BACHELOR OF ECONOMICS
THAMMASAT UNIVERSITY

Answer Sheet Cover Page
Final Examination Semester 2/2020

(Readable handwriting and printed version are acceptable)

Student Name: Harit Hatawong
Student ID: 6104640112

Course ID: EE426 Course Title: Econometrics 2
Lecturer: Assoc. Prof. Dr. Tatre Jantarakolica
Exam date: 28/05/2021 Time: 9.00-15.00

Total pages: 37

***Students are responsible for checking the number of pages before submitting the answer sheets. Incomplete submissions caused by carelessness will not be accepted as an excuse for resubmission**

Student Signature: Harit Hatawong
Date: 28/05/2021

Question 1

From the data set "`Final_q1-1_no.dta`":
Model for the degree of acceptance is stated as follows:

Prob(y_i=*j*|*X*) = *f*(x_1, x_2, x_3, x_4) (1) where: y_i is categorical data = 1 for disagree, = 2 for neutral, and = 3 for agree.

x_k is independent variable *k*. *k* = 1, 2, 3, 4.

(a) Estimate the model using multinomial logit model using y_i=0 is the base outcome. Make interpretation of your estimated results (sign & meaning (in term of rrr), overall test, goodness of fit, and individual test). Perform IIA test whether it is appropriated to apply multinomial logit in this case? Give explanation why IIA is important for multinomial logit. Determine whether multinomial logit is appropriated for this case? Why? (5 points).

Estimate the model using multinomial logit

```
mlogit y x1 x2 x3 x4, base(0) nolog

Multinomial logistic regression                 Number of obs     =        170
                                                LR chi2(8)        =      90.03
                                                Prob > chi2       =     0.0000
Log likelihood = -106.20802                     Pseudo R2         =     0.2977

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
-------------+----------------------------------------------------------------
1            |
          x1 |   -1.17921   .6174172    -1.91   0.056    -2.389325    .0309056
          x2 |  -.8208919   .7587919    -1.08   0.279    -2.308097    .6663129
          x3 |  -.3951391   .6225677    -0.63   0.526    -1.615349    .8250711
          x4 |   .6475044   .2489556     2.60   0.009     .1595603    1.135448
       _cons |  -7.756696   3.533794    -2.20   0.028     -14.6828   -.8305878
-------------+----------------------------------------------------------------
2            |
          x1 |  -1.040842   .6753533    -1.54   0.123     -2.36451    .2828262
          x2 |   -1.66206   .7889336    -2.11   0.035    -3.208341   -.1157786
          x3 |   .7053513   .6628839     1.06   0.287    -.5938773     2.00458
          x4 |   1.800799   .3022596     5.96   0.000     1.208381    2.393217
       _cons |    -25.869   4.507788    -5.74   0.000    -34.70411    -17.0339
------------------------------------------------------------------------------

. est store m1

. fitstat

Measures of Fit for mlogit of y

Log-Lik Intercept Only:      -151.224   Log-Lik Full Model:          -106.208
D(155):                       212.416   LR(8):                         90.032
                                        Prob > LR:                      0.000
McFadden's R2:                  0.298   McFadden's Adj R2:              0.198
Maximum Likelihood R2:          0.411   Cragg & Uhler's R2:             0.495
Count R2:                       0.259   Adj Count R2:                   0.023
AIC:                            1.426   AIC*n:                        242.416
BIC:                         -583.633   BIC':                         -48.946
```

Perform IIA test

```
mlogit y x1 x2 x3 x4 if y!=2 , base(0) nolog

Multinomial logistic regression                 Number of obs     =         62
                                                LR chi2(4)        =       9.36
                                                Prob > chi2       =     0.0526
Log likelihood = -35.008522                     Pseudo R2         =     0.1180

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
-------------+----------------------------------------------------------------
1            |
          x1 |  -.9900573   .5961108    -1.66   0.097    -2.158413    .1782984
          x2 |  -.5350665   .7317338    -0.73   0.465    -1.969238    .8991053
          x3 |  -.2076114    .643533    -0.32   0.747    -1.468913     1.05369
          x4 |   .5845398   .2453425     2.38   0.017     .1036773    1.065402
       _cons |  -7.222523   3.635066    -1.99   0.047    -14.34712   -.0979251
------------------------------------------------------------------------------

. est store m2

. hausman m1 m2, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       m1           m2         Difference          S.E.
-------------+----------------------------------------------------------------
          x1 |    -1.17921    -.9900573       -.1891526        .1607978
          x2 |   -.8208919    -.5350665       -.2858253        .2008254
          x3 |   -.3951391    -.2076114       -.1875277               .
          x4 |    .6475044     .5845398        .0629646        .0422605
       _cons |   -7.756696    -7.222523       -.5341731               .
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from mlogit
           B = inconsistent under Ha, efficient under Ho; obtained from mlogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        1.26
                Prob>chi2 =      0.9394
                (V_b-V_B is not positive definite)
```

Find rrr

```
. mlogit y x1 x2 x3 x4, base(0) rrr nolog

Multinomial logistic regression                 Number of obs     =        170
                                                LR chi2(8)        =      90.03
                                                Prob > chi2       =     0.0000
Log likelihood = -106.20802                     Pseudo R2         =     0.2977

------------------------------------------------------------------------------
           y |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
-------------+----------------------------------------------------------------
1            |
          x1 |   .3075216   .1898691    -1.91   0.056     .0916915    1.031388
          x2 |    .440039    .333898    -1.08   0.279     .0994504    1.947045
          x3 |   .6735864   .4193531    -0.63   0.526     .1988212    2.282043
          x4 |   1.910766    .475696     2.60   0.009     1.172995    3.112569
       _cons |   .0004279    .001512    -2.20   0.028     4.20e-07     .435793
-------------+----------------------------------------------------------------
2            |
          x1 |   .3531572   .2385059    -1.54   0.123     .0939953    1.326874
          x2 |   .1897477   .1496983    -2.11   0.035     .0404236    .8906724
          x3 |   2.024558   1.342047     1.06   0.287     .5521821    7.422975
          x4 |   6.054483   1.830026     5.96   0.000      3.34806    10.94866
       _cons |   5.82e-12   2.63e-11    -5.74   0.000     8.48e-16    4.00e-08
------------------------------------------------------------------------------
```

<u>Answer1A</u>: Intepretation

Sign&meaning (rrr)
From rrr stats all of the independent variables have the positive correlation on choosing boths choice y=1 and y=2
Which x1 impact more on choosing y=2
X2 impact more on choosing y=1
X3 impact more on choosing y=1
X4 impact more on choosing y=2

Overall test:
The model is jointly significant in overall test at 95% confidence (since p-value < 0.05)

Individual test:
Most of the parameters are insignificant z-test except x4

Goodness of fit
Pseudo R-square is quite low
As well as the counted R-square that lower than 0.5, So the model should be improved

Appropriateness
For the IIA test we accept the H0, which mean that the data is IIA and the MN logit is appropriated over the nested logit/ASM Probit model by the way the model has a lot of insignificant parameters.

The reason that we need IIA because if it's dependent the choice maker will choose differently when the some choice are limited which the MNlogit cannot capture that effect.

(b) Estimate the model using order probit model of y_i. Make interpretation of your estimated results (sign & meaning, overall test, goodness of fit, and individual test). Perform the test to determine whether order probit is appropriated in this case? Give explanation why? (5 points).

<u>Estimate the model using order probit model</u>

```
oprobit y x1 x2 x3 x4, nolog

Ordered probit regression                       Number of obs     =        170
                                                LR chi2(4)        =      81.97
                                                Prob > chi2       =     0.0000
Log likelihood = -110.23758                     Pseudo R2         =     0.2710

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.2304638   .2306873    -1.00   0.318    -.6826026    .2216749
          x2 |  -.4890222   .2370505    -2.06   0.039    -.9536326   -.0244118
          x3 |   .4885566   .2200936     2.22   0.026     .0571811    .9199322
          x4 |   .6858598   .0912415     7.52   0.000     .5070297      .86469
-------------+----------------------------------------------------------------
       /cut1 |   9.475686   1.391637                      6.748128    12.20324
       /cut2 |   10.68547   1.443344                      7.856569    13.51437
------------------------------------------------------------------------------
```

Perform the Order probit test

```
. gologit2 y x1 x2 x3 x4,pl sto(oprobit) link(p)

Generalized Ordered Probit Estimates              Number of obs     =        170
                                                  LR chi2(4)        =      81.97
                                                  Prob > chi2       =     0.0000
Log likelihood = -110.23758                       Pseudo R2         =     0.2710

 ( 1)  [0]x1 - [1]x1 = 0
 ( 2)  [0]x2 - [1]x2 = 0
 ( 3)  [0]x3 - [1]x3 = 0
 ( 4)  [0]x4 - [1]x4 = 0
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |
          x1 |  -.2304639   .2306873    -1.00   0.318    -.6826026    .2216749
          x2 |  -.4890222   .2370505    -2.06   0.039    -.9536326   -.0244119
          x3 |   .4885566   .2200936     2.22   0.026     .0571811    .9199322
          x4 |   .6858598   .0912415     7.52   0.000     .5070297      .86469
       _cons |  -9.475686   1.391637    -6.81   0.000    -12.20324   -6.748128
-------------+----------------------------------------------------------------
1            |
          x1 |  -.2304639   .2306873    -1.00   0.318    -.6826026    .2216749
          x2 |  -.4890222   .2370505    -2.06   0.039    -.9536326   -.0244119
          x3 |   .4885566   .2200936     2.22   0.026     .0571811    .9199322
          x4 |   .6858598   .0912415     7.52   0.000     .5070297      .86469
       _cons |  -10.68547   1.443344    -7.40   0.000    -13.51437   -7.856569
------------------------------------------------------------------------------

. gologit2 y x1 x2 x3 x4,npl sto(goprobit) link(p)

Generalized Ordered Probit Estimates              Number of obs     =        170
                                                  LR chi2(8)        =      87.15
                                                  Prob > chi2       =     0.0000
Log likelihood = -107.65138                       Pseudo R2         =     0.2881

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |
          x1 |    -.52637   .3142369    -1.68   0.094    -1.142263    .0895229
          x2 |  -.5121628   .3475867    -1.47   0.141     -1.19342    .1690946
          x3 |   .1676545   .3253398     0.52   0.606    -.4699997    .8053088
          x4 |   .5981086   .1208501     4.95   0.000     .3612467    .8349704
       _cons |  -7.775535   1.858973    -4.18   0.000    -11.41905   -4.132015
-------------+----------------------------------------------------------------
1            |
          x1 |  -.0834852   .2646317    -0.32   0.752    -.6021539    .4351834
          x2 |  -.5111958   .2583224    -1.98   0.048    -1.017498   -.0048931
          x3 |   .5913589   .2451609     2.41   0.016     .1108523    1.071865
          x4 |    .716643   .1072266     6.68   0.000     .5064828    .9268032
       _cons |  -11.31812   1.712532    -6.61   0.000    -14.67462   -7.961616
------------------------------------------------------------------------------

. lrtest oprobit goprobit, stats

Likelihood-ratio test                                 LR chi2(4)  =      5.17
(Assumption: oprobit nested in goprobit)              Prob > chi2 =    0.2701

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
     oprobit |        170 -151.2241  -110.2376       6    232.4752     251.29
    goprobit |        170 -151.2241  -107.6514      10    235.3028   266.6607
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.
```

Find the marginal effect

```
margins, dydx(*) predict(outcome(0))

Average marginal effects                        Number of obs     =        170
Model VCE    : OIM

Expression   : Pr(y==0), predict(outcome(0))
dy/dx w.r.t. : x1 x2 x3 x4

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0302728   .0301228     1.00   0.315    -.0287668    .0893124
          x2 |   .0642359   .0320269     2.01   0.045     .0014644    .1270075
          x3 |  -.0641748   .0293559    -2.19   0.029    -.1217112   -.0066383
          x4 |  -.0900917   .0132688    -6.79   0.000    -.1160981   -.0640853
------------------------------------------------------------------------------


margins, dydx(*) predict(outcome(1))

Average marginal effects                        Number of obs     =        170
Model VCE    : OIM

Expression   : Pr(y==1), predict(outcome(1))
dy/dx w.r.t. : x1 x2 x3 x4

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0287292   .0292413     0.98   0.326    -.0285827     .086041
          x2 |   .0609606   .0289968     2.10   0.036     .0041278    .1177934
          x3 |  -.0609026   .0269385    -2.26   0.024    -.1137011    -.008104
          x4 |   -.085498   .0123244    -6.94   0.000    -.1096533   -.0613427
------------------------------------------------------------------------------

. margins, dydx(*) predict(outcome(2))

Average marginal effects                        Number of obs     =        170
Model VCE    : OIM

Expression   : Pr(y==2), predict(outcome(2))
dy/dx w.r.t. : x1 x2 x3 x4

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   -.059002   .0589341    -1.00   0.317    -.1745107    .0565068
          x2 |  -.1251965   .0591178    -2.12   0.034    -.2410652   -.0093279
          x3 |   .1250773   .0542158     2.31   0.021     .0188164    .2313383
          x4 |   .1755897   .0141639    12.40   0.000     .1478289    .2033505
------------------------------------------------------------------------------
```

Answer1B: Intepretation

Sign&meaning (rrr)
From marginal effects
x1 has positive impact on choosing y= 0,1 but negative impact on choosing y=2
x2 has positive impact on choosing y= 0,1 but negative impact on choosing y=2
x3 has negative impact on choosing y= 0,1 but positive impact on choosing y=2
x4 has negative impact on choosing y= 0,1 but positive impact on choosing y=2

Overall test:
The model is jointly significant in overall test at 95% confidence (since p-value < 0.05)

Individual test:
Most of the parameters including cut, cut2 are significant z-test except x1

Goodness of fit
Pseudo R-square is quite low which mean that the model has the room to be improved

Appropriateness
For the order probit test we accept the H0, which mean that the ordered probit model is appropriated

From the data set "`Final_q1-2_no.dta`":
In the study of decision to subscribe mobile phone services, three choices have been studied

including AIS, DTAC, and TRUE. The probit model can be stated as:
$I = \beta + \beta x + \beta x + \beta x + \beta x + u$ (2)

and where:

$_{Ji}\ _{J0}\ _{J1}\ _{1i}\ _{J2}\ _{2i}\ _{J3}\ _{3i}\ _{J4}\ _{4i}\ _{Ji}$ $\Pr(Y_{Ji}=1)=\Phi(I_{Ji})$

I_{Ji} is index variables.
Y_{Ji} is decision to choose financial choice J, value equals to 1 if choosing

choice J or 0 if not. J = 1 for AIS, 2 for DTAC, 3 for TRUE. x_{ki} is independent variable k.

Φ (.) is multivariate normal probability distribution function. u_{Ji} is disturbance term.

(c) Estimate models for Y_{1i}, Y_{2i}, and Y_{3i} assuming that the probability functions follow separate normal distribution function. Should we use these three separate probit models? Why? (4 points)

Estimate by using seperate probit model

```
probit y1 x1 x2 x3 x4

Iteration 0:   log likelihood = -378.95508
Iteration 1:   log likelihood = -365.61869
Iteration 2:   log likelihood = -365.57585
Iteration 3:   log likelihood = -365.57585

Probit regression                               Number of obs     =        550
                                                LR chi2(4)        =      26.76
                                                Prob > chi2       =     0.0000
Log likelihood = -365.57585                     Pseudo R2         =     0.0353

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.319276   .3993005     3.30   0.001     .5366617    2.101891
          x2 |  -.7117926   .3853228    -1.85   0.065    -1.467011    .0434263
          x3 |   .6175273    .177546     3.48   0.001     .2695435    .9655112
          x4 |   .1600324   .1925305     0.83   0.406    -.2173205    .5373853
       _cons |  -.7849445   .3029845    -2.59   0.010    -1.378783   -.1911057
------------------------------------------------------------------------------
```

```
. probit y2 x1 x2 x3 x4

Iteration 0:   log likelihood = -368.47383
Iteration 1:   log likelihood = -362.24519
Iteration 2:   log likelihood = -362.23271
Iteration 3:   log likelihood = -362.23271

Probit regression                               Number of obs     =        550
                                                LR chi2(4)        =      12.48
                                                Prob > chi2       =     0.0141
Log likelihood = -362.23271                     Pseudo R2         =     0.0169

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.5130818   .3819133    -1.34   0.179    -1.261618    .2354544
          x2 |  -.7095044   .3776776    -1.88   0.060    -1.449739      .03073
          x3 |  -.2425533    .176711    -1.37   0.170    -.5889005    .1037938
          x4 |   .5990051   .1995009     3.00   0.003     .2079904    .9900197
       _cons |   .2977989   .3017738     0.99   0.324    -.2936668    .8892647
------------------------------------------------------------------------------

. probit y3 x1 x2 x3 x4

Iteration 0:   log likelihood = -333.39437
Iteration 1:   log likelihood =   -327.917
Iteration 2:   log likelihood =  -327.9036
Iteration 3:   log likelihood =  -327.9036

Probit regression                               Number of obs     =        550
                                                LR chi2(4)        =      10.98
                                                Prob > chi2       =     0.0268
Log likelihood =  -327.9036                     Pseudo R2         =     0.0165

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.6271044   .3935507    -1.59   0.111     -1.39845    .1442407
          x2 |   .8273167   .3913541     2.11   0.035     .0602768    1.594357
          x3 |  -.3804693   .1979142    -1.92   0.055    -.7683741    .0074355
          x4 |  -.2350291   .1921388    -1.22   0.221    -.6116142     .141556
       _cons |  -.4836421   .3169133    -1.53   0.127    -1.104781    .1374966
------------------------------------------------------------------------------
```

Answer 1C: in my opinion, we should not use the separate model since the choices are seemingly correlate because the customer must choose only one of them. As well as the models have very low R-square and insignificant in individual test

By the way we should try the multivariate model appropriateness test first before decision.

(d) Estimate models for Y_{1i}, Y_{2i}, and Y_{3i} assuming that the probability functions follow multivariate normal probability distribution function (MV Probit models). Determine whether MVProbit is appropriated. Why? How does the MV Probit models differ from three separate probit models? (6 points)

```
mvprobit (y1 x1 x2 x3 x4) (y2 x1 x2 x3 x4) (y3 x1 x2 x3 x4)

Iteration 0:   log likelihood = -1055.7122
Warning: cannot do Cholesky factorization of rho matrix
Warning: cannot do Cholesky factorization of rho matrix
Iteration 1:   log likelihood = -948.02163
Iteration 2:   log likelihood = -934.40808
Iteration 3:   log likelihood =  -932.8107
Iteration 4:   log likelihood = -932.80435
Iteration 5:   log likelihood = -932.80434

Multivariate probit (SML, # draws = 5)            Number of obs   =        550
                                                  Wald chi2(12)   =      37.21
Log likelihood = -932.80434                       Prob > chi2     =     0.0002
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
y1						
x1	1.171117	.3893354	3.01	0.003	.4080335	1.9342
x2	-.6025208	.3795038	-1.59	0.112	-1.346335	.141293
x3	.6046014	.1786074	3.39	0.001	.2545374	.9546655
x4	.1480072	.1915597	0.77	0.440	-.227443	.5234574
_cons	-.7106655	.3050131	-2.33	0.020	-1.30848	-.1128509
y2						
x1	-.5643036	.3755779	-1.50	0.133	-1.300423	.1718157
x2	-.6138367	.3747629	-1.64	0.101	-1.348359	.1206851
x3	-.2591482	.1782002	-1.45	0.146	-.608414	.0901177
x4	.5646381	.1983093	2.85	0.004	.175959	.9533172
_cons	.3151067	.2995943	1.05	0.293	-.2720874	.9023007
y3						
x1	-.5475637	.3858414	-1.42	0.156	-1.303799	.2086715
x2	.6638133	.3859124	1.72	0.085	-.0925611	1.420188
x3	-.2722606	.1869458	-1.46	0.145	-.6386676	.0941464
x4	-.1936979	.195082	-0.99	0.321	-.5760516	.1886558
_cons	-.4581402	.3194547	-1.43	0.152	-1.08426	.1679795
/atrho21	-.6213012	.0749254	-8.29	0.000	-.7681523	-.47445
/atrho31	-.4110659	.0740692	-5.55	0.000	-.5562389	-.2658929
/atrho32	-.3237628	.0710803	-4.55	0.000	-.4630777	-.1844479
rho21	-.5520333	.0520926	-10.60	0.000	-.6458538	-.4417878
rho31	-.3893774	.0628392	-6.20	0.000	-.5051815	-.2597991
rho32	-.3129053	.0641209	-4.88	0.000	-.4325893	-.1823843

```
Likelihood ratio test of  rho21 = rho31 = rho32 = 0:
             chi2(3) =   245.816    Prob > chi2 = 0.0000
.
```

Answer 1D: from the estimated result

We can reject H0 from Lr test which mean that there's correlation between dependent variable so,using the multivariate probit model is appropriated over separate model.

It's differenct because we run the model which assume that we have the correlated dependent variable instead of non-correlated. And in this case the data are following that assumption making mvprobit is appropriated.

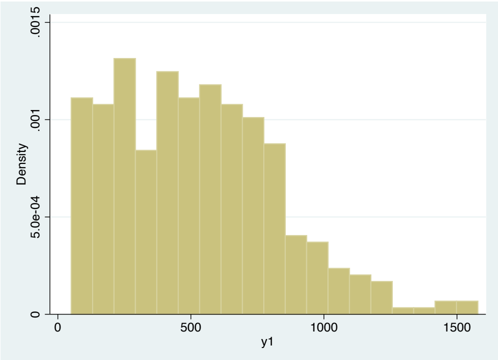
Question 2

From the data set "Final_q2_no.dta": According to the following model,

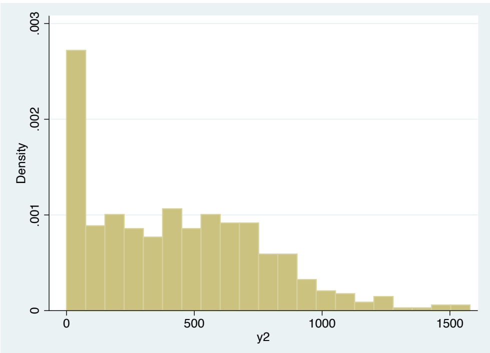
$y_{ki} = \beta_{k0} + \beta_{k1}x_i + u_{ki}$ (3) where: y_{ki} is Dependent variable, k=1, 2, 3. x_i is Independent variable u_i is Stochastic disturbance term

(a) Plot histogram of y_{1i}, y_{2i}, y_{3i}, compute descriptive statistics of these three variables. Determine limitations of these three dependent variables. (5 points)

```
hist y1
(bin=19, start=50.217594, width=80.436449)
```

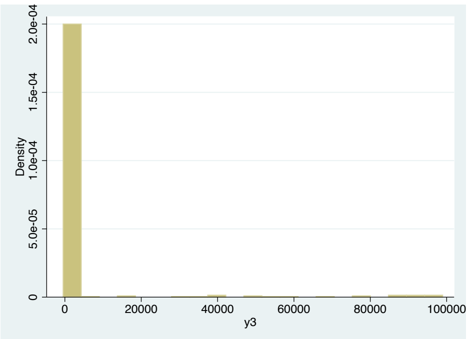



```
sum y1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          y1 |        369    522.0268    305.5502   50.21759    1578.51

hist y2
(bin=21, start=0, width=75.167149)
```




```
. sum y2

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          y2 |        450     429.054    340.6061          0    1578.51
```

```
hist y3
(bin=21, start=-450.62051, width=4733.4406)
```




```
. sum y3

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          y3 |        450    3648.754    15405.28  -450.6205   98951.63
```

Answer2A:

For y1 seem like we have the left truncated data limitation according to the histogram which lower limit is around 50
For y2 we have censored data limitation since we have a lot of 0 value in y2
For y3 we have outlier limitation

b) Estimate the model (3) for y_{1i} using OLS and truncated regression model. Determine the most appropriated model in this case. Why? What is the major problem in this case? What will happen if we ignore the problem? (5 points)

run the model using OLS

```
reg y1 x

      Source |       SS           df       MS      Number of obs   =       369
-------------+----------------------------------   F(1, 367)       =    152.11
       Model |  10067286.6         1  10067286.6   Prob > F        =    0.0000
    Residual |  24289539.5       367  66184.0312   R-squared       =    0.2930
-------------+----------------------------------   Adj R-squared   =    0.2911
       Total |  34356826.1       368  93360.9404   Root MSE        =    257.26

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |    171.428    13.8996    12.33   0.000     144.0952    198.7609
       _cons |  -32.15492   46.88712    -0.69   0.493     -124.356     60.0462
------------------------------------------------------------------------------

. est store m_y1hat
```

run the model using truncated regression

```
. truncreg y1 x, ll(50) nolog
(note: 0 obs. truncated)

Truncated regression
Limit:   lower =         50                     Number of obs     =        369
         upper =       +inf                     Wald chi2(1)      =     112.14
Log likelihood = -2532.9115                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   243.5473   22.99863    10.59   0.000     198.4709    288.6238
       _cons |  -350.6846   88.77073    -3.95   0.000    -524.6721   -176.6972
-------------+----------------------------------------------------------------
```

```
      /sigma |   307.6565   17.05061    18.04   0.000     274.2379    341.0751
------------------------------------------------------------------------------

. est store m_y1trunc
```

Perform the appropriateness test

```
. lrtest m_y1hat m_y1trunc ,force

Likelihood-ratio test                                 LR chi2(1)  =     75.32
(Assumption: m_y1hat nested in m_y1trunc)             Prob > chi2 =    0.0000
```

Answer2B:

According to sample selection bias test, we can reject the H0 which mean that the truncated regression is the appropriated model.
And in the case of using OLS, ignore the problem we will have Bias estimated results since the data is not normal distributed

c) Estimate the model (3) for y_{2i} using OLS and Tobit model. Determine the most appropriated model in this case. Why? What is the major problem in this case? What will happen if we ignore the problem? (5 points)

run the model using OLS

```
reg y2 x

      Source |       SS           df       MS      Number of obs   =       450
-------------+----------------------------------   F(1, 448)       =    276.04
       Model |  19859200.8         1  19859200.8   Prob > F        =    0.0000
    Residual |  32230409.6       448  71942.8785   R-squared       =    0.3813
-------------+----------------------------------   Adj R-squared   =    0.3799
       Total |  52089610.4       449  116012.495   Root MSE        =    268.22

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   202.2362   12.17228    16.61   0.000     178.3143     226.158
       _cons |  -183.0503   38.95095    -4.70   0.000    -259.5996   -106.5011
------------------------------------------------------------------------------

. est store m_y2hat
```

run the model using Tobit

```
. tobit y2 x, ll(0)

Tobit regression                                Number of obs     =        450
                                                LR chi2(1)        =     224.96
                                                Prob > chi2       =     0.0000
Log likelihood = -2793.2596                     Pseudo R2         =     0.0387

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   237.8077   14.46312    16.44   0.000     209.3839    266.2315
       _cons |   -320.499   47.07641    -6.81   0.000    -413.0165   -227.9816
-------------+----------------------------------------------------------------
      /sigma |   299.6877   11.07962                      277.9134    321.4621
------------------------------------------------------------------------------
           66  left-censored observations at y2 <= 0
          384     uncensored observations
            0 right-censored observations

. est store m_y2censor
```

Perform the appropriateness test

```
. lrtest m_y2hat m_y2censor, force

Likelihood-ratio test                                 LR chi2(1)  =    721.15
(Assumption: m_y2hat nested in m_y2censor)            Prob > chi2 =    0.0000

.
```

Answer2C:

According to test, we can reject the H0 which mean that the Tobit model is the appropriated model.
And in the case of using OLS, ignore the problem of censored data we will have Bias estimated results since the data is not normal distributed

(d) Estimate the model (3) for y_{3i} using OLS and Tobit model. Determine the most appropriated model in this case. Why? What is the major problem in this case? What will happen if we ignore the problem? (5 points)

run the model using OLS

```
reg y3 x

      Source |       SS           df       MS      Number of obs   =       450
-------------+----------------------------------   F(1, 448)       =     42.00
       Model |  9.1336e+09         1  9.1336e+09   Prob > F        =    0.0000
    Residual |  9.7424e+10       448   217464814   R-squared       =    0.0857
-------------+----------------------------------   Adj R-squared   =    0.0837
       Total |  1.0656e+11       449   237322670   Root MSE        =     14747

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   4337.105   669.2255     6.48   0.000     3021.894    5652.316
       _cons |  -9478.277   2141.503    -4.43   0.000    -13686.92   -5269.639
------------------------------------------------------------------------------

. est store m_y3hat
```

run the model using Tobit

```
. tobit y3 x, ul(2000)

Tobit regression                                Number of obs     =        450
                                                LR chi2(1)        =     197.15
                                                Prob > chi2       =     0.0000
Log likelihood = -3206.3445                     Pseudo R2         =     0.0298

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   294.0083    18.7774    15.66   0.000     257.1058    330.9108
       _cons |  -435.0138   59.90649    -7.26   0.000    -552.7457   -317.2818
-------------+----------------------------------------------------------------
      /sigma |   410.5882   14.36172                      382.3636    438.8127
------------------------------------------------------------------------------
              0  left-censored observations
            426     uncensored observations
             24 right-censored observations at y3 >= 2000

. est store m_y3outlier
```

Perform the appropriateness test

```
. lrtest m_y3hat m_y3outlier,force

Likelihood-ratio test                                 LR chi2(1)  =   3501.25
(Assumption: m_y3hat nested in m_y3outlier)           Prob > chi2 =    0.0000

.
```

Answer2D:

According to test, we can reject the H0 which mean that the Tobit model is the appropriated model.
And in the case of using OLS, ignore the problem of outlier we will have Bias estimated results since the data is not normal distributed by the way we can solve the problem with OLS with the limitation as well but the estimated data of tobit model will be still more accurate.

Question3

From the data set "Final_q3_no.dta":
The generalized linear regression model can be stated as:

and where:

$$I_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i \quad (4)$$

$$\Pr(Y_i = y_i) = f(I_i)$$

I_i is index variables.
y_i is counted number 0, 1, 2,... x_{ki} is independent variable k.

$f(.)$ is either Poisson or Negative Binomial probability distribution function. u_i is disturbance term.

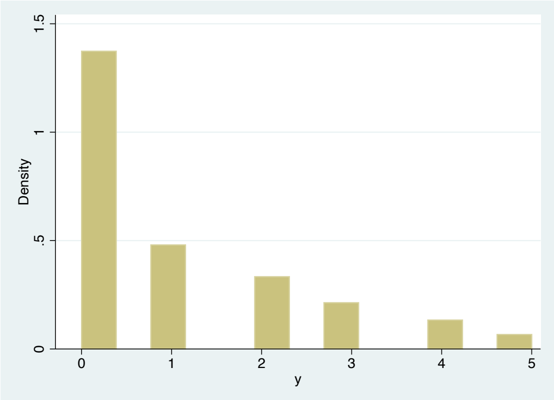
(a) Estimate models for Y_i assuming that the model is traditional linear regression model. Create histogram for Y_i. Determine whether there is limitation of dependent variable in this case. If yes, what type of limitation is it? (3 points)

Estimate the model using OLS

```
reg y x1 x2 x3 x4

      Source |       SS           df       MS      Number of obs   =       195
-------------+----------------------------------   F(4, 190)       =      5.75
       Model |  39.0914001         4  9.77285004   Prob > F        =    0.0002
    Residual |  322.826549       190   1.6990871   R-squared       =    0.1080
-------------+----------------------------------   Adj R-squared   =    0.0892
       Total |  361.917949       194  1.86555644   Root MSE        =    1.3035

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0850608   .0447949     1.90   0.059    -.0032985    .1734201
          x2 |   .1381494    .047924     2.88   0.004     .0436179    .2326809
          x3 |   .2165669   .0693506     3.12   0.002      .079771    .3533628
          x4 |  -.0356706   .0491106    -0.73   0.469    -.1325426    .0612014
       _cons |   .9640755   .1109524     8.69   0.000     .7452187    1.182932
------------------------------------------------------------------------------
```

Histogram of y

```
. hist y
(bin=13, start=0, width=.38461538)
```



Answer3A:

Yes, there's the limitation of dependent variable which is the count data
Which using OLS may causes in negative value prediction which make no sense

(b) Estimate models for Y_i assuming that the probability functions follow Poisson probability distribution. Perform GOF test and determine whether Poisson is appropriated in this case. Interpret the estimated result (sign and meaning (in term of incidence-rate ratios), overall test, individual test, pseudo R , marginal effects). (5 points)

Estimate the model using poisson

```
poisson y x1 x2 x3 x4, nolog

Poisson regression                              Number of obs     =        195
                                                LR chi2(4)        =      38.94
                                                Prob > chi2       =     0.0000
Log likelihood = -277.20544                     Pseudo R2         =     0.0656

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0839503   .0335816     2.50   0.012     .0181315     .149769
          x2 |   .1356394   .0359108     3.78   0.000     .0652555    .2060234
          x3 |   .2241043   .0545131     4.11   0.000     .1172605     .330948
          x4 |  -.0355363   .0372572    -0.95   0.340     -.108559    .0374865
       _cons |  -.1320593   .0924644    -1.43   0.153    -.3132863    .0491677
------------------------------------------------------------------------------
```

Perform GOF test

```
. estat gof

         Deviance goodness-of-fit =  319.1382
         Prob > chi2(190)         =    0.0000

         Pearson goodness-of-fit  =  332.0219
         Prob > chi2(190)         =    0.0000
```

Find the irr

```
. lincom x1, ir

 ( 1)  [y]x1 = 0

------------------------------------------------------------------------------
           y |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   1.087575   .0365225     2.50   0.012     1.018297    1.161566
------------------------------------------------------------------------------

. lincom x2, ir

 ( 1)  [y]x2 = 0

------------------------------------------------------------------------------
           y |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   1.145269   .0411276     3.78   0.000     1.067432    1.228782
------------------------------------------------------------------------------

. lincom x3, ir

 ( 1)  [y]x3 = 0

------------------------------------------------------------------------------
           y |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   1.251201   .0682069     4.11   0.000     1.124412    1.392287
------------------------------------------------------------------------------

. lincom x4, ir

 ( 1)  [y]x4 = 0

------------------------------------------------------------------------------
           y |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .9650877   .0359565    -0.95   0.340     .8971259    1.038198
```

Marginal effect from mean

```
mfx

Marginal effects after poisson
      y  = Predicted number of events (predict)
         =  .92331529
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .0775126       .0306    2.53   0.011   .017545   .13748   -.28246
      x2 |    .125238       .0322    3.89   0.000   .062134  .188342   .842848
      x3 |   .2069189      .04827    4.29   0.000   .112306  .301532  -.298846
      x4 |  -.0328112      .03432   -0.96   0.339  -.100086  .034463  -.805838
------------------------------------------------------------------------------

.
```

Answer3B: Intepretation

Appropriateness
From the GOF test we reject the H0 which mean that the Poisson model is not the appropriated one

Sign&meaning

IRR
From IRR all of the independent variables, x1, x2, x3 and x4, are having the positive impact on dependent variable, y
Marginal effect
the marginal effect from mean x1,x2,x3 have the positive impact on y value but x4 has the negative impact on y

Overall test:
The model is jointly significant in overall test at 95% confidence (since p-value < 0.05)

Individual test:
Most of the parameters including cut, cut2 are significant z-test except x1

Goodness of fit
Pseudo R-square is quite low which mean that the model has the room to be improved

(c) Estimate models for Y_i assuming that the probability functions follow Negative Binomial probability distribution. Determine whether Negative Binomial regression model is appropriated in this case. Interpret your estimated result (sign and meaning (in term of incidence-rate ratios), overall test, individual test, pseudo R , marginal effects). (5 points)

Estimate the model using negative binomial regression

```
. nbreg y x1 x2 x3 x4, nolog

Negative binomial regression                    Number of obs     =        195
                                                LR chi2(4)        =      20.86
Dispersion     = mean                           Prob > chi2       =     0.0003
Log likelihood = -262.50416                     Pseudo R2         =     0.0382

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .1119462   .0518511     2.16   0.031     .0103199    .2135725
          x2 |   .1546279   .0519008     2.98   0.003     .0529041    .2563516
          x3 |   .2156335   .0708465     3.04   0.002     .0767769    .3544901
          x4 |  -.0389698   .0499954    -0.78   0.436     -.136959    .0590193
       _cons |  -.1533718   .1221099    -1.26   0.209    -.3927028    .0859592
-------------+----------------------------------------------------------------
    /lnalpha |   -.201489   .2917749                     -.7733572    .3703792
-------------+----------------------------------------------------------------
       alpha |   .8175126   .2385296                      .4614612    1.448284
------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 29.40               Prob >= chibar2 = 0.000
```

Find the IRR

```
 nbreg y x1 x2 x3 x4,ir nolog

Negative binomial regression                    Number of obs     =        195
                                                LR chi2(4)        =      20.86
Dispersion     = mean                           Prob > chi2       =     0.0003
Log likelihood = -262.50416                     Pseudo R2         =     0.0382

------------------------------------------------------------------------------
           y |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.118453    .057993     2.16   0.031     1.010373    1.238093
          x2 |   1.167224   .0605799     2.98   0.003     1.054329    1.292207
          x3 |   1.240648   .0878956     3.04   0.002     1.079801    1.425454
          x4 |   .9617797   .0480845    -0.78   0.436      .872006    1.060796
       _cons |   .8578107   .1047472    -1.26   0.209     .6752294    1.089762
-------------+----------------------------------------------------------------
    /lnalpha |   -.201489   .2917749                     -.7733572    .3703792
-------------+----------------------------------------------------------------
       alpha |   .8175126   .2385296                      .4614612    1.448284
------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 29.40                Prob >= chibar2 = 0.000
```

Marginal effect

```
. mfx

Marginal effects after nbreg
      y  = Predicted number of events (predict)
         =  .91603373
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .1025465      .04734    2.17   0.030    .00976  .195333   -.28246
      x2 |   .1416443      .04734    2.99   0.003   .048867  .234422   .842848
      x3 |   .1975276      .06457    3.06   0.002   .070968  .324087  -.298846
      x4 |  -.0356977      .04579   -0.78   0.436  -.125448  .054053  -.805838
------------------------------------------------------------------------------
```

.

Answer3C: Intepretation

Appropriateness

From the overdispersion test we reject the H0 which mean that the negative binomial model is appropriated

Sign&meaning

IRR

From IRR all of the independent variables, x1, x2, x3 and x4, are having the positive impact on dependent variable, y

Marginal effect

the marginal effect from mean x1,x2,x3 have the positive impact on y value but x4 has the negative impact on y

Overall test:

The model is jointly significant in overall test at 95% confidence (since p-value < 0.05)

Individual test:

all the parameters are significant in z-test except x4

Goodness of fit

Pseudo R-square is quite low which mean that the model has the room to be improved

(d) Estimate models for Y_i assuming that the model is Zero Inflated Poisson (x_{1i}, x_{2i}, and x_{3i} are independent variables in Poisson model and x_{4i} is independent variable in Inflated (Logit) model). Interpret your estimated result. Determine which model (Linear regression model, Poisson, Negative Binomial, or ZIP) is the most appropriated model in this case? Why? (provide the tests) (7 points)

Estimate the model using zero inflated poisson

```
zip y x1 x2 x3 , inflate(x4) vuong

Fitting constant-only model:

Iteration 0:   log likelihood = -295.02298
Iteration 1:   log likelihood = -269.78844
Iteration 2:   log likelihood = -268.15628
Iteration 3:   log likelihood = -268.14904
Iteration 4:   log likelihood = -268.14904

Fitting full model:

Iteration 0:   log likelihood = -268.14904
Iteration 1:   log likelihood = -260.90347
Iteration 2:   log likelihood = -260.50346
Iteration 3:   log likelihood = -260.50243
Iteration 4:   log likelihood = -260.50243

Zero-inflated Poisson regression                Number of obs     =        195
                                                Nonzero obs       =         92
                                                Zero obs          =        103

Inflation model = logit                         LR chi2(3)        =      15.29
Log likelihood  = -260.5024                     Prob > chi2       =     0.0016

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
          x1 |   .0991548   .0425576     2.33   0.020     .0157434    .1825661
          x2 |   .1208637   .0400449     3.02   0.003     .0423772    .1993502
          x3 |   .1553467   .0562608     2.76   0.006     .0450775    .2656159
       _cons |   .3521057   .1147728     3.07   0.002     .1271551    .5770564
-------------+----------------------------------------------------------------
inflate      |
          x4 |   .0579183   .1008572     0.57   0.566    -.1397582    .2555947
       _cons |   -.511083   .2578789    -1.98   0.047    -1.016516   -.0056497
------------------------------------------------------------------------------
Vuong test of zip vs. standard Poisson:            z =     2.57  Pr>z = 0.0051

. mfx

Marginal effects after zip
      y  = Predicted number of events (predict)
         =  .92949079
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .0921635      .03931    2.34   0.019   .015117   .16921   -.28246
      x2 |   .1123417      .03649    3.08   0.002   .040819  .183864   .842848
      x3 |   .1443933      .05102    2.83   0.005   .044397  .244389  -.298846
      x4 |  -.0195994      .03441   -0.57   0.569  -.087044  .047846  -.805838
------------------------------------------------------------------------------
```

<u>Answer3D</u>:

Intepretation of ZIP

Appropriateness
From the vuong test we reject the H0 which mean that the zero inflated poisson is appropriated.

Sign&meaning (Marginal effect)
the marginal effect from mean x1,x2,x3 have the positive impact on y value but x4 has the negative impact on y

Overall test:
The model is jointly significant in overall test at 95% confidence (since p-value < 0.05)

Individual test:
all the parameters are significant in z-test except x4

Goodness of fit
Pseudo R-square is quite low which mean that the model has the room to be improved

The most appropriate model: according to the GOF test, overdispersion test and vuong test The appropriate models are negative binomial and zero-inflated poisson by the way we have to compare in detailed between them
1.sign&meaning => both are similar
2.overall test => both are jointly significant
3.individual test => both are having insignificant in x4
4.log-likelihood value => the negative binomial has the lowest log-likelihood value
Making it's the most appropriated model.

<u>Question4</u>
From the data set "`Final_q4_no.dta`":

(a) Test whether the series y_t and x_t are stationary series and determine their order of integration. Give explanation of the process of the test. Why is testing equation important? Why is stationary property important to time series analysis? (5 points)

<u>ADF test for y</u>

```
dfuller y, trend lags(1) regress

Augmented Dickey-Fuller test for unit root         Number of obs   =       498

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -15.067            -3.980            -3.420            -3.130
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

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

```
         L1. |  -.9740824   .0646506   -15.07   0.000    -1.101106   -.8470584
         LD. |  -.0594414   .0449986    -1.32   0.187    -.1478537     .028971
      _trend |   .0016711   .0021088     0.79   0.428    -.0024722    .0058144
       _cons |   .8261467   .6117395     1.35   0.177    -.3757854    2.028079
------------------------------------------------------------------------------
```

ADF test for x

```
. dfuller x, trend lags(1) regress

Augmented Dickey-Fuller test for unit root         Number of obs   =       498

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -15.219            -3.980            -3.420            -3.130
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
D.x          |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |
         L1. |  -.9857164    .064769   -15.22   0.000    -1.112973   -.8584597
         LD. |   -.050295   .0450108    -1.12   0.264    -.1387313    .0381413
      _trend |   .0026639   .0030113     0.88   0.377    -.0032527    .0085804
       _cons |   .4094375   .8696255     0.47   0.638    -1.299183    2.118058
------------------------------------------------------------------------------

.
```

Answer4A:

From the ADF test we can reject H0 in both variable which mean that we don't have the unit root in those variables. So, they are stationary data in the other word they're first order integration.

For the ADF test process we check the regression of 1st difference and lag term in the case whether we have any relationship left in them.
If we have the relationship between them, we don't have unit root
But if we have no relationship left, we have the unit root

The testing equation is important because misspecification error will lead to bias in the estimated result. So, we cannot summarize anything from the test.

In time series analysis, the stationary condition is important because
the impact of non-stationary will result in spurious regression leading to wrong conclusion.

(b) Estimate Autoregressive Integrated Moving Average (ARIMA(p,d,q)) model for y_t – determine the most appropriated order for p, d, and q using SBIC given the maximum lag equals 4. Make dynamic forecast for period time = 501 to 505. (5 points)

Answer4B:

Find the appropriated order of ARIMA

```
forvalues p = 1(1)4{
  2. forvalues q = 1(1)4{
  3.                 display " estimate arima`p'0`q'"
  4.                 qui arima y, arima(`p',0,`q') nolog
  5.                 est store arima`p'0`q'
  6. }
  7. }
 estimate arima101
 estimate arima102
 estimate arima103
 estimate arima104
 estimate arima201
 estimate arima202
 estimate arima203
 estimate arima204
 estimate arima301
 estimate arima302
 estimate arima303
 estimate arima304
 estimate arima401
 estimate arima402
 estimate arima403
 estimate arima404

.
end of do-file

. est table arima*,star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

------------------------------------------------------------------------------
    Variable |   arima101        arima102        arima103        arima104
-------------+----------------------------------------------------------------
y            |
       _cons |   1.301666***    1.3074418***    1.3031432***    1.3040346***
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. | -.86557047***    .81429377***   -.86427274***   -.82909595***
         L2. |
         L3. |
         L4. |
             |
          ma |
         L1. |  .82439723***   -.84805452***    .83578301***     .8009985***
         L2. |                  .06903768       .03064571       .02950302
         L3. |                                  .01809936       .04960661
         L4. |                                                   .0475399
-------------+----------------------------------------------------------------
sigma        |
       _cons |  6.7289966***    6.7299216***    6.7268287***    6.7198324***
-------------+----------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -1662.6926      -1662.7619       -1662.536      -1662.0178
        chi2 |  58.685883       22.152105       58.420897       49.186114
         aic |  3333.3852       3335.5239       3337.0719       3338.0356
         bic |  3350.2436       3356.5969       3362.3596       3367.5378
------------------------------------------------------------------------------
                                           legend: * p<.1; ** p<.05; *** p<.01


------------------------------------------------------------------------------
    Variable |   arima201        arima202        arima203        arima204
-------------+----------------------------------------------------------------
y            |
       _cons |  1.3073551***    1.3075824***    1.3076332***     1.308031***
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  .72833656***    .01069941      -.02391453      -.03883203
         L2. |  .07343696       .73210465***    .72886425***    .68853566**
```

```
         L3. |
         L4. |
             |
          ma |
         L1. | -.76420028***   -.02937305      -.00720022       .00857444
         L2. |                  -.6650084**    -.66251356**    -.63712588**
         L3. |                                  .02176552       .02159808
         L4. |                                                  .02591727
-------------+----------------------------------------------------------------
sigma        |
       _cons |  6.7292866***     6.717086***    6.7159767***    6.7140828***
-------------+----------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -1662.7092      -1661.8275      -1661.7334      -1661.6033
        chi2 |  19.097086       22.261236       23.402544       22.065999
         aic |  3335.4184       3335.6549       3337.4668       3339.2066
         bic |  3356.4915       3360.9426       3366.9691       3372.9234
------------------------------------------------------------------------------
                                           legend: * p<.1; ** p<.05; *** p<.01


------------------------------------------------------------------------------
    Variable |   arima301        arima302        arima303        arima304
-------------+----------------------------------------------------------------
y            |
       _cons |  1.3030581***      1.30767***    1.3222961***    1.3224072***
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. | -.85123464***   -.05144311       1.0589742***    1.0600046***
         L2. |  .03402859       .72972779***    .66956782**     .66743078*
         L3. |  .02022783       .02220492      -.84473861***   -.84357622***
         L4. |
             |
          ma |
         L1. |  .82424826***    .02123612      -1.1139831***   -1.1138522***
         L2. |                 -.66391706**    -.57543987*     -.57405983
         L3. |                                  .81620891***    .81322241***
         L4. |                                                  .00161096
-------------+----------------------------------------------------------------
sigma        |
       _cons |  6.7267628***    6.7160989***    6.6327376       6.6327349
-------------+----------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -1662.5206      -1661.7401      -1657.5933      -1657.5927
        chi2 |  59.442039       24.064744       55190.708       54913.937
         aic |  3337.0413       3337.4802       3329.1866       3331.1855
         bic |  3362.3289       3366.9825       3358.6889       3364.9024
------------------------------------------------------------------------------
                                           legend: * p<.1; ** p<.05; *** p<.01


------------------------------------------------------------------------------
    Variable |   arima401        arima402        arima403        arima404
-------------+----------------------------------------------------------------
y            |
       _cons |   1.304488***    1.3077626***     1.322444***    1.3085509***
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. | -.78736421***   -.07222081       1.0619352***     .06379198
         L2. |  .03406615       .65323856*      .66541253*     -.23017888
         L3. |  .05230227        .0238182      -.84491718***   -.06553786
         L4. |  .04891326       .02733522       .00165465       .70904435***
             |
          ma |
         L1. |  .75932045***    .04153469      -1.1157865      -.09628464
         L2. |                  -.6018471      -.57193745       .31341447
         L3. |                                  .81439767       .06673855
         L4. |                                                 -.65853127**
-------------+----------------------------------------------------------------
sigma        |
       _cons |  6.7195128***     6.714347***    6.6327306       6.6743326***
-------------+----------------------------------------------------------------
Statistics   |
```

```
           N |        500             500             500             500
          ll | -1661.9967      -1661.6088      -1657.5927      -1658.9217
        chi2 |  41.886103       21.434333       38010.909       1620.7592
         aic |  3337.9934       3339.2176       3331.1855       3337.8434
         bic |  3367.4956       3372.9345       3364.9024       3379.9895
------------------------------------------------------------------------------
                                           legend: * p<.1; ** p<.05; *** p<.01
```

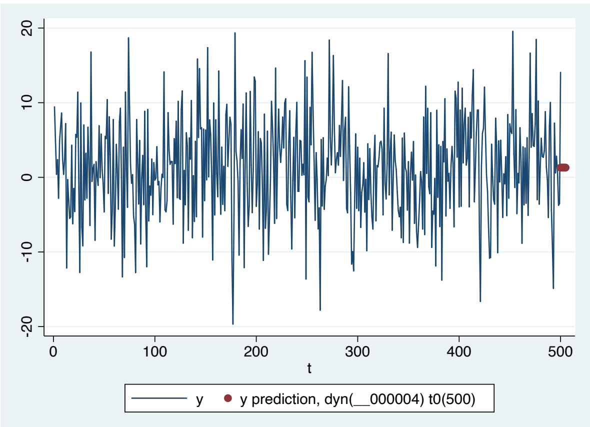
the most appropriated order is ARIMA(1,0,1) since it's the lowest BIC

Estimate the arima(1,0,1)

```
arima y, arima(1,0,1) nolog

ARIMA regression

Sample:  1 - 500                                Number of obs     =        500
                                                Wald chi2(2)      =      58.69
Log likelihood = -1662.693                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
       _cons |   1.301666   .2963809     4.39   0.000     .7207701    1.882562
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.8655705   .1555056    -5.57   0.000    -1.170356    -.560785
             |
          ma |
         L1. |   .8243972   .1763698     4.67   0.000     .4787187    1.170076
-------------+----------------------------------------------------------------
      /sigma |   6.728997   .2146963    31.34   0.000       6.3082    7.149794
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

Dynamics Forecasting

```
. set obs 505
number of observations (_N) was 500, now 505

. replace t=_n
(5 real changes made)

. predict y, y dynamic(.) t0(500)
variable y already defined
r(110);

. predict yhat, y dynamic(.) t0(500)
Note: beginning dynamic predictions in period         3.
(499 missing values generated)

. twoway(line y t,sort) (scatter yhat t, sort)
```



From the following GARCH model:

Mean Equation: VarianceEquation:

.

.

(c) Estimate model (5) using OLS by employing y_t as dependent variable and x_t as explanatory variable, determine whether ARCH-effect significantly occurs, and state null hypothesis of the test. Why the test is classified as LM test? (3 points)

Estimate model using OLS

```
reg y x

      Source |       SS           df       MS      Number of obs   =       500
-------------+----------------------------------   F(1, 498)       =  57285.96
       Model |  22596.1563         1  22596.1563   Prob > F        =    0.0000
    Residual |  196.433573       498  .394444925   R-squared       =    0.9914
-------------+----------------------------------   Adj R-squared   =    0.9914
       Total |  22792.5899       499  45.6765328   Root MSE        =    .62805

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .6977603   .0029153   239.34   0.000     .6920325    .7034881
       _cons |   .5175526    .028278    18.30   0.000     .4619937    .5731114
------------------------------------------------------------------------------
```

ARCH effect test

```
. estat archlm
LM test for autoregressive conditional heteroskedasticity (ARCH)
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |         38.506               1                   0.0000
---------------------------------------------------------------------------
         H0: no ARCH effects       vs.  H1: ARCH(p) disturbance
```

.

Answer4C

From the ARCH effect LM test, we can reject H0, H0: no ARCH effects, which mean that we do have the ARCH effect in our case.

The reason that its's LM test because the test is based on restricted model which is OLS

(d) Estimate GARCH(p,q) for y_t using x_t as explanatory variable for mean equation (model (5) and (6)) – determine the most appropriated order p and q for variance equation using SBIC given the maximum lag equals to 2. Then, predict the variance of y_t using the estimated result of GARCH(p,q) model with the most appropriated lag. (4 points)

Answer4D :

Find the most appropriated order

```
forvalues p = 1(1)2{
  2. forvalues q = 1(1)2{
  3.                 display " estimate garch`p'`q'"
  4.                 qui arch y x, garch(1/`p') arch(1/`q') nolog
  5.                 est store garch`p'`q'
  6. }
  7. }
 estimate garch11
 estimate garch12
 estimate garch21
 estimate garch22

.
end of do-file

. est table garch* , star(0.1 0.05 0.01) stat(N ll chi2 aic bic)
```

Variable	garch11	garch12	garch21	garch22
y				
x	.69788579***	.69787087***	.69787574***	.69777417***
_cons	.51826312***	.5182283***	.51823952***	.51597981***
ARCH				
arch				
L1.	.36201351***	.36039806***	.36077406***	.37189936***
L2.		.01483425		.30792554**
garch				
L1.	.31635697***	.28933812	.32663124*	-.39040448
L2.			-.00927871	.1071777
_cons	.12891287***	.13442533*	.12900707***	.24574265***
Statistics				
N	500	500	500	500
ll	-444.30123	-444.29481	-444.29646	-443.46007
chi2	72628.104	72713.019	72680.391	74301.891
aic	898.60246	900.58963	900.59293	900.92015
bic	919.67551	925.87728	925.88058	930.4224

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

based on our estimated result GARCH (1,1) is the most appropriated order since it's the lowest BIC among the others.

Run the model using GARCH(1,1)

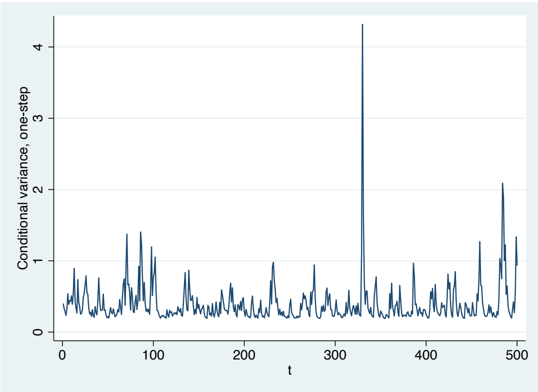
```
arch y x, garch(1/1) arch(1/1) nolog

ARCH family regression

Sample: 1 - 500                                 Number of obs     =        500
Distribution: Gaussian                          Wald chi2(1)      =   72628.10
Log likelihood = -444.3012                      Prob > chi2       =     0.0000
```

y	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
y						
x	.6978858	.0025896	269.50	0.000	.6928103	.7029613
_cons	.5182631	.02399	21.60	0.000	.4712436	.5652826

```
ARCH         |
        arch |
         L1. |   .3620135   .0745312     4.86   0.000      .215935     .508092
             |
       garch |
         L1. |    .316357   .1209539     2.62   0.009     .0792917    .5534223
             |
       _cons |   .1289129   .0367809     3.50   0.000     .0568237     .201002
------------------------------------------------------------------------------
```

Predict the Variance

```
. predict sigma2, v

. line sigma2 t
```




```
.
```

(e) Among these three models, ARCH(1), GARCH(1,1), or EGARCH(1,1,1), which model is the most appropriated in this case? Why? What are the differences among ARCH, GARCH, and EGARCH? (3 points)

```
arch y x, arch(1/1) nolog

ARCH family regression

Sample: 1 - 500                                 Number of obs     =        500
Distribution: Gaussian                          Wald chi2(1)      =   66110.08
Log likelihood = -449.3242                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
           x |   .6978037   .0027139   257.12   0.000     .6924844    .7031229
       _cons |   .5142993   .0246978    20.82   0.000     .4658924    .5627062
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |   .3814361   .0785313     4.86   0.000     .2275175    .5353547
             |
       _cons |   .2422225   .0222802    10.87   0.000     .1985541    .2858908
------------------------------------------------------------------------------

. est store arch1

. arch y x, garch(1/1) arch(1/1) nolog

ARCH family regression

Sample: 1 - 500                                 Number of obs     =        500
Distribution: Gaussian                          Wald chi2(1)      =   72628.10
Log likelihood = -444.3012                      Prob > chi2       =     0.0000

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

```
-------------+----------------------------------------------------------------
y            |
           x |   .6978858   .0025896   269.50   0.000     .6928103    .7029613
       _cons |   .5182631     .02399    21.60   0.000     .4712436    .5652826
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |   .3620135   .0745312     4.86   0.000      .215935     .508092
             |
       garch |
         L1. |    .316357   .1209539     2.62   0.009     .0792917    .5534223
             |
       _cons |   .1289129   .0367809     3.50   0.000     .0568237     .201002
------------------------------------------------------------------------------

. est store garch11

. arch y x, garch(1/1) arch(1/1) egarch(1/1) nolog

ARCH family regression

Sample: 1 - 500                                 Number of obs   =        500
Distribution: Gaussian                          Wald chi2(1)    =   66810.47
Log likelihood = -447.7399                      Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
           x |   .6980592   .0027007   258.48   0.000      .692766    .7033524
       _cons |   .5115844   .0250777    20.40   0.000      .462433    .5607358
-------------+----------------------------------------------------------------
ARCH         |
      egarch |
         L1. |   .3632161   .1241115     2.93   0.003     .1199621    .6064701
             |
        arch |
         L1. |   .6710909   .0925339     7.25   0.000     .4897278     .852454
             |
       garch |
         L1. |  -.0061248   .0106994    -0.57   0.567    -.0270953    .0148457
             |
       _cons |  -.9202979   .1570814    -5.86   0.000    -1.228172    -.612424
------------------------------------------------------------------------------

. est store egarch111

. est table arch1 garch11 egarch11, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)
estimation result egarch11 not found
r(111);

. est table arch1 garch11 egarch111, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

--------------------------------------------------------------
    Variable |    arch1          garch11        egarch111     
-------------+------------------------------------------------
y            |
           x |  .69780366***    .69788579***    .69805918***  
       _cons |  .51429931***    .51826312***    .51158439***  
-------------+------------------------------------------------
ARCH         |
        arch |
         L1. |  .38143612***    .36201351***    .67109089***  
             |
       garch |
         L1. |                  .31635697***   -.00612483     
             |
      egarch |
         L1. |                                  .36321612***  
             |
       _cons |  .24222246***    .12891287***   -.92029794***  
-------------+------------------------------------------------
Statistics   |
           N |        500             500             500     
          ll | -449.32421      -444.30123      -447.73993     
        chi2 |  66110.083       72628.104       66810.474     
```

```
        aic |  906.64842       898.60246       907.47987
        bic |  923.50685       919.67551       932.76752
--------------------------------------------------------------
                           legend: * p<.1; ** p<.05; *** p<.01
```

Answer4E:

According to the estimated result, the model GARCH (1,1) is the most appropriated among them since it's the lowest BIC

The differences among them:
ARCH(1) => the volatility is depend on the shocks only (similar to MA model)
GARCH(1,1) => the volatility is depend on the shocks and also itself from the past (similar to ARMA)
EGARCH(1,1,1) => similar to GARCH but using the exponential function.

Question5

From the data set "`Final_q5_no.dta`": The following VARs models:

$$Y_t = A_0 + A_1 Y_{t-1} + \in_t \qquad (7)$$

where: $Y_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix}$, $A_0 = \begin{pmatrix} a_{10} \\ a_{20} \end{pmatrix}$, $A_1 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, $\in_t = \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$

a) Estimate VARs models using y_t and x_t as endogenous variables and determine the most appropriated lags models using SBIC. (5 points).

Answer5A

Determine the appropriated lag

```
varsoc y x

   Selection-order criteria
   Sample:  5 - 500                            Number of obs      =       496
  +---------------------------------------------------------------------------+
  |lag |    LL      LR      df    p      FPE       AIC      HQIC      SBIC    |
  |----+----------------------------------------------------------------------|
  |  0 | -1366.82                      .855276   5.51942   5.52608   5.53638  |
  |  1 | -1319.99  93.661*   4  0.000  .719617*  5.34672*  5.36669*  5.3976*  |
  |  2 | -1319.24  1.4968    4  0.827  .729115   5.35983   5.39312   5.44464  |
  |  3 | -1318.29  1.8893    4  0.756  .738155   5.37215   5.41876   5.49088  |
  |  4 | -1316.19  4.2107    4  0.378  .743819   5.37979   5.43971   5.53245  |
  +---------------------------------------------------------------------------+
   Endogenous:  y x
    Exogenous:  _cons
```

According to the estimated result the appropriated lag is order 1 that has the lowest SBIC

Run the VAR model with lag = 1

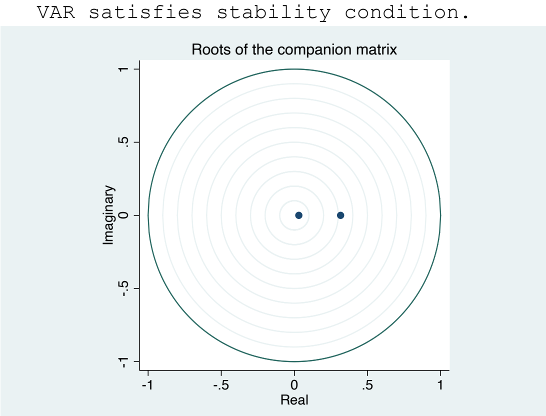
```
. var y x, lag(1/1)

Vector autoregression

Sample:  2 - 500                               Number of obs      =       499
Log likelihood =  -1330.499                    AIC                =   5.35671
FPE            =   .7268435                    HQIC               =  5.376587
```

```
Det(Sigma_ml)  =   .7095726                     SBIC              =   5.407362

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
y                    3      .993938   0.0154   7.794124   0.0203
x                    3      .999539   0.0854   46.59298   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
           y |
         L1. |   .1397088    .050078     2.79   0.005     .0415578    .2378598
             |
           x |
         L1. |   .0571404   .0479562     1.19   0.233     -.036852    .1511328
             |
       _cons |   .3347512   .0544186     6.15   0.000     .2280928    .4414097
-------------+----------------------------------------------------------------
x            |
           y |
         L1. |   .3367226   .0503601     6.69   0.000     .2380185    .4354266
             |
           x |
         L1. |   .2072735   .0482264     4.30   0.000     .1127514    .3017955
             |
       _cons |   .1268942   .0547252     2.32   0.020     .0196347    .2341536
------------------------------------------------------------------------------
```

(b) Perform stability test and Granger exogeneity test. Determine whether assumptions of VARs are satisfied, explain your evaluation criteria. If the stability assumption is unsatisfied, what will happen? (5 points)

Stability test

```
varstable ,graph

   Eigenvalue stability condition
  +----------------------------------------+
  |        Eigenvalue         |   Modulus  |
  |---------------------------+------------|
  |   .3162556                |   .316256  |
  |  .03072664                |   .030727  |
  +----------------------------------------+
   All the eigenvalues lie inside the unit circle.
   VAR satisfies stability condition.
```



Granger exogeneity test

```
vargranger

   Granger causality Wald tests
  +------------------------------------------------------------------+
  |          Equation           Excluded |   chi2     df Prob > chi2 |
  |--------------------------------------+---------------------------|
  |                 y                  x |  1.4197     1    0.233    |
  |                 y                ALL |  1.4197     1    0.233    |
  |--------------------------------------+---------------------------|
```

```
|               x                 y |  44.707     1    0.000    |
|               x               ALL |  44.707     1    0.000    |
+------------------------------------------------------------------+
```

Answer5B

From the stability test, all the eigenvalues are in the unit circle so, the system is stable

From the Granger exogeneity test, both y are endogenous variable since we accept the H0
But x is seemed to be exogenous variable

If the system is not stable it will lead to the spurious problem.

(c) Perform Impulse response function analysis (irf), Orthogonal impulse response function analysis (oirf), Cumulative impulse response function analysis (coirf), make interpretation of the analysis, and determine which variable has more impact (using Cholesky order – $y_t x_t$). (5 points)

IRF

```
irf create order1, o(y x) step(10) set(irf01)
(file irf01.irf created)
(file irf01.irf now active)
(file irf01.irf updated)

. irf table irf, impulse(y x) response(y x)
```

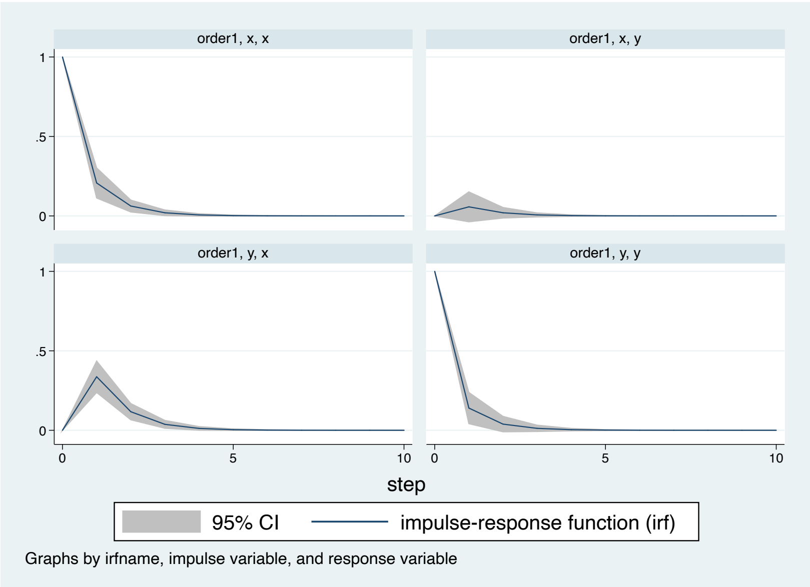
Results from order1

step	(1) irf	(1) Lower	(1) Upper	(2) irf	(2) Lower	(2) Upper
0	1	1	1	0	0	0
1	.139709	.041558	.23786	.336723	.238019	.435427
2	.038759	-.009591	.087109	.116837	.065895	.167779
3	.012091	-.007742	.031924	.037268	.012762	.061774
4	.003819	-.003751	.011389	.011796	.000589	.023003
5	.001208	-.001597	.004012	.003731	-.001027	.008489
6	.000382	-.000636	.0014	.00118	-.000724	.003084
7	.000121	-.000243	.000485	.000373	-.000358	.001104
8	.000038	-.00009	.000167	.000118	-.000154	.00039
9	.000012	-.000033	.000057	.000037	-.000062	.000137
10	3.8e-06	-.000012	.000019	.000012	-.000024	.000047

step	(3) irf	(3) Lower	(3) Upper	(4) irf	(4) Lower	(4) Upper
0	0	0	0	1	1	1
1	.05714	-.036852	.151133	.207273	.112751	.301795
2	.019827	-.013254	.052907	.062203	.024891	.099515
3	.006324	-.006018	.018666	.019569	.001938	.0372
4	.002002	-.00257	.006574	.006186	-.001426	.013798
5	.000633	-.00103	.002296	.001956	-.001118	.00503
6	.0002	-.000395	.000796	.000619	-.000568	.001805
7	.000063	-.000147	.000274	.000196	-.000248	.000639
8	.00002	-.000054	.000094	.000062	-.0001	.000224
9	6.3e-06	-.000019	.000032	.00002	-.000039	.000078
10	2.0e-06	-6.8e-06	.000011	6.2e-06	-.000014	.000027

```
95% lower and upper bounds reported
(1) irfname = order1, impulse = y, and response = y
(2) irfname = order1, impulse = y, and response = x
```

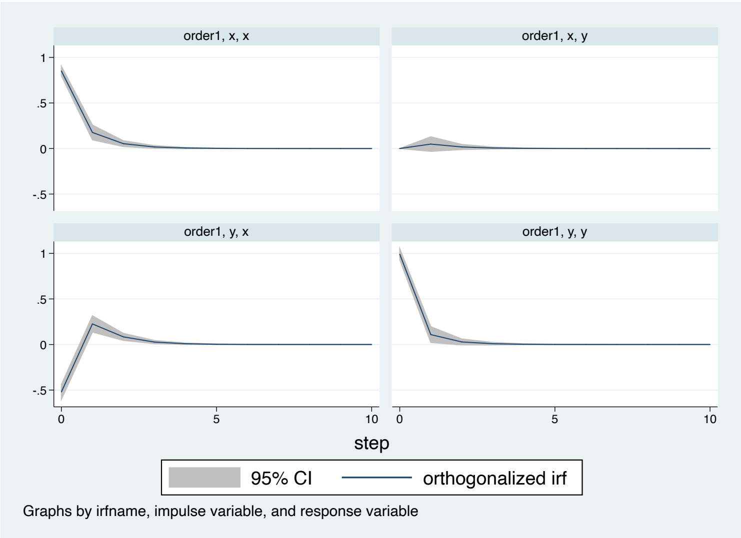
```
(3) irfname = order1, impulse = x, and response = y
(4) irfname = order1, impulse = x, and response = x

. irf graph irf, impulse(y x) response(y x)
```



OIRF

```
irf table oirf, impulse(y x) response(y x)

                              Results from order1

+--------------------------------------------------------------------------------+
|        |    (1)         (1)         (1)     |    (2)         (2)         (2)     |
|  step  |   oirf        Lower       Upper    |   oirf        Lower       Upper    |
|--------+-----------------------------------+-----------------------------------|
|0       | .990946     .929466     1.05243   | -.520071    -.601335    -.438807  |
|1       | .108727     .021924     .19553    | .225877     .136485     .315269   |
|2       | .028097     -.004189    .060382   | .083429     .04468      .122178   |
|3       | .008693     -.004787    .022172   | .026753     .009658     .043849   |
|4       | .002743     -.002439    .007925   | .008472     .000914     .01603    |
|5       | .000867     -.001062    .002797   | .00268      -.000528    .005887   |
|6       | .000274     -.000429    .000978   | .000847     -.000444    .002139   |
|7       | .000087     -.000166    .000339   | .000268     -.000231    .000767   |
|8       | .000027     -.000062    .000117   | .000085     -.000102    .000272   |
|9       | 8.7e-06     -.000023    .00004    | .000027     -.000042    .000095   |
|10      | 2.7e-06     -8.1e-06    .000014   | 8.5e-06     -.000016    .000033   |
+--------------------------------------------------------------------------------+

+--------------------------------------------------------------------------------+
|        |    (3)         (3)         (3)     |    (4)         (4)         (4)     |
|  step  |   oirf        Lower       Upper    |   oirf        Lower       Upper    |
|--------+-----------------------------------+-----------------------------------|
|0       | 0           0           0         | .850058     .797319     .902796   |
|1       | .048573     -.031383    .128528   | .176194     .095105     .257284   |
|2       | .016854     -.011286    .044993   | .052876     .02099      .084762   |
|3       | .005376     -.005121    .015873   | .016635     .001612     .031657   |
|4       | .001702     -.002186    .00559    | .005258     -.001221    .011737   |
|5       | .000538     -.000876    .001952   | .001663     -.000952    .004278   |
|6       | .00017      -.000336    .000677   | .000526     -.000483    .001535   |
|7       | .000054     -.000125    .000233   | .000166     -.000211    .000544   |
|8       | .000017     -.000046    .00008    | .000053     -.000085    .00019    |
|9       | 5.4e-06     -.000016    .000027   | .000017     -.000033    .000066   |
|10      | 1.7e-06     -5.8e-06    9.2e-06   | 5.3e-06     -.000012    .000023   |
+--------------------------------------------------------------------------------+
95% lower and upper bounds reported
(1) irfname = order1, impulse = y, and response = y
(2) irfname = order1, impulse = y, and response = x
(3) irfname = order1, impulse = x, and response = y
(4) irfname = order1, impulse = x, and response = x
```

```
. irf graph oirf, impulse(y x) response(y x)
```



COIRF

```
. irf table coirf, impulse(y x) response(y x)
```

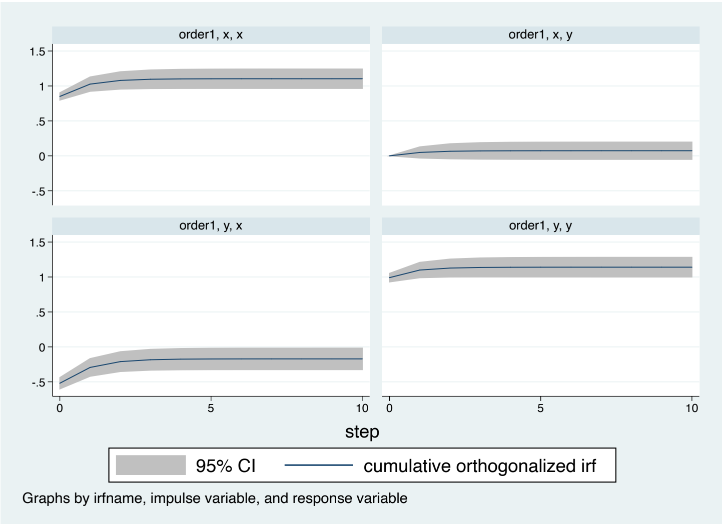
Results from order1

step	(1) coirf	(1) Lower	(1) Upper	(2) coirf	(2) Lower	(2) Upper
0	.990946	.929466	1.05243	-.520071	-.601335	-.438807
1	1.09967	.989473	1.20987	-.294194	-.420671	-.167717
2	1.12777	.999326	1.25621	-.210765	-.352446	-.069084
3	1.13646	1.00031	1.27262	-.184012	-.33288	-.035143
4	1.13921	1.00004	1.27837	-.175539	-.327399	-.023679
5	1.14007	.999798	1.28035	-.17286	-.325878	-.019841
6	1.14035	.999677	1.28102	-.172012	-.325459	-.018565
7	1.14043	.999625	1.28124	-.171744	-.325345	-.018144
8	1.14046	.999604	1.28132	-.171659	-.325314	-.018005
9	1.14047	.999596	1.28134	-.171633	-.325306	-.017959
10	1.14047	.999593	1.28135	-.171624	-.325304	-.017944

step	(3) coirf	(3) Lower	(3) Upper	(4) coirf	(4) Lower	(4) Upper
0	0	0	0	.850058	.797319	.902796
1	.048573	-.031383	.128528	1.02625	.923734	1.12877
2	.065426	-.042226	.173079	1.07913	.955049	1.20321
3	.070802	-.047081	.188686	1.09576	.962683	1.22884
4	.072504	-.049137	.194145	1.10102	.964313	1.23773
5	.073042	-.049954	.196038	1.10268	.964594	1.24077
6	.073212	-.050265	.19669	1.10321	.964614	1.24181
7	.073266	-.05038	.196912	1.10338	.964599	1.24215
8	.073283	-.050422	.196988	1.10343	.964588	1.24227
9	.073289	-.050436	.197014	1.10345	.964583	1.24231
10	.07329	-.050442	.197022	1.10345	.96458	1.24232

95% lower and upper bounds reported
(1) irfname = order1, impulse = y, and response = y
(2) irfname = order1, impulse = y, and response = x
(3) irfname = order1, impulse = x, and response = y
(4) irfname = order1, impulse = x, and response = x

```
. irf graph coirf, impulse(y x) response(y x)
```



<u>Answer5C</u>

IRF analysis

Direction:
x has the positive impact on x itself and y
y has the negative impact on x and positive impact on y itself

Magnitude:
<u>The most impact is x on x and y on y themselves,</u>
followed by y on x (negative impact) and x on y
which mean that y has more impact on x than y

Duration:
All of them have the short duration in order to self-adjust to the equilibrium

(d) Perform Forecast error variance decomposition (fevd) and determine variable that has more impact on each endogenous variable. (3 points)

<u>FEVD</u>

```
irf table fevd, impulse(y x) response(y x)
```

Results from order1

step	(1) fevd	(1) Lower	(1) Upper	(2) fevd	(2) Lower	(2) Upper
0	0	0	0	0	0	0
1	1	1	1	.272361	.205724	.338998
2	.997632	.989845	1.00542	.299027	.234158	.363895
3	.997349	.988646	1.00605	.302753	.23842	.367086
4	.997321	.988511	1.00613	.303136	.238863	.367408
5	.997318	.988495	1.00614	.303174	.238908	.36744
6	.997317	.988494	1.00614	.303178	.238912	.367444
7	.997317	.988493	1.00614	.303178	.238913	.367444
8	.997317	.988493	1.00614	.303178	.238913	.367444

```
|9        | .997317      .988493      1.00614   | .303178      .238913      .367444   |
|10       | .997317      .988493      1.00614   | .303178      .238913      .367444   |
+-------------------------------------------------------------------------------------+

+-------------------------------------------------------------------------------------+
|         |    (3)          (3)          (3)      |    (4)          (4)          (4)     |
|  step   |   fevd         Lower        Upper     |   fevd         Lower        Upper    |
|---------+-----------------------------------------+---------------------------------|
|0        | 0            0            0         | 0            0            0         |
|1        | 0            0            0         | .727639      .661002      .794276   |
|2        | .002368      -.005418     .010155   | .700973      .636105      .765842   |
|3        | .002651      -.006052     .011354   | .697247      .632914      .76158    |
|4        | .002679      -.006131     .011489   | .696864      .632592      .761137   |
|5        | .002682      -.00614      .011505   | .696826      .63256       .761092   |
|6        | .002683      -.006141     .011506   | .696822      .632556      .761088   |
|7        | .002683      -.006141     .011507   | .696822      .632556      .761087   |
|8        | .002683      -.006141     .011507   | .696822      .632556      .761087   |
|9        | .002683      -.006141     .011507   | .696822      .632556      .761087   |
|10       | .002683      -.006141     .011507   | .696822      .632556      .761087   |
+-------------------------------------------------------------------------------------+
95% lower and upper bounds reported
(1) irfname = order1, impulse = y, and response = y
(2) irfname = order1, impulse = y, and response = x
(3) irfname = order1, impulse = x, and response = y
(4) irfname = order1, impulse = x, and response = x
```

Answer5D

According to forecast error variance decomposition, y has more impact on x

(e) Determine whether changing Cholesky order from – “$y_t x_t$” to “$x_t y_t$” will change the results of irf, oirf, coirf, and fevd. Why? or why not? (2 points)

Changing the Cholesky order

```
irf create order2, o(x y) step(10) set(irf02)
(file irf02.irf created)
(file irf02.irf now active)
(file irf02.irf updated)

. irf graph irf, impulse(y x) response(y x)

. irf table irf, impulse(y x) response(y x)

                                Results from order2

+-------------------------------------------------------------------------------------+
|         |    (1)          (1)          (1)      |    (2)          (2)          (2)     |
|  step   |   irf          Lower        Upper     |   irf          Lower        Upper    |
|---------+-----------------------------------------+---------------------------------|
|0        | 1            1            1         | 0            0            0         |
|1        | .139709      .041558      .23786    | .336723      .238019      .435427   |
|2        | .038759      -.009591     .087109   | .116837      .065895      .167779   |
|3        | .012091      -.007742     .031924   | .037268      .012762      .061774   |
|4        | .003819      -.003751     .011389   | .011796      .000589      .023003   |
|5        | .001208      -.001597     .004012   | .003731      -.001027     .008489   |
|6        | .000382      -.000636     .0014     | .00118       -.000724     .003084   |
|7        | .000121      -.000243     .000485   | .000373      -.000358     .001104   |
|8        | .000038      -.00009      .000167   | .000118      -.000154     .00039    |
|9        | .000012      -.000033     .000057   | .000037      -.000062     .000137   |
|10       | 3.8e-06      -.000012     .000019   | .000012      -.000024     .000047   |
+-------------------------------------------------------------------------------------+

+-------------------------------------------------------------------------------------+
|         |    (3)          (3)          (3)      |    (4)          (4)          (4)     |
|  step   |   irf          Lower        Upper     |   irf          Lower        Upper    |
|---------+-----------------------------------------+---------------------------------|
|0        | 0            0            0         | 1            1            1         |
|1        | .05714       -.036852     .151133   | .207273      .112751      .301795   |
|2        | .019827      -.013254     .052907   | .062203      .024891      .099515   |
|3        | .006324      -.006018     .018666   | .019569      .001938      .0372     |
```

4	.002002	-.00257	.006574	.006186	-.001426	.013798
5	.000633	-.00103	.002296	.001956	-.001118	.00503
6	.0002	-.000395	.000796	.000619	-.000568	.001805
7	.000063	-.000147	.000274	.000196	-.000248	.000639
8	.00002	-.000054	.000094	.000062	-.0001	.000224
9	6.3e-06	-.000019	.000032	.00002	-.000039	.000078
10	2.0e-06	-6.8e-06	.000011	6.2e-06	-.000014	.000027

95% lower and upper bounds reported
(1) irfname = order2, impulse = y, and response = y
(2) irfname = order2, impulse = y, and response = x
(3) irfname = order2, impulse = x, and response = y
(4) irfname = order2, impulse = x, and response = x

```
. irf table oirf, impulse(y x) response(y x)
```

Results from order2

step	(1) oirf	(1) Lower	(1) Upper	(2) oirf	(2) Lower	(2) Upper
0	.845295	.792851	.897738	0	0	0
1	.118095	.034806	.201384	.28463	.199347	.369912
2	.032763	-.008158	.073683	.098761	.055267	.142256
3	.010221	-.006556	.026997	.031503	.010696	.052309
4	.003228	-.003174	.00963	.009971	.000478	.019464
5	.001021	-.001351	.003392	.003154	-.000873	.00718
6	.000323	-.000538	.001184	.000997	-.000614	.002608
7	.000102	-.000206	.00041	.000315	-.000303	.000934
8	.000032	-.000076	.000141	.0001	-.000131	.00033
9	.00001	-.000028	.000048	.000032	-.000052	.000115
10	3.2e-06	-9.9e-06	.000016	1.0e-05	-.00002	.00004

step	(3) oirf	(3) Lower	(3) Upper	(4) oirf	(4) Lower	(4) Upper
0	-.517157	-.597966	-.436348	.99653	.934704	1.05836
1	-.015309	-.099451	.068832	.032416	-.055209	.12004
2	-.000287	-.013162	.012589	.001564	-.024512	.02764
3	.000049	-.003268	.003367	.000228	-.009287	.009743
4	.00002	-.001001	.001041	.000064	-.003023	.003151
5	6.4e-06	-.000317	.00033	.00002	-.000964	.001004
6	2.0e-06	-.000101	.000105	6.3e-06	-.000306	.000319
7	6.4e-07	-.000032	.000033	2.0e-06	-.000097	.000101
8	2.0e-07	-.00001	.000011	6.3e-07	-.000031	.000032
9	6.4e-08	-3.2e-06	3.4e-06	2.0e-07	-9.8e-06	.00001
10	2.0e-08	-1.0e-06	1.1e-06	6.3e-08	-3.1e-06	3.3e-06

95% lower and upper bounds reported
(1) irfname = order2, impulse = y, and response = y
(2) irfname = order2, impulse = y, and response = x
(3) irfname = order2, impulse = x, and response = y
(4) irfname = order2, impulse = x, and response = x

```
. irf graph oirf, impulse(y x) response(y x)

. irf table coirf, impulse(y x) response(y x)
```

Results from order2

step	(1) coirf	(1) Lower	(1) Upper	(2) coirf	(2) Lower	(2) Upper
0	.845295	.792851	.897738	0	0	0
1	.96339	.861135	1.06564	.28463	.199347	.369912
2	.996152	.862544	1.12976	.383391	.265721	.501062
3	1.00637	.859272	1.15347	.414894	.283031	.546757
4	1.0096	.857305	1.1619	.424865	.287155	.562575
5	1.01062	.856413	1.16483	.428019	.288027	.56801
6	1.01094	.856049	1.16584	.429016	.288172	.56986
7	1.01105	.855909	1.16618	.429331	.288179	.570484
8	1.01108	.855856	1.1663	.429431	.288169	.570694
9	1.01109	.855837	1.16634	.429463	.288162	.570764

10	1.01109	.85583	1.16635	.429473	.288158	.570787

step	(3) coirf	(3) Lower	(3) Upper	(4) coirf	(4) Lower	(4) Upper
0	-.517157	-.597966	-.436348	.99653	.934704	1.05836
1	-.532466	-.655786	-.409147	1.02895	.920552	1.13734
2	-.532753	-.666326	-.39918	1.03051	.913081	1.14794
3	-.532704	-.668918	-.39649	1.03074	.910047	1.15143
4	-.532684	-.669709	-.395658	1.0308	.908966	1.15264
5	-.532677	-.669959	-.395395	1.03082	.908609	1.15303
6	-.532675	-.670039	-.395312	1.03083	.908494	1.15316
7	-.532675	-.670064	-.395285	1.03083	.908458	1.1532
8	-.532674	-.670072	-.395277	1.03083	.908446	1.15321
9	-.532674	-.670075	-.395274	1.03083	.908443	1.15322
10	-.532674	-.670076	-.395273	1.03083	.908441	1.15322

95% lower and upper bounds reported
(1) irfname = order2, impulse = y, and response = y
(2) irfname = order2, impulse = y, and response = x
(3) irfname = order2, impulse = x, and response = y
(4) irfname = order2, impulse = x, and response = x

```
. irf graph coirf, impulse(y x) response(y x)

. irf table fevd, impulse(y x) response(y x)
```

Results from order2

step	(1) fevd	(1) Lower	(1) Upper	(2) fevd	(2) Lower	(2) Upper
0	0	0	0	0	0	0
1	.727639	.661002	.794276	0	0	0
2	.731281	.664826	.797736	.075352	.032679	.118026
3	.73157	.665099	.798041	.083665	.036757	.130573
4	.731598	.665124	.798073	.084503	.037097	.131908
5	.731601	.665126	.798076	.084587	.037126	.132048
6	.731601	.665126	.798076	.084595	.037128	.132062
7	.731601	.665126	.798076	.084596	.037128	.132064
8	.731601	.665126	.798076	.084596	.037128	.132064
9	.731601	.665126	.798076	.084596	.037128	.132064
10	.731601	.665126	.798076	.084596	.037128	.132064

step	(3) fevd	(3) Lower	(3) Upper	(4) fevd	(4) Lower	(4) Upper
0	0	0	0	0	0	0
1	.272361	.205724	.338998	1	1	1
2	.268719	.202264	.335174	.924648	.881974	.967321
3	.26843	.201959	.334901	.916335	.869427	.963243
4	.268402	.201927	.334876	.915497	.868092	.962903
5	.268399	.201924	.334874	.915413	.867952	.962874
6	.268399	.201924	.334874	.915405	.867938	.962872
7	.268399	.201924	.334874	.915404	.867936	.962872
8	.268399	.201924	.334874	.915404	.867936	.962872
9	.268399	.201924	.334874	.915404	.867936	.962872
10	.268399	.201924	.334874	.915404	.867936	.962872

95% lower and upper bounds reported
(1) irfname = order2, impulse = y, and response = y
(2) irfname = order2, impulse = y, and response = x
(3) irfname = order2, impulse = x, and response = y
(4) irfname = order2, impulse = x, and response = x

.

<u>Answer5E</u>

Yes, it will change and according to the printout
especially on FEVD, the y has no impact on x instead which is a big change from the previous case

this thing happened because the Cholesky decomposition used to fix the under-identification problem of reduced from by putting the restriction on structural form assume some variable to have no contemporanion effect on the other which will impact the FEVD value to be 0 and by changing the order it's mean that we change the restriction we put from one to other one. So, our estimated result on IRF,OIRF,COIRF and FEVD changed.