BACHELOR OF ECONOMICS
THAMMASAT UNIVERSITY

Final Examination: 2/2020

Subject: EE426
Instructor: Tatre Jantarakolica
Date: May 28, 2021
Time: 9:00-15:00

Instructions

1. This is a take home examination.
2. In this exam paper, there are 5 main problems (5 pages, including this cover sheet), with the total marks of 100. Each question is worth 20 points.
3. You are required to estimate the model of each question based on the provided data set. You will have your own data set. Please download your data set (#.zip) based on your assigned number on the list. Copy estimated results from STATA program to your answer of each question (using Font-Courier New).
4. Time: 6 Hours (9:00 – 15:00 hr.) (approximately one hour per question)
5. Submit your answer before 3:00 pm. in ONE pdf-file by setting the file-name as No_Yourname_Student ID, **e.g. 1_6104640021_ THAMCHANOK.pdf**

No.	Student ID	Name- Surname	Files	Zip-file Download
1	6104640021	Mr THAMCHANOK PIANMUEAN	Final_q*_1.dta	01.zip
2	6104640112	Mr HARIT HATAWONG	Final_q*_2.dta	02.zip
3	6104640328	Miss WIRINRATCH KIRIRAK	Final_q*_3.dta	03.zip
4	6104641086	Mr NATTHAPHONG TISAVIPAT	Final_q*_4.dta	04.zip
5	6104641300	Mr NUNTAYOD TULAKARNWONG	Final_q*_5.dta	05.zip
6	6104641318	Miss PHATAKAN KANCHANAPRADIST	Final_q*_6.dta	06.zip
7	6204640087	Mr THANAKRIT METHASATE	Final_q*_7.dta	07.zip
9	6204640178	Miss MAI SURINTRABOON	Final_q*_9.dta	09.zip
10	6204641218	Miss PORNWARAT FOONGPIRIYA	Final_q*_10.dta	10.zip
11	6204641267	Mr PONGPANOT CHALOEMVISETPHON	Final_q*_11.dta	11.zip
12	6204641473	Mr WIRAPHAT LIN	Final_q*_12.dta	12.zip

Take Home Final Exam
EE426
Semester 2/2020

There are five questions - answer all questions.
Each question is worth 20 points.

1. From the data set "`Final_q1-1_no.dta`":

   Model for the degree of acceptance is stated as follows:

$$Prob(y_i=j|X) = f(x_1, x_2, x_3, x_4) \quad (1)$$

   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).

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

Multinomial logistic regression                 Number of obs     =        170
                                                LR chi2(8)        =      93.45
                                                Prob > chi2       =     0.0000
Log likelihood = -105.45005                     Pseudo R2         =     0.3071

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
-------------+----------------------------------------------------------------
1            |
          x1 |  -1.139582   .6146586    -1.85   0.064    -2.344291    .0651262
          x2 |  -.9083221   .7703074    -1.18   0.238    -2.418097    .6014525
          x3 |  -.6564347    .622678    -1.05   0.292    -1.876861    .5639918
          x4 |   .6609733   .2527119     2.62   0.009     .1656672    1.156279
       _cons |  -7.687864   3.555853    -2.16   0.031    -14.65721   -.7185191
-------------+----------------------------------------------------------------
2            |
          x1 |  -1.072935   .6766848    -1.59   0.113    -2.399213    .2533424
          x2 |  -1.787461   .8023606    -2.23   0.026    -3.360059   -.2148629
          x3 |   .5315965   .6657183     0.80   0.425    -.7731874     1.83638
          x4 |    1.84457   .3067994     6.01   0.000     1.243254    2.445885
       _cons |  -26.32819   4.539073    -5.80   0.000    -35.22461   -17.43177
------------------------------------------------------------------------------

. est store m1

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

```
Multinomial logistic regression                 Number of obs     =        170
                                                LR chi2(8)        =      93.45
                                                Prob > chi2       =     0.0000
Log likelihood = -105.45005                     Pseudo R2         =     0.3071

------------------------------------------------------------------------------
           y |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
-------------+----------------------------------------------------------------
1            |
          x1 |   .3199526   .1966616    -1.85   0.064     .0959152    1.067294
          x2 |   .4032002   .3105881    -1.18   0.238      .089091    1.824767
          x3 |   .5186974   .3229814    -1.05   0.292     .1530698    1.757675
          x4 |   1.936676   .4894211     2.62   0.009      1.18018    3.178087
       _cons |   .0004584   .0016298    -2.16   0.031     4.31e-07    .4874736
-------------+----------------------------------------------------------------
2            |
          x1 |   .3420031   .2314283    -1.59   0.113     .0907893    1.288324
          x2 |   .1673847   .1343029    -2.23   0.026     .0347332     .806652
          x3 |   1.701647   1.132817     0.80   0.425     .4615396    6.273788
          x4 |   6.325377   1.940622     6.01   0.000     3.466876    11.54076
       _cons |   3.68e-12   1.67e-11    -5.80   0.000     5.04e-16    2.69e-08
------------------------------------------------------------------------------
```

- Based on rrr, x1, x2, and x3 are negatively related with the degree of acceptance since their rrr are less than 1 while x4 is positively related with the degree of acceptance since its rrr is greater than 1 in case that y=1. And , x1, x2 are negatively related with the degree of acceptance since their rrr are less than 1 while x3, x4 is positively related with the degree of acceptance since its rrr is greater than 1 in case that y=2.
- Overall test (LR chi2) > reject H0 meaning that all independent variables in the model can be used together to significantly explain the variation of the degree of acceptance.
- Pseudo R-square is moderate level meaning that there is still a room for improvement in this model.
- Individual tests show that all independent variables cannot make the distinction between the disagree and neutral level except for x4. And x2, x4 can make the distinction between the disagree and agree level.

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

Multinomial logistic regression                 Number of obs     =         63
                                                LR chi2(4)        =       9.62
                                                Prob > chi2       =     0.0474
Log likelihood =  -35.29146                     Pseudo R2         =     0.1199

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
```

```
-------------+----------------------------------------------------------------
1            |
          x1 |  -.9328384   .5918786    -1.58   0.115    -2.092899    .2272224
          x2 |  -.5901403   .7442161    -0.79   0.428    -2.048777    .8684965
          x3 |  -.4733217    .639715    -0.74   0.459     -1.72714    .7804967
          x4 |   .5904273   .2497463     2.36   0.018     .1009335    1.079921
       _cons |   -7.06389   3.667131    -1.93   0.054    -14.25134     .123555
------------------------------------------------------------------------------

. est store m2

. . hausman m1 m2

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       m1           m2         Difference          S.E.
-------------+----------------------------------------------------------------
          x1 |   -1.139582    -.9328384        -.206744        .1657855
          x2 |   -.9083221    -.5901403       -.3181818        .1987858
          x3 |   -.6564347    -.4733217        -.183113               .
          x4 |    .6609733     .5904273        .0705461        .0386014
------------------------------------------------------------------------------
                           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(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        1.21
                Prob>chi2 =      0.8758
                (V_b-V_B is not positive definite)
```

According to the p-value of Hausman test which is greater than 0.05, we accept H0. So, IIA assumption is held, multinomial logit model is appropriated.

IIA suggests that the decision between two alternatives is independent from the existence of more alternatives. And IIA is important for multinomial logit because it assumes that error terms are independently distributed over all alternatives.

However, in this case if we look at the dependent variable in this model. It is ordered data (*yi* is categorical data = 1 for disagree, = 2 for neutral, and = 3 for agree), so order probit model should be considered.

(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).

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

```
Ordered probit regression                         Number of obs     =        170
                                                  LR chi2(4)        =      84.19
                                                  Prob > chi2       =     0.0000
Log likelihood =  -110.0814                       Pseudo R2         =     0.2766

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.2342737   .2282562    -1.03   0.305    -.6816477    .2131003
          x2 |   -.513981   .2386795    -2.15   0.031    -.9817843   -.0461777
          x3 |   .4692108   .2168354     2.16   0.030     .0442213    .8942003
          x4 |    .696647    .091764     7.59   0.000      .516793    .8765011
-------------+----------------------------------------------------------------
       /cut1 |   9.593414   1.390157                      6.868757    12.31807
       /cut2 |   10.83561   1.444039                      8.005343    13.66587
------------------------------------------------------------------------------
est store o1

. fitstat

Measures of Fit for oprobit of y

Log-Lik Intercept Only:       -152.176   Log-Lik Full Model:           -110.081
D(164):                        220.163   LR(4):                          84.189
                                         Prob > LR:                       0.000
McFadden's R2:                   0.277   McFadden's Adj R2:               0.237
ML (Cox-Snell) R2:               0.391   Cragg-Uhler(Nagelkerke) R2:      0.469
McKelvey & Zavoina's R2:         0.513
Variance of y*:                  2.053   Variance of error:               1.000
Count R2:                        0.718   Adj Count R2:                    0.238
AIC:                             1.366   AIC*n:                         232.163
BIC:                          -622.108   BIC':                          -63.646
BIC used by Stata:             250.978   AIC used by Stata:             232.163
```

- Coefficient of x1 and x2 have a negative sign, but x3 and x4 have a positive sign.
- Overall test (LR chi2) shows that all independent variables in the model can be used together to explain the variation of the degree of acceptance.
- Pseudo R-square is moderate level, which means that there is still a room for improvement in the model.
- Individual tests illustrate that all independent variables can significantly explain the variation of the degree of acceptance except for x1.

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

Generalized Ordered Probit Estimates              Number of obs     =        170
                                                  LR chi2(3)        =      83.13
                                                  Prob > chi2       =     0.0000
Log likelihood = -110.61304                       Pseudo R2         =     0.2731

 ( 1)  [0]x2 - [1]x2 = 0
```

```
 ( 2)  [0]x3 - [1]x3 = 0
 ( 3)  [0]x4 - [1]x4 = 0
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |
          x2 |  -.5326495   .2373104    -2.24   0.025    -.9977693   -.0675296
          x3 |   .4320436   .2133152     2.03   0.043     .0139535    .8501336
          x4 |   .6731946    .088072     7.64   0.000     .5005767    .8458125
       _cons |  -9.275096   1.345163    -6.90   0.000    -11.91157   -6.638624
-------------+----------------------------------------------------------------
1            |
          x2 |  -.5326495   .2373104    -2.24   0.025    -.9977693   -.0675296
          x3 |   .4320436   .2133152     2.03   0.043     .0139535    .8501336
          x4 |   .6731946    .088072     7.64   0.000     .5005767    .8458125
       _cons |  -10.50344   1.396104    -7.52   0.000    -13.23975   -7.767121
------------------------------------------------------------------------------

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

Generalized Ordered Probit Estimates             Number of obs     =        170
                                                 LR chi2(8)        =      90.21
                                                 Prob > chi2       =     0.0000
Log likelihood = -107.07164                      Pseudo R2         =     0.2964

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |
          x1 |  -.5088087    .312306    -1.63   0.103    -1.120917    .1032998
          x2 |  -.5509522   .3519411    -1.57   0.117    -1.240744    .1388396
          x3 |   .0485293   .3248555     0.15   0.881    -.5881757    .6852343
          x4 |   .6027095   .1214707     4.96   0.000     .3646313    .8407877
       _cons |  -7.727312   1.857501    -4.16   0.000    -11.36795   -4.086678
-------------+----------------------------------------------------------------
1            |
          x1 |  -.1061387   .2636141    -0.40   0.687    -.6228128    .4105353
          x2 |  -.5386005   .2597115    -2.07   0.038    -1.047626   -.0295753
          x3 |   .6135726   .2432453     2.52   0.012     .1368206    1.090325
          x4 |   .7285438   .1080621     6.74   0.000     .5167459    .9403417
       _cons |  -11.51174   1.717689    -6.70   0.000    -14.87835   -8.145136
------------------------------------------------------------------------------

. lrtest oprobit goprobit, stats

Likelihood-ratio test                                 LR chi2(5)  =      7.08
(Assumption: oprobit nested in goprobit)              Prob > chi2 =    0.2146

Akaike's information criterion and Bayesian information criterion

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

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
oprobit	170	-152.176	-110.613	5	231.2261	246.9051
goprobit	170	-152.176	-107.0716	10	234.1433	265.5013

Note: N=Obs used in calculating BIC; see [R] BIC note.

According to p-value of LR test which is greater than 0.05, we accept Ho. This means that the estimated coefficients of Order Probit is consistent to all degrees of acceptance. Hence, Order Probit model is appropriated for this case.

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:

$$I_{Ji} = \beta_{J0} + \beta_{J1}x_{1i} + \beta_{J2}x_{2i} + \beta_{J3}x_{3i} + \beta_{J4}x_{4i} + u_{Ji} \quad (2)$$

and

$$\Pr(Y_{Ji} = 1) = \Phi(I_{Ji})$$

where: 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.
$\Phi(.)$ 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)

```
. probit y1 x*, nolog

Probit regression                               Number of obs     =        550
                                                LR chi2(4)        =      30.78
                                                Prob > chi2       =     0.0000
Log likelihood = -363.91698                     Pseudo R2         =     0.0406
```

y1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	1.390201	.3998738	3.48	0.001	.6064626	2.173939
x2	-.7740057	.3808589	-2.03	0.042	-1.520475	-.027536
x3	.6267583	.176273	3.56	0.000	.2812695	.9722471
x4	.2469299	.1846926	1.34	0.181	-.1150609	.6089207
_cons	-.8363226	.3024071	-2.77	0.006	-1.42903	-.2436156

```
. probit y2 x*, nolog

Probit regression                               Number of obs     =        550
```

```
                                                  LR chi2(4)        =      14.82
                                                  Prob > chi2       =     0.0051
Log likelihood = -360.62385                       Pseudo R2         =     0.0201

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.7722348   .3788359    -2.04   0.042    -1.514739   -.0297301
          x2 |  -.5403149   .3689094    -1.46   0.143    -1.263364    .1827343
          x3 |  -.2302021   .1742739    -1.32   0.187    -.5717727    .1113684
          x4 |   .6077808    .190547     3.19   0.001     .2343156     .981246
       _cons |   .3704869   .3010353     1.23   0.218    -.2195315    .9605053
------------------------------------------------------------------------------

. probit y3 x*, nolog

Probit regression                                 Number of obs     =        550
                                                  LR chi2(4)        =      11.67
                                                  Prob > chi2       =     0.0199
Log likelihood = -326.67955                       Pseudo R2         =     0.0176

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.4807416   .3902546    -1.23   0.218    -1.245627    .2841434
          x2 |   .7517781    .383604     1.96   0.050     -.000072    1.503628
          x3 |  -.3835021   .1970231    -1.95   0.052    -.7696603    .0026561
          x4 |  -.3269474   .1833785    -1.78   0.075    -.6863627    .0324679
       _cons |  -.4945789   .3160064    -1.57   0.118     -1.11394    .1247822
------------------------------------------------------------------------------
```

In this situation, the mobile phone services is our dependent variables, so they should have some relationships between each other since selecting one choice of service could have an effect on the probability of the other choices. Hence, these three separate probit models should not be used.

(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 x*) (y2 x*) (y3 x*), nolog

Multivariate probit (MSL, # draws = 5)              Number of obs   =        550
                                                    Wald chi2(12)   =      42.84
Log likelihood = -930.46418                         Prob > chi2     =     0.0000

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

```
          x1 |   1.239993   .3899615     3.18   0.001     .4756821    2.004303
          x2 |  -.6395128   .3746254    -1.71   0.088    -1.373765    .0947395
          x3 |   .6117551   .1767944     3.46   0.001     .2652444    .9582659
          x4 |   .2244387   .1837265     1.22   0.222    -.1356586    .5845361
       _cons |  -.7730431   .3042983    -2.54   0.011    -1.369457   -.1766295
-------------+----------------------------------------------------------------
y2           |
          x1 |  -.8460207   .3747646    -2.26   0.024    -1.580546   -.1114956
          x2 |  -.4642576   .3669165    -1.27   0.206    -1.183401    .2548855
          x3 |  -.2452939   .1756833    -1.40   0.163    -.5896268     .099039
          x4 |   .5858579   .1896389     3.09   0.002     .2141726    .9575433
       _cons |    .407426   .2996414     1.36   0.174    -.1798603    .9947124
-------------+----------------------------------------------------------------
y3           |
          x1 |  -.4356133    .383309    -1.14   0.256    -1.186885    .3156585
          x2 |   .5905746   .3787825     1.56   0.119    -.1518254    1.332975
          x3 |  -.2661106   .1852372    -1.44   0.151    -.6291688    .0969475
          x4 |  -.2912631    .187216    -1.56   0.120    -.6581997    .0756735
       _cons |  -.4418587   .3177148    -1.39   0.164    -1.064568    .1808509
-------------+----------------------------------------------------------------
    /atrho21 |  -.5823545   .0733326    -7.94   0.000    -.7260838   -.4386252
-------------+----------------------------------------------------------------
    /atrho31 |  -.4493911   .0744662    -6.03   0.000    -.5953422   -.3034401
-------------+----------------------------------------------------------------
    /atrho32 |    -.32037   .0708034    -4.52   0.000    -.4591422   -.1815979
-------------+----------------------------------------------------------------
       rho21 |  -.5243746   .0531684    -9.86   0.000    -.6206636   -.4125042
-------------+----------------------------------------------------------------
       rho31 |  -.4213984   .0612427    -6.88   0.000    -.5337269   -.2944576
-------------+----------------------------------------------------------------
       rho32 |  -.3098415   .0640062    -4.84   0.000    -.4293848   -.1796277
------------------------------------------------------------------------------
Likelihood ratio test of  rho21 = rho31 = rho32 = 0:
             chi2(3) =  241.512   Prob > chi2 = 0.0000
```

Since p-value of LR test of rhoij=0 is less than 0.05, Ho is rejected. It means that there exists a significant correlation among the disturbance terms of these three models. Therefore, The MVProbit is more appropriated than three separate probit models.

How does the MV Probit models differ from three separate probit models?

- Run as a system.

2. From the data set “Final_q2_no.dta”:

According to the following model,

$$y_{ki} = \beta_{k0} + \beta_{k1} x_i + u_{ki} \qquad (3)$$

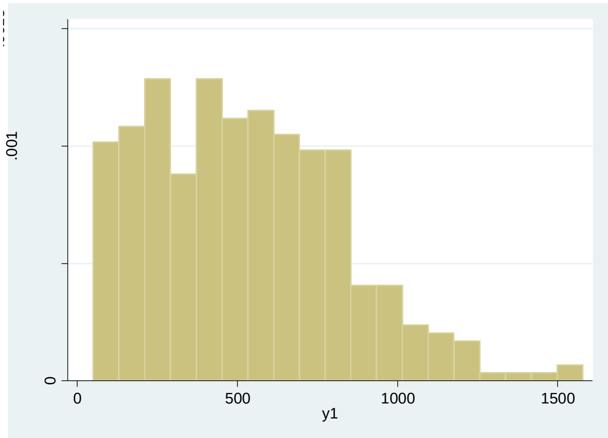
where: y_{ki} is Dependent variable, k=1, 2, 3.

x_i is Independent variable

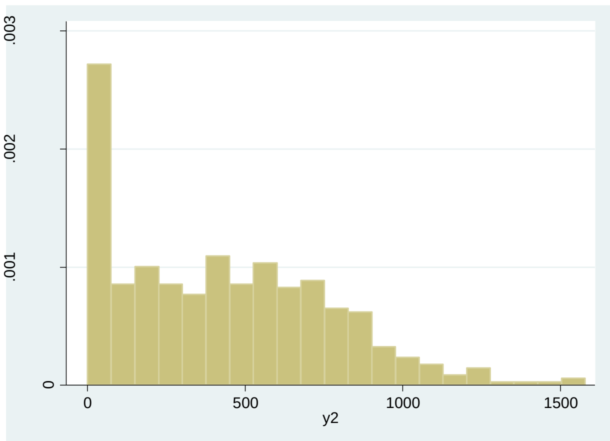
u_i is Stochastic disturbance term

(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)

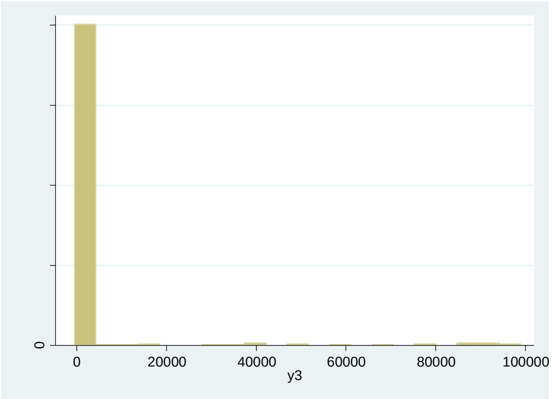
```
sum y1 y2 y3

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          y1 |        367    525.2579    301.6184   50.21759    1578.51
          y2 |        450    429.3814    339.0001          0    1578.51
          y3 |        450    3341.138    14545.28  -466.0042   98951.63

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



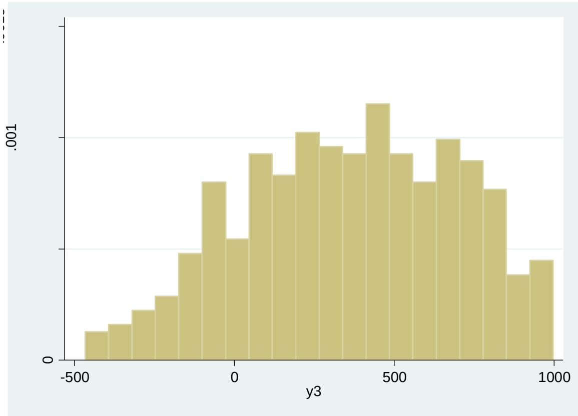

```
. histogram y2
(bin=21, start=0, width=75.167149)
```



```
. histogram y3
(bin=21, start=-466.00415, width=4734.1732)
```




```
histogram y3 if y3<2000
(bin=20, start=-466.00415, width=73.209753)
```



According to the descriptive statistics and histogram, y1 has truncated problem at around 50, y2 has censored problem at 0, and y3 has outlier problem.

(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)

```
. reg y1 x

      Source |       SS           df       MS      Number of obs   =       367
-------------+----------------------------------   F(1, 365)       =    135.47
       Model |   9012725.4         1   9012725.4   Prob > F        =    0.0000
    Residual |  24283633.1       365  66530.5016   R-squared       =    0.2707
-------------+----------------------------------   Adj R-squared   =    0.2687
       Total |  33296358.5       366  90973.6571   Root MSE        =    257.94

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   165.1361   14.18811    11.64   0.000     137.2354    193.0368
       _cons |  -5.977279   47.58695    -0.13   0.900    -99.55628    87.60172
-----------------------------------------------------------------------------
```

```
. est store m_y1

. sum y1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          y1 |        367    525.2579    301.6184   50.21759    1578.51

. scalar miny1=round(r(min))

. truncreg y1 x, ll(miny1) nolog
(note: 0 obs. truncated)

Truncated regression
Limit:   lower =         50                     Number of obs     =        367
         upper =       +inf                     Wald chi2(1)      =     103.19
Log likelihood =  -2523.232                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   230.6668   22.70729    10.16   0.000     186.1613    275.1723
       _cons |   -295.795   86.40175    -3.42   0.001    -465.1393   -126.4507
-------------+----------------------------------------------------------------
      /sigma |   306.5489   16.91234    18.13   0.000     273.4013    339.6965
------------------------------------------------------------------------------

. predict truncated, e(0,.)

. est store m_y1t

. lrtest m_y1 m_y1t, force

Likelihood-ratio test                                 LR chi2(1)  =     68.72
(Assumption: m_y1 nested in m_y1t)                    Prob > chi2 =    0.0000
```

Ho is rejected, it concludes that Truncated regression model is more appropriated than OLS because the data below 50.21759 is truncated or there are missing data.

Since In OLS, it assume normal distribution, while in this case we have truncated normal distribution. so, if we ignore the problem, the estimated result will be biased.

(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)

```
. reg y2 x

      Source |       SS           df       MS      Number of obs   =       450
-------------+----------------------------------   F(1, 448)       =    261.62
       Model |  19023316.5         1  19023316.5   Prob > F        =    0.0000
    Residual |  32576249.4       448  72714.8425   R-squared       =    0.3687
-------------+----------------------------------   Adj R-squared   =    0.3673
       Total |  51599565.9       449  114921.082   Root MSE        =    269.66

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   199.0755   12.30796    16.17   0.000     174.8869     223.264
       _cons |  -168.4949   39.08876    -4.31   0.000     -245.315    -91.6748
------------------------------------------------------------------------------

.  est store m_y2

. tobit y2 x, ll(0)

Tobit regression                                Number of obs     =        450
                                                LR chi2(1)        =     218.41
                                                Prob > chi2       =     0.0000
Log likelihood = -2789.4105                     Pseudo R2         =     0.0377

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   236.6343   14.70838    16.09   0.000     207.7285    265.5401
       _cons |  -311.9509   47.54813    -6.56   0.000    -405.3954   -218.5064
-------------+----------------------------------------------------------------
      /sigma |    301.912    11.1803                      279.9398    323.8842
------------------------------------------------------------------------------
            67  left-censored observations at y2 <= 0
           383     uncensored observations
             0 right-censored observations

. est store m_y2c

. lrtest m_y2 m_y2c, force

Likelihood-ratio test                                 LR chi2(1)  =    733.65
    (Assumption: m_y2 nested in m_y2c)                    Prob > chi2 =    0.0000
```

Ho is rejected, it concludes that Tobit regression model is more appropriated than OLS.

Since In OLS, it assume normal distribution, while in this case, we have censored normal distribution at zero. So if we ignore the problem, the estimated result will be biased.

(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)

```
. reg y3 x

      Source |       SS           df       MS      Number of obs   =       450
-------------+----------------------------------   F(1, 448)       =     37.28
       Model |  7.2978e+09         1  7.2978e+09   Prob > F        =    0.0000
    Residual |  8.7695e+10       448   195747673   R-squared       =    0.0768
-------------+----------------------------------   Adj R-squared   =    0.0748
       Total |  9.4993e+10       449   211565203   Root MSE        =     13991

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   3899.161   638.5912     6.11   0.000     2644.154    5154.167
       _cons |  -8369.074   2028.096    -4.13   0.000    -12354.84    -4383.31
------------------------------------------------------------------------------

. predict y3_hat
(option xb assumed; fitted values)

. est store m_y3

. reg y3 x if y3<=1000

      Source |       SS           df       MS      Number of obs   =       427
-------------+----------------------------------   F(1, 425)       =    204.87
       Model |  15589460.5         1  15589460.5   Prob > F        =    0.0000
    Residual |  32339580.2       425    76093.13   R-squared       =    0.3253
-------------+----------------------------------   Adj R-squared   =    0.3237
       Total |  47929040.8       426  112509.485   Root MSE        =    275.85

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   193.8728   13.54485    14.31   0.000     167.2496     220.496
       _cons |  -205.1155   41.81688    -4.91   0.000    -287.3091   -122.9218
------------------------------------------------------------------------------

. predict y3_hat_o
(option xb assumed; fitted values)

. est store m_y3o

.  tobit y3 x, ul(1000) nolog

Tobit regression                                Number of obs     =        450
                                                LR chi2(1)        =     218.30
                                                Prob > chi2       =     0.0000
Log likelihood = -3053.3781                     Pseudo R2         =     0.0345

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   224.3366   13.58925    16.51   0.000     197.6302     251.043
       _cons |  -270.6592   42.82189    -6.32   0.000    -354.8155    -186.503
-------------+----------------------------------------------------------------
      /sigma |   291.6248   10.07836                      271.8182    311.4314
------------------------------------------------------------------------------
             0  left-censored observations
           427     uncensored observations
            23 right-censored observations at y3 >= 1000

. est store m_y3o_t

. lrtest m_y3 m_y3o_t, force

Likelihood-ratio test                                 LR chi2(1)  =   3759.84
```

```
(Assumption: m_y3 nested in m_y3o_t)                    Prob > chi2 =     0.0000
```

Ho is rejected, it concludes that Tobit regression model is more appropriated than OLS in this case.

Since In OLS, it assumes normal distribution while in this case we have censored normal distribution (at 1,000). so, if we ignore the problem, the estimated result will be biased.

3. From the data set "Final_q3_no.dta":

The generalized linear regression model can be stated as:

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

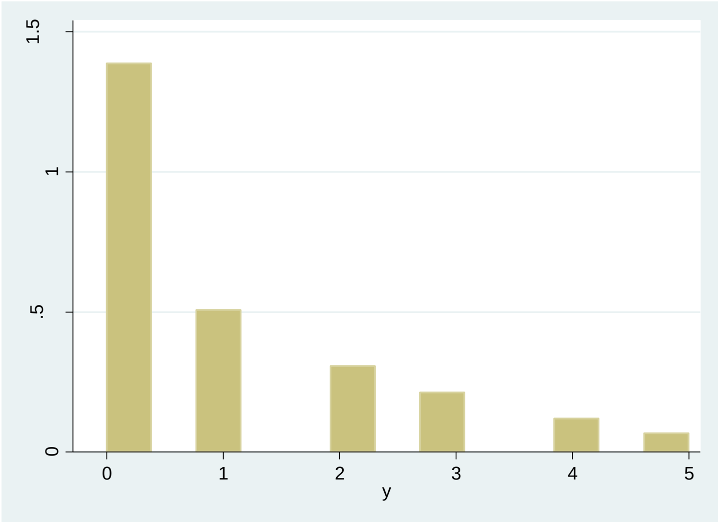
and

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

where: 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)

```
. reg y x1 x2 x3 x4

      Source |       SS           df       MS      Number of obs   =       195
-------------+----------------------------------   F(4, 190)       =      5.25
       Model |  35.0234447         4  8.75586118   Prob > F        =    0.0005
    Residual |  316.956042       190   1.6681897   R-squared       =    0.0995
-------------+----------------------------------   Adj R-squared   =    0.0805
       Total |  351.979487       194  1.81432725   Root MSE        =    1.2916

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0819159   .0445871     1.84   0.068    -.0060333    .1698651
          x2 |   .1304033   .0476292     2.74   0.007     .0364533    .2243533
          x3 |   .1974294   .0678561     2.91   0.004     .0635814    .3312775
          x4 |  -.0431584   .0485082    -0.89   0.375    -.1388423    .0525255
       _cons |    .928118   .1093498     8.49   0.000     .7124225    1.143813
------------------------------------------------------------------------------


. est store linear


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



From Histogram, there is a limitation of dependent variable in this case, and it seems like the distribution of dependent variable follows Poisson distribution.

(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^2, marginal effects). (5 points)

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

Poisson regression                              Number of obs     =        195
                                                LR chi2(4)        =      35.91
                                                Prob > chi2       =     0.0000
Log likelihood = -274.18362                     Pseudo R2         =     0.0615

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0830985   .0341682     2.43   0.015       .01613    .1500669
          x2 |   .1313239   .0365182     3.60   0.000     .0597497    .2028982
          x3 |   .2111245   .0547201     3.86   0.000     .1038751    .3183739
          x4 |  -.0443485   .0376632    -1.18   0.239    -.1181669      .02947
       _cons |  -.1643941   .0935358    -1.76   0.079    -.3477208    .0189327
------------------------------------------------------------------------------

. estat gof

         Deviance goodness-of-fit =  317.5877
         Prob > chi2(190)         =    0.0000

         Pearson goodness-of-fit  =  334.4442
         Prob > chi2(190)         =    0.0000
```

From GOF test, the Ho is rejected, therefore, Poisson regression model is inappropriated.

```
. poisson y x1 x2 x3 x4, ir nolog

Poisson regression                              Number of obs     =        195
                                                LR chi2(4)        =      35.91
                                                Prob > chi2       =     0.0000
Log likelihood = -274.18362                     Pseudo R2         =     0.0615

------------------------------------------------------------------------------
           y |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.086649   .0371288     2.43   0.015     1.016261    1.161912
          x2 |   1.140337    .041643     3.60   0.000     1.061571    1.224948
          x3 |   1.235066   .0675829     3.86   0.000     1.109462     1.37489
          x4 |   .9566206   .0360294    -1.18   0.239     .8885477    1.029909
       _cons |   .8484076   .0793565    -1.76   0.079      .706296    1.019113
------------------------------------------------------------------------------

.  est store poisso

. mfx

Marginal effects after poisson
      y  = Predicted number of events (predict)
         =  .89995812
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .0747852      .03035    2.46   0.014   .015297  .134273  -.277678
      x2 |    .118186      .03197    3.70   0.000   .055524  .180848   .827623
      x3 |   .1900032      .04739    4.01   0.000   .097129  .282877  -.293624
      x4 |  -.0399118      .03378   -1.18   0.237   -.10612  .026297  -.797462
```

- correct sign (positive mfx sign and irr>1 for x1, x2, and x3 and negative mfx sign and irr<1 for x4. Meaning that x1, x2, and x3 have a positive relationship while x4 has a negative relationship.
- For mfx of x1, we can interpret that if x1 changes by 1 unit from -.277678, y will change correspondingly by .0747852
- For mfx of x2, we can interpret that if x2 changes by 1 unit from .827623, y will change correspondingly by .118186
- For mfx of x3, we can interpret that if x3 changes by 1 unit from -.293624 , y will change correspondingly by .1900032
- For mfx of x4, we can interpret that if x4 changes by 1 unit from -.797462, y will change reversely by -.0399118
- The overall test (LR chi2) is significant at 0.05 level of significance.
- Pseudo R-squared is quite low. The model has a low level of goodness of fit.
- Individual tests (Z-test) are all significant except for x4.

(c) Estimate models for Yi 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 2 , marginal effects). (5 points)

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

Negative binomial regression                      Number of obs     =        195
                                                  LR chi2(4)        =      19.27
Dispersion     = mean                             Prob > chi2       =     0.0007
Log likelihood = -259.14063                       Pseudo R2         =     0.0358

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .1100377   .0526137     2.09   0.036     .0069167    .2131587
          x2 |   .1485162   .0523808     2.84   0.005     .0458517    .2511807
          x3 |   .2027863   .0707621     2.87   0.004     .0640952    .3414774
          x4 |  -.0469895   .0506778    -0.93   0.354    -.1463161    .0523371
       _cons |  -.1826632    .123203    -1.48   0.138    -.4241366    .0588102
-------------+----------------------------------------------------------------
    /lnalpha |  -.1672277   .2899995                     -.7356162    .4011609
-------------+----------------------------------------------------------------
       alpha |    .846007   .2453416                      .4792101    1.493558
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =   30.09 Prob>=chibar2 = 0.000
```

H0 is rejected, so the distribution of dependent variable follows Negative Binomial distribution, therefore, Negative Binomial regression model is more appropriated than Poisson regression model.

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

Negative binomial regression                      Number of obs     =        195
                                                  LR chi2(4)        =      19.27
Dispersion     = mean                             Prob > chi2       =     0.0007
Log likelihood = -259.14063                       Pseudo R2         =     0.0358

------------------------------------------------------------------------------
           y |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |    1.11632   .0587337     2.09   0.036     1.006941    1.237581
          x2 |   1.160112   .0607676     2.84   0.005     1.046919    1.285542
          x3 |   1.224811   .0866701     2.87   0.004     1.066194    1.407025
          x4 |   .9540974   .0483515    -0.93   0.354     .8638846    1.053731
       _cons |   .8330487   .1026341    -1.48   0.138     .6543345    1.060574
-------------+----------------------------------------------------------------
    /lnalpha |  -.1672277   .2899995                     -.7356162    .4011609
-------------+----------------------------------------------------------------
       alpha |    .846007   .2453416                      .4792101    1.493558
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =   30.09 Prob>=chibar2 = 0.000

mfx

Marginal effects after nbreg
      y  = Predicted number of events (predict)
         =   .8937104
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
```

```
      x1 |   .0983418       .04688     2.10    0.036    .006456  .190227  -.277678
      x2 |   .1327305       .04665     2.85    0.004    .041302  .224159   .827623
      x3 |   .1812322       .06295     2.88    0.004     .05785  .304614  -.293624
      x4 |   -.041995       .04528    -0.93    0.354   -.130742  .046752  -.797462
------------------------------------------------------------------------------
```

- correct sign (positive mfx sign and irr>1 for x1, x2, and x3 and negative mfx sign and irr<1 for x4. Meaning that x1, x2, and x3 have a positive relationship while x4 has a negative relationship.
- The overall test (LR chi2) is significant at 0.05 level of significance.
- Pseudo R-squared is quite low . The model has a low level of goodness of fit.
- Individual tests (Z-test) are all significant except for x4.

(d) Estimate models for Yi assuming that the model is Zero Inflated Poisson (x1i, x2i, and x3i are independent variables in Poisson model and x4i 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)

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

Zero-inflated Poisson regression                Number of obs     =        195
                                                Nonzero obs       =         91
                                                Zero obs          =        104

Inflation model = logit                         LR chi2(3)        =      14.01
Log likelihood  =  -258.198                     Prob > chi2       =     0.0029

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
          x1 |   .0957282   .0433527     2.21   0.027     .0107586    .1806979
          x2 |   .1169135   .0405778     2.88   0.004     .0373826    .1964445
          x3 |   .1462367   .0559321     2.61   0.009     .0366118    .2558616
       _cons |   .3354217   .1159017     2.89   0.004     .1082584    .5625849
-------------+----------------------------------------------------------------
inflate      |
          x4 |   .0722899   .1018143     0.71   0.478    -.1272625    .2718423
       _cons |  -.4812057   .2568159    -1.87   0.061    -.9845556    .0221441
------------------------------------------------------------------------------
Vuong test of zip vs. standard Poisson:            z =     2.52  Pr>z = 0.0059

. est store zip
```

From Vuong test, the null hypothesis is rejected, therefore, Zero Inflated Poisson regression model is more appropriated than Poisson regression model.

```
. mfx

Marginal effects after zip
      y  = Predicted number of events (predict)
         =  .90762583
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .0868854       .03909    2.22   0.026   .010267  .163504  -.277678
      x2 |   .1061138       .03613    2.94   0.003   .035298   .17693   .827623
      x3 |   .1327282       .04963    2.67   0.007   .035459  .229997  -.293624
      x4 |  -.0241751       .03437   -0.70   0.482  -.091536  .043186  -.797462
------------------------------------------------------------------------------
```

- correct sign (positive mfx sign and irr>1 for x1, x2, and x3 and negative mfx sign and irr<1 for x4. Meaning that x1, x2, and x3 have a positive relationship while x4 has a negative relationship.
- The overall test (LR chi2) is significant at 0.05 level of significance.
- Individual tests (Z-test) are all significant except for x4.
- To compare with other models, we have log likelihood value, which is equal to -258.198

```
est table linear poisso nb zip, star(.1 .05 .01) stat(N ll chi2 chi2_c vuong)

------------------------------------------------------------------------------
    Variable |    linear          poisso            nb              zip
-------------+----------------------------------------------------------------
_            |
          x1 |  .08191593*
          x2 |  .13040328***
          x3 |  .19742943***
          x4 | -.04315841
       _cons |  .92811798***
-------------+----------------------------------------------------------------
y            |
          x1 |                 .08309849**     .11003771**     .09572825**
          x2 |                 .13132394***    .14851618***    .11691355***
          x3 |                 .21112452***    .20278629***    .14623674***
          x4 |                -.04434846      -.0469895
       _cons |                -.16439406*     -.18266319       .33542166***
-------------+----------------------------------------------------------------
lnalpha      |
       _cons |                                -.16722766
-------------+----------------------------------------------------------------
inflate      |
          x4 |                                                 .07228991
       _cons |                                                -.48120574*
-------------+----------------------------------------------------------------
Statistics   |
           N |        195             195             195             195
          ll | -324.05496      -274.18362      -259.14063      -258.19798
        chi2 |                  35.913136       19.269785       14.012339
      chi2_c |                                   30.08598
       vuong |                                                  2.5155458
------------------------------------------------------------------------------
                                          legend: * p<.1; ** p<.05; *** p<.01
```

.

From GOF test, the null hypothesis is rejected, therefore, Poisson regression model is inappropriated.

From LR test of alpha, the null hypothesis is rejected. This implies that the distribution of dependent variable follows Negative Binomial distribution, therefore, Negative Binomial regression model is more appropriated than Poisson regression model.

From Vuong test, the null hypothesis is rejected, therefore, Zero Inflated Poisson regression model is more appropriated than Poisson regression model.

Lastly, Histogram shows zero inflated distribution of dependent variable, therefore, Zero Inflated Poisson regression model should be applied.

4. 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)

```
. use "C:\Users\acer\Desktop\06 (4)\Final_q4_6.dta", clear


. tsset t
        time variable:  t, 1 to 500
                delta:  1 unit
```

```
.  dfuller y, trend lag(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)            -15.001            -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. |  -.9710315   .0647326   -15.00   0.000    -1.098217   -.8438462
         LD. |  -.0620008   .0449836    -1.38   0.169    -.1503836    .0263819
      _trend |    .001482   .0021087     0.70   0.483    -.0026611     .005625
       _cons |   .8850897   .6115951     1.45   0.148    -.3165588    2.086738
------------------------------------------------------------------------------


. dfuller x, trend lag(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)            -15.169            -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. |  -.9837783   .0648541   -15.17   0.000    -1.111202   -.8563544
         LD. |  -.0521873   .0449925    -1.16   0.247    -.1405875     .036213
      _trend |   .0024391   .0030117     0.81   0.418    -.0034782    .0083564
       _cons |   .4820483   .8695552     0.55   0.580    -1.226434    2.190531
------------------------------------------------------------------------------
```

Y and X is stationary series with 0 order of integration.

Testing equation is important because we can distinguish between stationary and nonstationary.

The stationary property is also very important to time series analysis because if x and y are uncorrelated nonstationary series, OLS estimated result can lead to spurious problem.

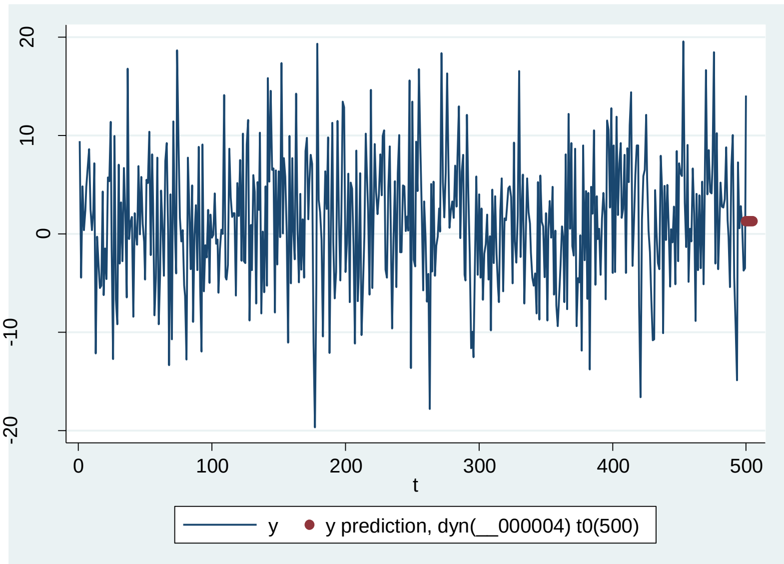
(b) Estimate Autoregressive Integrated Moving Average (ARIMA(p,d,q)) model for y_t – determine the most appropriated order for p, d, and q using SBIC given the maximum lag equals 4. Make dynamic forecast for period time = 501 to 505. (5 points)

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

.  est store arima101

. qui arima y, arima(1,0,2) nolog

. est store arima102

. qui arima y, arima(1,0,3) nolog

. est store arima103

. qui arima y, arima(1,0,4) nolog

. est store arima104

. qui arima y, arima(2,0,1) nolog

. est store arima201

. qui arima y, arima(2,0,2) nolog

. est store arima202

. qui arima y, arima(2,0,3) nolog

. est store arima203

. qui arima y, arima(2,0,4) nolog

. est store arima204

. qui arima y, arima(3,0,1) nolog

. est store arima301

. qui arima y, arima(3,0,2) nolog

.  est store arima302

. qui arima y, arima(3,0,3) nolog

. est store arima303

.  qui arima y, arima(3,0,4) nolog
```

```
. est store arima304

. qui arima y, arima(4,0,1) nolog

. est store arima401

. qui arima y, arima(4,0,2) nolog

. est store arima402

. qui arima y, arima(4,0,3) nolog

. est store arima403

. qui arima y, arima(4,0,4) nolog

. est store arima404

. est table arima10*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

------------------------------------------------------------------------------
    Variable |   arima101        arima102        arima103        arima104
-------------+----------------------------------------------------------------
y            |
       _cons |  1.2976234***    1.3017904***    1.2984793***    1.2990195***
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. | -.85339435***    .80471486***   -.85468396***   -.81369849***
             |
          ma |
         L1. |   .8100097***   -.84140844***    .82228359***     .7816922***
         L2. |                  .07274513        .0306972       .02996439
         L3. |                                  .02098039       .05220677
         L4. |                                                  .04870085
-------------+----------------------------------------------------------------
sigma        |
       _cons |   6.731176***    6.7309467***    6.7288012***    6.7214207***
-------------+----------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -1662.8474      -1662.8331      -1662.6765      -1662.1331
        chi2 |  50.892425        20.51703       51.035225       42.475594
         aic |  3333.6948       3335.6663        3337.353       3338.2663
         bic |  3350.5532       3356.7393       3362.6407       3367.7685
------------------------------------------------------------------------------
                                              legend: * p<.1; ** p<.05; *** p<.01


. est table arima20*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)
------------------------------------------------------------------------------
```

```
    Variable |   arima201       arima202       arima203       arima204
-------------+------------------------------------------------------------
y            |
       _cons |  1.3018413***   1.3019576***   1.3017336***   1.3017428***
-------------+------------------------------------------------------------
ARMA         |
          ar |
         L1. |  .71409244***   .02468243     -.01811561     -.03322983
         L2. |  .07757072      .71766646***   .71547163***   .67485622**
             |
          ma |
         L1. | -.75319804***  -.04626585     -.01664366     -.00061583
         L2. |                -.64829034**   -.64745466**   -.61994551*
         L3. |                               .02423351      .02374469
         L4. |                                              .02366394
-------------+------------------------------------------------------------
sigma        |
       _cons |  6.7299527***   6.7190432***   6.7177443***   6.7163847***
-------------+------------------------------------------------------------
Statistics   |
           N |        500            500            500            500
          ll | -1662.7696     -1661.9778     -1661.8638     -1661.7575
        chi2 |  17.704828      20.073698      21.169536      19.961376
         aic |  3335.5393      3335.9556      3337.7276      3339.5149
         bic |  3356.6123      3361.2433      3367.2298      3373.2318
--------------------------------------------------------------------------
                                          legend: * p<.1; ** p<.05; *** p<.01

. est table arima30*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

--------------------------------------------------------------------------
    Variable |   arima301       arima302       arima303       arima304
-------------+------------------------------------------------------------
y            |
       _cons |  1.2986188***   1.3017002***    1.294539***   1.3172638***
-------------+------------------------------------------------------------
ARMA         |
          ar |
         L1. | -.84511426**   -.04959244      1.0083066***   1.0723736***
         L2. |  .03426642      .71676019***   .63479491**    .64325628*
         L3. |  .02318444      .02495846     -.83116631***  -.83130545***
             |
          ma |
         L1. |  .81452378**    .01568792     -1.0775891     -1.1290034
         L2. |                -.64941261**   -.55116705     -.54534196
         L3. |                               .81821177      .79923642
         L4. |                                              .0009893
-------------+------------------------------------------------------------
sigma        |
       _cons |  6.7286239***    6.717953***   6.6718578      6.6376001
-------------+------------------------------------------------------------
Statistics   |
           N |        500            500            500            500
          ll | -1662.6604     -1661.8715     -1660.0652     -1657.9328
        chi2 |  52.638834       21.81691      13844.474      56115.793
         aic |  3337.3208       3337.743      3334.1304      3331.8656
         bic |  3362.6084      3367.2453      3363.6327      3365.5824
--------------------------------------------------------------------------
                                          legend: * p<.1; ** p<.05; *** p<.01

. est table arima40*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

--------------------------------------------------------------------------
    Variable |   arima401       arima402       arima403       arima404
-------------+------------------------------------------------------------
y            |
       _cons |  1.2995213***   1.3017757***   1.3172469***   1.3025988***
-------------+------------------------------------------------------------
ARMA         |
          ar |
         L1. | -.76992642**   -.07039727      1.0736203***   .07331874
         L2. |  .03472584      .64022156      .64192325     -.23964631
         L3. |  .05473538      .02622256     -.83211544***  -.05449533
         L4. |  .05036241      .02580944      .00104042      .69542805***
             |
          ma |
         L1. |  .73813557**    .03616772     -1.1302345     -.10895666
         L2. |                -.58600182     -.54395494      .32521929
         L3. |                               .79991478      .05450831
         L4. |                                             -.64403588**
-------------+------------------------------------------------------------
sigma        |
       _cons |  6.7209213***   6.7162027***   6.6375942      6.6759466***
```

```
-------------+----------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -1662.1069      -1661.7584      -1657.9328      -1659.0192
        chi2 |  35.804182       19.402517        38105.66       1412.2428
         aic |  3338.2137       3339.5168       3331.8656       3338.0384
         bic |   3367.716       3373.2337       3365.5824       3380.1845
------------------------------------------------------------------------------
                                          legend: * p<.1; ** p<.05; *** p<.01

.
The most appropriated order for y is ARIMA(1,0,1) because this model has the lowest BIC value
comparing to others.

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

ARIMA regression

Sample:  1 - 500                                Number of obs     =        500
                                                Wald chi2(2)      =      50.89
Log likelihood = -1662.847                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
       _cons |   1.297623   .2959697     4.38   0.000     .7175334    1.877713
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.8533943   .1664072    -5.13   0.000    -1.179546   -.5272423
             |
          ma |
         L1. |   .8100097   .1879745     4.31   0.000     .4415865    1.178433
-------------+----------------------------------------------------------------
      /sigma |   6.731176   .2150212    31.30   0.000     6.309742     7.15261
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.

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

. replace t=_n
(5 real changes made)

.  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: $y_t = \alpha + \beta x_t + \varepsilon_t$ (5)

Variance Equation: $\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \delta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$ (6)

where: y_t is Dependent variable.

x_t is Independent variable

ε_t is Stochastic disturbance term

σ_t^2 is Conditional variance of stochastic disturbance term.

(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)

```
.  reg y x


      Source |       SS           df       MS      Number of obs   =       500
-------------+----------------------------------   F(1, 498)       =  57542.43
       Model |  22613.3561         1  22613.3561   Prob > F        =    0.0000
    Residual |  195.706923       498  .39298579    R-squared       =    0.9914
-------------+----------------------------------   Adj R-squared   =    0.9914
       Total |   22809.063       499  45.7095451   Root MSE        =    .62689


------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .6976068   .0029081   239.88   0.000      .691893    .7033205
       _cons |   .5202284   .0282226    18.43   0.000     .4647784    .5756784
------------------------------------------------------------------------------


. . estat archlm
LM test for autoregressive conditional heteroskedasticity (ARCH)
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |         38.734               1                   0.0000
---------------------------------------------------------------------------
         H0: no ARCH effects      vs.  H1: ARCH(p) disturbance
```

There exist significant ARCH effects since p-value of the ARCHLM-test is less than 0.05.

And The null hypothesis is there is no ARCH effects.

The test is classified as LM test because it applied only the restricted model.

(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)

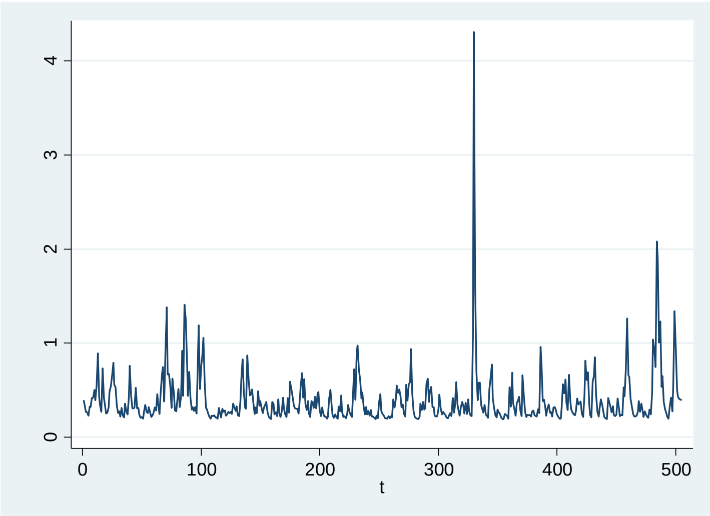
```
. qui arch y x, arch(1) garch(1) nolog

. est store garch11

. qui arch y x, arch(1) garch(1/2) nolog

. est store garch21

. qui arch y x, arch(1/2) garch(1) nolog

. est store garch12

. qui arch y x, arch(1/2) garch(1/2) nolog

. est store garch22

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

------------------------------------------------------------------------------
    Variable |    garch11         garch21         garch12         garch22
-------------+----------------------------------------------------------------
y            |
           x |  .69778823***    .69777546***    .69748956***     .6976637***
       _cons |  .52211886***    .52208818***    .51847827***    .51953722***
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |  .36209437***     .3604112***    .37129708***    .37147301***
         L2. |                                  .20249581       .30914074**
             |
       garch |
         L1. |  .31747404***    .33098321*     -.13894879      -.38917515
         L2. |                 -.01197256                       .10891045
             |
       _cons |  .12790338***    .12795545***    .22682787***    .24348238***
-------------+----------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -443.07673      -443.06878      -442.90866      -442.22038
        chi2 |  73316.172        73392.74       73791.769       75006.387
         aic |  896.15345       898.13756       897.81731       898.44076
         bic |  917.22649       923.42521       923.10496       927.94302
```

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

.

If we use BIC, the most appropriated lags order in this case is GARCH(1,1)

```
. qui arch y x, arch(1) garch(1) nolog

. predict sigmahat, v

. line sigmahat 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) garch(1) egarch(1) nolog

ARCH family regression

Sample: 1 - 500                                 Number of obs     =        500
Distribution: Gaussian                          Wald chi2(1)      =   67421.52
Log likelihood = -446.5533                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
           x |   .6979331   .0026879   259.66   0.000     .6926649    .7032013
       _cons |   .5153192   .0250464    20.57   0.000     .4662293    .5644092
-------------+----------------------------------------------------------------
ARCH         |
      egarch |
         L1. |   .3625097   .1231862     2.94   0.003     .1210692    .6039502
             |
        arch |
         L1. |   .6746068   .0926285     7.28   0.000     .4930583    .8561553
             |
       garch |
         L1. |  -.0059931    .010524    -0.57   0.569    -.0266197    .0146335
             |
       _cons |  -.9245405   .1563816    -5.91   0.000    -1.231043   -.6180382
------------------------------------------------------------------------------

. est store egarch111
```

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

ARCH family regression

Sample: 1 - 500                                 Number of obs   =        500
Distribution: Gaussian                          Wald chi2(1)    =   66650.90
Log likelihood = -448.1987                      Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
           x |   .6976727   .0027024   258.17   0.000     .6923761    .7029692
       _cons |   .5179213   .0246652    21.00   0.000     .4695785    .5662641
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |   .3813052   .0783018     4.87   0.000     .2278365    .5347738
             |
       _cons |   .2412147   .0221462    10.89   0.000     .1978089    .2846206
------------------------------------------------------------------------------

. est store arch1

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

--------------------------------------------------------------
    Variable |     arch1          garch11        egarch111
-------------+------------------------------------------------
y            |
           x |  .69767265***    .69778823***    .69793313***
       _cons |   .5179213***    .52211886***    .51531924***
-------------+------------------------------------------------
ARCH         |
        arch |
         L1. |  .38130516***    .36209437***    .67460682***
             |
       garch |
         L1. |                  .31747404***    -.0059931
             |
      egarch |
         L1. |                                  .36250967***
             |
       _cons |  .24121474***    .12790338***   -.92454047***
-------------+------------------------------------------------
Statistics   |
           N |        500             500             500
          ll | -448.19868      -443.07673      -446.55329
        chi2 |  66650.901       73316.172       67421.522
         aic |  904.39735       896.15345       905.10657
         bic |  921.25579       917.22649       930.39422
--------------------------------------------------------------
                           legend: * p<.1; ** p<.05; *** p<.01
```

the most appropriated model is GARCH(1,1) since it has the lowest BIC value.

To distinguish, conditional variance of ARCH model only depends on error term while conditional variance of GARCH model depends on both error terms and variances. And EGARCH like GARCH but it adds the exponential term.

5. From the data set "`Final_q5_no.dta`":

The following VARs models:

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

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

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

```
use "C:\Users\acer\Desktop\06 (4)\Final_q5_6.dta", clear

. dfuller y, trend lag(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)            -14.146            -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. |  -.8517762   .0602112   -14.15   0.000    -.9700779   -.7334746
         LD. |  -.0426613   .0451472    -0.94   0.345    -.1313655    .0460429
      _trend |   .0000273   .0003107     0.09   0.930    -.0005831    .0006378
       _cons |   .3428503   .0928616     3.69   0.000      .160398    .5253027
------------------------------------------------------------------------------

. dfuller x, trend lag(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)            -15.634            -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. |  -.9649704   .0617243   -15.63   0.000    -1.086245   -.8436958
```

```
         LD. |   .0252997   .0450268     0.56   0.574    -.0631679    .1137674
      _trend |   .0004297   .0003257     1.32   0.188    -.0002101    .0010696
       _cons |   .2134465   .0947913     2.25   0.025     .0272026    .3996904
------------------------------------------------------------------------------
```

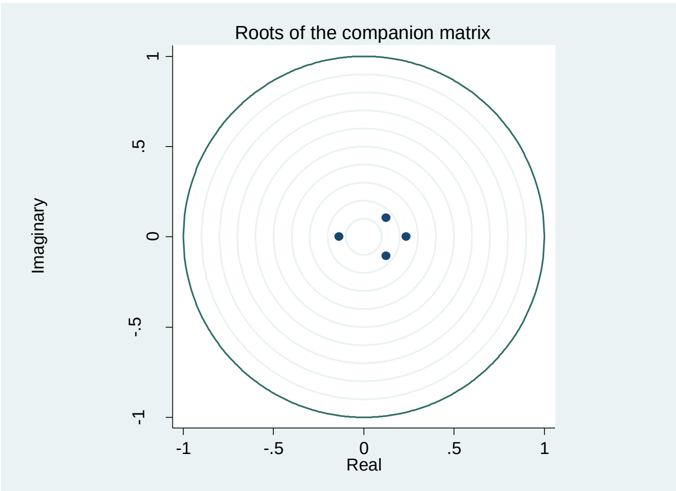
```
.  varsoc x y

   Selection-order criteria
   Sample:   5 - 500                             Number of obs      =       496
  +---------------------------------------------------------------------------+
  |lag |    LL      LR      df    p      FPE       AIC      HQIC      SBIC     |
  |----+----------------------------------------------------------------------|
  |  0 | -1365.52                      .850806   5.51418   5.52084   5.53114   |
  |  1 | -1319.44  92.163*   4  0.000  .718022*   5.3445*  5.36447*  5.39539*  |
  |  2 | -1318.57  1.7272    4  0.786  .727161   5.35715   5.39044   5.44196   |
  |  3 | -1317.41  2.3175    4  0.678  .735542    5.3686   5.41521   5.48734   |
  |  4 | -1315.17  4.4807    4  0.345  .740783    5.3757   5.43562   5.52836   |
  +---------------------------------------------------------------------------+
   Endogenous:  x y
    Exogenous:  _cons
```

According to SBIC, the most appropriated lag order is 1.

(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)

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

Vector autoregression

Sample:  2 - 500                                 Number of obs     =       499
Log likelihood =  -1329.569                      AIC               =  5.352981
FPE            =   .7241384                      HQIC              =  5.372859
Det(Sigma_ml)  =   .7069318                      SBIC              =  5.403634

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
x                     3     .996525   0.0882   48.27445   0.0000
y                     3     .994722   0.0140   7.071343   0.0291
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
x            |
           x |
         L1. |   .2131469    .048179     4.42   0.000     .1187178    .3075759
             |
           y |
         L1. |   .3412956   .0502324     6.79   0.000     .2428419    .4397493
```

```
             |
       _cons |   .1215588   .0545529     2.23   0.026     .0146372    .2284805
-------------+----------------------------------------------------------------
y            |
           x |
         L1. |   .0466198   .0480918     0.97   0.332    -.0476384     .140878
             |
           y |
         L1. |   .1325819   .0501415     2.64   0.008     .0343063    .2308574
             |
       _cons |   .3415196   .0544542     6.27   0.000     .2347914    .4482479
------------------------------------------------------------------------------


. varstable, graph

   Eigenvalue stability condition
  +----------------------------------------+
  |        Eigenvalue         |   Modulus   |
  |--------------------------+-------------|
  |   .2347851                |   .234785   |
  |   .1231458 +   .105778i   |   .162339   |
  |   .1231458 -   .105778i   |   .162339   |
  |  -.1381894                |   .138189   |
  +----------------------------------------+
   All the eigenvalues lie inside the unit circle.
   VAR satisfies stability condition.
```



```
vargranger

   Granger causality Wald tests
  +------------------------------------------------------------------+
  |          Equation           Excluded |   chi2     df  Prob > chi2 |
  |--------------------------------------+---------------------------|
  |                x                   y |  46.338     2    0.000     |
  |                x                 ALL |  46.338     2    0.000     |
  |--------------------------------------+---------------------------|
  |                y                   x |  .53897     2    0.764     |
  |                y                 ALL |  .53897     2    0.764     |
  +------------------------------------------------------------------+
```

According to stability test, the system is stable since all the eigenvalues lie inside the unit circle.

According to Granger exogeneity, x is endogenous variable since the tests are significant while y seems not to be endogenous.

- If the stability assumption is unsatisfied, what will happen?

If there is a violation of this assumption, irf will not go back to equilibrium.

(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)

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

Vector autoregression

Sample:  2 - 500                                Number of obs     =        499
Log likelihood =  -1329.569                     AIC               =   5.352981
FPE            =   .7241384                     HQIC              =   5.372859
Det(Sigma_ml)  =   .7069318                     SBIC              =   5.403634

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
x                     3     .996525   0.0882   48.27445   0.0000
y                     3     .994722   0.0140   7.071343   0.0291
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
x            |
           x |
         L1. |   .2131469    .048179     4.42   0.000     .1187178    .3075759
             |
           y |
         L1. |   .3412956   .0502324     6.79   0.000     .2428419    .4397493
             |
       _cons |   .1215588   .0545529     2.23   0.026     .0146372    .2284805
```

```
-------------+----------------------------------------------------------------
y            |
           x |
         L1. |   .0466198   .0480918     0.97   0.332    -.0476384     .140878
             |
           y |
         L1. |   .1325819   .0501415     2.64   0.008     .0343063    .2308574
             |
       _cons |   .3415196   .0544542     6.27   0.000     .2347914    .4482479
------------------------------------------------------------------------------

. . irf create order1, o(y x) step(10) set(irf04)
(file irf04.irf created)
(file irf04.irf now active)
(file irf04.irf updated)

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

                                         Results from order1

+--------------------------------------------------------------------------------+
|        |    (1)         (1)         (1)     |    (2)         (2)         (2)     |
|  step  |   irf        Lower       Upper     |   irf        Lower       Upper     |
|--------+------------------------------------+------------------------------------|
|0       | 1           1           1          | 0           0           0          |
|1       | .125686     .022795     .228577    | .343395     .240325     .446465    |
|2       | .055733     -.043435    .154901    | .113116     .009836     .216397    |
|3       | .01144      -.036681    .059561    | .032708     -.019473    .08489     |
|4       | .003107     -.019368    .025582    | .007343     -.01674     .031426    |
|5       | .000667     -.007523    .008856    | .001617     -.008785    .012018    |
|6       | .000159     -.00255     .002868    | .000343     -.003703    .004389    |
|7       | .000036     -.000783    .000854    | .000077     -.001328    .001482    |
|8       | 8.4e-06     -.000226    .000243    | .000018     -.000419    .000455    |
|9       | 2.0e-06     -.000063    .000066    | 4.3e-06     -.000121    .00013     |
|10      | 4.7e-07     -.000017    .000018    | 1.0e-06     -.000033    .000035    |
+--------------------------------------------------------------------------------+

+--------------------------------------------------------------------------------+
|        |    (3)         (3)         (3)     |    (4)         (4)         (4)     |
|  step  |   irf        Lower       Upper     |   irf        Lower       Upper     |
|--------+------------------------------------+------------------------------------|
|0       | 0           0           0          | 1           1           1          |
|1       | .038455     -.064216    .141127    | .217201     .11435      .320052    |
|2       | .003651     -.091337    .09864     | .030045     -.06917     .12926     |
|3       | .000571     -.032003    .033146    | .001013     -.038787    .040812    |
|4       | -.000078    -.010574    .010418    | -.000512    -.016496    .015472    |
|5       | -.000024    -.002733    .002685    | -.000171    -.006177    .005834    |
|6       | -6.8e-06    -.000677    .000663    | -.00003     -.001819    .00176     |
|7       | -1.0e-06    -.000153    .000151    | -3.4e-06    -.000444    .000437    |
```

8	-1.6e-07	-.000034	.000034	-1.6e-07	-.000093	.000093
9	-2.0e-08	-7.3e-06	7.3e-06	2.0e-08	-.000018	.000018
10	-4.4e-09	-1.6e-06	1.6e-06	3.1e-09	-3.4e-06	3.4e-06

95% lower and upper bounds reported

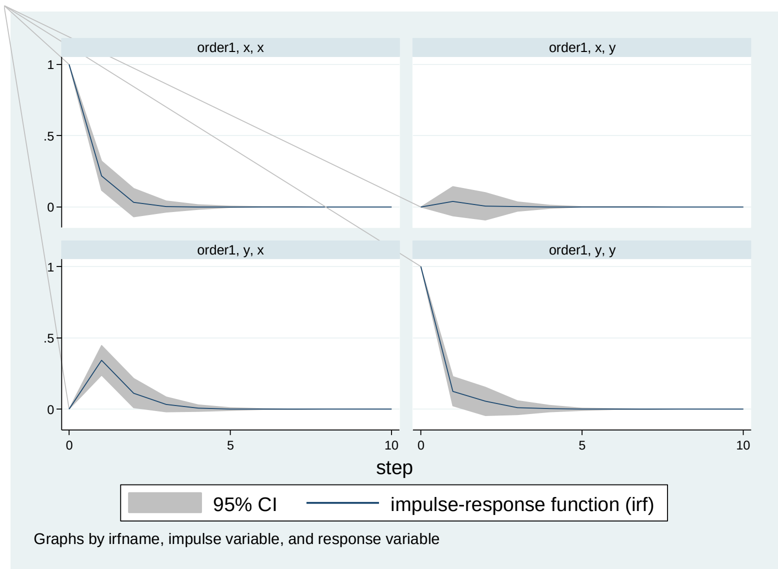
(1) irfname = order1, impulse = y, and response = y

(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)
```




```
. 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	.992088	.930476	1.0537	-.519819	-.600915	-.438723
1	.104702	.017296	.192107	.227773	.137901	.317645
2	.053394	-.034177	.140965	.096603	.005536	.187671
3	.011053	-.021652	.043758	.031923	-.007671	.071517
4	.003123	-.014403	.020649	.007551	-.011288	.026391
5	.000674	-.006119	.007467	.001693	-.006504	.00989
6	.000161	-.002199	.002522	.000356	-.002916	.003627
7	.000036	-.0007	.000772	.000078	-.00112	.001276
8	8.4e-06	-.000208	.000225	.000018	-.000373	.000408
9	2.0e-06	-.000058	.000062	4.2e-06	-.000112	.00012
10	4.6e-07	-.000016	.000017	1.0e-06	-.000031	.000033

	(3)	(3)	(3)	(4)	(4)	(4)
step	oirf	Lower	Upper	oirf	Lower	Upper

0	0	0	0	.847034	.79443	.899638
1	.032573	-.054417	.119563	.183977	.096112	.271841
2	.003093	-.077366	.083551	.025449	-.058604	.109503
3	.000484	-.027108	.028076	.000858	-.032854	.034569
4	-.000066	-.008956	.008824	-.000434	-.013973	.013105
5	-.00002	-.002315	.002274	-.000145	-.005232	.004941
6	-5.8e-06	-.000573	.000562	-.000025	-.001541	.001491
7	-8.4e-07	-.00013	.000128	-2.9e-06	-.000376	.00037
8	-1.3e-07	-.000029	.000028	-1.4e-07	-.000079	.000079
9	-1.7e-08	-6.2e-06	6.2e-06	1.7e-08	-.000015	.000015
10	-3.7e-09	-1.4e-06	1.3e-06	2.7e-09	-2.9e-06	2.9e-06

95% lower and upper bounds reported

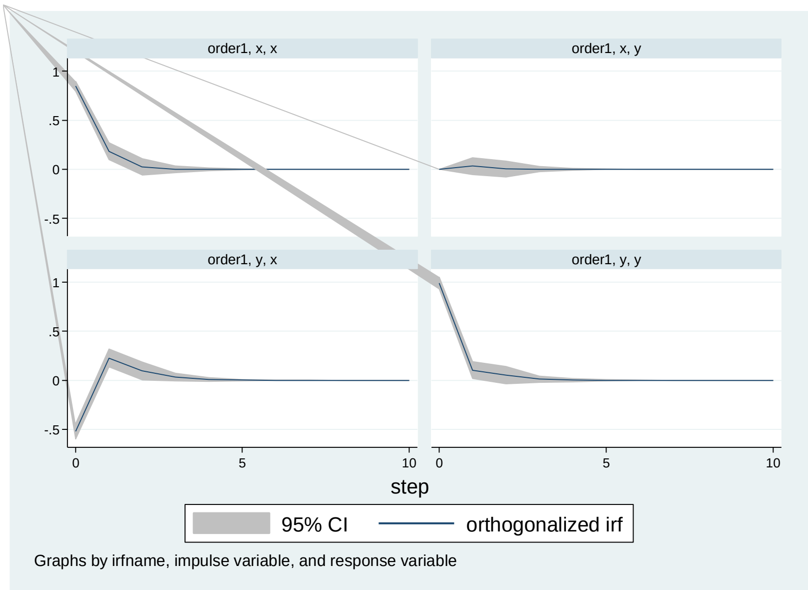
(1) irfname = order1, impulse = y, and response = y

(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)
```




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

Results from order1

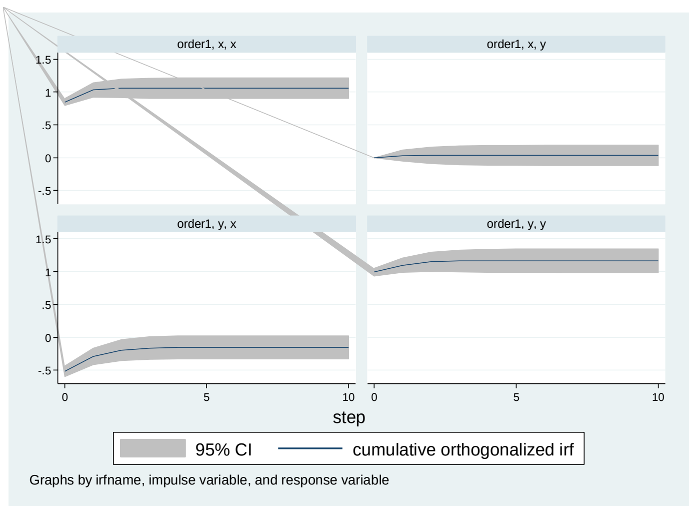
	(1)	(1)	(1)	(2)	(2)	(2)
step	coirf	Lower	Upper	coirf	Lower	Upper
0	.992088	.930476	1.0537	-.519819	-.600915	-.438723
1	1.09679	.986169	1.20741	-.292046	-.419106	-.164986
2	1.15018	1.00199	1.29838	-.195443	-.354197	-.036689
3	1.16124	.995419	1.32705	-.16352	-.334172	.007133
4	1.16436	.988689	1.34003	-.155968	-.332116	.020179

5	1.16503	.985834	1.34423	-.154275	-.332038	.023488
6	1.1652	.984775	1.34562	-.15392	-.332079	.024239
7	1.16523	.984446	1.34602	-.153841	-.332071	.024388
8	1.16524	.984349	1.34613	-.153824	-.332061	.024414
9	1.16524	.984322	1.34616	-.153819	-.332057	.024418
10	1.16524	.984315	1.34617	-.153818	-.332055	.024418

	(3)	(3)	(3)	(4)	(4)	(4)
step	coirf	Lower	Upper	coirf	Lower	Upper
0	0	0	0	.847034	.79443	.899638
1	.032573	-.054417	.119563	1.03101	.922893	1.13913
2	.035666	-.089374	.160706	1.05646	.916714	1.19621
3	.03615	-.108973	.181273	1.05732	.903733	1.2109
4	.036083	-.115706	.187873	1.05688	.900944	1.21282
5	.036063	-.117449	.189576	1.05674	.900873	1.2126
6	.036057	-.117878	.189993	1.05671	.901019	1.21241
7	.036057	-.117974	.190087	1.05671	.901079	1.21234
8	.036056	-.117995	.190108	1.05671	.901094	1.21233
9	.036056	-.118	.190113	1.05671	.901096	1.21232
10	.036056	-.118001	.190114	1.05671	.901097	1.21232

```
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)
```



According to IRF analysis, y has more impact on x.

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

```
. 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	.273583	.206842	.340324
2	.998935	.993251	1.00462	.300066	.235285	.364847
3	.998928	.993183	1.00467	.305915	.240649	.371181
4	.998928	.99318	1.00468	.306568	.24128	.371855
5	.998928	.99318	1.00468	.306604	.241316	.371891
6	.998928	.99318	1.00468	.306606	.241319	.371893
7	.998928	.99318	1.00468	.306606	.241319	.371893
8	.998928	.99318	1.00468	.306606	.241319	.371893
9	.998928	.99318	1.00468	.306606	.241319	.371893
10	.998928	.99318	1.00468	.306606	.241319	.371893

step	(3) fevd	(3) Lower	(3) Upper	(4) fevd	(4) Lower	(4) Upper
0	0	0	0	0	0	0
1	0	0	0	.726417	.659676	.793158
2	.001065	-.004619	.006749	.699934	.635153	.764715
3	.001072	-.004674	.006817	.694085	.628819	.759351
4	.001072	-.004677	.00682	.693432	.628145	.75872
5	.001072	-.004677	.00682	.693396	.628109	.758684
6	.001072	-.004677	.00682	.693394	.628107	.758681
7	.001072	-.004677	.00682	.693394	.628107	.758681
8	.001072	-.004677	.00682	.693394	.628107	.758681
9	.001072	-.004677	.00682	.693394	.628107	.758681
10	.001072	-.004677	.00682	.693394	.628107	.758681

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

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)

Yes, it can change the result since Different order sets up different var-cov matrix