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)        =      90.03
                                                Prob > chi2       =     0.0000
Log likelihood = -106.20802                     Pseudo R2         =     0.2977

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

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

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

------------------------------------------------------------------------------
           y |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
-------------+----------------------------------------------------------------
1            |
          x1 |   .3075216   .1898691    -1.91   0.056     .0916915    1.031388
          x2 |    .440039    .333898    -1.08   0.279     .0994504    1.947045
```

```
          x3 |   .6735864   .4193531    -0.63   0.526     .1988212    2.282043
          x4 |   1.910766    .475696     2.60   0.009     1.172995    3.112569
       _cons |   .0004279    .001512    -2.20   0.028     4.20e-07     .435793
-------------+----------------------------------------------------------------
2            |
          x1 |   .3531572   .2385059    -1.54   0.123     .0939953    1.326874
          x2 |   .1897477   .1496983    -2.11   0.035     .0404236    .8906724
          x3 |   2.024558   1.342047     1.06   0.287     .5521821    7.422975
          x4 |   6.054483   1.830026     5.96   0.000      3.34806    10.94866
       _cons |   5.82e-12   2.63e-11    -5.74   0.000     8.48e-16    4.00e-08
------------------------------------------------------------------------------
Note: _cons estimates baseline relative risk for each outcome.

. est store my

. fitstat, saving(my)

Measures of Fit for mlogit of y

Log-Lik Intercept Only:       -151.224   Log-Lik Full Model:            -106.208
D(155):                        212.416   LR(8):                           90.032
                                         Prob > LR:                        0.000
McFadden's R2:                   0.298   McFadden's Adj R2:                0.198
ML (Cox-Snell) R2:               0.411   Cragg-Uhler(Nagelkerke) R2:       0.495
Count R2:                        0.735   Adj Count R2:                     0.274
AIC:                             1.426   AIC*n:                          242.416
BIC:                          -583.633   BIC':                           -48.946
BIC used by Stata:             263.774   AIC used by Stata:              232.416

(Indices saved in matrix fs_my)
```

From the estimated result, in tern if RRR, x1 x2 x3 have negative effect on dependent variable y. Only x4 has positive effect. LR-chi-squared test is significant which is indicative of overall significance of the model. For individual test, only x4 and constant term is significant in the level of 0.05. The rest are insignificant.

```
. mlogit y x1 x2 x3 x4 if y!=1, base(0) nolog


Multinomial logistic regression                   Number of obs     =        129
                                                  LR chi2(4)        =      58.57
                                                  Prob > chi2       =     0.0000
Log likelihood = -28.025327                       Pseudo R2         =     0.5110

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
-------------+----------------------------------------------------------------
2            |
          x1 |  -1.636463   .8818933    -1.86   0.064    -3.364942    .0920161
          x2 |   -2.31497   1.045126    -2.22   0.027     -4.36338   -.2665601
          x3 |   .6298355   .7253067     0.87   0.385    -.7917395    2.051411
          x4 |   2.038249   .4485343     4.54   0.000     1.159138     2.91736
       _cons |  -28.81833   6.458478    -4.46   0.000    -41.47671   -16.15995
------------------------------------------------------------------------------
```

```
. est store my1

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

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

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

. est store my2

. hausman my my1, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       my          my1         Difference          S.E.
-------------+----------------------------------------------------------------
          x1 |   -1.040842    -1.636463         .595621               .
          x2 |    -1.66206     -2.31497        .6529099               .
          x3 |    .7053513     .6298355        .0755157               .
          x4 |    1.800799     2.038249         -.23745               .
       _cons |     -25.869    -28.81833        2.949327               .
------------------------------------------------------------------------------
                          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)
                          =    -3.35    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test

. drop _est_my

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

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

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

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

. est store my

. hausman my my1, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       my          my1         Difference          S.E.
-------------+----------------------------------------------------------------
          x1 |   -1.040842    -1.636463         .595621               .
          x2 |    -1.66206     -2.31497        .6529099               .
          x3 |    .7053513     .6298355        .0755157               .
          x4 |    1.800799     2.038249         -.23745               .
       _cons |     -25.869    -28.81833        2.949327               .
------------------------------------------------------------------------------
                          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)
                          =    -3.35    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test

. hausman my my2, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       my          my2         Difference          S.E.
-------------+----------------------------------------------------------------
          x1 |    -1.17921    -.9900573       -.1891526        .1607978
          x2 |   -.8208919    -.5350665       -.2858253        .2008254
          x3 |   -.3951391    -.2076114       -.1875277               .
          x4 |    .6475044     .5845398        .0629646        .0422605
       _cons |   -7.756696    -7.222523       -.5341731               .
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from mlogit
           B = inconsistent under Ha, efficient under Ho; obtained from mlogit
```

```
    Test:  Ho:  difference in coefficients not systematic

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

. hausman my1 my, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      my1           my         Difference          S.E.
-------------+----------------------------------------------------------------
          x1 |   -1.636463    -1.040842        -.595621        .5671276
          x2 |    -2.31497     -1.66206       -.6529099        .6854726
          x3 |    .6298355     .7053513       -.0755157        .2943717
          x4 |    2.038249     1.800799          .23745        .3313943
       _cons |   -28.81833      -25.869       -2.949327        4.625125
------------------------------------------------------------------------------
                           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)
                          =        3.35
                Prob>chi2 =      0.6459

.
```

From Hausman test between dropping y=1, y=2 and normal case, all tests fail to reject H0. IIA hypothesis is not rejected. Hence, IIA is satisfied in the model.

IIA test is important because it makes the estimation unbiased. That is, with or without certain outcome, the probability of others outcome occurring is still the same.

(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)        =      81.97
                                                  Prob > chi2       =     0.0000
Log likelihood = -110.23758                       Pseudo R2         =     0.2710

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

```
       /cut1 |   9.475686   1.391637                      6.748128    12.20324
       /cut2 |   10.68547   1.443344                      7.856569    13.51437
------------------------------------------------------------------------------

      . probit y2 x1 x2 x3 x4, nolog

Probit regression                                 Number of obs     =        170
                                                  LR chi2(4)        =      78.68
                                                  Prob > chi2       =     0.0000
Log likelihood = -72.191808                       Pseudo R2         =     0.3527

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.1050855   .2671418    -0.39   0.694    -.6286739    .4185029
          x2 |  -.5287732   .2635885    -2.01   0.045    -1.045397   -.0121492
          x3 |   .6235435   .2521941     2.47   0.013      .129252    1.117835
          x4 |   .7467914   .1119396     6.67   0.000     .5273939     .966189
       _cons |  -11.81003   1.785646    -6.61   0.000    -15.30983    -8.31023
------------------------------------------------------------------------------

. est store probit2

. suest probit12 probit2

Simultaneous results for probit12, probit2

                                                  Number of obs     =        170

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
probit12_y12 |
          x1 |  -.6064434    .342389    -1.77   0.077    -1.277514    .0646268
          x2 |  -.6159647   .3636737    -1.69   0.090    -1.328752    .0968227
          x3 |   .1118567   .3168044     0.35   0.724    -.5090686    .7327819
          x4 |   .6486668   .1082616     5.99   0.000      .436478    .8608556
       _cons |  -8.396524    1.57024    -5.35   0.000    -11.47414   -5.318909
-------------+----------------------------------------------------------------
probit2_y2   |
          x1 |  -.1050855    .280877    -0.37   0.708    -.6555942    .4454232
          x2 |  -.5287732   .2956178    -1.79   0.074    -1.108173    .0506271
          x3 |   .6235435   .2639601     2.36   0.018     .1061911    1.140896
          x4 |   .7467914   .1223592     6.10   0.000     .5069719     .986611
       _cons |  -11.81003   1.939521    -6.09   0.000    -15.61142    -8.00864
------------------------------------------------------------------------------

. test [probit12_y12]x1=[probit2_y2]x1

 ( 1)  [probit12_y12]x1 - [probit2_y2]x1 = 0

           chi2(  1) =    1.72
         Prob > chi2 =    0.1891

. test [probit12_y12]x2=[probit2_y2]x2

 ( 1)  [probit12_y12]x2 - [probit2_y2]x2 = 0
```

```
          chi2(  1) =    0.05
        Prob > chi2 =    0.8310

. test [probit12_y12]x3=[probit2_y2]x3

 ( 1)  [probit12_y12]x3 - [probit2_y2]x3 = 0

          chi2(  1) =    2.05
        Prob > chi2 =    0.1521

. test [probit12_y12]x4=[probit2_y2]x4

 ( 1)  [probit12_y12]x4 - [probit2_y2]x4 = 0

          chi2(  1) =    0.39
        Prob > chi2 =    0.5310

.
```

From the model, x1 and x2 have negative effect on the probability of y while x3 and x4 have positive effect. LR-chi-squared test is significant which is indicative of overall significance of the model. Pseudo R squared is 0.271; not quite fit the data. All individual tests are significant except x1.

From the test, we accept all the tests. Hence, the model is ordered.

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:

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

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)

```
. use "C:\Users\Jilllin\OneDrive\Desktop\Thammasat\EE426\12_final\Final_q1-2_12.dta"

. probit y1 x*, nolog

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

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

. probit y2 x*, nolog

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

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

. probit y3 x*, nolog

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

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

. mvprobit (y1 x*) (y2 x*) (y3 x*), nolog

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

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y1           |
          x1 |   1.171117   .3893354     3.01   0.003     .4080335      1.9342
          x2 |  -.6025208   .3795038    -1.59   0.112    -1.346335     .141293
          x3 |   .6046014   .1786074     3.39   0.001     .2545374    .9546655
          x4 |   .1480072   .1915597     0.77   0.440     -.227443    .5234574
```

```
       _cons |  -.7106655   .3050131    -2.33   0.020     -1.30848   -.1128509
-------------+----------------------------------------------------------------
y2           |
          x1 |  -.5643036   .3755779    -1.50   0.133    -1.300423    .1718157
          x2 |  -.6138367   .3747629    -1.64   0.101    -1.348359    .1206851
          x3 |  -.2591482   .1782002    -1.45   0.146     -.608414    .0901177
          x4 |   .5646381   .1983093     2.85   0.004      .175959    .9533172
       _cons |   .3151067   .2995943     1.05   0.293    -.2720874    .9023007
-------------+----------------------------------------------------------------
y3           |
          x1 |  -.5475637   .3858414    -1.42   0.156    -1.303799    .2086715
          x2 |   .6638133   .3859124     1.72   0.085    -.0925611    1.420188
          x3 |  -.2722606   .1869458    -1.46   0.145    -.6386676    .0941464
          x4 |  -.1936979    .195082    -0.99   0.321    -.5760515    .1886558
       _cons |  -.4581402   .3194547    -1.43   0.152     -1.08426    .1679795
-------------+----------------------------------------------------------------
    /atrho21 |  -.6213012   .0749254    -8.29   0.000    -.7681523     -.47445
-------------+----------------------------------------------------------------
    /atrho31 |  -.4110659   .0740692    -5.55   0.000    -.5562389   -.2658929
-------------+----------------------------------------------------------------
    /atrho32 |  -.3237628   .0710803    -4.55   0.000    -.4630777   -.1844479
-------------+----------------------------------------------------------------
       rho21 |  -.5520333   .0520926   -10.60   0.000    -.6458538   -.4417878
-------------+----------------------------------------------------------------
       rho31 |  -.3893774   .0628392    -6.20   0.000    -.5051815   -.2597991
-------------+----------------------------------------------------------------
       rho32 |  -.3129053   .0641209    -4.88   0.000    -.4325893   -.1823843
------------------------------------------------------------------------------
Likelihood ratio test of  rho21 = rho31 = rho32 = 0:
             chi2(3) =  245.816   Prob > chi2 = 0.0000
```

We should estimate the models separately if we think that the decision is independent. In this case, from LR test in mvprobit model, H0 is rejected. It suggests that the model has correlation among disturbance term. Hence, separate probit model should not be employed.

(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)

From mvprobit above and its LR test, we can see that H0 is rejected which is indicative of correlation among disturbance terms. Hence, mvprobit should be employed in this case.

The main difference between mvprobit and normal probit is that the former model assumes that there exists correlation among error terms.

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

According to the following model,

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

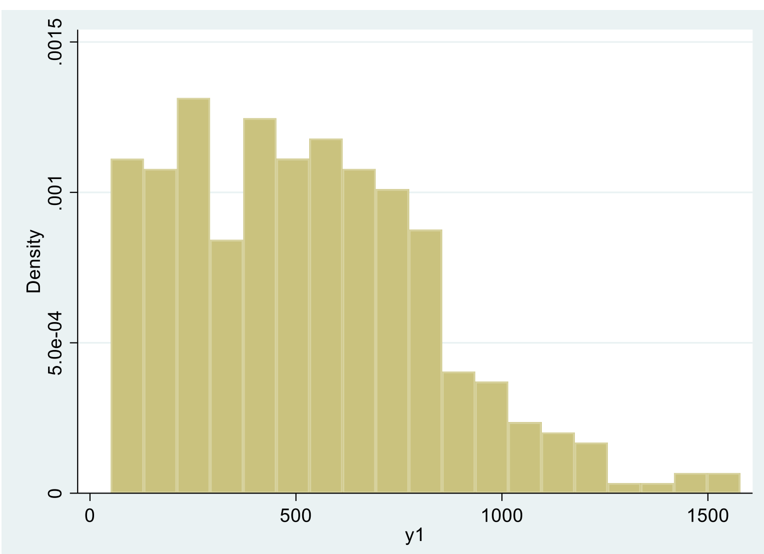
where: y_{ki} is Dependent variable, k=1, 2, 3.

x_i is Independent variable

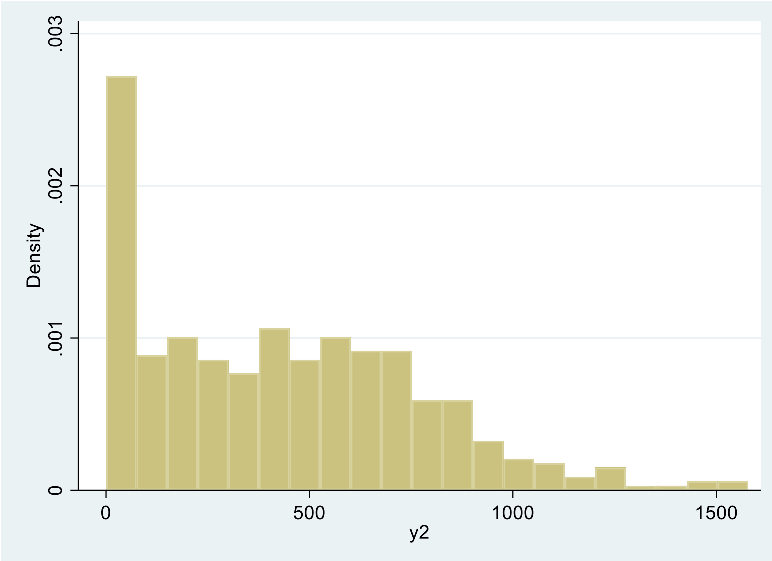
u_i is Stochastic disturbance term

(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)

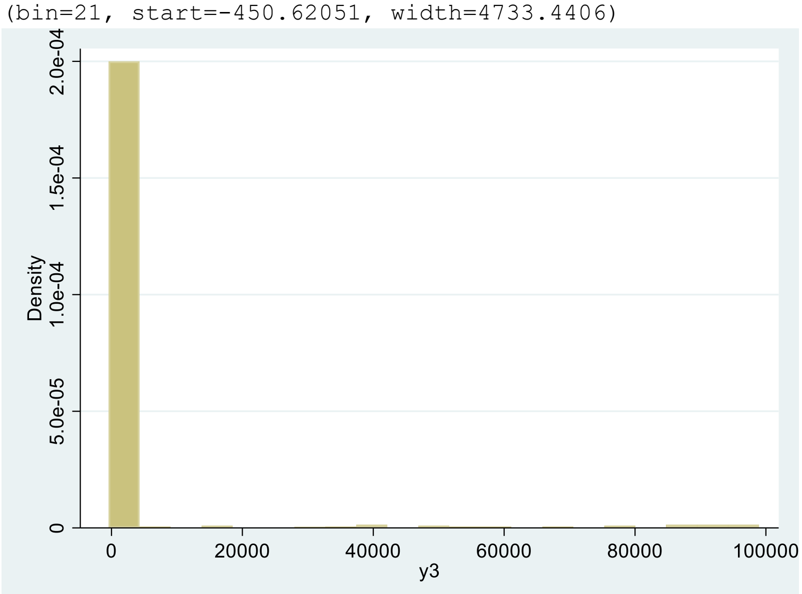
```
. use "C:\Users\Jilllin\OneDrive\Desktop\Thammasat\EE426\12_final\Final_q2_12.dta"

. tabstat y1 y2 y3

   stats |        y1        y2        y3
---------+------------------------------
    mean |  522.0268   429.054  3648.754
----------------------------------------

. sum y1 y2 y3

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          y1 |        369    522.0268    305.5502   50.21759    1578.51
          y2 |        450     429.054    340.6061          0    1578.51
          y3 |        450    3648.754    15405.28  -450.6205   98951.63

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




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




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



From the histograms and descriptive statistics, we can conclude that y1 has truncated problem around 50. Y2 has censored problem at 0. 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   =       369
-------------+----------------------------------   F(1, 367)       =    152.11
       Model |  10067286.6         1  10067286.6   Prob > F        =    0.0000
```

```
    Residual |  24289539.5       367  66184.0312   R-squared       =    0.2930
-------------+----------------------------------   Adj R-squared   =    0.2911
       Total |  34356826.1       368  93360.9404   Root MSE        =    257.26

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

. est store olsy1

. sum y1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          y1 |        369    522.0268    305.5502   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     =        369
         upper =       +inf                     Wald chi2(1)      =     112.14
Log likelihood = -2532.9115                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   243.5473   22.99863    10.59   0.000     198.4709    288.6238
       _cons |  -350.6846   88.77073    -3.95   0.000    -524.6721   -176.6972
-------------+----------------------------------------------------------------
      /sigma |   307.6565   17.05061    18.04   0.000     274.2379    341.0751
------------------------------------------------------------------------------

. est store truncy1

. lrtest olsy1 truncy1, force

Likelihood-ratio test                                 LR chi2(1)  =     75.32
(Assumption: olsy1 nested in truncy1)                 Prob > chi2 =    0.0000
```

From the estimated result and LR test, H0 is rejected. It can be concluded that truncated regression is more appropriated model in this case.

The major problem in this case is truncated problem; that is there exists unobservable data whether by survey design or incidental truncation.

If we ignored the problem, it would cause the model to be biased. This is because we ignore the existed data which are not observed in the database.

(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)       =    276.04
       Model |  19859200.8         1  19859200.8   Prob > F        =    0.0000
    Residual |  32230409.6       448  71942.8785   R-squared       =    0.3813
-------------+----------------------------------   Adj R-squared   =    0.3799
       Total |  52089610.4       449  116012.495   Root MSE        =    268.22

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

.  est store olsy2

. tobit y2 x, ll(0)

Refining starting values:

Grid node 0:   log likelihood = -2803.0826

Fitting full model:

Iteration 0:   log likelihood = -2803.0826
Iteration 1:   log likelihood = -2793.4686
Iteration 2:   log likelihood = -2793.2604
Iteration 3:   log likelihood = -2793.2596
Iteration 4:   log likelihood = -2793.2596

Tobit regression                                Number of obs     =        450
                                                   Uncensored     =        384
Limits: lower = 0                                  Left-censored  =         66
        upper = +inf                               Right-censored =          0

                                                LR chi2(1)        =     224.96
                                                Prob > chi2       =     0.0000
Log likelihood = -2793.2596                     Pseudo R2         =     0.0387

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   237.8077   14.46312    16.44   0.000     209.3839    266.2315
       _cons |   -320.499   47.07643    -6.81   0.000    -413.0165   -227.9815
-------------+----------------------------------------------------------------
    var(e.y2)|   89812.73   6640.855                      77665.64    103859.6
------------------------------------------------------------------------------

. est store toby2

. lrtest olsy2 toby2, force

Likelihood-ratio test                                 LR chi2(1)  =    721.15
(Assumption: olsy2 nested in toby2)                   Prob > chi2 =    0.0000
```

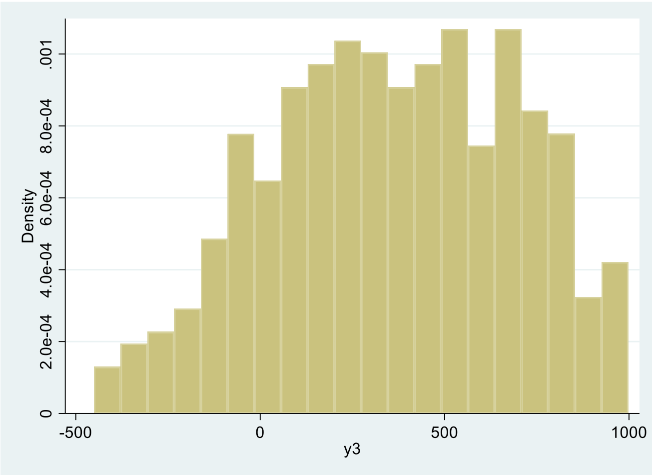
From significant LR test, it is indicative of tobit model being appropriated.

The main problem in this model is the censored problem, meaning that some data is not equal to zero, but got censored to be zero.

If we ignored this issue, it would cause the model to be biased since we take some data as zero when they are other values.

(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)

```
. hist y3 if y3<4000
(bin=20, start=-450.62051, width=72.440572)
```




```
reg y3 x

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

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

. est store olsy3

. tobit y3 x, ul(1000) nolog

Tobit regression                                Number of obs     =        450
                                                   Uncensored     =        426
Limits: lower = -inf                               Left-censored  =          0
        upper = 1,000                              Right-censored =         24

                                                LR chi2(1)        =     224.64
                                                Prob > chi2       =     0.0000
Log likelihood =  -3041.355                     Pseudo R2         =     0.0356
```

```
------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   223.9965   13.33998    16.79   0.000     197.7799     250.213
       _cons |  -274.6075   42.33736    -6.49   0.000    -357.8115   -191.4036
-------------+----------------------------------------------------------------
   var(e.y3)|   82756.79   5729.129                      72229.89     94817.9
------------------------------------------------------------------------------

. est store toby3

. lrtest olsy3 toby3, force

Likelihood-ratio test                                 LR chi2(1)  =   3831.23
(Assumption: olsy3 nested in toby3)                   Prob > chi2 =    0.0000
```

From the histogram, we can see that most of the data are accumulated around -500 to 1000. Hence, the problem in this question is clearly an outlier problem. With outlier problem, if we used all the data points including outlier, it would cause the model to be biased since the outlier will carry so much weight from its huge squared-value. Along with significant LR test, it can be concluded that tobit model is more appropriated than OLS method.

As mentioned above, if we ignored the outliers and took them as part of the model, it would cause the model to be biased due to their huge squared values.

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

The generalized linear regression model can be stated as:

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

and

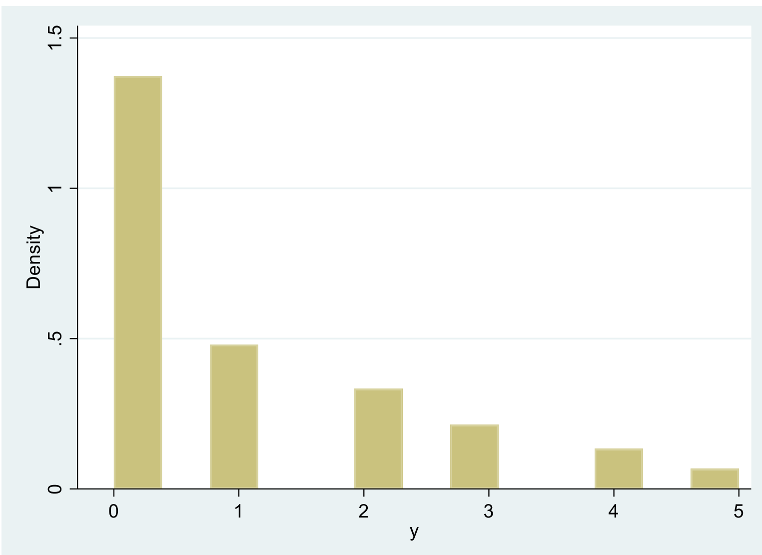
$$\Pr(Y_i = y_i) = f(I_i)$$

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 x*

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

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

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

. est store ols

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



.

From the histogram, it looks like the distribution follows poisson distribution. Hence, Poisson regression should be employed in this case.

(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 x*, nolog

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

------------------------------------------------------------------------------
            y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0839503   .0335816     2.50   0.012     .0181315     .149769
```

```
          x2 |   .1356394   .0359108     3.78   0.000     .0652555    .2060234
          x4 |  -.0355363   .0372572    -0.95   0.340     -.108559    .0374865
          x3 |   .2241043   .0545131     4.11   0.000     .1172605     .330948
       _cons |  -.1320593   .0924644    -1.43   0.153    -.3132863    .0491677
------------------------------------------------------------------------------

. estat gof

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

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

. poisson y x*,ir nolog

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

------------------------------------------------------------------------------
           y |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.087575   .0365225     2.50   0.012     1.018297    1.161566
          x2 |   1.145269   .0411276     3.78   0.000     1.067432    1.228782
          x4 |   .9650877   .0359565    -0.95   0.340     .8971259    1.038198
          x3 |   1.251201   .0682069     4.11   0.000     1.124412    1.392287
       _cons |    .876289   .0810256    -1.43   0.153     .7310406    1.050396
------------------------------------------------------------------------------
Note: _cons estimates baseline incidence rate.

. mfx

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

From GOF test, H0 is rejected. Hence, Poisson regression model is not appropriated in this case.

From the estimated result, sign and meaning are correct. For x1, x2, and x3, IR is more than 1 and marginal effect is positive. For x4, IR is less than 1 and marginal effect is negative. Overall LR test is significant. Pseudo R2 is 0.0656; not quite fit the data. Individual z test is all significant except x4.

(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

```
nbreg y x*, nolog

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

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

. est store nbreg

. nbreg y x*,ir nolog

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

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

. mfx

Marginal effects after nbreg
      y  = Predicted number of events (predict)
         =  .91603373
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .1025465      .04734    2.17   0.030    .00976  .195333   -.28246
      x2 |   .1416443      .04734    2.99   0.003   .048867  .234422   .842848
```

```
      x4 |  -.0356977      .04579   -0.78   0.436  -.125448  .054053  -.805838
      x3 |   .1975276      .06457    3.06   0.002   .070968  .324087  -.298846
```

From the estimated result, sign and meaning are correct. For x1, x2, and x3, IR is more than 1 and marginal effect is positive. For x4, IR is less than 1 and marginal effect is negative. Overall LR test is significant. Pseudo R2 is 0.0656; not quite fit the data. Individual z test is all 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)

```
. zip y x1 x2 x3, inflate(x4) vuong nolog
Vuong test is not appropriate for testing zero-inflation.  Specify option forcevuong
to perform the test anyway.
r(498);

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

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

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

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
          x1 |   .0991548   .0425576     2.33   0.020     .0157434    .1825661
          x2 |   .1208637   .0400449     3.02   0.003     .0423772    .1993502
          x3 |   .1553467   .0562608     2.76   0.006     .0450775    .2656159
       _cons |   .3521057   .1147728     3.07   0.002     .1271551    .5770564
-------------+----------------------------------------------------------------
inflate      |
          x4 |   .0579183   .1008572     0.57   0.566    -.1397582    .2555947
       _cons |   -.511083   .2578789    -1.98   0.047    -1.016516   -.0056497
------------------------------------------------------------------------------
Vuong test of zip vs. standard Poisson:            z =     2.57  Pr>z = 0.0051
Warning: The Vuong test is not appropriate for testing zero-inflation.

. mfx

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

From Vuong test, H0 is rejected. Hence, Zero-inflated-Poisson regression is more appropriated than normal Poisson regression in this case.

Furthermore, the histogram of y also indicates zero inflated problem. This means that ZIP model should be applied in this case.

Regarding interpretation, x1, x2, and x3 have positive marginal effect. X4 has negative marginal effect. Overall LR test is significant.

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\Jilllin\OneDrive\Desktop\Thammasat\EE426\12_final\Final_q4_12.dta"

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

. 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)            -15.219            -3.980            -3.420            -3.130
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
D.x          |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |
         L1. |  -.9857164    .064769   -15.22   0.000    -1.112973   -.8584597
         LD. |   -.050295   .0450108    -1.12   0.264    -.1387313    .0381413
      _trend |   .0026639   .0030113     0.88   0.377    -.0032527    .0085804
       _cons |   .4094375   .8696255     0.47   0.638    -1.299183    2.118058
------------------------------------------------------------------------------
. 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)            -15.067            -3.980            -3.420            -3.130
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

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

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

From Dfuller test, H0 is rejected. X and Y are stationary or I(0). The process is, when we test if there is a unit root or not, there must be unit root.

The importance of stationary process is that it makes the estimation unbiased. Just like OLS, we require the data to have constant mean, variance, and covariance. That is, the process of generating data does not change overtime.

(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 arima1*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

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

. est table arima2*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

------------------------------------------------------------------------------
    Variable |   arima201        arima202        arima203        arima204     
```

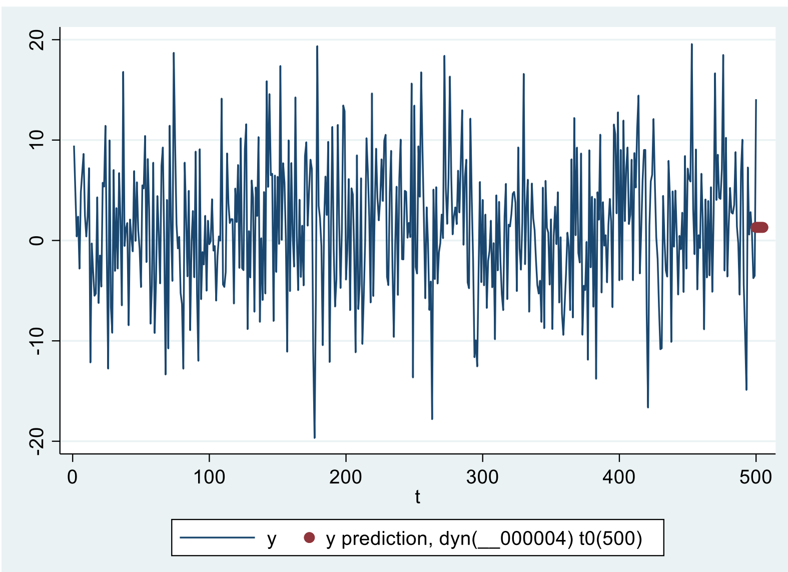
```
-------------+--------------------------------------------------------------
y            |
       _cons |  1.3073551***    1.3075824***    1.3076332***     1.308031***
-------------+--------------------------------------------------------------
ARMA         |
          ar |
         L1. |  .72833656***    .01069941      -.02391453      -.03883203
         L2. |  .07343696       .73210465***    .72886425***    .68853566**
             |
          ma |
         L1. | -.76420028***   -.02937305      -.00720022       .00857444
         L2. |                 -.6650084**     -.66251356**    -.63712588**
         L3. |                                  .02176552       .02159808
         L4. |                                                  .02591727
-------------+--------------------------------------------------------------
sigma        |
       _cons |  6.7292866***     6.717086***    6.7159767***    6.7140828***
-------------+--------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -1662.7092      -1661.8275      -1661.7334      -1661.6033
        chi2 |  19.097086       22.261236       23.402544       22.065999
         aic |  3335.4184       3335.6549       3337.4668       3339.2066
         bic |  3356.4915       3360.9426       3366.9691       3372.9234
----------------------------------------------------------------------------
                                           legend: * p<.1; ** p<.05; *** p<.01

. est table arima3*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

----------------------------------------------------------------------------
    Variable |   arima301        arima302        arima303        arima304
-------------+--------------------------------------------------------------
y            |
       _cons |  1.3030581***      1.30767***     1.322265***    1.3224088***
-------------+--------------------------------------------------------------
ARMA         |
          ar |
         L1. | -.85123464***   -.05144311       1.0589369***    1.0600047***
         L2. |  .03402859       .72972779***    .66963573**      .6674315*
         L3. |  .02022783       .02220492      -.84477131***   -.84357701***
             |
          ma |
         L1. |  .82424826***    .02123612      -1.1139372       -1.113851
         L2. |                 -.66391706**    -.57552789      -.57406071
         L3. |                                  .81625432       .81322054
         L4. |                                                  .00161264
-------------+--------------------------------------------------------------
sigma        |
       _cons |  6.7267628***    6.7160989***    6.6327409       6.6327335
-------------+--------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -1662.5206      -1661.7401      -1657.5933      -1657.5927
        chi2 |  59.442039       24.064744       38039.814       37857.088
         aic |  3337.0413       3337.4802       3329.1866       3331.1855
         bic |  3362.3289       3366.9825       3358.6889       3364.9024
----------------------------------------------------------------------------
                                           legend: * p<.1; ** p<.05; *** p<.01
```

```
. est table arima4*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

----------------------------------------------------------------------------
    Variable |   arima401        arima402        arima403        arima404
-------------+--------------------------------------------------------------
y            |
       _cons |   1.304488***    1.3077626***    1.3224016***    1.3085509***
-------------+--------------------------------------------------------------
ARMA         |
          ar |
         L1. | -.78736421***   -.07222081       1.0619698***    .06379198
         L2. |  .03406615       .65323856*      .66535028*     -.23017888
         L3. |  .05230227        .0238182      -.84489349***   -.06553786
         L4. |  .04891326       .02733522       .00166106       .70904435***
             |
          ma |
         L1. |  .75932045***    .04153469      -1.1158233      -.09628464
         L2. |                 -.6018471      -.57186605       .31341447
         L3. |                                 .81436075       .06673855
         L4. |                                                -.65853127**
-------------+--------------------------------------------------------------
sigma        |
       _cons |  6.7195128***     6.714347***    6.6327178       6.6743326***
-------------+--------------------------------------------------------------
Statistics   |
           N |        500             500             500             500
          ll | -1661.9967      -1661.6088      -1657.5927      -1658.9217
        chi2 |  41.886103       21.434333       38010.977       1620.7592
         aic |  3337.9934       3339.2176       3331.1855       3337.8434
         bic |  3367.4956       3372.9345       3364.9024       3379.9895
----------------------------------------------------------------------------
                                         legend: * p<.1; ** p<.05; *** p<.01
```

From the result, the most appropriate order for y is ARIMA(1,0,1).

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

ARIMA regression

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

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
       _cons |   1.301666   .2963809     4.39   0.000     .7207701    1.882562
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.8655705   .1555056    -5.57   0.000    -1.170356    -.560785
             |
          ma |
         L1. |   .8243972   .1763698     4.67   0.000     .4787187    1.170076
-------------+----------------------------------------------------------------
      /sigma |   6.728997   .2146963    31.34   0.000       6.3082    7.149794
```

```
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.

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

. replace time=_n
variable time not found
r(111);

. 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: $$y_t = \alpha + \beta x_t + \varepsilon_t \tag{5}$$

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

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)       =  57285.96
       Model |  22596.1563         1  22596.1563   Prob > F        =    0.0000
    Residual |  196.433573       498  .394444925   R-squared       =    0.9914
-------------+----------------------------------   Adj R-squared   =    0.9914
       Total |  22792.5899       499  45.6765328   Root MSE        =    .62805

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

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

.
```

From the LM test, H0 is rejected. Hence, there exists arch effect.
The test is classified as LM test because we only use OLS (restricted) model to test whether there exists ARCH effect or not.

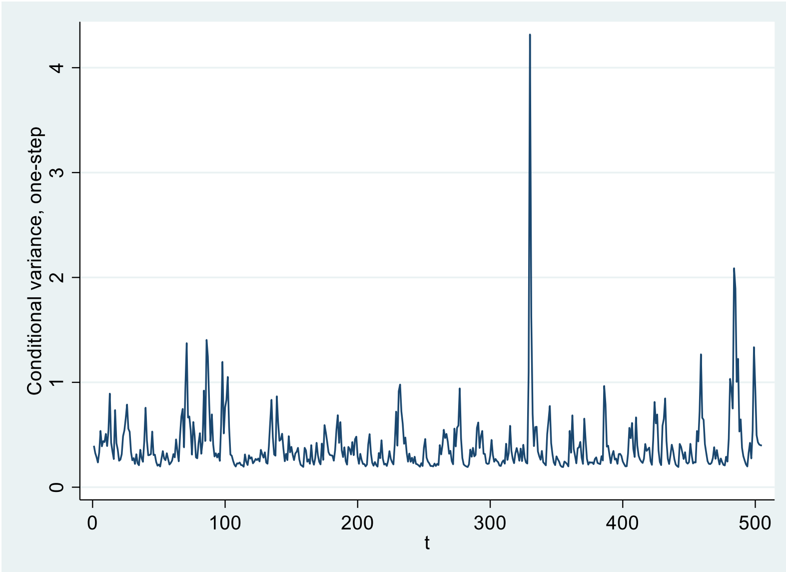
(d) Estimate GARCH(p,q) for y_t using x_t as explanatory variable for mean equation (model (5) and (6)) – determine the most appropriated order p and q for variance equation using SBIC given the maximum lag equals to 2. Then, predict the variance of y_t using the estimated result of GARCH(p,q) model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

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

. est store garch11

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

. est store garch12

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

. est store garch21

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

. est store garch22

. est table garch1* garch2*, star(0.1 0.05 0.01) stat(N ll chi2 aic bic)

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

```
    Variable |    garch11         garch12         garch21         garch22     
-------------+----------------------------------------------------------------
y            |
           x |  .69788579***    .69787087***    .69787574***    .69777417***  
       _cons |  .51826312***     .5182283***    .51823952***    .51597981***  
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |  .36201351***    .36039806***    .36077406***    .37189936***  
         L2. |                  .01483425                       .30792554**   
             |
       garch |
         L1. |  .31635697***    .28933812       .32663124*     -.39040448     
         L2. |                                 -.00927871        .1071777     
             |
       _cons |  .12891287***    .13442533*      .12900707***    .24574265***  
-------------+----------------------------------------------------------------
Statistics   |
           N |        500             500             500             500     
          ll | -444.30123      -444.29481      -444.29646      -443.46007     
        chi2 |  72628.104       72713.019       72680.391       74301.891     
         aic |  898.60246       900.58963       900.59293       900.92015     
         bic |  919.67551       925.87728       925.88058        930.4224     
------------------------------------------------------------------------------
                                      legend: * p<.1; ** p<.05; *** p<.01
```

From the fact that garch11 has the lowest BIC value, the most appropriated lag is 1.

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

ARCH family regression

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

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
           x |   .6978858   .0025896   269.50   0.000     .6928103    .7029613
       _cons |   .5182631     .02399    21.60   0.000     .4712436    .5652826
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |   .3620135   .0745312     4.86   0.000      .215935     .508092
             |
       garch |
         L1. |    .316357   .1209539     2.62   0.009     .0792917    .5534223
             |
       _cons |   .1289129   .0367809     3.50   0.000     .0568237     .201002
------------------------------------------------------------------------------

. predict variance, v

. line variance 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) nolog

ARCH family regression

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

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

. est store arch1

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

ARCH family regression

Sample: 1 - 500                                 Number of obs     =        500
```

```
Distribution: Gaussian                          Wald chi2(1)    =   72628.10
Log likelihood = -444.3012                      Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
           x |   .6978858   .0025896   269.50   0.000     .6928103    .7029613
       _cons |   .5182631     .02399    21.60   0.000     .4712436    .5652826
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |   .3620135   .0745312     4.86   0.000      .215935     .508092
             |
       garch |
         L1. |    .316357   .1209539     2.62   0.009     .0792917    .5534223
             |
       _cons |   .1289129   .0367809     3.50   0.000     .0568237     .201002
------------------------------------------------------------------------------

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

ARCH family regression

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

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

. est store egarch111

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

----------------------------------------------------------------
    Variable |    arch1          garch11        egarch111
-------------+--------------------------------------------------
y            |
```

```
            x |  .69780366***    .69788579***    .69805918***
        _cons |  .51429931***    .51826312***    .51158439***
-------------+------------------------------------------------
ARCH          |
         arch |
          L1. |  .38143612***    .36201351***    .67109089***
              |
        garch |
          L1. |                  .31635697***   -.00612483
              |
       egarch |
          L1. |                                  .36321612***
              |
        _cons |  .24222246***    .12891287***   -.92029794***
-------------+------------------------------------------------
Statistics    |
            N |        500             500             500
           ll | -449.32421      -444.30123      -447.73993
         chi2 |  66110.083       72628.104       66810.474
          aic |  906.64842       898.60246       907.47987
          bic |  923.50685       919.67551       932.76752
--------------------------------------------------------------
                           legend: * p<.1; ** p<.05; *** p<.01
```

From the table, GARCH11 is the most appropriated model with the lowest BIC value.

The differences among these models are how each model is constructed. In ARCH, the conditional variance is specified as a linear function of past sample only. GARCH also allows lagged conditional variance. EGARCH is used in asymmetric GARCH model.

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

The following VARs models:

$$Y_t = A_0 + A_1 Y_{t-1} + \epsilon_t \tag{7}$$

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).

```
varsoc y x

   Selection-order criteria
   Sample:  5 - 500                              Number of obs      =       496
  +---------------------------------------------------------------------------+
  |lag |    LL      LR      df    p      FPE       AIC      HQIC      SBIC    |
  |----+----------------------------------------------------------------------|
  |  0 | -1366.29                      .853444   5.51728   5.52394   5.53424  |
  |  1 | -1318.39  95.784*   4  0.000   .71501*  5.34029*  5.36027*  5.39118* |
  |  2 | -1317.46  1.8564    4  0.762  .723922   5.35268   5.38597   5.43749  |
```

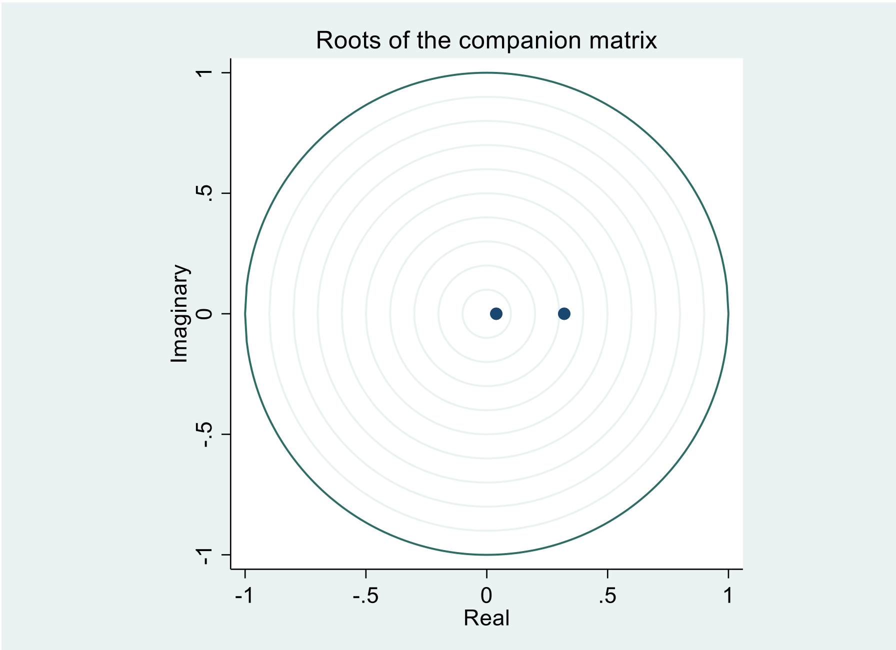
```
  |  3 | -1316.48  1.9613    4  0.743  .732791   5.36486   5.41146   5.48359  |
  |  4 | -1314.52  3.9321    4  0.415  .738829   5.37306   5.43298   5.52572  |
  +---------------------------------------------------------------------------+
   Endogenous:  y x
    Exogenous:  _cons
```

From the result, lag order of 1 has the lowest SBIC, so it is the most appropriate one.

(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 =  -1328.462                    AIC               =   5.348544
FPE            =   .7209326                    HQIC              =   5.368422
Det(Sigma_ml)  =   .7038021                    SBIC              =   5.399197

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
y                     3     .994409   0.0147   7.431103   0.0243
x                     3     .995581   0.0913   50.11181   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
           y |
         L1. |    .136194   .0500873     2.72   0.007     .0380247    .2343633
             |
           x |
         L1. |    .051822   .0480067     1.08   0.280    -.0422694    .1459134
             |
       _cons |   .3382272   .0545227     6.20   0.000     .2313646    .4450898
-------------+----------------------------------------------------------------
x            |
           y |
         L1. |   .3449878   .0501463     6.88   0.000     .2467027    .4432728
             |
           x |
         L1. |   .2233362   .0480633     4.65   0.000     .1291339    .3175386
             |
       _cons |   .1202379    .054587     2.20   0.028     .0132493    .2272265
------------------------------------------------------------------------------

. varstable, graph

   Eigenvalue stability condition
  +----------------------------------------+
  |        Eigenvalue         |   Modulus  |
  |---------------------------+------------|
  |   .3203937                |   .320394  |
```

```
|  .03913659             |   .039137   |
+----------------------------------------+
 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 |  1.1653     1    0.280    |
  |                 y                ALL |  1.1653     1    0.280    |
  |--------------------------------------+---------------------------|
  |                 x                  y |  47.329     1    0.000    |
  |                 x                ALL |  47.329     1    0.000    |
  +------------------------------------------------------------------+
```

Since all eigenvalues lie inside a unit circle, the system is stable. However, in Granger exogeneity test, we cannot conclude that there exists relationship from y to y and y to x since we fail to reject H0 in the test. We can conclude, on the other hand, that there exists relationship between x to x and x to y.

If the stability assumption is not satisfied, the model will be biased.

(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(irf02)
(file irf02.irf created)
(file irf02.irf now active)
(file irf02.irf updated)
```

```
. 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
0	.991415	.929906	1.05292	-.518824	-.599746	-.437901	0	0	0
1	.108138	.021289	.194988	.226154	.136865	.315442	.043851	-.035815	.123517
2	.026448	-.005744	.058639	.087815	.048408	.127222	.015766	-.013169	.044701
3	.008153	-.00547	.021775	.028736	.010809	.046663	.005118	-.005643	.015879
4	.0026	-.002657	.007857	.00923	.001177	.017284	.001643	-.002346	.005632
5	.000832	-.001131	.002795	.002958	-.000517	.006434	.000526	-.000931	.001984
6	.000267	-.000453	.000986	.000948	-.000476	.002371	.000169	-.000357	.000694
7	.000085	-.000174	.000345	.000304	-.000255	.000862	.000054	-.000133	.000241
8	.000027	-.000065	.00012	.000097	-.000115	.00031	.000017	-.000049	.000083
9	8.8e-06	-.000024	.000042	.000031	-.000048	.00011	5.5e-06	-.000018	.000029
10	2.8e-06	-8.7e-06	.000014	1.0e-05	-.000019	.000039	1.8e-06	-6.2e-06	9.8e-06

step	(4) oirf	(4) Lower	(4) Upper
0	.846194	.793695	.898693
1	.188986	.108415	.269557
2	.057336	.023939	.090732
3	.018244	.002138	.034351
4	.00584	-.001247	.012928
5	.001871	-.001039	.004781
6	.000599	-.000541	.00174
7	.000192	-.000241	.000625
8	.000062	-.000099	.000222
9	.00002	-.000039	.000078
10	6.3e-06	-.000015	.000027

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

.

```
. irf 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
0	.991415	.929906	1.05292	-.518824	-.599746	-.437901	0	0	0
1	1.09955	.989319	1.20979	-.29267	-.419476	-.165864	.043851	-.035815	.123517
2	1.126	.997655	1.25435	-.204855	-.347748	-.061962	.059617	-.048638	.167872
3	1.13415	.998274	1.27003	-.176119	-.326612	-.025626	.064736	-.054054	.183525
4	1.13675	.99792	1.27559	-.166888	-.320598	-.013179	.066378	-.056284	.189041
5	1.13759	.997658	1.27751	-.16393	-.318908	-.008953	.066905	-.057161	.190971
6	1.13785	.997531	1.27817	-.162982	-.318436	-.007528	.067073	-.057495	.191641
7	1.13794	.997478	1.2784	-.162679	-.318306	-.007051	.067127	-.057618	.191873
8	1.13796	.997456	1.27847	-.162581	-.318271	-.006892	.067145	-.057663	.191953
9	1.13797	.997448	1.2785	-.16255	-.318261	-.006839	.06715	-.057679	.19198
10	1.13798	.997445	1.27851	-.16254	-.318259	-.006821	.067152	-.057685	.191989

step	(4) coirf	(4) Lower	(4) Upper
0	.846194	.793695	.898693
1	1.03518	.932813	1.13755
2	1.09252	.966996	1.21803
3	1.11076	.97537	1.24615
4	1.1166	.977149	1.25605
5	1.11847	.977442	1.2595
6	1.11907	.977453	1.26069
7	1.11926	.977431	1.26109
8	1.11932	.977416	1.26123
9	1.11934	.977409	1.26128

10	1.11935	.977406	1.26129

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

.

From the table, 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	(3) fevd	(3) Lower	(3) Upper
0	0	0	0	0	0	0	0	0	0
1	1	1	1	.273216	.206552	.339879	0	0	0
2	.99807	.991067	1.00507	.298786	.23423	.363342	.00193	-.005074	.008933
3	.997823	.989931	1.00571	.302872	.238883	.36686	.002177	-.005715	.010069
4	.997797	.989799	1.00579	.30331	.239392	.367227	.002203	-.005794	.010201
5	.997794	.989784	1.0058	.303355	.239445	.367265	.002206	-.005804	.010216
6	.997794	.989783	1.00581	.30336	.23945	.367269	.002206	-.005805	.010217
7	.997794	.989782	1.00581	.30336	.239451	.367269	.002206	-.005805	.010218
8	.997794	.989782	1.00581	.30336	.239451	.367269	.002206	-.005805	.010218
9	.997794	.989782	1.00581	.30336	.239451	.367269	.002206	-.005805	.010218
10	.997794	.989782	1.00581	.30336	.239451	.367269	.002206	-.005805	.010218

step	(4) fevd	(4) Lower	(4) Upper
0	0	0	0
1	.726784	.660121	.793448
2	.701214	.636658	.76577
3	.697128	.63314	.761117
4	.69669	.632773	.760608
5	.696645	.632735	.760555
6	.69664	.632731	.76055
7	.69664	.632731	.760549
8	.69664	.632731	.760549
9	.69664	.632731	.760549
10	.69664	.632731	.760549

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

From the table, 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 create order1, o(x y) step(10) set(irf03)
(file irf03.irf created)
(file irf03.irf now active)
(file irf03.irf updated)

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

Results from order1

step	(1) oirf	(1) Lower	(1) Upper	(2) oirf	(2) Lower	(2) Upper	(3) oirf	(3) Lower	(3) Upper
0	.992584	.931002	1.05416	-.518213	-.59904	-.437386	0	0	0
1	.042903	-.044366	.130171	-.01914	-.103052	.064772	.291583	.206566	.3766
2	.002979	-.024158	.030115	-.000383	-.013842	.013075	.104833	.06021	.149455
3	.000533	-.009788	.010854	.000102	-.003223	.003428	.034034	.012111	.055957
4	.000154	-.003292	.0036	.000042	-.000979	.001062	.010922	.00073	.021114
5	.000049	-.001073	.00117	.000014	-.000314	.000341	.0035	-.0009	.0079
6	.000016	-.000348	.000379	4.4e-06	-.000102	.00011	.001121	-.000668	.002911
7	5.0e-06	-.000113	.000123	1.4e-06	-.000033	.000036	.000359	-.000338	.001057
8	1.6e-06	-.000037	.00004	4.5e-07	-.000011	.000012	.000115	-.000149	.000379
9	5.1e-07	-.000012	.000013	1.4e-07	-3.5e-06	3.7e-06	.000037	-.000061	.000134
10	1.6e-07	-3.8e-06	4.2e-06	4.6e-08	-1.1e-06	1.2e-06	.000012	-.000024	.000047

step	(4) oirf	(4) Lower	(4) Upper
0	.845198	.79276	.897635
1	.115111	.031832	.19839
2	.030788	-.010417	.071992
3	.009626	-.007428	.02668
4	.003075	-.003449	.009598
5	.000985	-.001439	.003409
6	.000315	-.000569	.0012
7	.000101	-.000217	.000419
8	.000032	-.000081	.000146
9	.00001	-.00003	.00005
10	3.3e-06	-.000011	.000017

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

step	(1) coirf	(1) Lower	(1) Upper	(2) coirf	(2) Lower	(2) Upper	(3) coirf	(3) Lower	(3) Upper
0	.992584	.931002	1.05416	-.518213	-.59904	-.437386	0	0	0
1	1.03549	.927154	1.14382	-.537353	-.660432	-.414273	.291583	.206566	.3766
2	1.03847	.919453	1.15748	-.537736	-.671507	-.403965	.396416	.277965	.514866
3	1.039	.916189	1.16181	-.537634	-.674053	-.401215	.43045	.296941	.563959
4	1.03915	.914997	1.16331	-.537592	-.674819	-.400366	.441372	.301514	.58123
5	1.0392	.914596	1.16381	-.537579	-.675064	-.400094	.444872	.302486	.587258
6	1.03922	.914464	1.16397	-.537574	-.675143	-.400006	.445993	.302645	.589342
7	1.03922	.914422	1.16402	-.537573	-.675169	-.399977	.446353	.30265	.590056
8	1.03922	.914408	1.16404	-.537573	-.675177	-.399968	.446468	.302637	.590299
9	1.03922	.914403	1.16404	-.537572	-.675179	-.399965	.446505	.302628	.590381
10	1.03922	.914402	1.16405	-.537572	-.67518	-.399964	.446516	.302624	.590409

step	(4) coirf	(4) Lower	(4) Upper
0	.845198	.79276	.897635
1	.960308	.858161	1.06246
2	.991096	.857561	1.12463
3	1.00072	.853612	1.14783
4	1.0038	.85144	1.15615
5	1.00478	.850485	1.15908
6	1.0051	.850099	1.16009
7	1.0052	.849951	1.16045
8	1.00523	.849895	1.16057
9	1.00524	.849875	1.16061
10	1.00524	.849868	1.16062

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

step	(1) fevd	(1) Lower	(1) Upper	(2) fevd	(2) Lower	(2) Upper	(3) fevd	(3) Lower	(3) Upper
0	0	0	0	0	0	0	0	0	0
1	1	1	1	.273216	.206552	.339879	0	0	0
2	.920696	.877113	.964279	.26985	.203302	.336398	.079304	.035721	.122887
3	.911354	.86319	.959519	.269594	.203026	.336161	.088646	.040481	.13681
4	.910381	.861653	.959109	.269569	.202997	.33614	.089619	.040891	.138347
5	.910281	.861488	.959074	.269566	.202994	.336138	.089719	.040926	.138512
6	.91027	.86147	.959071	.269566	.202994	.336138	.08973	.040929	.13853
7	.910269	.861468	.959071	.269566	.202994	.336138	.089731	.040929	.138532
8	.910269	.861468	.959071	.269566	.202994	.336138	.089731	.040929	.138532
9	.910269	.861468	.959071	.269566	.202994	.336138	.089731	.040929	.138532
10	.910269	.861468	.959071	.269566	.202994	.336138	.089731	.040929	.138532

step	(4) fevd	(4) Lower	(4) Upper
0	0	0	0
1	.726784	.660121	.793448
2	.73015	.663602	.796698
3	.730406	.663839	.796974
4	.730431	.66386	.797003
5	.730434	.663862	.797006
6	.730434	.663862	.797006
7	.730434	.663862	.797006
8	.730434	.663862	.797006
9	.730434	.663862	.797006
10	.730434	.663862	.797006

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

.

Yes, because order is matter in matrix multiplication.