bar}) + (\varepsilon_{it} - \varepsilon_i\text{bar})$$

This method is also called “Within Group Mean Difference”, and it can be applied in the case of unbalanced panel as well.

For the fixed effect model, we assume α_i to be correlated with the error term ε_{it}, and the random effect is kind of between the OLS and Fixed effect that is there exists the variance of α_i in between 0 to infinity.

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)

$$\text{FE}: \quad (y_{it} - \bar{y}) = \beta_2 (x_{2it} - \bar{x}_{2i}) + \ldots + \beta_k (x_{kit} - \bar{x}_{ki}) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

$$y^* = \beta_2 x^*_{2it} + \ldots + \beta_k x^*_{kit} + \varepsilon^*_{it}$$

$$\boxed{\star \ \text{Overall } R^2 = \frac{\Sigma (\hat{y}_i - \bar{y})^2}{\Sigma (y - \bar{y})^2}}$$

$$\text{Within-}R^2 = \frac{\Sigma (\hat{y}^*_{it} - \bar{y}^*)^2}{\Sigma (y^*_{it} - \bar{y}^*)^2}$$

$$= \frac{\Sigma \left(\widehat{(y_{it} - \bar{y}_i)} - (\overline{y_{it} - \bar{y}_i}) \right)^2}{\Sigma \left((y_{it} - \bar{y}) - (\overline{y_{it} - \bar{y}}) \right)^2}$$

$$\text{Between } R^2 = \frac{\Sigma (\hat{\bar{y}}_i - g)^2}{\Sigma (\bar{y}_i - \bar{y})^2}$$

In this case, we are interested in Overall R^2 because it tells us the variation in the predicted y.

While The between R^2 is "How much of the variance between seperate panel units does my model account for".

The within R^2 is "How much of the variance within the panel units does my model account for"