Assignment 12
Poisson – Negative Binomial – Zero Inflated Poisson

The model

The generalized linear regression model can be stated as:

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

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.

Requirements (using Data set file – `assign12.dta`):

1 Estimate models for y_i assuming that the model is traditional linear regression model. Interpret your estimated result. (3)
2 Create histogram for y_i. Determine whether there is limitation of dependent variable in this case. If yes, what type of limitation is it? (4)
3 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)-(8)
4 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 effects).
5 Estimate models for y_i assuming that the model is Zero Inflated Poisson (x_{1i}, x_{2i}, and x_{3j} 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)
6 According to the above (1-5), determine the most appropriated model for this case. Give explanation why? ZIP ->

By Tatre Jantarakolica

```
      name:  <unnamed>
       log:  C:\Users\Kang\Documents\ASS12.smcl
  log type:  smcl
 opened on:  22 Apr 2021, 01:08:09

. use "C:\Users\Kang\Downloads\assign12 (1).dta", clear

. reg y x1 x2 x3 x4

      Source |       SS           df       MS      Number of obs   =       232
-------------+----------------------------------   F(4, 227)       =      5.96
       Model |  44.7298499         4  11.1824625   Prob > F        =    0.0001
    Residual |  425.748598       227  1.87554449   R-squared       =    0.0951
-------------+----------------------------------   Adj R-squared   =    0.0791
       Total |  470.478448       231  2.03670324   Root MSE        =    1.3695

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .1016201   .0435073     2.34   0.020     .0158904    .1873499
          x2 |   .1345044   .0462142     2.91   0.004     .0434407     .225568
          x3 |  -.0748194   .0480457    -1.56   0.121    -.1694919    .0198531
          x4 |   .1684563   .0688243     2.45   0.015     .0328401    .3040725
       _cons |   .9568064    .107007     8.94   0.000     .7459523     1.16766
------------------------------------------------------------------------------
```

(2)

```
. histogram y
(bin=15, start=0, width=.33333333)
```

The limitation is the highly skewness which violate normality assumption

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

(3)

```
Poisson regression                              Number of obs     =        232
                                                LR chi2(4)        =      43.33
                                                Prob > chi2       =     0.0000
Log likelihood = -342.88107                     Pseudo R2         =     0.0594

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0971474   .0306762     3.17   0.002     .0370231    .1572717
          x2 |   .1293024   .0330916     3.91   0.000      .064444    .1941607
          x3 |  -.0715533   .0342177    -2.09   0.037    -.1386187   -.0044879
          x4 |   .1734482   .0507707     3.42   0.001     .0739395    .2729569
       _cons |  -.1284876   .0849064    -1.51   0.130    -.294901     .0379259
------------------------------------------------------------------------------

. estat gof

         Deviance goodness-of-fit =  409.4921
         Prob > chi2(227)         =    0.0000

         Pearson goodness-of-fit  =  423.3541
         Prob > chi2(227)         =    0.0000
```

reject the null hypothesis which implies that it is not poisson distributed.

. poisson y x1 x2 x3 x4, ir nolog

Poisson regression

Log likelihood = **-342.88107**

Number of obs	=	**232**
LR chi2(**4**)	=	**43.33**
Prob > chi2	=	**0.0000**
Pseudo R2	=	**0.0594**

y	IRR	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	1.102023	.0338059	3.17	0.002	1.037717	1.170314
x2	1.138034	.0376594	3.91	0.000	1.066566	1.214291
x3	.9309467	.0318548	-2.09	0.037	.8705599	.9955222
x4	1.189399	.0603866	3.42	0.001	1.076742	1.313844
_cons	.8794245	.0746687	-1.51	0.130	.7446053	1.038654

. mfx

Marginal effects after poisson
y = Predicted number of events (predict)
= **.95621703**

variable	dy/dx	Std. Err.	z	P>|z|	[95%	C.I.]	X
x1	.092894	.02882	3.22	0.001	.036403	.149385	-.317697
x2	.1236411	.0308	4.01	0.000	.063277	.184005	.812709
x3	-.0684205	.03246	-2.11	0.035	-.13204	-.004801	-.818103
x4	.1658541	.04736	3.50	0.000	.073022	.258686	-.28275

4

. nbreg y x1 x2 x3 x4, nolog

Negative binomial regression

Dispersion = **mean**
Log likelihood = **-317.49278**

Number of obs	=	**232**
LR chi2(**4**)	=	**21.24**
Prob > chi2	=	**0.0003**
Pseudo R2	=	**0.0324**

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	.1285534	.0506934	2.54	0.011	.0291962	.2279106
x2	.151011	.0506477	2.98	0.003	.0517434	.2502785
x3	-.0672859	.0481376	-1.40	0.162	-.1616339	.0270621
x4	.1726312	.0707035	2.44	0.015	.034055	.3112075
_cons	-.1435596	.1177204	-1.22	0.223	-.3742874	.0871682
/lnalpha	.0479945	.2389531			-.4203449	.5163339
alpha	1.049165	.2507012			.6568202	1.675872

Likelihood-ratio test of alpha=0: chibar2(01) = **50.78** Prob>=chibar2 = **0.000**

Overdispersion Test

: H_0 : $\alpha = 0$

reject the null hypothesis which implies that it is negative binomial model

However, the individuals test P-value, some are not significant which means that it is not negative binomial model.

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

Negative binomial regression                      Number of obs     =        232
                                                  LR chi2(4)        =      21.24
Dispersion     = mean                             Prob > chi2       =     0.0003
Log likelihood = -317.49278                       Pseudo R2         =     0.0324

------------------------------------------------------------------------------
           y |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.137182   .0576476     2.54   0.011     1.029627    1.255973
          x2 |   1.163009   .0589037     2.98   0.003     1.053105    1.284383
          x3 |   .9349279   .0450052    -1.40   0.162     .8507526    1.027432
          x4 |   1.188428    .084026     2.44   0.015     1.034641    1.365072
       _cons |   .8662692   .1019776    -1.22   0.223     .6877792     1.09108
-------------+----------------------------------------------------------------
    /lnalpha |   .0479945   .2389531                     -.4203449    .5163339
-------------+----------------------------------------------------------------
       alpha |   1.049165   .2507012                      .6568202    1.675872
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =   50.78 Prob>=chibar2 = 0.000

. mfx

Marginal effects after nbreg
      y  = Predicted number of events (predict)
         =  .94607122
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .1216207      .04796    2.54   0.011    .02763  .215611  -.317697
      x2 |   .1428671      .04796    2.98   0.003   .048862  .236872   .812709
      x3 |  -.0636573      .04556   -1.40   0.162  -.152956  .025642  -.818103
      x4 |   .1633214      .06686    2.44   0.015   .032269  .294374   -.28275
------------------------------------------------------------------------------
```

(5)

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

Fitting constant-only model:

Iteration 0:   log likelihood = -355.48956
Iteration 1:   log likelihood =  -321.8304
Iteration 2:   log likelihood = -317.80147
Iteration 3:   log likelihood = -317.79274
Iteration 4:   log likelihood = -317.79274

Fitting full model:

Iteration 0:   log likelihood = -317.79274
Iteration 1:   log likelihood = -311.40478
Iteration 2:   log likelihood = -311.22574
Iteration 3:   log likelihood = -311.22548
Iteration 4:   log likelihood = -311.22548

Zero-inflated Poisson regression                  Number of obs     =        232
                                                  Nonzero obs       =        106
                                                  Zero obs          =        126

Inflation model  = logit                          LR chi2(4)        =      13.13
Log likelihood   = -311.2255                      Prob > chi2       =     0.0106
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
y						
x1	.0962618	.0402556	2.39	0.017	.0173622	.1751613
x2	.0954378	.0373253	2.56	0.011	.0222815	.168594
x3	-.0672218	.0358764	-1.87	0.061	-.1375383	.0030946
x4	.0877817	.0527928	1.66	0.096	-.0156904	.1912537
_cons	.4349837	.1120917	3.88	0.000	.2152881	.6546793
inflate						
x4	-.2232387	.1282354	-1.74	0.082	-.4745754	.028098
_cons	-.3696851	.1945938	-1.90	0.057	-.7510819	.0117117

Vuong test of zip vs. standard Poisson: z = **3.71** Pr>z = **0.0001**

It reject the null hypothesis. It implies that zero-inflated poisson model.

. mfx

Marginal effects after zip
y = Predicted number of events (predict)
= **.96133901**

variable	dy/dx	Std. Err.	z	P>|z|	[95%	C.I.]	X
x1	.0925402	.03814	2.43	0.015	.017789	.167291	-.317697
x2	.0917481	.03511	2.61	0.009	.022937	.160559	.812709
x3	-.064623	.03433	-1.88	0.060	-.131909	.002663	-.818103
x4	.1753718	.06213	2.82	0.005	.053599	.297145	-.28275

. log close
name: **<unnamed>**
log: **C:\Users\Kang\Documents\ASS12.smcl**
log type: **smcl**
closed on: **22 Apr 2021, 01:50:41**

(6) Zero-inflated proisson model is the most appropriated model as it has significant z-test on all variables while others model fail to meet evalution.

(2)