Thanakrit Methasate 6204640087

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:

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

   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 x*, b(0)

Iteration 0:   log likelihood = -149.55771

. . .

Iteration 5:   log likelihood = -101.82217


Multinomial logistic regression                 Number of obs     =        170

                                                LR chi2(8)        =      95.47

                                                Prob > chi2       =     0.0000

Log likelihood = -101.82217                     Pseudo R2         =     0.3192


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

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

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

0            |  (base outcome)

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

1            |

          x1 |  -1.132993   .6280843    -1.80   0.071    -2.364015      .09803

          x2 |  -.7710116   .7633309    -1.01   0.312    -2.267113    .7250896

          x3 |  -.6175655   .6301033    -0.98   0.327    -1.852545    .6174143
```

```
          x4 |   .6370348   .2509491     2.54   0.011     .1451836    1.128886
       _cons |  -7.436673   3.561582    -2.09   0.037    -14.41725   -.4561011
-------------+----------------------------------------------------------------
2            |
          x1 |  -.9945724   .6972202    -1.43   0.154    -2.361099    .3719541
          x2 |  -1.527706    .797026    -1.92   0.055    -3.089848    .0344361
          x3 |   .6192554   .6740561     0.92   0.358    -.7018704    1.940381
          x4 |   1.885854    .314478     6.00   0.000     1.269488    2.502219
       _cons |  -27.28322   4.709975    -5.79   0.000     -36.5146   -18.05184
------------------------------------------------------------------------------

. mlogit y x*, rrr b(0)

Iteration 0:   log likelihood = -149.55771
. . .
Iteration 5:   log likelihood = -101.82217

Multinomial logistic regression                 Number of obs     =        170
                                                LR chi2(8)        =      95.47
                                                Prob > chi2       =     0.0000
Log likelihood = -101.82217                     Pseudo R2         =     0.3192

------------------------------------------------------------------------------
           y |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
-------------+----------------------------------------------------------------
1            |
          x1 |    .322068   .2022858    -1.80   0.071     .0940419    1.102996
          x2 |   .4625449   .3530749    -1.01   0.312     .1036109    2.064916
          x3 |   .5392557   .3397868    -0.98   0.327     .1568375    1.854128
          x4 |   1.890866    .474511     2.54   0.011     1.156252     3.09221
       _cons |   .0005892   .0020986    -2.09   0.037     5.48e-07    .6337498
-------------+----------------------------------------------------------------
```

```
2            |
          x1 |   .3698816   .2578889    -1.43   0.154     .0943165    1.450566
          x2 |    .217033   .1729809    -1.92   0.055     .0455089    1.035036
          x3 |   1.857544   1.252089     0.92   0.358     .4956574    6.961404
          x4 |   6.591981   2.073033     6.00   0.000     3.559031    12.20956
       _cons |   1.42e-12   6.67e-12    -5.79   0.000     1.39e-16    1.45e-08
------------------------------------------------------------------------------


. est store mlogit


. fitstat

Measures of Fit for mlogit of y


Log-Lik Intercept Only:       -149.558   Log-Lik Full Model:            -101.822
D(155):                        203.644   LR(8):                           95.471
                                         Prob > LR:                        0.000
McFadden's R2:                   0.319   McFadden's Adj R2:                0.219
ML (Cox-Snell) R2:               0.430   Cragg-Uhler(Nagelkerke) R2:       0.519
Count R2:                        0.735   Adj Count R2:                     0.262
AIC:                             1.374   AIC*n:                          233.644
BIC:                          -592.404   BIC':                           -54.385
BIC used by Stata:             255.002   AIC used by Stata:              223.644
```

According to the estimated results, comparing with the base outcome (y=0), x4 has a positive effect on the choice y=1 and y=2 (it seems like x4 has a larger (positive) effect on choice y = 2 since (RRR > 1) while x3 has the positive effect only on choice y=2. The other x (x1 x2) have negative effects on both choices since RRR < 1.

From the overall test, the null hypothesis is rejected since p-value = 0.000 < 0.05; thereby, all independent variables(x1 x2 x3 x4) altogether significantly explain the dependent variable (y).

In term of Goodness-of-fit, the pseudo R-squared is 0.319 (there are rooms for improvements) and counted R-squared = 0.735 meaning that this model can correctly predict the outcome 73.5%

The individual test suggests that only x4 is significant, and the rest are not.

IIA Test

```
. qui mlogit y x* if y!=1, b(0)

. est store mlogit1

. qui mlogit y x* if y!=2, b(0)

. est store mlogit2

. hausman mlogit1 mlogit,alleqs constant


                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    mlogit1       mlogit       Difference          S.E.
-------------+----------------------------------------------------------------
          x1 |   -1.510394    -.9945724       -.5158214        .5647965
          x2 |   -2.032272    -1.527706        -.504566        .6287449
          x3 |    .5052412     .6192554       -.1140142        .3330594
          x4 |    2.119195     1.885854        .2333412        .3529509
       _cons |   -30.33564    -27.28322       -3.052417        5.065843
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from mlogit
           B = inconsistent under Ha, efficient under Ho; obtained from mlogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       12.45
                Prob>chi2 =      0.0291

. hausman mlogit mlogit2,alleqs constant


                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     mlogit      mlogit2       Difference          S.E.
-------------+----------------------------------------------------------------
          x1 |   -1.132993    -.9574483       -.1755444        .1691332
          x2 |   -.7710116    -.5193718       -.2516398        .1573762
```

```
          x3 |   -.6175655    -.4413165       -.1762489               .
          x4 |    .6370348     .5667076        .0703272        .0475769
       _cons |   -7.436673    -6.747037       -.6896369               .
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from mlogit
           B = inconsistent under Ha, efficient under Ho; obtained from mlogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        1.20
                Prob>chi2 =      0.9446
                (V_b-V_B is not positive definite)
```

From the Hausman test, the IIA does not be held since y not equal to 1 case rejects the null hypothesis. Therefore, it could lead to a Red-Bus Blue-Bus Problem causing the estimated results to be biased.

IIA property is important since, as mentioned above, it could lead to a biasness in the estimated result. Moreover, Multinomial Logit ,the result, compares with the base case. That is why IIA must be held.

Since IIA cannot be held, Multinomial Logit is not appropriated in this case.

(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 x*
Iteration 0:   log likelihood = -149.55771
. . .
Iteration 4:   log likelihood = -106.87931

Ordered probit regression                       Number of obs     =        170
                                                LR chi2(4)        =      85.36
                                                Prob > chi2       =     0.0000
Log likelihood = -106.87931                     Pseudo R2         =     0.2854
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	-.2072486	.2327348	-0.89	0.373	-.6634004	.2489032
x2	-.4271198	.239373	-1.78	0.074	-.8962822	.0420427
x3	.5026001	.2192698	2.29	0.022	.0728392	.9323611
x4	.7067322	.0937676	7.54	0.000	.522951	.8905133
/cut1	9.831223	1.423997			7.04024	12.62221
/cut2	11.08689	1.480594			8.184983	13.9888

```
. fitstat
Measures of Fit for oprobit of y

Log-Lik Intercept Only:       -149.558   Log-Lik Full Model:          -106.879
D(164):                        213.759   LR(4):                         85.357
                                                Prob > LR:                     0.000
McFadden's R2:                   0.285   McFadden's Adj R2:              0.245
ML (Cox-Snell) R2:               0.395   Cragg-Uhler(Nagelkerke) R2:     0.477
McKelvey & Zavoina's R2:         0.519
Variance of y*:                  2.080   Variance of error:              1.000
Count R2:                        0.735   Adj Count R2:                   0.262
AIC:                             1.328   AIC*n:                        225.759
BIC:                          -628.512   BIC':                         -64.814
BIC used by Stata:             244.573   AIC used by Stata:            225.759
```

From the estimated results,x1 and x2 have the negative effects on y, but x3 and x4 have the positive effects on the dependent variable.

From the overall test, it suggests that all independent variables can significantly explain the dependent variable since the null hypothesis is rejected p-value = 0.00 < 0.05

From the Pseudo R-squared = 0.285, it is quite low, therefore there exists rooms for improvements. The Counted R-squared = 0.735 meaning that this model can correctly predict the outcome 73.5%.

The individual test suggests that x3, x4, cut1, and cut2 are individually significant, but x1 and x2 are not.

Thanakrit Methasate 6204640087

<u>Test Ordered Probit</u>

```
. gen y01 = y >0

. gen y12 = y>1

. qui probit y01 x*

. est store probit01

. qui probit y12 x*

. est store probit12


. suest probit01 probit12

Simultaneous results for probit01, probit12


                                                Number of obs     =        170


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

             |               Robust

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

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

probit01_y01 |

          x1 |  -.5759762   .3517281    -1.64   0.102    -1.265351    .1133982

          x2 |  -.5610958   .3561682    -1.58   0.115    -1.259173    .1369811

          x3 |   .0353658    .322031     0.11   0.913    -.5958035     .666535

          x4 |   .6493143   .1099608     5.90   0.000     .4337951    .8648336

       _cons |  -8.379872   1.588201    -5.28   0.000    -11.49269   -5.267055

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

probit12_y12 |

          x1 |  -.0859675   .2917692    -0.29   0.768    -.6578247    .4858897

          x2 |  -.4665854   .3010255    -1.55   0.121    -1.056584    .1234137

          x3 |   .6657033   .2613174     2.55   0.011     .1535307    1.177876

          x4 |   .7898094   .1308796     6.03   0.000     .5332902    1.046329

       _cons |   -12.5982   2.071075    -6.08   0.000    -16.65743   -8.538965

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

```
. test [probit01_y01]x1 = [probit12_y12]x1

 ( 1)  [probit01_y01]x1 - [probit12_y12]x1 = 0

           chi2(  1) =    1.53
         Prob > chi2 =    0.2162

. test [probit01_y01]x2 = [probit12_y12]x2

 ( 1)  [probit01_y01]x2 - [probit12_y12]x2 = 0

           chi2(  1) =    0.05
         Prob > chi2 =    0.8162

. test [probit01_y01]x3 = [probit12_y12]x3

 ( 1)  [probit01_y01]x3 - [probit12_y12]x3 = 0

           chi2(  1) =    3.00
         Prob > chi2 =    0.0835

. test [probit01_y01]x4 = [probit12_y12]x4

 ( 1)  [probit01_y01]x4 - [probit12_y12]x4 = 0

           chi2(  1) =    0.72
         Prob > chi2 =    0.3954
```

From the test, all null hypothesizes are failed to reject. So, Ordered Probit is appropriated.

From the data set "`Final_q1-2_no.dta`":

In the study of decision to subscribe mobile phone services, three choices have been studied including AIS, DTAC, and TRUE. The probit model can be stated as:

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

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)        =      33.18
                                                Prob > chi2       =     0.0000
Log likelihood = -362.36305                     Pseudo R2         =     0.0438

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.394585   .4001617     3.49   0.000     .6102824    2.178887
          x2 |  -.8054803   .3810398    -2.11   0.035    -1.552305    -.058656
          x3 |   .6878752   .1775705     3.87   0.000     .3398435    1.035907
          x4 |   .2237994   .1847999     1.21   0.226    -.1384017    .5860006
       _cons |  -.8169487   .3025723    -2.70   0.007     -1.40998   -.2239178
------------------------------------------------------------------------------

. probit y2 x*, nolog

Probit regression                               Number of obs     =        550
                                                LR chi2(4)        =      14.83
                                                Prob > chi2       =     0.0051
Log likelihood = -360.17388                     Pseudo R2         =     0.0202
```

```
------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   -.739561   .3784533    -1.95   0.051    -1.481316    .0021938
          x2 |  -.5164994   .3686219    -1.40   0.161    -1.238985    .2059862
          x3 |  -.2704721   .1752924    -1.54   0.123    -.6140389    .0730947
          x4 |   .6061985   .1902349     3.19   0.001      .233345     .979052
       _cons |   .3278563   .3006511     1.09   0.275    -.2614089    .9171216
------------------------------------------------------------------------------

. probit y3 x*, nolog

Probit regression                                 Number of obs     =        550
                                                  LR chi2(4)        =      11.92
                                                  Prob > chi2       =     0.0180
Log likelihood = -327.43565                       Pseudo R2         =     0.0179

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.5188997   .3896565    -1.33   0.183    -1.282612     .244813
          x2 |   .7601524    .383251     1.98   0.047     .0089943    1.511311
          x3 |  -.3898287   .1970154    -1.98   0.048    -.7759717   -.0036856
          x4 |  -.3178066   .1831805    -1.73   0.083    -.6768338    .0412206
       _cons |   -.471871   .3152587    -1.50   0.134    -1.089767    .1460246
------------------------------------------------------------------------------
```

To determine whether these separate Probit models are appropriated, we should be able to know that there is no any correlation among the error terms, i.e. Tetrachoric Correlation between the error term ε_i and ε_j are equal to zero. If not, then we should not apply the separate models. To clarify, the test is required.

(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*)


Iteration 0:   log likelihood = -1049.9726

. . .

Iteration 5:   log likelihood =   -929.435


Multivariate probit (MSL, # draws = 5)             Number of obs   =        550

                                                   Wald chi2(12)   =      44.62

Log likelihood =   -929.435                        Prob > chi2     =     0.0000


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

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

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

y1           |

          x1 |   1.242071   .3903115     3.18   0.001      .477074    2.007067

          x2 |  -.6719538   .3750584    -1.79   0.073    -1.407055    .0631472

          x3 |   .6696143   .1786955     3.75   0.000     .3193775    1.019851

          x4 |    .204277   .1839031     1.11   0.267    -.1561664    .5647204

       _cons |  -.7530149   .3046773    -2.47   0.013    -1.350172   -.1558583

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

y2           |

          x1 |  -.8113389   .3748761    -2.16   0.030    -1.546083   -.0765952

          x2 |  -.4461635   .3665143    -1.22   0.223    -1.164518    .2721913

          x3 |  -.2892671   .1757756    -1.65   0.100     -.633781    .0552469

          x4 |   .5873489   .1894209     3.10   0.002     .2160909     .958607

       _cons |   .3688367   .2992854     1.23   0.218    -.2177519    .9554253

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

y3           |

          x1 |  -.4616851   .3829211    -1.21   0.228    -1.212197    .2888264

          x2 |   .5866839   .3783605     1.55   0.121    -.1548892    1.328257
```

```
          x3 |  -.2848337   .1854174    -1.54   0.124    -.6482451    .0785778
          x4 |   -.275354   .1869964    -1.47   0.141    -.6418603    .0911522
       _cons |   -.418268    .317282    -1.32   0.187    -1.040129    .2035933
-------------+----------------------------------------------------------------
    /atrho21 |  -.5774284   .0738682    -7.82   0.000    -.7222074   -.4326494
-------------+----------------------------------------------------------------
    /atrho31 |  -.4353952   .0750685    -5.80   0.000    -.5825268   -.2882636
-------------+----------------------------------------------------------------
    /atrho32 |  -.3384142    .071239    -4.75   0.000      -.47804   -.1987883
-------------+----------------------------------------------------------------
       rho21 |  -.5207938   .0538332    -9.67   0.000    -.6182747    -.407533
-------------+----------------------------------------------------------------
       rho31 |  -.4098203   .0624606    -6.56   0.000    -.5244996   -.2805358
-------------+----------------------------------------------------------------
       rho32 |  -.3260609   .0636652    -5.12   0.000    -.4446726   -.1962105
------------------------------------------------------------------------------
Likelihood ratio test of  rho21 = rho31 = rho32 = 0:
             chi2(3) =  241.075   Prob > chi2 = 0.0000
```

From the LR-test, the null hypothesis that is all ρ(rho) are equal to zero, i.e. there is no correlations among the error terms, is rejected. Therefore, MV Probit us more appropriated than the separate ones. The main difference between the separate Probit and MV Probit is that MV Probit is like the model is run as a system (run together) because there exists Tetrachoric Correlation between the error term ε_i and ε_j

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

According to the following model,

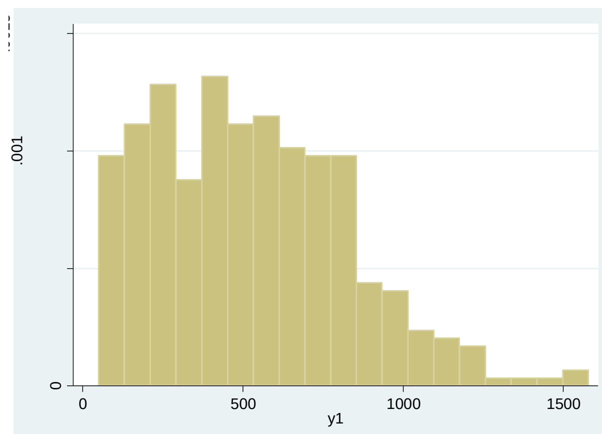
$$y_{ki} = \beta_{k0} + \beta_{k1} x_i + u_{ki} \tag{3}$$

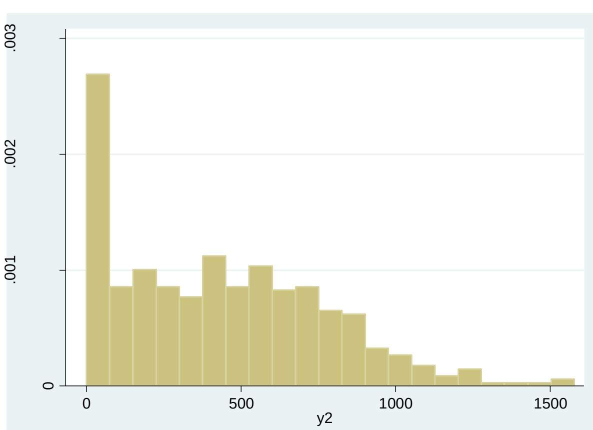
where: y_{ki} is Dependent variable, k=1, 2, 3.
x_i is Independent variable
u_i is Stochastic disturbance term

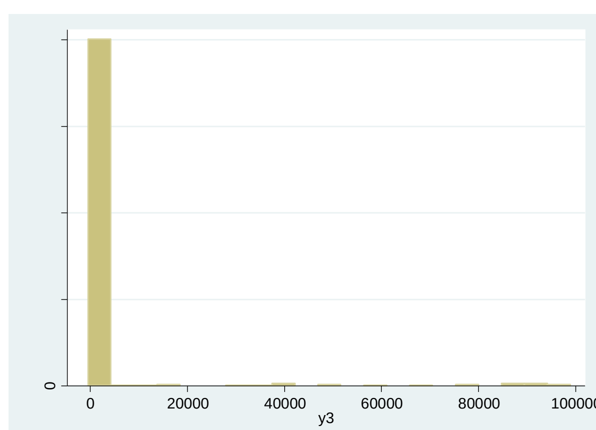
(a) Plot histogram of y_{1i}, y_{2i}, y_{3i}, compute descriptive statistics of these three variables. Determine limitations of these three dependent variables. (5 points)

```
. sum y*

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          y1 |        368    525.8013    301.8327   50.21759    1578.51
          y2 |        450     430.993    339.0019          0    1578.51
          y3 |        450    3343.313    14544.82  -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)
```

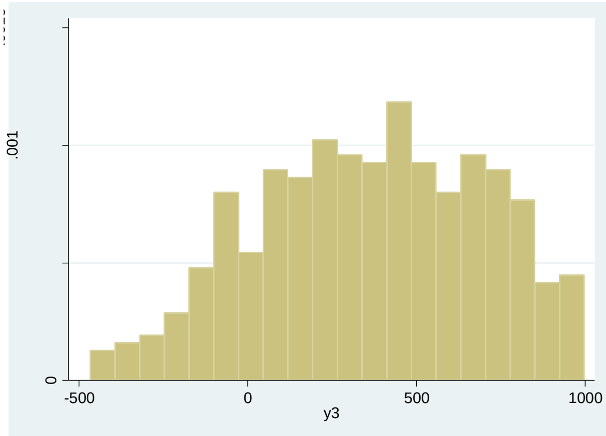






```
. histogram y3 if y3 < 4000
(bin=20, start=-466.00415, width=73.209753)
```

According to the summarized statistics table and histograms, it might be the case that there exists the truncated problem for y1 (the value of y below 50 cannot be observed). For y2, we can clearly observe that the dependent variable can be fully observed but the value equal to zero has significantly high density. Therefore, this can be the case that y2 has a censored data problem (censored at 0). For y3, it is highly likely to be the case of 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)

Run y1 with OLS

```
. reg y1 x

      Source |       SS           df       MS      Number of obs   =       368
-------------+----------------------------------   F(1, 366)       =    137.65
       Model |   9137792.8         1   9137792.8   Prob > F        =    0.0000
    Residual |  24297002.7       366  66385.2534   R-squared       =    0.2733
-------------+----------------------------------   Adj R-squared   =    0.2713
       Total |  33434795.5       367  91102.9851   Root MSE        =    257.65

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |    165.724    14.1254    11.73   0.000     137.9469    193.5011
       _cons |  -9.272247    47.5433    -0.20   0.845    -102.7646    84.22007
-----------------------------------------------------------------------------

. est store y1_ols
```

<u>Run y1 with Truncated Regression</u>

```
. truncreg y1 x, ll(50)

(note: 0 obs. truncated)

Fitting full model:

. . .

Iteration 4:   log likelihood = -2529.8335

Truncated regression

Limit:   lower =         50                    Number of obs     =        368

         upper =       +inf                    Wald chi2(1)      =     104.94

Log likelihood = -2529.8335                    Prob > chi2       =     0.0000


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

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

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

           x |   230.7558   22.52558    10.24   0.000     186.6065    274.9052

       _cons |  -297.5609   85.92795    -3.46   0.001    -465.9766   -129.1453

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

      /sigma |   305.8521   16.81074    18.19   0.000     272.9036    338.8005

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

. est store y1_trunc
```

<u>Test whether Truncated Regression is appropriated</u>

```
. lrtest y1_ols y1_trunc, force


Likelihood-ratio test                                 LR chi2(1)  =     68.66

(Assumption: y1_ols nested in y1_trunc)               Prob > chi2 =    0.0000
```

Since the LR-test rejects the null hypothesis, the Truncated Regression is appropriated for y1. The major problem is that some observations are not available to observe; however, those observation exists. Therefore, the sample is selected based on the value of the dependent variable (y), i.e. Sample Selection Bias. If the problem is ignored and continue to use the traditional OLS, the estimated results is 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)

Run y2 with OLS

```
. reg y2 x

      Source |       SS           df       MS      Number of obs   =       450
-------------+----------------------------------   F(1, 448)       =    264.88
       Model |  19172494.6         1  19172494.6   Prob > F        =    0.0000
    Residual |  32427611.1       448  72383.0604   R-squared       =    0.3716
-------------+----------------------------------   Adj R-squared   =    0.3702
       Total |  51600105.7       449  114922.284   Root MSE        =    269.04

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   199.1064   12.23388    16.27   0.000     175.0635    223.1493
       _cons |  -169.2913   39.00344    -4.34   0.000    -245.9437   -92.63885
------------------------------------------------------------------------------

. est store y2_ols
```

Run y2 with Tobit Regression

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

Tobit regression                                Number of obs     =        450
                                                LR chi2(1)        =     220.85
                                                Prob > chi2       =     0.0000
Log likelihood = -2793.8497                     Pseudo R2         =     0.0380

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   236.3091   14.58958    16.20   0.000     207.6367    264.9814
       _cons |  -311.5844   47.33943    -6.58   0.000    -404.6188   -218.5501
-------------+----------------------------------------------------------------
      /sigma |   300.5597   11.10914                      278.7274    322.3921
```

```
------------------------------------------------------------------------------
            66  left-censored observations at y2 <= 0
           384     uncensored observations
             0 right-censored observations

. est store y2_tobit
```

Test whether Tobit Regression is appropriated

```
. lrtest y2_ols y2_tobit, force

Likelihood-ratio test                                 LR chi2(1)  =    722.72
(Assumption: y2_ols nested in y2_tobit)               Prob > chi2 =    0.0000
```

Since the LR-test rejects the null hypothesis, the Tobit Regression is more appropriated for y2. In this case, all observations can be observed. However, the major problem is that some of the dependent variable (y) are censored at 0. That is the real values might be lower than 0, but according to what is observable, they are all 0. If the problem is ignored and OLS is applied, the estimated result is biased because of the censored data.

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

Run y3 with OLS

```
. reg y3 x

      Source |       SS           df       MS      Number of obs   =       450
-------------+----------------------------------   F(1, 448)       =     36.36
       Model |  7.1301e+09         1  7.1301e+09   Prob > F        =    0.0000
    Residual |  8.7857e+10       448   196108662   R-squared       =    0.0751
-------------+----------------------------------   Adj R-squared   =    0.0730
       Total |  9.4987e+10       449   211551878   Root MSE        =     14004

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
           x |   3839.672   636.7872     6.03   0.000     2588.211    5091.133
       _cons |  -8232.881   2030.172    -4.06   0.000    -12222.72   -4243.038
------------------------------------------------------------------------------

. est store y3_ols
```

<u>Run y3 with Truncated Regression</u>

```
. tobit y3 x, ul(4000)

Tobit regression                                Number of obs     =        450
                                                LR chi2(1)        =     123.88
                                                Prob > chi2       =     0.0000
Log likelihood = -3495.9514                     Pseudo R2         =     0.0174

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   429.1709    36.0279    11.91   0.000     358.3667    499.9752
       _cons |  -733.6497   114.6576    -6.40   0.000    -958.9819   -508.3176
-------------+----------------------------------------------------------------
      /sigma |   788.7746   27.59472                      734.5438    843.0055
------------------------------------------------------------------------------

            0  left-censored observations
          427     uncensored observations
           23 right-censored observations at y3 >= 4000

. est store y3_tobit
```

<u>Test whether Tobit Regression is appropriated</u>

```
. lrtest y3_ols y3_tobit, force

Likelihood-ratio test                                 LR chi2(1)  =   2875.52
(Assumption: y3_ols nested in y3_tobit)               Prob > chi2 =    0.0000
```

Since the LR-test rejects the null hypothesis, the Tobit Regression is more appropriated for y3. In this case all observations can be fully observed, but there exists the outlier problem on the dependent variable (y). Of course, if OLS is used, the estimated result is biased because of those extreme values. So, they should be removed by setting up the upper limit such that the outliers are eliminated. This is likely to the censored data problem in which the traditional OLS leads to a biasness. Therefore, Tobit should be applied instead.

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

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

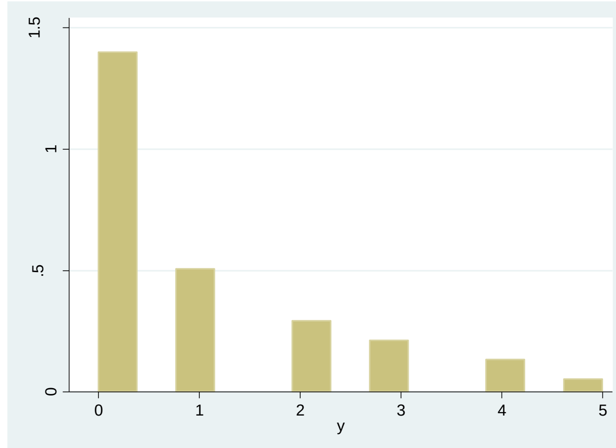
where: I_i is index variables.
y_i is counted number 0, 1, 2,...
x_{ki} is independent variable k.
$f(.)$ is either Poisson or Negative Binomial probability distribution function.
u_i is disturbance term.

(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)

Run y with OLS

```
. reg y x*

      Source |       SS           df       MS      Number of obs   =       195
-------------+----------------------------------   F(4, 190)       =      5.61
       Model |  36.4072626         4  9.10181564   Prob > F        =    0.0003
    Residual |  308.464532       190  1.62349754   R-squared       =    0.1056
-------------+----------------------------------   Adj R-squared   =    0.0867
       Total |  344.871795       194  1.77768966   Root MSE        =    1.2742

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0847709   .0439189     1.93   0.055    -.0018604    .1714022
          x2 |   .1428018    .046958     3.04   0.003     .0501758    .2354279
          x4 |  -.0365707   .0475964    -0.77   0.443     -.130456    .0573146
          x3 |   .1901753   .0668986     2.84   0.005     .0582159    .3221347
       _cons |   .9012497   .1080894     8.34   0.000     .6880403    1.114459
------------------------------------------------------------------------------
```

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



There exists a limitation. It is obvious that the dependent variable is not normally distributed. It looks like Poisson distribution instead. Moreover, it seems to have the excess zero problem.

(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)

<u>Poisson Regression</u>

```
. poisson y x*

Iteration 0:   log likelihood = -270.78875
Iteration 1:   log likelihood = -270.78874

Poisson regression                              Number of obs     =        195
                                                LR chi2(4)        =      37.99
                                                Prob > chi2       =     0.0000
Log likelihood = -270.78874                     Pseudo R2         =     0.0655

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0876226   .0344264     2.55   0.011     .0201482     .155097
          x2 |   .1470317    .036948     3.98   0.000     .0746151    .2194484
          x4 |  -.0378213   .0376442    -1.00   0.315    -.1116026      .03596
          x3 |   .2078186   .0552623     3.76   0.000     .0995065    .3161308
       _cons |  -.2000234   .0955572    -2.09   0.036    -.3873121   -.0127347
```

```
. poisson y x*, ir

Iteration 0:   log likelihood = -270.78875

Iteration 1:   log likelihood = -270.78874


Poisson regression                              Number of obs     =        195

                                                LR chi2(4)        =      37.99

                                                Prob > chi2       =     0.0000

Log likelihood = -270.78874                     Pseudo R2         =     0.0655


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

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

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

          x1 |   1.091576    .037579     2.55   0.011     1.020353    1.167771

          x2 |   1.158391   .0428002     3.98   0.000     1.077469     1.24539

          x4 |    .962885   .0362471    -1.00   0.315     .8943996    1.036614

          x3 |    1.23099   .0680273     3.76   0.000     1.104626     1.37181

       _cons |   .8187116   .0782338    -2.09   0.036     .6788792    .9873461

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

. mfx


Marginal effects after poisson

      y  = Predicted number of events (predict)

         =  .87962845

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

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

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

      x1 |   .0770753      .02985    2.58   0.010   .018573  .135577  -.269304

      x2 |   .1293333      .03141    4.12   0.000   .067781  .190886   .865611

      x4 |  -.0332687      .03303   -1.01   0.314  -.098006  .031469  -.778541

      x3 |   .1828032      .04685    3.90   0.000   .090985  .274622  -.295223

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

```
. estat gof

         Deviance goodness-of-fit =  313.6265
         Prob > chi2(190)         =    0.0000


         Pearson goodness-of-fit  =  332.5726
         Prob > chi2(190)         =    0.0000
```

From the estimated result, it suggests that each independent variables has a positive effect on the dependent variable (Marginal effect > 0 and IRR > 1) except x4 that negatively affects the dependent variable (Marginal effect < 0 and IRR < 1).

The overall test shows that all independent variables can jointly explain the dependent variable.

The pseudo R-squared = 0.0655 is quite low. Therefore, it can be improved.

The individual test reveals that each of the independent variables is significant except x4.

However, from the goodness of fit test, the null hypothesis is rejected (P-value < 0.05). It means that Poisson Regression should not be applied.

(c) Estimate models for Y_i assuming that the probability functions follow Negative Binomial probability distribution. Determine whether Negative Binomial regression model is appropriated in this case. Interpret your estimated result (sign and meaning (in term of incidence-rate ratios), overall test, individual test, pseudo R^2, marginal

Negative Binomial Probability Distribution

```
. nbreg y x*, nolog

Negative binomial regression                    Number of obs     =        195
                                                LR chi2(4)        =      20.53
Dispersion     = mean                           Prob > chi2       =     0.0004
Log likelihood = -256.37137                     Pseudo R2         =     0.0385

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .1157622   .0526251     2.20   0.028     .0126188    .2189055
          x2 |   .1653749   .0528556     3.13   0.002     .0617799    .2689699
          x4 |  -.0395441    .050271    -0.79   0.432    -.1380735    .0589852
          x3 |   .1980496   .0707578     2.80   0.005     .0593669    .3367323
```

```
       _cons |  -.2205923   .1249006    -1.77   0.077    -.4653931    .0242084
-------------+----------------------------------------------------------------
    /lnalpha |  -.1768216   .2946078                     -.7542424    .4005991
-------------+----------------------------------------------------------------
       alpha |   .8379292   .2468605                      .4703668    1.492719
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =   28.83 Prob>=chibar2 = 0.000

. nbreg y x*, ir nolog

Negative binomial regression                    Number of obs     =        195
                                                LR chi2(4)        =      20.53
Dispersion     = mean                           Prob > chi2       =     0.0004
Log likelihood = -256.37137                     Pseudo R2         =     0.0385

------------------------------------------------------------------------------
           y |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.122729   .0590837     2.20   0.028     1.012699    1.244714
          x2 |   1.179835   .0623609     3.13   0.002     1.063728    1.308616
          x4 |   .9612275   .0483219    -0.79   0.432     .8710347     1.06076
          x3 |   1.219023   .0862553     2.80   0.005     1.061164    1.400364
       _cons |   .8020436   .1001758    -1.77   0.077     .6278882    1.024504
-------------+----------------------------------------------------------------
    /lnalpha |  -.1768216   .2946078                     -.7542424    .4005991
-------------+----------------------------------------------------------------
       alpha |   .8379292   .2468605                      .4703668    1.492719
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =   28.83 Prob>=chibar2 = 0.000

. mfx
Marginal effects after nbreg
      y  = Predicted number of events (predict)
         =  .87258142
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
x1	.1010119	.04575	2.21	0.027	.011349	.190675	-.269304
x2	.144303	.04587	3.15	0.002	.054393	.234213	.865611
x4	-.0345055	.04386	-0.79	0.431	-.120469	.051458	-.778541
x3	.1728144	.06142	2.81	0.005	.05243	.293199	-.295223

The overdispersion test rejects the null hypothesis that is α = 0 meaning that the Negative Binomial Model is more appropriated compared with the Poisson Model.

From the estimated result, it suggests that each independent variables has a positive effect on the dependent variable (Marginal effect > 0 and IRR > 1) except x4 that negatively affects the dependent variable (Marginal effect < 0 and IRR < 1).

The overall test shows that all independent variables can jointly explain the dependent variable since p-value = 0.0004 < 0.05.

The pseudo R-squared =0.0385 is quite low. Therefore, it can be improved.

The individual test reveals that each of the independent variables is significant except x4.

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

Zero Inflated Poisson

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

Zero-inflated Poisson regression                 Number of obs    =        195
                                                 Nonzero obs      =         90
                                                 Zero obs         =        105

Inflation model = logit                          LR chi2(3)       =      15.67
Log likelihood  = -255.2978                      Prob > chi2      =     0.0013

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

```
          x1 |   .1029852   .0431638     2.39   0.017     .0183857    .1875848
          x2 |   .1310613   .0411813     3.18   0.001     .0503474    .2117753
          x3 |   .1464086   .0561832     2.61   0.009     .0362916    .2565256
       _cons |    .289415   .1204087     2.40   0.016     .0534183    .5254117
-------------+----------------------------------------------------------------
inflate      |
          x4 |    .041464   .1015959     0.41   0.683    -.1576603    .2405883
       _cons |  -.5217272   .2645918    -1.97   0.049    -1.040318   -.0031368
------------------------------------------------------------------------------
Vuong test of zip vs. standard Poisson:            z =     2.46  Pr>z = 0.0070
```

```
. mfx

Marginal effects after zip
      y  = Predicted number of events (predict)
         =  .88504026
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .0911461      .03794    2.40   0.016   .016781  .165511  -.269304
      x2 |   .1159945      .03564    3.25   0.001   .046136  .185853   .865611
      x3 |   .1295775      .04864    2.66   0.008   .034248  .224907  -.295223
      x4 |  -.0133921      .03302   -0.41   0.685  -.078107  .051323  -.778541
------------------------------------------------------------------------------
```

The Vuong test rejects the null hypothesis (p-value = 0.007 < 0.05) meaning that the Zero-Inflated Poisson Model is more appropriated compared with the Poisson Model.

From the estimated result, it suggests that each independent variables has a positive effect on the dependent variable (Marginal effect > 0) except x4 that negatively affects the dependent variable (Marginal effect < 0).

The overall test shows that all independent variables can jointly explain the dependent variable.

The individual test reveals that each of the independent variables is significant except x4.

To determine which model is the most appropriated, *Firstly, it cannot be the Poisson Regression since the Goodness of Fit Test (GOF Test) rejects the null hypothesis (implying that the Poisson is inappropriate).*

Next, the Negative Binomial Model seems to be good because the Overdispersion also confirms that the Poisson Model is inappropriate; however, there exists an insignificant variable in this model.

Lastly, look at the Zero-Inflated Poisson Model, the Vuong test shows that this model is appropriated; however, x4 is still insignificant as same as in the Negative Binomial case.

In conclusion, the Zero-Inflated Poisson Model is the most appropriated model because of the reason provided above; plus, from the histogram at the beginning, it illustrates that the dependent variable follows zero inflated distribution. Therefore, Zero Inflated Poisson Regression model should be applied in this case.

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)

Test whether y and x are stationary by Augmented Dickey-Fuller Test

```
. 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)            -15.058            -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. |  -.9760353   .0648188   -15.06   0.000     -1.10339   -.8486808
         LD. |  -.0594166   .0449897    -1.32   0.187    -.1478114    .0289782
      _trend |   .0016656   .0021092     0.79   0.430    -.0024786    .0058098
       _cons |   .8305426   .6112434     1.36   0.175    -.3704148      2.0315
------------------------------------------------------------------------------


. 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)            -15.212            -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. |  -.9874688   .0649129   -15.21   0.000    -1.115008   -.8599295
         LD. |  -.0501877   .0449965    -1.12   0.265    -.1385958    .0382203
      _trend |   .0026561   .0030122     0.88   0.378    -.0032622    .0085743
       _cons |   .4140897   .8693345     0.48   0.634    -1.293959    2.122139
------------------------------------------------------------------------------
```

According to **Augmented Dickey-Fuller Test, it suggests that both y and x are stationary since the null hypothesis that is $\beta = 1$ is rejected (p-value = 0.00< 0.05).**

Therefore, y and x are stationary at level, i.e. y and x are integrated series of order 0.

The process of the test is that, from,

$$\Delta y_t = \alpha_0 + \alpha_1 t + \gamma y_{t-1} + \sum \beta_i \, \Delta y_{t-1} + \varepsilon_t \quad \text{where } \gamma = \beta - 1$$

We test whether H_0: $\gamma=0$ ($\beta=1$), if it is rejected, then there is no Unit Root i.e. Stationary.

However, if the null hypothesis is failed to reject, the significance of testing equation comes to play a role, the null hypothesis cannot be rejected might be because we have an excess variable in the model. So, we have to look at other coefficients (α_1 and α_0 which are trend and constant respectively), if it is insignificant, drop it and test again.

Moreover, Stationary is very important for time series because if it is Nonstationary, it will lead to a Spurious Problem (False Relationship)

(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)

According to (a), parameter "d" = 0 since they are integrated order 0.

```
. qui arima y, arima(1,0,1)
. est store arima101
. qui arima y, arima(1,0,2)
. est store arima102
. qui arima y, arima(1,0,3)
. est store arima103
```

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

. est store arima104

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

. est store arima201

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

. est store arima202

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

. est store arima203

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

. est store arima204

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

. est store arima301

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

. est store arima302

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

. est store arima303

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

. est store arima304

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

. est store arima401

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

. est store arima402

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

. est store arima403

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

. 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 |  1.2822737***   1.2852717***   1.2830207***   1.2834182***
-------------+----------------------------------------------------------------
```

```
ARMA         |
          ar |
         L1. | -.85017938***    .81704179***   -.85050658***   -.80668976***
             |
          ma |
         L1. |  .80583931***   -.85594029***    .81624866***    .77253015***
         L2. |                  .07309734       .02708951       .02664907
         L3. |                                  .01783289       .04904404
         L4. |                                                  .04879605
-------------+----------------------------------------------------------------
sigma        |
       _cons |  6.7312771***    6.7316582***    6.7295342***    6.7221303***
-------------+----------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -1662.8601      -1662.8819      -1662.7286      -1662.1835
        chi2 |    49.8428       22.748584       49.803311       40.869619
         aic |  3333.7202       3335.7639       3337.4571       3338.3669
         bic |  3350.5786       3356.8369       3362.7448       3367.8692
------------------------------------------------------------------------------
                                            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 |  1.2852742***    1.2849413***    1.2850517***    1.2848488***
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  .72718848***    .03434887      -.0071867       -.01900249
         L2. |  .07767849       .72045376***    .72060757***    .67835837**
```

```
             |
          ma |
         L1. |  -.7684233***   -.05809605       -.0297673      -.01700001
         L2. |                 -.65119293**     -.6531575**    -.62523011*
         L3. |                                  .02414018       .02280762
         L4. |                                                  .02572472
-------------+---------------------------------------------------------------
sigma        |
       _cons |  6.7307631***    6.7199691***    6.7183825***    6.7169304***
-------------+---------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -1662.8217      -1662.0242      -1661.9107      -1661.7824
        chi2 |  19.588959       20.181824       21.261155       19.918424
         aic |  3335.6434       3336.0484       3337.8214       3339.5648
         bic |  3356.7165       3361.3361       3367.3236       3373.2817
-----------------------------------------------------------------------------
                                         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 |  1.2829597***    1.2851221***    1.3010793***    1.3010576***
-------------+---------------------------------------------------------------
ARMA         |
          ar |
         L1. | -.84116634**    -.03809733        1.074127***    1.0736076***
         L2. |   .0303744       .72200193***    .63826377**     .63932718
         L3. |  .01991732       .02474792      -.82783544***   -.82840766***
             |
          ma |
```

```
         L1. |  .80847988**     .00208541       -1.133384      -1.1333014
         L2. |                -.65521465**    -.53766529      -.53847758
         L3. |                                 .79668804       .79801641
         L4. |                                                -.00064372
-------------+----------------------------------------------------------------
sigma        |
       _cons |  6.7293725***    6.7185376***     6.635098       6.6350719
-------------+----------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -1662.7154       -1661.919      -1657.7457      -1657.7457
        chi2 |   50.95581       21.823841       50151.443       35434.925
         aic |  3337.4309       3337.8381       3329.4915       3331.4913
         bic |  3362.7185       3367.3403       3358.9937       3365.2082
------------------------------------------------------------------------------
                                           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 |   1.283736***    1.2851361***    1.3010538***    1.2857328***
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. | -.75902834**    -.05431084       1.072809***     .08134479
         L2. |  .03140494       .64109765*      .64018188      -.23331753
         L3. |  .05089759       .02507798      -.82789454***   -.04574194
         L4. |  .05080793       .02801024      -.0006621        .69819982***
             |
          ma |
         L1. |  .72522412**     .01787627      -1.132502      -.11980729
```

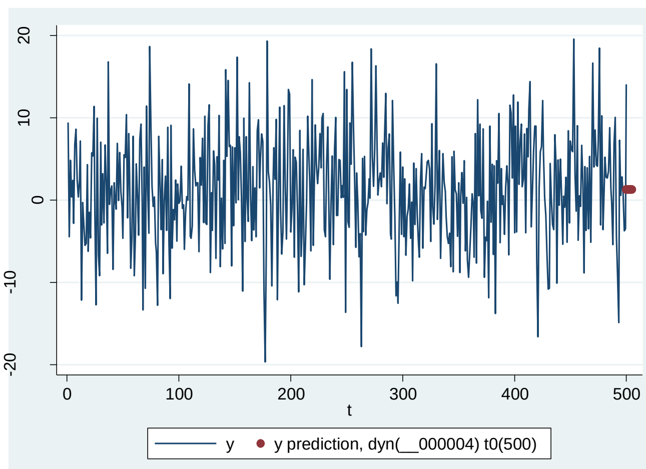
```
         L2. |                -.58863871      -.53937619       .31872652
         L3. |                                 .79757316       .04504616
         L4. |                                                -.64787648**
-------------+--------------------------------------------------------------
sigma        |
       _cons |  6.7218422***     6.716615***     6.635093      6.6766313***
-------------+--------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -1662.1618      -1661.7838      -1657.7457      -1659.0454
        chi2 |  33.841275       19.156218       35372.689       1213.8068
         aic |  3338.3237       3339.5677       3331.4913       3338.0908
         bic |  3367.8259       3373.2845       3365.2082       3380.2368
----------------------------------------------------------------------------
                                           legend: * p<.1; ** p<.05; *** p<.01
```

From the result, the best ARIMA model (ARIMA(p,d,q)) for y is ARIMA(1,0,1) since it gives the lowest value of BIC (SBIC) among any other combinations of (p,0,q), $1 \leq p,q \leq$

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


(setting optimization to BHHH)
. . .
Iteration 7:   log likelihood = -1662.8601
ARIMA regression


Sample:  1 - 500                                Number of obs     =        500
                                                Wald chi2(2)      =      49.84
Log likelihood =  -1662.86                      Prob > chi2       =     0.0000


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

```
       _cons |   1.282274   .2957515     4.34   0.000     .7026113    1.861936
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.8501794   .1677986    -5.07   0.000    -1.179059   -.5213001
             |
          ma |
         L1. |   .8058393    .189501     4.25   0.000     .4344241    1.177255
-------------+----------------------------------------------------------------
      /sigma |   6.731277   .2150805    31.30   0.000     6.309727    7.152827
------------------------------------------------------------------------------
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
(499 missing values generated)

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



From the following GARCH model:

Mean Equation: $$y_t = \alpha + \beta x_t + \varepsilon_t \qquad (5)$$

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

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)       =  57380.10
       Model |  22617.2475         1  22617.2475   Prob > F        =    0.0000
    Residual |  196.294343       498  .394165348   R-squared       =    0.9914
-------------+----------------------------------   Adj R-squared   =    0.9914
       Total |  22813.5419       499  45.7185208   Root MSE        =    .62783


------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .6976501   .0029124   239.54   0.000     .6919279    .7033723
       _cons |   .5173437    .028259    18.31   0.000     .4618221    .5728653
------------------------------------------------------------------------------

. estat archlm
LM test for autoregressive conditional heteroskedasticity (ARCH)
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |         38.724               1                   0.0000
---------------------------------------------------------------------------
         H0: no ARCH effects      vs.  H1: ARCH(p) disturbance
```

From the test, there exists a significant ARCH effect since p-value of the test =0.00 < 0.05 (The null hypothesis of the test is that H_0: $\alpha_1 = \alpha_2 = \alpha_q = 0$). This test uses LM test since uses only 1 restricted model which is the OLS one.

(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)

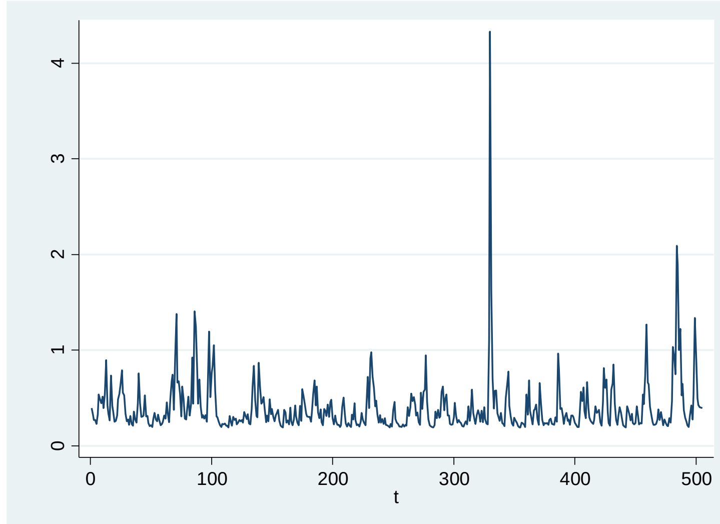
```
. qui arch y x, garch(1/1) arch(1/1)

. est store garch11

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

. est store garch12

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

. est store garch21

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

. est store garch22

. estimates table garch11 garch12 garch21 garch22, star(0.1 0.05 0.01) stat(aic bic
ll)


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

    Variable |    garch11         garch12         garch21         garch22

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

y            |

           x |  .69776959***     .6977543***    .69775887***    .69763896***

       _cons |  .51862389***    .51858003***    .51860028***    .51612708***

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

ARCH         |

        arch |

         L1. |  .36335397***    .36174929***    .36211559***    .37334452***

         L2. |                  .01489233                       .30932813**

             |

       garch |

         L1. |  .31446618***    .28733881       .32471553*     -.39406428

         L2. |                                 -.00925651       .10779269

             |

       _cons |  .12902835***    .13455126*      .12912145***    .24581943***

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

```
Statistics   |
         aic |  897.80341       899.79061       899.79388       900.15062
         bic |  918.87645       925.07826       925.08153       929.65288
          ll |  -443.9017       -443.8953      -443.89694      -443.07531
------------------------------------------------------------------------------
                                           legend: * p<.1; ** p<.05; *** p<.01
```

According to BIC(SBIC), the most appropriated lags order in this case is GARCH(1,1)

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

ARCH family regression

Sample: 1 - 500                                 Number of obs     =        500
Distribution: Gaussian                          Wald chi2(1)      =   72809.92
Log likelihood = -443.9017                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
           x |   .6977696   .0025859   269.83   0.000     .6927013    .7028379
       _cons |   .5186239   .0239338    21.67   0.000     .4717145    .5655333
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |    .363354   .0747846     4.86   0.000     .2167788    .5099291
             |
       garch |
         L1. |   .3144662    .120527     2.61   0.009     .0782376    .5506948
             |
       _cons |   .1290283   .0366217     3.52   0.000     .0572511    .2008056
------------------------------------------------------------------------------
```

```
. 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)

The differences among ARCH, GARCH, and EGARCH are that

Autoregressive Conditional Heteroskedasticity, or ARCH, is a method that explicitly models the change in variance over time in a time series.

Generalized Autoregressive Conditional Heteroskedasticity, or GARCH, is an extension of the ARCH model that incorporates a moving average component together with the autoregressive component.

While, the Exponential GARCH (EGARCH) Model is a subset of GARCH model in which the conditional variance process are done in exponential form.

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

. est store arch1

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

. est store archgarch

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

. est store archgarchegarch


. estimates table arch1 archgarch archgarchegarch , star(0.1 0.05 0.01) stat(bic)


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

    Variable |     arch1          archgarch      archgarcheg~h

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

y            |

           x |  .69766735***    .69776959***    .69793118***
```

```
       _cons |  .51465166***    .51862389***    .51184646***
-------------+------------------------------------------------
ARCH         |
        arch |
         L1. |  .38285474***    .36335397***    .67472926***
             |
       garch |
         L1. |                  .31446618***   -.00592512
             |
      egarch |
         L1. |                                  .36128496***
             |
       _cons |   .2414657***    .12902835***   -.92456597***
-------------+------------------------------------------------
Statistics   |
         bic |  922.68979       918.87645       932.02784
--------------------------------------------------------------
                           legend: * p<.1; ** p<.05; *** p<.01

.
```

Based on BIC, GARCH(1,1) is the most appropriated one since it provides the lowest BIC.

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

The following VARs models:

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

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

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

```
. varsoc y x

  Selection-order criteria

  Sample:  5 - 500                              Number of obs      =       496

 +---------------------------------------------------------------------------+
 |lag |    LL       LR       df     p      FPE       AIC      HQIC      SBIC    |
 |----+----------------------------------------------------------------------|
 |  0 | -1363.83                         .845032   5.50737   5.51403   5.52434  |
 |  1 | -1317.45  92.757*    4  0.000  .712297*  5.33649*  5.35647*  5.38738* |
 |  2 | -1316.57  1.7619     4  0.779  .721313   5.34907   5.38236   5.43388  |
 |  3 | -1315.39  2.3605     4  0.670  .729562   5.36044   5.40705   5.47917  |
 |  4 | -1313.09  4.5974     4  0.331  .734588    5.3673   5.42722   5.51996  |
 +---------------------------------------------------------------------------+

  Endogenous:  y x
  Exogenous:  _cons
```

Based on SBIC, the most appropriated lag order is 1 since it gives the lowest SBIC.

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

```
Vector autoregression

Sample:  2 - 500                               Number of obs     =        499
Log likelihood =  -1327.573                    AIC               =   5.344982
FPE            =   .7183688                    HQIC              =   5.364859
Det(Sigma_ml)  =   .7012993                    SBIC              =   5.395634

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
y                     3     .990399   0.0132   6.670976   0.0356
x                     3     .997289   0.0899   49.27269   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
           y |
         L1. |   .1292078   .0501774     2.58   0.010     .0308619    .2275536
             |
           x |
         L1. |   .0483716   .0478178     1.01   0.312    -.0453495    .1420928
             |
       _cons |   .3388104   .0542276     6.25   0.000     .2325263    .4450946
-------------+----------------------------------------------------------------
x            |
           y |
         L1. |   .3475582   .0505264     6.88   0.000     .2485282    .4465882
             |
           x |
         L1. |   .2127216   .0481504     4.42   0.000     .1183485    .3070947
             |
       _cons |   .1226802   .0546048     2.25   0.025     .0156567    .2297037
```

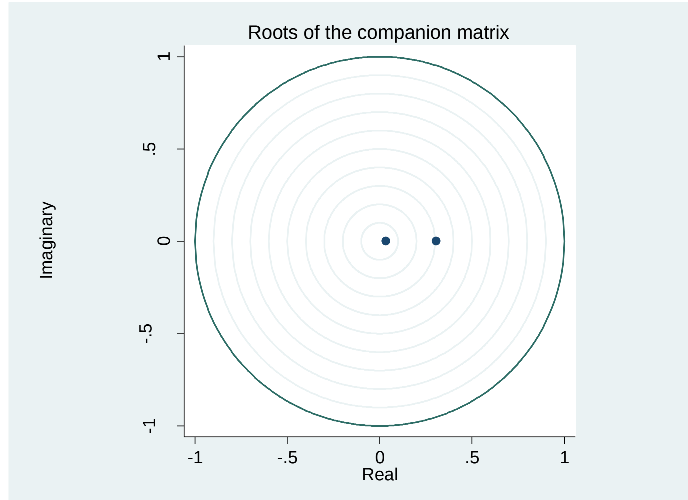
```
. varstable, graph
```

Eigenvalue stability condition

Eigenvalue	Modulus
.3071836	.307184
.03474578	.034746

All the eigenvalues lie inside the unit circle.

VAR satisfies stability condition.




```
. vargranger
```

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
y	x	1.0233	1	0.312
y	ALL	1.0233	1	**0.312**
x	y	47.317	1	0.000
x	ALL	47.317	1	**0.000**

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

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

The stability is required because the this is the dynamic system. The effect of this period will affect the next period, i.e. it should be convergent. If this assumption is violated, the effect will diverge (and not die down).

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

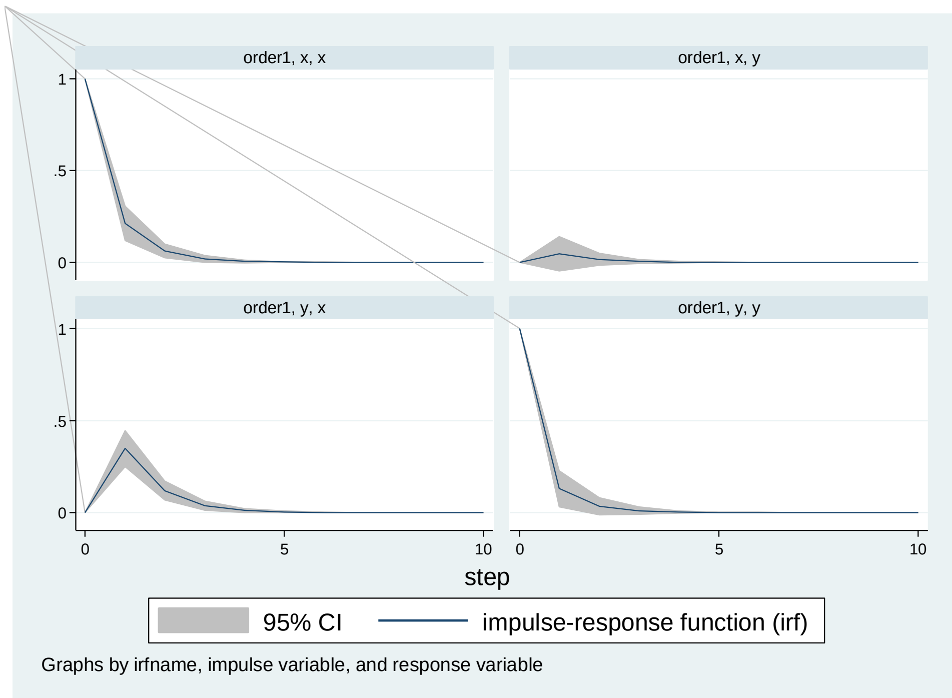
```
. irf create order1, o(y x) step(10) set(irf001)
(file irf001.irf created)
(file irf001.irf now active)
(file irf001.irf updated)

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

Results from order1

step	(1) irf	(1) Lower	(1) Upper	(2) irf	(2) Lower	(2) Upper	(3) irf	(3) Lower	(3) Upper	(4) irf	(4) Lower	(4) Upper
0	1	1	1	0	0	0	0	0	0	1	1	1
1	.129208	.030862	.227554	.347558	.248528	.446588	.048372	-.04535	.142093	.212722	.118349	.307095
2	.033507	-.014372	.081385	.11884	.067187	.170494	.01654	-.015834	.048914	.062062	.024204	.099921
3	.010078	-.008978	.029133	.036925	.012175	.061676	.005139	-.006431	.016709	.018951	.001157	.036744
4	.003088	-.003914	.01009	.011357	.000172	.022543	.001581	-.002539	.0057	.005817	-.001731	.013366
5	.000948	-.00155	.003447	.003489	-.001173	.008152	.000486	-.000959	.00193	.001787	-.001192	.004766
6	.000291	-.000584	.001167	.001072	-.000753	.002896	.000149	-.00035	.000649	.000549	-.000572	.00167
7	.000089	-.000213	.000392	.000329	-.000354	.001013	.000046	-.000125	.000216	.000169	-.000239	.000577
8	.000027	-.000076	.000131	.000101	-.000147	.000349	.000014	-.000044	.000072	.000052	-.000093	.000197
9	8.4e-06	-.000026	.000043	.000031	-.000057	.000119	4.3e-06	-.000015	.000024	.000016	-.000035	.000067
10	2.6e-06	-9.1e-06	.000014	9.5e-06	-.000021	.00004	1.3e-06	-5.1e-06	7.8e-06	4.9e-06	-.000013	.000022

```
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

step	(1) oirf	(1) Lower	(1) Upper	(2) oirf	(2) Lower	(2) Upper	(3) oirf	(3) Lower	(3) Upper	(4) oirf	(4) Lower	(4) Upper
0	.987418	.926157	1.04868	-.51896	-.60004	-.437881	0	0	0	.848107	.795489	.900725
1	.102479	.015961	.188998	.232791	.143376	.322206	.041024	-.038502	.120551	.180411	.099593	.261228
2	.024502	-.00738	.056384	.085137	.045814	.12446	.014027	-.013443	.041498	.052636	.020362	.084909
3	.007284	-.005717	.020285	.026626	.009405	.043848	.004359	-.005458	.014175	.016072	.000948	.031196
4	.002229	-.00259	.007048	.008196	.000669	.015722	.001341	-.002154	.004835	.004934	-.001476	.011343
5	.000684	-.001044	.002413	.002518	-.000624	.00566	.000412	-.000814	.001637	.001515	-.001013	.004043
6	.00021	-.000398	.000818	.000774	-.000466	.002013	.000127	-.000297	.00055	.000466	-.000485	.001417
7	.000065	-.000146	.000275	.000238	-.00023	.000705	.000039	-.000106	.000184	.000143	-.000203	.000489
8	.00002	-.000052	.000092	.000073	-.000098	.000244	.000012	-.000037	.000061	.000044	-.000079	.000167
9	6.1e-06	-.000018	.000031	.000022	-.000038	.000083	3.7e-06	-.000013	.00002	.000013	-.000029	.000056
10	1.9e-06	-6.3e-06	.00001	6.9e-06	-.000014	.000028	1.1e-06	-4.3e-06	6.6e-06	4.1e-06	-.000011	.000019

95% lower and upper bounds reported

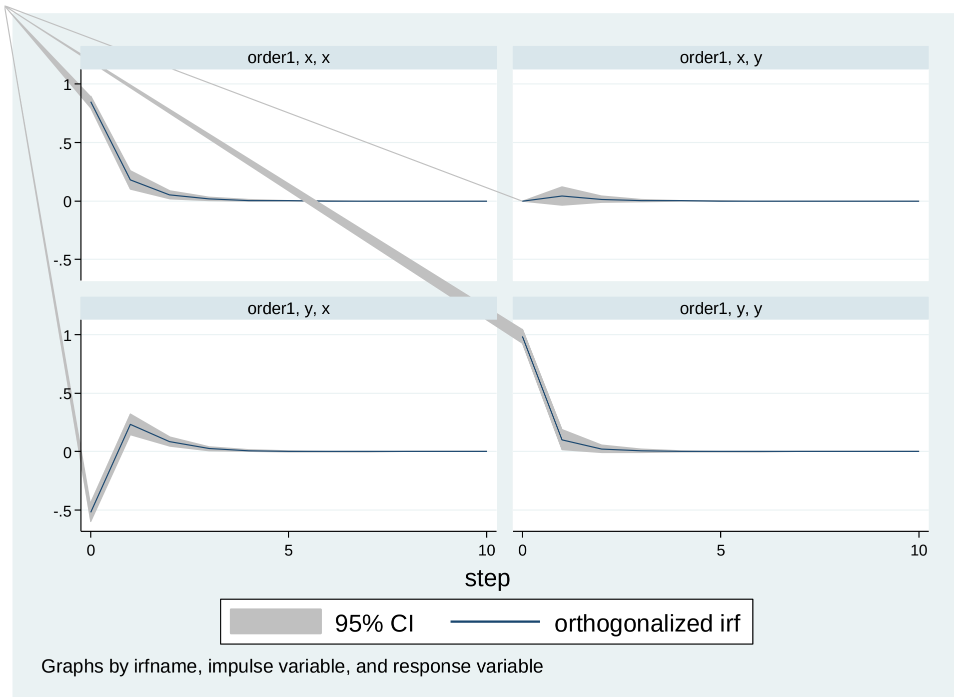
(1) irfname = order1, impulse = y, and response = y

(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)
```



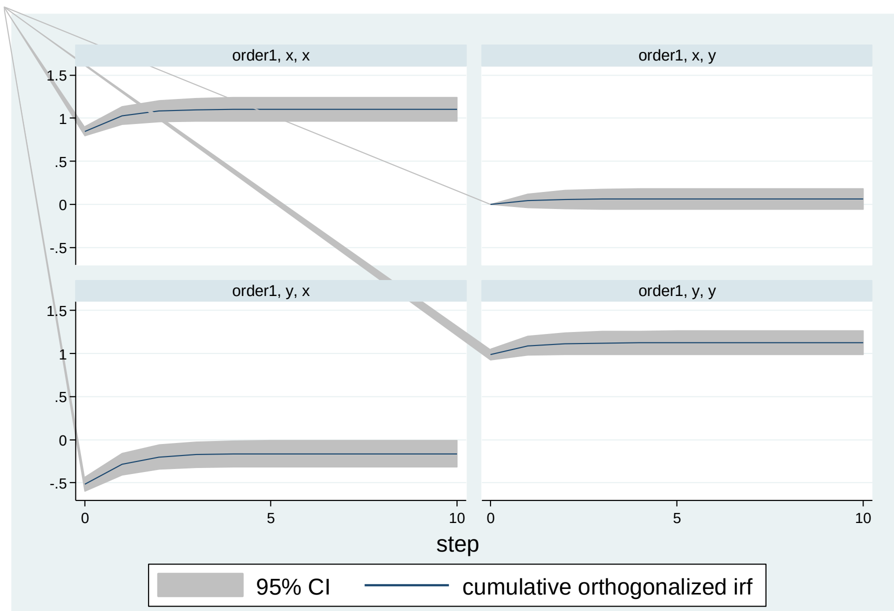

```
. irf table coirf, impulse(y x) response(y x)
```

Results from order1

step	(1) coirf	(1) Lower	(1) Upper	(2) coirf	(2) Lower	(2) Upper	(3) coirf	(3) Lower	(3) Upper	(4) coirf	(4) Lower	(4) Upper
0	.987418	.926157	1.04868	-.51896	-.60004	-.437881	0	0	0	.848107	.795489	.900725
1	1.0899	.980273	1.19952	-.286169	-.412639	-.159699	.041024	-.038502	.120551	1.02852	.926156	1.13088
2	1.1144	.987392	1.2414	-.201032	-.342581	-.059483	.055052	-.051621	.161725	1.08115	.957139	1.20517
3	1.12168	.987583	1.25578	-.174406	-.322962	-.02585	.05941	-.056878	.175699	1.09723	.964314	1.23014
4	1.12391	.987135	1.26069	-.16621	-.317625	-.014796	.060751	-.058934	.180436	1.10216	.965731	1.23859
5	1.1246	.986867	1.26232	-.163692	-.31619	-.011195	.061163	-.059704	.18203	1.10367	.965939	1.24141
6	1.12481	.986749	1.26286	-.162919	-.315806	-.010031	.061289	-.059983	.182562	1.10414	.965937	1.24234
7	1.12487	.986702	1.26304	-.162681	-.315705	-.009657	.061328	-.060082	.182738	1.10428	.965918	1.24265
8	1.12489	.986684	1.2631	-.162608	-.315678	-.009538	.06134	-.060116	.182796	1.10433	.965906	1.24275
9	1.1249	.986678	1.26312	-.162586	-.315672	-.009499	.061344	-.060128	.182816	1.10434	.965901	1.24278
10	1.1249	.986676	1.26312	-.162579	-.31567	-.009487	.061345	-.060132	.182822	1.10434	.965899	1.24279

```
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)
```



Graphs by irfname, impulse variable, and response variable

According to IRF Analysis, y has a positive effect on x with a greater magnitude compared with what x does on y. Moreover, the impact of y on x exists longer (longer duration) until it dies down.

In conclusion, 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

	(1)	(1)	(1)	(2)	(2)	(2)	(3)	(3)	(3)	(4)	(4)	(4)
step	fevd	Lower	Upper	fevd	Lower	Upper	fevd	Lower	Upper	fevd	Lower	Upper

0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	.272424	.205785	.339063	0	0	0	.727576	.660937	.794215
2	.998295	.991692	1.0049	.300844	.236036	.365652	.001705	-.004898	.008308	.699156	.634348	.763964
3	.998097	.990736	1.00546	.304746	.240486	.369005	.001903	-.005459	.009264	.695254	.630995	.759514
4	.998078	.990633	1.00552	.305127	.240929	.369324	.001922	-.005523	.009367	.694873	.630676	.759071
5	.998076	.990623	1.00553	.305163	.240971	.369354	.001924	-.00553	.009377	.694837	.630646	.759029
6	.998076	.990622	1.00553	.305166	.240975	.369357	.001924	-.005531	.009378	.694834	.630643	.759025
7	.998076	.990621	1.00553	.305167	.240976	.369358	.001924	-.005531	.009379	.694833	.630642	.759024
8	.998076	.990621	1.00553	.305167	.240976	.369358	.001924	-.005531	.009379	.694833	.630642	.759024
9	.998076	.990621	1.00553	.305167	.240976	.369358	.001924	-.005531	.009379	.694833	.630642	.759024
10	.998076	.990621	1.00553	.305167	.240976	.369358	.001924	-.005531	.009379	.694833	.630642	.759024

```
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 the 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)

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

Results from order1

step	(1) irf	(1) Lower	(1) Upper	(2) irf	(2) Lower	(2) Upper	(3) irf	(3) Lower	(3) Upper	(4) irf	(4) Lower	(4) Upper
0	1	1	1	0	0	0	0	0	0	1	1	1
1	.212722	.118349	.307095	.048372	-.04535	.142093	.347558	.248528	.446588	.129208	.030862	.227554
2	.062062	.024204	.099921	.01654	-.015834	.048914	.11884	.067187	.170494	.033507	-.014372	.081385
3	.018951	.001157	.036744	.005139	-.006431	.016709	.036925	.012175	.061676	.010078	-.008978	.029133
4	.005817	-.001731	.013366	.001581	-.002539	.0057	.011357	.000172	.022543	.003088	-.003914	.01009
5	.001787	-.001192	.004766	.000486	-.000959	.00193	.003489	-.001173	.008152	.000948	-.00155	.003447
6	.000549	-.000572	.00167	.000149	-.00035	.000649	.001072	-.000753	.002896	.000291	-.000584	.001167
7	.000169	-.000239	.000577	.000046	-.000125	.000216	.000329	-.000354	.001013	.000089	-.000213	.000392
8	.000052	-.000093	.000197	.000014	-.000044	.000072	.000101	-.000147	.000349	.000027	-.000076	.000131
9	.000016	-.000035	.000067	4.3e-06	-.000015	.000024	.000031	-.000057	.000119	8.4e-06	-.000026	.000043
10	4.9e-06	-.000013	.000022	1.3e-06	-5.1e-06	7.8e-06	9.5e-06	-.000021	.00004	2.6e-06	-9.1e-06	.000014

```
95% lower and upper bounds reported
```

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

(2) irfname = order1, impulse = x, and response = y

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

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

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

Results from order1

	(1)	(1)	(1)	(2)	(2)	(2)	(3)	(3)	(3)	(4)	(4)	(4)
step	oirf	Lower	Upper	oirf	Lower	Upper	oirf	Lower	Upper	oirf	Lower	Upper
0	.994287	.932599	1.05597	-.515375	-.595895	-.434856	0	0	0	.842248	.789994	.894503
1	.032383	-.055074	.119841	-.018495	-.10205	.06506	.29273	.207368	.378093	.108825	.025719	.191931
2	.00046	-.026231	.027152	-.000823	-.012926	.01128	.100093	.056147	.144039	.028221	-.012143	.068584
3	-.000188	-.009805	.009429	-.000084	-.002898	.00273	.0311	.010165	.052035	.008488	-.00757	.024546
4	-.000069	-.003091	.002953	-.00002	-.000837	.000797	.009566	.000126	.019006	.002601	-.003298	.008501
5	-.000022	-.000948	.000905	-5.9e-06	-.000253	.000242	.002939	-.000992	.00687	.000799	-.001306	.002904
6	-6.7e-06	-.00029	.000277	-1.8e-06	-.000077	.000074	.000903	-.000635	.002441	.000245	-.000492	.000983
7	-2.0e-06	-.000089	.000084	-5.6e-07	-.000024	.000022	.000277	-.000298	.000853	.000075	-.000179	.00033
8	-6.3e-07	-.000027	.000026	-1.7e-07	-7.2e-06	6.9e-06	.000085	-.000124	.000294	.000023	-.000064	.00011
9	-1.9e-07	-8.3e-06	7.9e-06	-5.3e-08	-2.2e-06	2.1e-06	.000026	-.000048	.0001	7.1e-06	-.000022	.000036
10	-5.9e-08	-2.5e-06	2.4e-06	-1.6e-08	-6.7e-07	6.4e-07	8.0e-06	-.000018	.000034	2.2e-06	-7.7e-06	.000012

95% lower and upper bounds reported

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

(2) irfname = order1, impulse = x, and response = y

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

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

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

Results from order1

	(1)	(1)	(1)	(2)	(2)	(2)	(3)	(3)	(3)	(4)	(4)	(4)
step	coirf	Lower	Upper	coirf	Lower	Upper	coirf	Lower	Upper	coirf	Lower	Upper
0	.994287	.932599	1.05597	-.515375	-.595895	-.434856	0	0	0	.842248	.789994	.894503
1	1.02667	.918494	1.13485	-.533871	-.656139	-.411603	.29273	.207368	.378093	.951073	.849374	1.05277
2	1.02713	.909734	1.14453	-.534694	-.666579	-.402808	.392823	.275251	.510396	.979294	.8471	1.11149
3	1.02694	.906338	1.14755	-.534778	-.668901	-.400655	.423924	.292377	.555471	.987782	.842865	1.1327
4	1.02687	.905186	1.14856	-.534798	-.669568	-.400028	.43349	.296281	.570698	.990383	.840743	1.14002
5	1.02685	.904821	1.14888	-.534804	-.66977	-.399838	.436428	.297057	.575799	.991182	.839866	1.1425
6	1.02684	.904709	1.14898	-.534806	-.669831	-.39978	.437331	.297172	.577491	.991427	.83953	1.14333

7	1.02684	.904674	1.14901	-.534806	-.66985	-.399762	.437609	.297171	.578046	.991503	.839406	1.1436
8	1.02684	.904664	1.14902	-.534806	-.669856	-.399757	.437694	.29716	.578228	.991526	.839362	1.14369
9	1.02684	.90466	1.14902	-.534806	-.669858	-.399755	.43772	.297153	.578287	.991533	.839347	1.14372
10	1.02684	.904659	1.14902	-.534806	-.669858	-.399755	.437728	.29715	.578306	.991535	.839342	1.14373

95% lower and upper bounds reported

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

(2) irfname = order1, impulse = x, and response = y

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

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

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

Results from order1

	(1)	(1)	(1)	(2)	(2)	(2)	(3)	(3)	(3)	(4)	(4)	(4)
step	fevd	Lower	Upper	fevd	Lower	Upper	fevd	Lower	Upper	fevd	Lower	Upper
0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	.272424	.205785	.339063	0	0	0	.727576	.660937	.794215
2	.920313	.876547	.964079	.269408	.202867	.335949	.079687	.035921	.123453	.730592	.664051	.797133
3	.911818	.863879	.959757	.269191	.202631	.335751	.088182	.040243	.136121	.730809	.664249	.797369
4	.911006	.862601	.959411	.269172	.202608	.335735	.088994	.040589	.137399	.730828	.664265	.797392
5	.910929	.862476	.959383	.26917	.202606	.335734	.089071	.040617	.137524	.73083	.664266	.797394
6	.910922	.862464	.959381	.26917	.202606	.335733	.089078	.040619	.137536	.73083	.664267	.797394
7	.910921	.862462	.959381	.26917	.202606	.335733	.089079	.040619	.137538	.73083	.664267	.797394
8	.910921	.862462	.959381	.26917	.202606	.335733	.089079	.040619	.137538	.73083	.664267	.797394
9	.910921	.862462	.959381	.26917	.202606	.335733	.089079	.040619	.137538	.73083	.664267	.797394
10	.910921	.862462	.959381	.26917	.202606	.335733	.089079	.040619	.137538	.73083	.664267	.797394

95% lower and upper bounds reported

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

(2) irfname = order1, impulse = x, and response = y

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

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

Now, the result changes. With the order (y x), we have that y has more impact on x. However, when to order changes to (x y), the x, now, has more impact on x. Since we use the Cholesky Order, order matters.