1.

a) The models indicate the endogeneity bias because dependent variable simultaneously determine independent variable as well. Therefore, the OLS estimated results would be biased, inconsistent, and inefficient.

```
. reg y1 y2 x3

      Source |       SS           df       MS      Number of obs   =        5
> 0
-------------+----------------------------------   F(2, 47)        =    101.6
> 1
       Model |  51624.5487         2  25812.2743   Prob > F        =    0.000
> 0
    Residual |  11939.1313        47  254.024071   R-squared       =    0.812
> 2
-------------+----------------------------------   Adj R-squared   =    0.804
> 2
       Total |    63563.68        49  1297.21796   Root MSE        =    15.93
> 8

------------------------------------------------------------------------------
> —
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
> ]
-------------+----------------------------------------------------------------
> —
          y2 |   .0009735    .000209     4.66   0.000      .000553     .00139
> 4
          x3 |    .007247   .0010994     6.59   0.000     .0050353    .009458
> 7
       _cons |   45.97379   16.46146     2.79   0.008      12.8576    79.0899
> 8
------------------------------------------------------------------------------
```

```
. reg y2 y1 x1 x2

      Source |       SS           df       MS      Number of obs   =        50
-------------+----------------------------------   F(3, 46)        =     38.30
       Model |  7.8620e+09         3  2.6207e+09   Prob > F        =    0.0000
    Residual |  3.1474e+09        46  68421406.7   R-squared       =    0.7141
-------------+----------------------------------   Adj R-squared   =    0.6955
       Total |  1.1009e+10        49   224682127   Root MSE        =    8271.7

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          y1 |    274.9763   41.52252     6.62   0.000     191.3958    358.5568
          x1 |    285.7591   117.7665     2.43   0.019     48.70732    522.8108
          x2 |   -.0011751   .0003508    -3.35   0.002    -.0018812   -.0004691
       _cons |   -30624.39   7905.142    -3.87   0.000    -46536.62   -14712.17
------------------------------------------------------------------------------
```

b) Reduced Form : $y_{1t} = \pi_{10} + \pi_{11}x_{1t} + \pi_{12}x_{2t} + \pi_{13}x_{3t} + w_{1t}$

$$y_{2t} = \pi_{20} + \pi_{21}x_{1t} + \pi_{22}x_{2t} + \pi_{23}x_{3t} + w_{2t}$$

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

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

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**	**15.95613**	**0.8117**	**94.18**	**0.0000**
y2	**50**	**3**	**8282.287**	**0.7134**	**33.49**	**0.0000**

	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
y1						
y2	**.0010416**	**.0003875**	**2.69**	**0.009**	**.0002721**	**.001811**
x3	**.0070009**	**.0016124**	**4.34**	**0.000**	**.0037991**	**.0102027**
_cons	**47.52584**	**18.0776**	**2.63**	**0.010**	**11.62731**	**83.42436**
y2						
y1	**289.2138**	**53.16393**	**5.44**	**0.000**	**183.6408**	**394.7868**
x1	**261.0591**	**131.1819**	**1.99**	**0.050**	**.5577525**	**521.5604**
x2	**-.0011448**	**.0003582**	**-3.20**	**0.002**	**-.0018562**	**-.0004334**
_cons	**-32444.1**	**8976.941**	**-3.61**	**0.000**	**-50270.53**	**-14617.68**

```
Endogenous variables:  y1 y2
Exogenous variables:   x1 x2 x3
```

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

Source	SS	df	MS	Number of obs	=	**50**
				F(3, 46)	=	**47.89**
Model	**48148.6633**	**3**	**16049.5544**	Prob > F	=	**0.0000**
Residual	**15415.0167**	**46**	**335.109059**	R-squared	=	**0.7575**
				Adj R-squared	=	**0.7417**
Total	**63563.68**	**49**	**1297.21796**	Root MSE	=	**18.306**

y1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x1	**.5641608**	**.2478611**	**2.28**	**0.028**	**.0652422**	**1.063079**
x2	**-1.35e-06**	**7.57e-07**	**-1.79**	**0.081**	**-2.87e-06**	**1.72e-07**
x3	**.0094773**	**.0011136**	**8.51**	**0.000**	**.0072357**	**.011719**
_cons	**17.1494**	**18.36329**	**0.93**	**0.355**	**-19.81399**	**54.11279**

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

```
. reg y2 x1 x2 x3
```

Source	SS	df	MS	Number of obs	=	50
				F(3, 46)	=	25.66
Model	6.8914e+09	3	2.2971e+09	Prob > F	=	0.0000
Residual	4.1180e+09	46	89521588.5	R-squared	=	0.6260
				Adj R-squared	=	0.6016
Total	1.1009e+10	49	224682127	Root MSE	=	9461.6

y2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x1	424.2221	128.1089	3.31	0.002	166.3523	682.092
x2	-.0015356	.0003911	-3.93	0.000	-.0023228	-.0007483
x3	2.74097	.5755936	4.76	0.000	1.58236	3.89958
_cons	-27484.26	9491.205	-2.90	0.006	-46589.06	-8379.452

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

```
. reg y1 y2hat x3
```

Source	SS	df	MS	Number of obs	=	50
				F(2, 47)	=	72.20
Model	47954.9378	2	23977.4689	Prob > F	=	0.0000
Residual	15608.7422	47	332.100899	R-squared	=	0.7544
				Adj R-squared	=	0.7440
Total	63563.68	49	1297.21796	Root MSE	=	18.224

y1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
y2hat	.0010416	.0004425	2.35	0.023	.0001513	.0019319
x3	.0070009	.0018415	3.80	0.000	.0032963	.0107055
_cons	47.52584	20.64658	2.30	0.026	5.990273	89.06141

```
. reg y2 y1hat x1 x2
```

Source	SS	df	MS	Number of obs	=	50
				F(3, 46)	=	25.66
Model	6.8914e+09	3	2.2971e+09	Prob > F	=	0.0000
Residual	4.1180e+09	46	89521587.2	R-squared	=	0.6260
				Adj R-squared	=	0.6016
Total	1.1009e+10	49	224682127	Root MSE	=	9461.6

y2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
y1hat	289.2138	60.73383	4.76	0.000	166.9629	411.4646
x1	261.0591	149.8606	1.74	0.088	-40.59475	562.7129
x2	-.0011448	.0004092	-2.80	0.007	-.0019686	-.000321
_cons	-32444.1	10255.15	-3.16	0.003	-53086.65	-11801.56

C)

```
. 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	15.47004	0.8117	200.38	0.0000
y2	50	3	7952.619	0.7128	109.65	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
y1						
y2	.0010416	.0003757	2.77	0.006	.0003053	.0017779
x3	.0070009	.0015632	4.48	0.000	.003937	.0100648
_cons	47.52584	17.52688	2.71	0.007	13.17378	81.87789
y2						
y1	276.9252	49.12981	5.64	0.000	180.6325	373.2178
x1	305.6173	115.6688	2.64	0.008	78.91072	532.3239
x2	-.001085	.0003371	-3.22	0.001	-.0017457	-.0004243
_cons	-32771.52	8602.686	-3.81	0.000	-49632.47	-15910.56

```
Endogenous variables:  y1 y2
Exogenous variables:   x1 x2 x3
```

```
.
```

The differences among these 3 estimation methods are:
When there is endogeneity Bias, OLS estimators would be biased, inconsistent, and inefficient. For the 2SLS, the estimators would still be biased but consistent and asymptotically efficient. So we are fine with endogeneity bias by using 2SLS as long as n approaches infinity.

For 3SLS, the estimators would be biased, “may be” consistent, and more asymptotically efficient. 3SLS estimators can be inconsistent if we have specification error since 3SLS is the combination between 2SLS and system of equations. So if there is specification error in 1 equation, it will spread throughout all the equations. And this problem is very concerning. So this is like we might trade unbiasedness for efficiency which is not good. Also, we would have more asymptotic efficiency if and only if there actually exist endogeneity bias.

2.
A)

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

Iteration 0:  residual SS =  320.6977
Iteration 1:  residual SS =  301.8529
Iteration 2:  residual SS =  301.6507
Iteration 3:  residual SS =  301.6506
Iteration 4:  residual SS =  301.6506
Iteration 5:  residual SS =  301.6506
Iteration 6:  residual SS =  301.6506
```

Source	SS	df	MS		
Model	54.484926	3	18.1616418	Number of obs =	250
Residual	301.65058	246	1.22622188	R-squared =	0.1530
				Adj R-squared =	0.1427
Total	356.13551	249	1.43026308	Root MSE =	1.107349
				Res. dev. =	756.4214

lnC	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
/lnlamda	1.920144	.7400844	2.59	0.010	.4624332	3.377854
/beta	.8625099	.1355566	6.36	0.000	.5955103	1.129509
/alpha	-.8267275	.4666301	-1.77	0.078	-1.745827	.0923725
/theta	.2624028	.1786237	1.47	0.143	-.0894242	.6142297

Parameter lnlamda taken as constant term in model & ANOVA table

value of efficiency parameter (Lamda) = 6.8219

Distribution parameter(theta) = 0.2624

Parameter(beta) = 0.8625

Substitution Parameter = -0.8267

```
. 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) =   39.58
            Prob > F =    0.0000
```

P-value is less than 0.05 so therefore, Ho is rejected. The model is significant.

B)

```
.  nl (lnC = {lnlamda}-(({beta}/{alpha})*(log({theta}*(R^(-{alpha}))+(1-{theta})*W^(-{al
> pha})))))), init(lnlamda 0.5 theta 0.1 beta 0.1 alpha -0.1)
(obs = 250)

Iteration 0:  residual SS =  4392.861
Iteration 1:  residual SS =  3727.718
Iteration 2:  residual SS =  357.4753
Iteration 3:  residual SS =  348.7848
Iteration 4:  residual SS =  324.1563
Iteration 5:  residual SS =  324.0814
Iteration 6:  residual SS =  323.8134
Iteration 7:  residual SS =  323.4938
Iteration 8:  residual SS =  323.1047
Iteration 9:  residual SS =  322.4563
Iteration 10:  residual SS =  321.2419
Iteration 11:  residual SS =  318.1517
Iteration 12:  residual SS =   314.016
Iteration 13:  residual SS =  301.7794
Iteration 14:  residual SS =  301.6506
Iteration 15:  residual SS =  301.6506
Iteration 16:  residual SS =  301.6506
Iteration 17:  residual SS =  301.6506
```

Source	SS	df	MS		
				Number of obs =	250
Model	54.484926	3	18.1616418	R-squared =	0.1530
Residual	301.65058	246	1.22622188	Adj R-squared =	0.1427
				Root MSE =	1.107349
Total	356.13551	249	1.43026308	Res. dev. =	756.4214

lnC	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
/lnlamda	1.920144	.7400844	2.59	0.010	.4624332	3.377854
/beta	.8625098	.1355565	6.36	0.000	.5955103	1.129509
/alpha	-.8267271	.4666279	-1.77	0.078	-1.745823	.0923684
/theta	.2624029	.1786237	1.47	0.143	-.0894241	.6142298

Parameter lnlamda taken as constant term in model & ANOVA table

.

```
. est store nls2

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

Variable	nls1	nls2
lnlamda		
_cons	1.9201435**	1.9201436**
beta		
_cons	.86250986***	.86250983***
alpha		
_cons	-.82672747*	-.82672714*
theta		
_cons	.26240276	.26240288
Statistics		
N	250	250
rss	301.65058	301.65058
r2	.15298931	.15298931
r2_a	.14265991	.14265991

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

The results are slightly different than in a). This is because the different set of initial values would give the different estimated results. However, we change in initial values that we have done here are only a bit difference so this might make the estimated results to be different only a tiny bit. Also, changing the initial values to b) would have more iteration times because this set of initial values have the program to iterate more and it would stop when estimator of the new round is equal to the last round.

C)

```
.  nl (lnC = {lnlamda}-(({beta}/{alpha})*(log({theta}*(R^(-{alpha}))+(1-{theta})*W^(-{al
> pha})))))), init(lnlamda 0.5 theta 0.1 beta 0.1 alpha -0.1) eps(1e-1)
(obs = 250)

Iteration 0:  residual SS =  4392.861
Iteration 1:  residual SS =  3727.718
Iteration 2:  residual SS =  357.4753
Iteration 3:  residual SS =  348.7848
Iteration 4:  residual SS =  324.1563
Iteration 5:  residual SS =  324.0814
Iteration 6:  residual SS =  323.8134
Iteration 7:  residual SS =  323.4938
Iteration 8:  residual SS =  323.1047
Iteration 9:  residual SS =  322.4563
Iteration 10:  residual SS =  321.2419
Iteration 11:  residual SS =  318.1517
Iteration 12:  residual SS =   314.016
Iteration 13:  residual SS =  301.7794
Iteration 14:  residual SS =  301.6506
```

Source	SS	df	MS		
Model	54.484891	3	18.1616304	Number of obs =	250
Residual	301.65062	246	1.22622202	R-squared =	0.1530
				Adj R-squared =	0.1427
Total	356.13551	249	1.43026308	Root MSE =	1.107349
				Res. dev. =	756.4214

lnC	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
/lnlamda	1.92033	.7402206	2.59	0.010	.4623511	3.378308
/beta	.8624709	.1355743	6.36	0.000	.5954365	1.129505
/alpha	-.8255423	.4737031	-1.74	0.083	-1.758574	.1074889
/theta	.2626096	.1824575	1.44	0.151	-.0967685	.6219877

Parameter beta taken as constant term in model & ANOVA table

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

Iteration 0:  residual SS =  4392.861
Iteration 1:  residual SS =  3727.718
Iteration 2:  residual SS =  357.4753
Iteration 3:  residual SS =  348.7848
Iteration 4:  residual SS =  324.1563
Iteration 5:  residual SS =  324.0814
Iteration 6:  residual SS =  323.8134
Iteration 7:  residual SS =  323.4938
Iteration 8:  residual SS =  323.1047
Iteration 9:  residual SS =  322.4563
Iteration 10:  residual SS =  321.2419
Iteration 11:  residual SS =  318.1517
Iteration 12:  residual SS =   314.016
Iteration 13:  residual SS =  301.7794
Iteration 14:  residual SS =  301.6506
Iteration 15:  residual SS =  301.6506
Iteration 16:  residual SS =  301.6506
Iteration 17:  residual SS =  301.6506
Iteration 18:  residual SS =  301.6506
Iteration 19:  residual SS =  301.6506
Iteration 20:  residual SS =  301.6506
Iteration 21:  residual SS =  301.6506
Iteration 22:  residual SS =  301.6506
Iteration 23:  residual SS =  301.6506
Iteration 24:  residual SS =  301.6506
Iteration 25:  residual SS =  301.6506
Iteration 26:  residual SS =  301.6506
Iteration 27:  residual SS =  301.6506
Iteration 28:  residual SS =  301.6506
Iteration 29:  residual SS =  301.6506
Iteration 30:  residual SS =  301.6506
Iteration 31:  residual SS =  301.6506
Iteration 32:  residual SS =  301.6506
Iteration 33:  residual SS =  301.6506
Iteration 34:  residual SS =  301.6506
Iteration 35:  residual SS =  301.6506
Iteration 36:  residual SS =  301.6506
Iteration 37:  residual SS =  301.6506
Iteration 38:  residual SS =  301.6506
Iteration 39:  residual SS =  301.6506

      Source |       SS           df       MS      Number of obs =       250
-------------+----------------------------------   R-squared     =    0.9743
       Model |  11449.247          4  2862.31181   Adj R-squared =    0.9739
    Residual |  301.65058        246  1.22622188   Root MSE      =  1.107349
-------------+----------------------------------   Res. dev.     =  756.4214
       Total |  11750.898        250  47.0035913

------------------------------------------------------------------------------
         lnC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    /lnlamda |    1.920143   .7400844     2.59   0.010     .4624331    3.377854
       /beta |    .8625099   .1355566     6.36   0.000     .5955103    1.129509
      /alpha |   -.8267275   .4666305    -1.77   0.078    -1.745828    .0923732
      /theta |    .2624027   .1786237     1.47   0.143    -.0894242    .6142297
------------------------------------------------------------------------------
convergence not achieved
r(430);
```

```
. est table nls2 nls_c nls3, star(.1 .05 .01) stat(N rss r2 r2_a)
```

Variable	nls2	nls_c	nls3
lnlamda			
_cons	1.9201436**	1.9203297**	1.9201435**
beta			
_cons	.86250983***	.86247093***	.86250987***
alpha			
_cons	-.82672714*	-.82554235*	-.82672753*
theta			
_cons	.26240288	.26260962	.26240273
Statistics			
N	250	250	250
rss	301.65058	301.65062	301.65058
r2	.15298931	.15298921	.97432957
r2_a	.14265991	.14265981	.97391217

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

We have different estimated results from b) because the convergence values are different. The lower convergence value will stay driving STATA to find the estimated results. However, if we set the maximum iteration times, it might report convergence not achieved since it stops at the rounds we set before it can find the answer.

D) Because it's faster. Let's say there are 10,000 of them to run but with log-form, we would have log(10^5) which mean we won't have to run all of the 10,000 just like in the original form.

3.
A)

```
. 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 = -272.63079
rescale:       log likelihood = -271.87577
Iteration 0:   log likelihood = -271.87577
Iteration 1:   log likelihood = -229.84968
Iteration 2:   log likelihood = -229.48827
Iteration 3:   log likelihood = -229.48797
Iteration 4:   log likelihood = -229.48797

                                                Number of obs     =        400
                                                Wald chi2(2)      =      61.78
Log likelihood = -229.48797                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .5449328   .1230867     4.43   0.000     .3036873    .7861782
          x2 |  -.5155686   .0715727    -7.20   0.000    -.6558485   -.3752887
       _cons |   .0717861   .1992166     0.36   0.719    -.3186713    .4622436
------------------------------------------------------------------------------

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

. ml maximize

initial:       log likelihood = -277.25887
alternative:   log likelihood = -272.63079
rescale:       log likelihood = -271.87577
Iteration 0:   log likelihood = -271.87577
Iteration 1:   log likelihood = -229.85671
Iteration 2:   log likelihood =  -229.4898
Iteration 3:   log likelihood = -229.48797
Iteration 4:   log likelihood = -229.48797

                                                Number of obs     =        400
                                                Wald chi2(2)      =      62.12
Log likelihood = -229.48797                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .5449256   .1221639     4.46   0.000     .3054887    .7843625
          x2 |  -.5155846   .0723328    -7.13   0.000    -.6573542    -.373815
       _cons |   .0718396   .2050023     0.35   0.726    -.3299575    .4736366
------------------------------------------------------------------------------

. est store bhhh
```

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

. ml maximize

initial:       log likelihood = -277.25887
alternative:   log likelihood = -272.63079
rescale:       log likelihood = -271.87577
Iteration 0:   log likelihood = -271.87577
Iteration 1:   log likelihood = -250.87722  (backed up)
Iteration 2:   log likelihood = -230.14003  (backed up)
Iteration 3:   log likelihood = -229.92959
Iteration 4:   log likelihood = -229.48942
Iteration 5:   log likelihood =   -229.488
Iteration 6:   log likelihood = -229.48797
Iteration 7:   log likelihood = -229.48797

                                                Number of obs     =        400
                                                Wald chi2(2)      =      61.78
Log likelihood = -229.48797                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .5449327   .1230867     4.43   0.000     .3036873    .7861782
          x2 |  -.5155686   .0715727    -7.20   0.000    -.6558484   -.3752887
       _cons |   .0717861   .1992166     0.36   0.719    -.3186714    .4622435
------------------------------------------------------------------------------

. est store bfgs

. est table nr bhhh bfgs, star(.1 .05 .01) stat(N ll chi2 p)

-------------------------------------------------------
    Variable |      nr           bhhh          bfgs
-------------+-----------------------------------------
          x1 |  .54493275***   .54492561***   .54493275***
          x2 |  -.5155686***  -.51558458***  -.51556859***
       _cons |  .07178611      .07183955      .07178607
-------------+-----------------------------------------
           N |       400            400            400
          ll |  -229.48797     -229.48797     -229.48797
        chi2 |  61.779312      62.124104      61.77931
           p |  3.844e-14      3.235e-14      3.844e-14
-------------------------------------------------------
                       legend: * p<.1; ** p<.05; *** p<.01
```

We have different estimated results because we use the different algorithms.

B)
Wald Test

```
. test x1 x2

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

           chi2(  2) =   61.78
         Prob > chi2 =    0.0000
```

P-value<0.05, Ho is rejected. We have overall significance according to Wald Test

LR Test

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

. ml maximize

initial:       log likelihood = -277.25887
alternative:   log likelihood = -272.63079
rescale:       log likelihood = -271.87577
Iteration 0:   log likelihood = -271.87577
Iteration 1:   log likelihood = -271.45072
Iteration 2:   log likelihood =  -271.4507

                                                Number of obs     =        400
                                                Wald chi2(0)      =          .
Log likelihood =  -271.4507                     Prob > chi2       =          .

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -.3433333   .1014771    -3.38   0.001    -.5422247   -.1444419
------------------------------------------------------------------------------

. est store res

. lrtest nr res

Likelihood-ratio test                                 LR chi2(2)  =     83.93
(Assumption: res nested in nr)                        Prob > chi2 =    0.0000

.
```

P-value<0.05, Ho is rejected. We have overall significance according to LR test.
We prefer LR test because it uses 2 models which are unrestricted and restricted and the results are optimal. For Wald test, it only uses unrestricted model just case the restricted is hard to find. The results from Wald would depend on paramiterization; sometimes you may have many parameters. It can be overestimate.

C) It uses Chi2 test because it has Chi2 distribution. We can't employ F test since it doesn't have F distribution.

D)

```
. 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

. do "/Users/mien/Downloads/03/Midterm_q3 - ml_probit_het.do"

. 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



. ml model lf ml_probit (y=x1 x2)

. ml maximize

initial:       log likelihood = -277.25887
alternative:   log likelihood = -281.53481
rescale:       log likelihood = -271.60653
Iteration 0:   log likelihood = -271.60653
Iteration 1:   log likelihood = -229.86554
Iteration 2:   log likelihood = -229.59367
Iteration 3:   log likelihood =  -229.5935
Iteration 4:   log likelihood =  -229.5935

                                                Number of obs     =        400
                                                Wald chi2(2)      =      69.64
Log likelihood =  -229.5935                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .3244745   .0723427     4.49   0.000     .1826855    .4662635
          x2 |  -.3082745   .0408367    -7.55   0.000     -.388313   -.2282361
       _cons |   .0469691   .1205919     0.39   0.697    -.1893866    .2833248
------------------------------------------------------------------------------
```

```
. est store probit

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

. ml maximize

initial:       log likelihood = -277.25887
alternative:   log likelihood = -281.46023
rescale:       log likelihood = -271.58192
rescale eq:    log likelihood = -271.36437
Iteration 0:   log likelihood = -271.36437
Iteration 1:   log likelihood = -236.90428
Iteration 2:   log likelihood =  -228.1736
Iteration 3:   log likelihood = -227.35327
Iteration 4:   log likelihood =  -227.3515
Iteration 5:   log likelihood =  -227.3515

                                                Number of obs     =        400
                                                Wald chi2(2)      =      66.09
Log likelihood =  -227.3515                     Prob > chi2       =     0.0000
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
eq1						
x1	.2970344	.072702	4.09	0.000	.1545411	.4395277
x2	-.315029	.0411731	-7.65	0.000	-.3957267	-.2343312
_cons	.0907444	.1164622	0.78	0.436	-.1375174	.3190061
eq2						
x3	-.2484118	.1178887	-2.11	0.035	-.4794693	-.0173543

```
. est store hetprob

. lrtest hetprob probit

Likelihood-ratio test                                 LR chi2(1)  =      4.48
(Assumption: probit nested in hetprob)                Prob > chi2 =    0.0342
```

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

Variable	probit	hetprob
eq1		
x1	.32447448***	.29703439***
x2	-.30827454***	-.31502896***
_cons	.0469691	.09074437
eq2		
x3		-.2484118**
Statistics		
N	400	400
ll	-229.5935	-227.3515
chi2	69.635238	66.088837
p	7.567e-16	4.456e-15

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

P-value of LR test <0.05; Ho is rejected. Therefore, there exists heteroskedasticity.
We can't perform Wald Test and LM test for heteroskedaticity test because to test this, it has to use both restricted and unrestricted model.

E) Since we assume asymptotically normal distribution where n goes to infinity. Therefore, we don't lose df. So the test would Z-test, not T-test.
We don't have R2 in MLE because it does not minimize RSS; as a result in maximizing R2. So it would not make sense to look at R2 as the goodness of fit
To compare 2 estimated results, we should log likelihood function where the higher, the better.

4.

```
. 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) =  8.702e-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

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   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
------------------------------------------------------------------------------

. estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(0) = 5.9e-17 (p =     .)

Note: test cannot be performed because there are no
      overidentifying restrictions.

. est store unres
```

```
. 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) =  4.145e-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
```

```
end of do-file

. estat overid

  Test of overidentifying restriction:

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

. est store Merton

. 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) =  3.158e-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

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   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

. estat overid

  Test of overidentifying restriction:

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

. est store Varsicek
```

```
. est table unres Merton Varsicek, star(0.1 0.05 0.01) stat(N J)
```

Variable	unres	Merton	Varsicek
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

A) Merton model ; we have coefficients of beta equal to - 0.0008, and sigma^2 = 0.0004 but they both are insignificant. Looking at the J-test, p-value is less than 0.05 so we reject Ho which is not desirable. Since that means some of the moment conditions shouldn't be used (they are correlated with error terms).
B) Vasicek might be the most appropriated because individual tests are all significant and its p-value of J-test is > 0.05 (Ho is not rejected) which means all moments of conditions in Vasicek should be included.

4, part 2)
C) The estimated results will be biased since we have x1 and x2 correlate with error term. Or we can call it endogeneity bias.

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

Instrumental variables (GMM) regression          Number of obs   =        500
                                                 Wald chi2(4)    =     158.10
                                                 Prob > chi2     =     0.0000
                                                 R-squared       =     0.6001
GMM weight matrix: Robust                        Root MSE        =     20.261

------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   3.140592   .9631846     3.26   0.001     1.252784    5.028399
          x2 |   2.055367   .5094418     4.03   0.000      1.05688    3.053855
          x3 |   1.012003   .3166248     3.20   0.001     .3914301    1.632576
          x4 |   1.348025   .1766068     7.63   0.000     1.001882    1.694168
       _cons |   2.813565   7.556667     0.37   0.710    -11.99723    17.62436
------------------------------------------------------------------------------
Instrumented:  x1 x2
Instruments:   x3 x4 z1 z2 z3

. estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(1) = .003956 (p = 0.9499)

. est store gmm
```

D)

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

Instrumental variables (2SLS) regression         Number of obs   =        500
                                                 Wald chi2(4)    =     144.77
                                                 Prob > chi2     =     0.0000
                                                 R-squared       =     0.6001
                                                 Root MSE        =      20.26

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   3.139613   .9465968     3.32   0.001     1.284317    4.994909
          x2 |   2.056345   .5126418     4.01   0.000     1.051586    3.061105
          x3 |   1.013019   .3157263     3.21   0.001     .3942064    1.631831
          x4 |   1.347801   .1810136     7.45   0.000     .9930214    1.702582
       _cons |   2.798488   8.263154     0.34   0.735      -13.397    18.99397
------------------------------------------------------------------------------
Instrumented:  x1 x2
Instruments:   x3 x4 z1 z2 z3

. est store 2sls
```

```
. estat endogenous x1 x2

  Tests of endogeneity
  Ho: variables are exogenous

  Durbin (score) chi2(2)          =  458.832  (p = 0.0000)
  Wu-Hausman F(2,493)             =  2747.35  (p = 0.0000)
```

According to the test above, p-value < 0.05; Ho is rejected. Therefore, x1 and x2 are endogenous variables. GMM estimated result using *z1, z2, and z3* as instrumental variables is the most appropriated estimated result.

E)
In 2sls, we use predicted endogenous variables (x1hat, x2hat) to form the moment of conditions and fix the endogeneity bias. So it'd be exactly identified. But in GMM, we use instruments (z1, z2, z3) to fix the problem. And we use the actual endogenous variables not the predicted ones to form moment conditions. Now it's like we lose 2 x's but get 3 more z's instead. This'd be over-identified. So the 2 approaches are not exactly the same and this may cause the different estimated results.

5.
A)
Probit

```
. probit y x1 x2 x3

Iteration 0:   log likelihood = -112.90429
Iteration 1:   log likelihood =  -106.7958
Iteration 2:   log likelihood = -106.72267
Iteration 3:   log likelihood = -106.72264
Iteration 4:   log likelihood = -106.72264

Probit regression                                 Number of obs     =        215
                                                  LR chi2(3)        =      12.36
                                                  Prob > chi2       =     0.0062
Log likelihood = -106.72264                       Pseudo R2         =     0.0548

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0032535   .0014392     2.26   0.024     .0004327    .0060742
          x2 |   .0108366   .0082199     1.32   0.187    -.0052741    .0269473
          x3 |   .0917118   .0427879     2.14   0.032      .007849    .1755745
       _cons |   .5060492   .1236126     4.09   0.000      .263773    .7483254
------------------------------------------------------------------------------
```

No we cant compare the estimated coefficients because they are 2 different likelihood functional forms. So the results are surely different.
Interpret Logit : For the overall significance, P-value < 0.05 so Ho is rejected.
Pseudo R2 = 5.46%. Individual Test : All of them except x2 are individually significant.
Counted r2 = 78.1%

B)

```
. fitstat
. mfx

Marginal effects after logit
      y  = Pr(y) (predict)
         =  .80256657
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
x1	.0009079	.0004	2.29	0.022	.000129	.001687	60.8628
x2	.0033204	.0024	1.38	0.167	-.001385	.008026	-1.66444
x3	.0265759	.01221	2.18	0.029	.002653	.050499	1.63431

```
. mfx, at(median)

Marginal effects after logit
      y  = Pr(y) (predict)
         =  .70199392
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
x1	.0011986	.0006	2.01	0.044	.000031	.002366	.770154
x2	.0043837	.00328	1.34	0.181	-.002044	.010812	.414689
x3	.0350869	.01768	1.98	0.047	.000434	.06974	.174391

interpretation of marginal effects at mean of x_{1i} : If x1 increases by 1 unit (from mean = 0.7701 to 1.7701), probability that y =1 will increase by 0.0012 point or 0.12%, holding x2 and x3 at their means.

C)

```
. constraint 1 x1= x2

. logit y x1 x2 x3, constraint(1)

Iteration 0:   log likelihood = -112.90429
Iteration 1:   log likelihood = -107.44136
Iteration 2:   log likelihood =  -107.2562
Iteration 3:   log likelihood =  -107.2558
Iteration 4:   log likelihood =  -107.2558

Logistic regression                             Number of obs     =        215
                                                Wald chi2(2)      =       9.46
Log likelihood =  -107.2558                     Prob > chi2       =     0.0088

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

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0055491   .0025394     2.19   0.029     .0005719    .0105263
          x2 |   .0055491   .0025394     2.19   0.029     .0005719    .0105263
          x3 |   .1438469   .0763761     1.88   0.060    -.0058475    .2935413
       _cons |   .8186871   .2048395     4.00   0.000     .4172089    1.220165
------------------------------------------------------------------------------

. est store logit_c

. lrtest unres_log logit_c

Likelihood-ratio test                                 LR chi2(1)  =      1.02
(Assumption: logit_c nested in unres_log)             Prob > chi2 =    0.3122

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   1.273816   .1650121     7.72   0.000     .9503987    1.597234
------------------------------------------------------------------------------

. est store res_log

. lrtest unres_log res_log

Likelihood-ratio test                                 LR chi2(3)  =     12.32
(Assumption: res_log nested in unres_log)             Prob > chi2 =    0.0064
```

We have p-value < 0.05. Ho is rejected. Therefore, at least one of the betas is significant(not equal to 0)

P-value > 0.05. Ho is not rejected. Therefore, beta1 = beta2

D)

Yes if we change the threshold, counted R2 would also change. This is because counted R2 is computed from # of correct prediction/# of total observations. If we change the threshold, the # of correct predictions would also change.

6.

A)

```
. 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) nolog


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 het

. xtgls y x1 x2 x3 nolog
variable nolog not found
r(111);

.  xtgls y x1 x2 x3, nolog

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 het, df(`df')

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

value > 0.05. Ho is not rejected. No heteroskedaticity. If there is heteroskedaticity, we have problem with var-cov matrix(var is not constant). So the results would still be Unbiased but bot Best. And the tests are not reliable.

B)

```
. 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 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
```

Fixed Effect Test : Ho : alpha1 = alpha2 = alpha3

we have p-value in the red box < 0.05. Ho is rejected. So there exists fixed effects.

Hausman Test : Ho : $E(\alpha_1 x_1) = 0,\ E(\alpha_2 x_2) = 0, E(\alpha_3 x_3) = 0$ (Random Effect)

P-value < 0.05. Ho is rejected. So we have fixed effects.

The most appropriated model is the FE because firstly, we can't use the PGLS according to the FE test. Each of them has specific characteristics that won't change overtime. Now according Hausman Test, we shouldn't use RE because the correlation of xi and αi is big enough to use FE model.

Interpretation : If x1 increases by 1 unit, predicted y would increase by 0.1014 or 10.14%, holding other things constant. If x2 increases by 1 unit, predicted y would increase by 0.6951 or 69.51%, holding other things constant. If x1 increases by 1 unit, predicted y would decrease by 0.4779 or 47.79%, holding other things constant.

c)

FE estimation method is can be done with both balanced and unbalanced panel data but First difference method can be done only with balanced panel data.

FE model is used when correlation of xi and αi is big but RE model is used when correlation of xi and αi is small enough and we can get rid of the correlated part.

D)

Within R-squares is used when we would like to look at the FE. Between R-squares is computed based on ybar,hat which we don't really care about it since we care about yhat. Overall R-squares is what we care about because it's computed based on yhat.