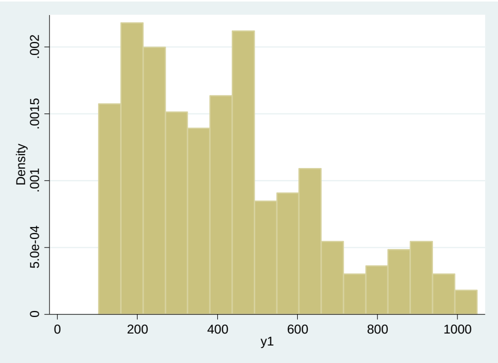
Mai Surintraboon
6204640178

Assignment 11: Truncated Regression & Tobit Model

1. Plot histogram of y1i, y2i, y3i, compute descriptive statistics of these three variables, then determine limitation of these three dependent variables.

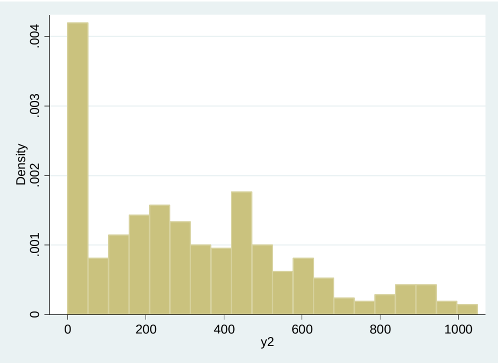
. histogram y1

(bin=17, start=103.46626, width=55.638129)



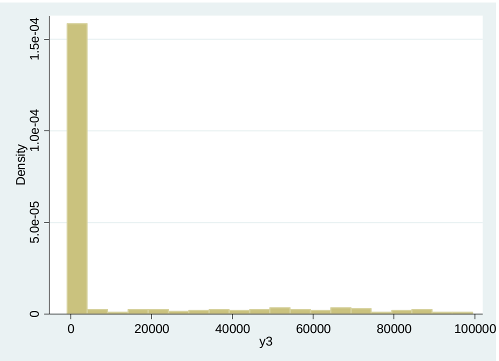
. histogram y2

(bin=20, start=0, width=52.465723)



. histogram y3

(bin=20, start=-866.33484, width=5018.7203)



. summarize y1 y2 y3

Variable	Obs	Mean	Std. Dev.	Min	Max
y1	297	423.1683	227.8976	103.4663	1049.314

```
          y2 |        400    317.5914    266.3419          0   1049.314

          y3 |        400     10446.1    23186.57  -866.3348   99508.07
```

The limitation in y1 is that 103 observations cannot be observed. y1 is truncated.

The limitation in y2 is that even all observations can be observed, some observations of y2 are censored at the value of 0.

The limitation in y3 is that they contain many extreme values(outliers), which make the mean very large. The distribution is condense in one area at the left hand side.

2. Estimate the model (1) for y1i, y2i, y3i using OLS, using truncated regression model, Tobit model, determine the most appropriated models for each yki

. *OLS

. reg y1 x

```
      Source |       SS           df       MS      Number of obs   =       297
-------------+----------------------------------   F(1, 295)       =     30.30
       Model |  1431905.59         1  1431905.59   Prob > F        =    0.0000
    Residual |  13941534.7       295  47259.4396   R-squared       =    0.0931
-------------+----------------------------------   Adj R-squared   =    0.0901
       Total |  15373440.3       296  51937.2982   Root MSE        =    217.39

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   123.1693   22.37637     5.50   0.000     79.13176    167.2069
       _cons |   69.01677   65.56422     1.05   0.293    -60.01611    198.0496
------------------------------------------------------------------------------
```

. est store m_y1_ols

. reg y2 x

```
      Source |       SS           df       MS      Number of obs   =       400
-------------+----------------------------------   F(1, 398)       =     76.58
       Model |  4567087.34         1  4567087.34   Prob > F        =    0.0000
    Residual |  23737174.2       398  59641.1413   R-squared       =    0.1614
-------------+----------------------------------   Adj R-squared   =    0.1592
       Total |  28304261.6       399  70937.9989   Root MSE        =    244.22
```

```
------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   179.8325   20.55046     8.75   0.000     139.4315    220.2335
       _cons |  -179.3679   58.08821    -3.09   0.002     -293.566   -65.1698
------------------------------------------------------------------------------

. est store m_y2_ols
. reg y3 x

      Source |       SS           df       MS      Number of obs   =       400
-------------+----------------------------------   F(1, 398)       =     26.43
       Model |  1.3359e+10         1  1.3359e+10   Prob > F        =    0.0000
    Residual |  2.0115e+11       398   505402183   R-squared       =    0.0623
-------------+----------------------------------   Adj R-squared   =    0.0599
       Total |  2.1451e+11       399   537616802   Root MSE        =     22481

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   9726.042   1891.765     5.14   0.000     6006.941    13445.14
       _cons |  -16431.39   5347.288    -3.07   0.002    -26943.85   -5918.932
------------------------------------------------------------------------------
. est store m_y3_ols
. *Truncated
. truncreg y1 x, ll(0) nolog
(note: 0 obs. truncated)

Truncated regression
Limit:   lower =          0                     Number of obs     =        297
         upper =       +inf                     Wald chi2(1)      =      28.37
Log likelihood = -2005.4436                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |    156.374   29.35911     5.33   0.000     98.83124    213.9168
```

Mai Surintraboon
6204640178

```
       _cons |  -58.70702   89.47948    -0.66   0.512    -234.0836    116.6695
-------------+----------------------------------------------------------------
      /sigma |   243.4921   13.69212    17.78   0.000      216.656    270.3282
------------------------------------------------------------------------------
. est store m_y1_trunc
. truncreg y2 x, ll(0) nolog
(note: 72 obs. truncated)

Truncated regression
Limit:   lower =          0                     Number of obs     =        328
         upper =       +inf                     Wald chi2(1)      =      33.45
Log likelihood = -2218.6432                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   215.6174   37.28024     5.78   0.000     142.5495    288.6853
       _cons |  -309.4683   118.8119    -2.60   0.009    -542.3353   -76.60125
-------------+----------------------------------------------------------------
      /sigma |   285.8084   18.56435    15.40   0.000     249.4229    322.1939
------------------------------------------------------------------------------
. est store m_y2_trunc
. truncreg y3 x, ll(0) nolog
(note: 72 obs. truncated)
cannot compute an improvement -- discontinuous region encountered
r(430);

. *Test the appropriateness of the truncated model
. lrtest m_y1_ols m_y1_trunc, force

Likelihood-ratio test                                 LR chi2(1)  =     26.69
(Assumption: m_y1_ols nested in m_y1_trunc)           Prob > chi2 =    0.0000

. lrtest m_y2_ols m_y2_trunc, force

Likelihood-ratio test                                 LR chi2(1)  =   1094.30
(Assumption: m_y2_ols nested in m_y2_trunc)           Prob > chi2 =    0.0000
```

Mai Surintraboon
6204640178

```
. *Heckit Model

. g s=(y1!=.)

. heckman y1 x, select(s=x) twostep

note: two-step estimate of rho = 1.3338082 is being truncated to 1


Heckman selection model -- two-step estimates   Number of obs     =        400

(regression model with sample selection)        Censored obs      =        103

                                                Uncensored obs    =        297


                                                Wald chi2(1)      =       1.76

                                                Prob > chi2       =     0.1848


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

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

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

y1           |

           x |   320.4132   241.6413     1.33   0.185    -153.1951    794.0215

       _cons |  -724.8884   955.5554    -0.76   0.448    -2597.743    1147.966

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

s            |

           x |    .796607   .1279218     6.23   0.000     .5458848    1.047329

       _cons |  -1.481547   .3442645    -4.30   0.000    -2.156293   -.8068005

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

mills        |

      lambda |   578.2076   675.5096     0.86   0.392    -745.7668    1902.182

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

         rho |    1.00000

       sigma |   578.2076

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


. *Tobit

. tobit y1 x, ll(0)

Tobit regression                                Number of obs     =        297

                                                LR chi2(1)        =      29.04

                                                Prob > chi2       =     0.0000

Log likelihood = -2018.7874                     Pseudo R2         =     0.0071
```

Mai Surintraboon
6204640178

```
------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   123.1693   22.30089     5.52   0.000     79.28092    167.0577
       _cons |   69.01677   65.34305     1.06   0.292    -59.57905    197.6126
-------------+----------------------------------------------------------------
      /sigma |   216.6592    8.88962                      199.1643     234.154
------------------------------------------------------------------------------
             0  left-censored observations
           297     uncensored observations
             0 right-censored observations

. est store m_y1_censor
. tobit y2 x, ll(0)
Tobit regression                                Number of obs     =        400
                                                LR chi2(1)        =      73.17
                                                Prob > chi2       =     0.0000
Log likelihood = -2389.3077                     Pseudo R2         =     0.0151

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   219.1562   24.86468     8.81   0.000     170.2741    268.0383
       _cons |  -318.6265   70.93087    -4.49   0.000    -458.0714   -179.1816
-------------+----------------------------------------------------------------
      /sigma |   284.9073   11.56717                      262.1671    307.6475
------------------------------------------------------------------------------
            72  left-censored observations at y2 <= 0
           328     uncensored observations
             0 right-censored observations

. est store m_y2_censor
. tobit y3 x, ll(0)
Tobit regression                                Number of obs     =        400
                                                LR chi2(1)        =      39.23
                                                Prob > chi2       =     0.0000
Log likelihood = -3842.6041                     Pseudo R2         =     0.0051
```

```
------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   14101.09   2236.663     6.30   0.000     9703.972    18498.21
       _cons |  -32047.13   6399.513    -5.01   0.000    -44628.11   -19466.15
-------------+----------------------------------------------------------------
      /sigma |    25293.9   1000.184                      23327.61    27260.18
------------------------------------------------------------------------------
            72  left-censored observations at y3 <= 0
           328     uncensored observations
             0 right-censored observations

. est store m_y3_censor

. *Test the appropriateness of the tobit model

. lrtest m_y1_ols m_y1_censor, force
Likelihood-ratio test                                 LR chi2(1)  =     -0.00
(Assumption: m_y1_ols nested in m_y1_censor)          Prob > chi2 =    1.0000

. lrtest m_y2_ols m_y2_censor, force
Likelihood-ratio test                                 LR chi2(1)  =    752.97
(Assumption: m_y2_ols nested in m_y2_censor)          Prob > chi2 =    0.0000

. lrtest m_y3_ols m_y3_censor, force
Likelihood-ratio test                                 LR chi2(1)  =   1464.28
(Assumption: m_y3_ols nested in m_y3_censor)          Prob > chi2 =    0.0000

. est table m_y1_ols m_y1_trunc m_y1_trunc2 m_y1_censor, star(.1 .05 .01) stat(N
N_cens rss ll rms
> e F chi2 r2 r2_a)
------------------------------------------------------------------------------
    Variable |   m_y1_ols       m_y1_trunc      m_y1_trunc2     m_y1_censor
-------------+----------------------------------------------------------------
_            |
           x |  123.16931***
```

_cons	69.01677			
eq1				
x		156.37404***		
_cons		-58.707017		
sigma				
_cons		243.49209***		216.65917***
y1				
x			320.41319	
_cons			-724.88837	
s				
x			.79660697***	
_cons			-1.4815466***	
mills				
lambda			578.2076	
model				
x				123.16931***
_cons				69.01677
Statistics				
N	297	297	400	297
N_cens			103	
rss	13941535			
ll	-2018.7874	-2005.4436		-2018.7874
rmse	217.39236			
F	30.298827			
chi2		28.368968	1.7582406	29.037355
r2	.09314152			
r2_a	.09006742			

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

Mai Surintraboon
6204640178

```
. est table m_y2_ols m_y2_trunc m_y2_trunc2 m_y2_censor, star(.1 .05 .01) stat(N
N_cens rss ll rms
> e F chi2 r2 r2_a)

------------------------------------------------------------------------------
    Variable |   m_y2_ols       m_y2_trunc      m_y2_trunc2     m_y2_censor
-------------+----------------------------------------------------------------
_            |
           x |  179.83253***
       _cons | -179.36788***
-------------+----------------------------------------------------------------
eq1          |
           x |                   215.6174***
       _cons |                 -309.46827***
-------------+----------------------------------------------------------------
sigma        |
       _cons |                   285.8084***                    284.90726***
-------------+----------------------------------------------------------------
y2           |
           x |                                  320.41319
       _cons |                                 -724.88837
-------------+----------------------------------------------------------------
s            |
           x |                                  .79660697***
       _cons |                                 -1.4815466***
-------------+----------------------------------------------------------------
mills        |
      lambda |                                   578.2076
-------------+----------------------------------------------------------------
model        |
           x |                                                   219.1562***
       _cons |                                                  -318.6265***
-------------+----------------------------------------------------------------
Statistics   |
           N |        400             328             400             400
      N_cens |                                        103
         rss |   23737174
```

```
          ll | -2765.7931      -2218.6432                        -2389.3077

        rmse |  244.21536

           F |  76.576122

        chi2 |                  33.451092       1.7582406       73.173596

          r2 |  .16135688

        r2_a |  .15924974

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

```
. est table m_y3_ols m_y3_censor, star(.1 .05 .01) stat(N N_cens rss ll rmse F chi2
r2 r2_a)


----------------------------------------------
    Variable |   m_y3_ols       m_y3_censor
-------------+--------------------------------
_            |
           x |  9726.0419***
       _cons | -16431.393***
-------------+--------------------------------
model        |
           x |                  14101.089***
       _cons |                 -32047.129***
-------------+--------------------------------
sigma        |
       _cons |                  25293.896***
-------------+--------------------------------
Statistics   |
           N |        400             400
      N_cens |
         rss |  2.012e+11
          ll | -4574.7459      -3842.6041
        rmse |  22481.152
           F |  26.432484
        chi2 |                  39.233422
          r2 |  .06227724
        r2_a |  .05992115
---------------------------------------------- legend: * p<.1; ** p<.05; *** p<.01
```

Mai Surintraboon
6204640178

According to the irtest of the truncated regression model, it shows that model y1 and y2 are better in using the truncated regression than the normal least square regression as the null hypothesis can be rejected.

According to the test of the significant of lambda in the heckit model, it shows that none of the model should employ the heckit model.

According to the irtest of the tobit model, it shows that model y2 and y3 are better in using the tobit model than the normal least square regression as the null hypothesis can be rejected.

I think the most appropriate model for y1 is the truncated model and for y2 and y3 is the tobit model.

Though y2 passes both truncated and tobit model appropriateness test, it contains more observation in the tobit model (400 in tobit and 328 in truncated).