

## Assignment 12 Poisson – Negative Binomial – Zero Inflated Poisson

### The model

The generalized linear regression model can be stated as:

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

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  $k$ .

$f(\cdot)$  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.

```
. reg y x*
```

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

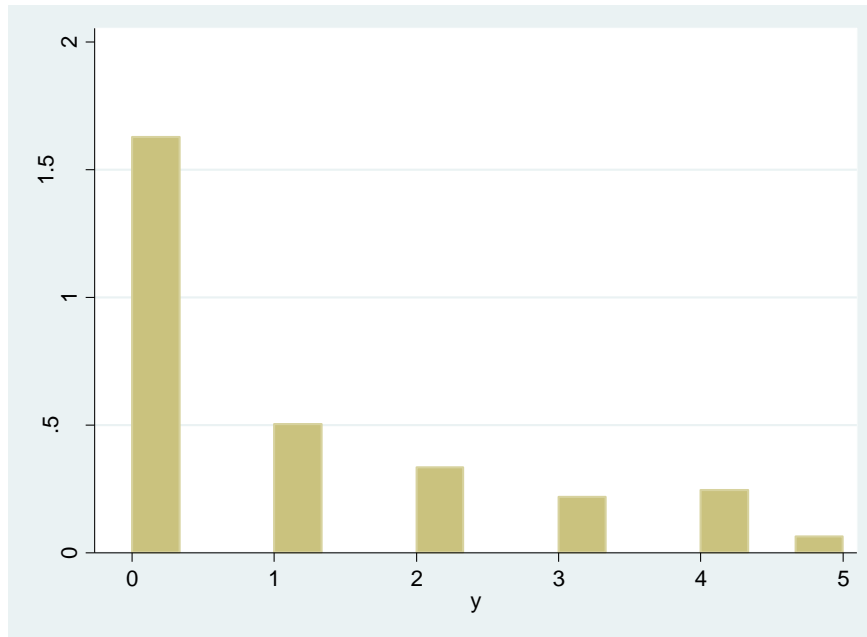
*According to the estimated result, all of the independent variables are positively affect the dependent variable except x3 which has a negative effect.*

*From the overall test, it suggests that all independent variables are jointly significant.*

*R-squared is quite low; therefore, there are rooms for improvements. Individual test suggests that each independent variable can significantly explain the dependent variable except x3.*

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?

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



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

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).

```
. poisson y x*
Iteration 0: log likelihood = -342.88107
Iteration 1: log likelihood = -342.88107
```

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
-----

```

. poisson y x\*, ir

Iteration 0: log likelihood = -342.88107

Iteration 1: log likelihood = -342.88107

```

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

```

```

-----
. estat gof

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

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

```

From the estimated result, it suggests that each independent variables has a positive effect on the dependent variable (Marginal effect > 0 and IRR > 1) **except x3** that negatively affects the dependent variable (Marginal effect < 0 and IRR < 1).

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

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

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

**However**, from the goodness of fit test, the null hypothesis is rejected (P-value < 0.05). It means that the data are not poisson distributed. So, **the estimated result provided above cannot be used.**

- 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).

```
. nbreg y x* 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 |      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

. nbreg y x\*, 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 = 0.00

. 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
-----
    
```

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

*From the estimated result, it suggests that each independent variables has a positive effect on the dependent variable (Marginal effect > 0 and IRR > 1) except x3 that negatively affects the dependent variable (Marginal effect < 0 and IRR < 1).*

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

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

*The individual test reveals that each of the independent variables is significant except x3.*

5 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

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

```

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

Inflation model = logit                LR chi2(3)      =      10.35
Log likelihood = -312.6158              Prob > chi2     =      0.0158
    
```

```

-----
          y |      Coef.   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
y          |
    x1 |   .0805446   .0398159     2.02   0.043    .0025068   .1585824
    x2 |   .0857883   .0372107     2.31   0.021    .0128567   .1587199
    x3 |  -.0672468   .0357098    -1.88   0.060   -0.1372367   .002743
    _cons |   .4589728   .1106031     4.15   0.000    .2421947   .6757508
-----+-----
inflate    |
    x4 |  -.2738532   .1212311    -2.26   0.024   -0.5114618  -.0362446
    _cons |  -.3379298   .1908217    -1.77   0.077   -0.7119334   .0360738
-----

```

Vuong test of zip vs. standard Poisson:                    z =        3.92   Pr>z = 0.0000

. mfx

Marginal effects after zip

y = Predicted number of events (predict)  
 = .9868501

```

-----
variable |      dy/dx   Std. Err.      z    P>|z|    [ 95% C.I. ]      X
-----+-----
    x1 |   .0794855   .03893     2.04   0.041    .003191   .15578  -.317697
    x2 |   .0846602   .03606     2.35   0.019    .013992   .155329  .812709
    x3 |  -.0663626   .03507    -1.89   0.058   -0.135102  .002377  -.818103
    x4 |   .1176249   .05326     2.21   0.027    .01323   .222019  -.28275
-----

```

*The Vuong test rejects the null hypothesis meaning that the Zero-Inflated Poisson Model is more appropriated compared with the Poisson Model.*

*From the estimated result, it suggests that each independent variables has a positive effect on the dependent variable (Marginal effect > 0) **except** x3 that negatively affects the dependent variable (Marginal effect < 0).*

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

*The individual test reveals that each of the independent variables is significant*

- 6 According to the above (1-5), determine the most appropriated model for this case.  
Give explanation why?

*Firstly, it cannot be the Poisson Model since the Goodness of Fit Test rejects the null hypothesis (implying that the Poisson is unappropriated). Next, the Negative Binomial Model seems to be good because the Overdispersion also confirms that the Poisson Model is unappropriated; however, there exists an insignificant variable in this model. Lastly, look at the Zero-Inflated Poisson Model, the Vuong test shows that this model is appropriated; moreover, every single independent variables is significant (unlike in the Negative Binomial Model)*

*In conclusion, the Zero-Inflated Poisson Model is the most appropriated model because of the reason provided above.*