Sep 15, intro econometrics

Count Model: The Poisson Regression Model

- Name : Jikhan Jeong
- Ref: Using R for Introductory Econometrics
  1)**chunk n: Ctrl + Alt + I**
  2)**knit: Ctrl + Shift + k**
  3)**run: Ctrl + Enter**

Count data: the poisson regression model $ p(y=h|x)= , h{ 0,1,2,… } $

The conditional mean of y is $E(y|x) = e^{x\beta}$

The coefficient has the interpretation of a semi elasticity $\beta_j = \frac{1}{E(y|x)} \cdot \frac{\partial E(y|x)}{\partial x_j}$ if β_j increases by 1 unit, $E(y|x)$ will inrease roughly by 100·β_j

```
data(crime1, package='wooldridge')
class(crime1)
```

```
## [1] "data.frame"
```

```
colnames(crime1)
```

```
##  [1] "narr86"  "nfarr86" "nparr86" "pcnv"    "avgsen"  "tottime" "ptime86"
##  [8] "qemp86"  "inc86"   "durat"   "black"   "hispan"  "born60"  "pcnvsq"
## [15] "pt86sq"  "inc86sq"
```

```
summary(crime1)
```

```
##      narr86          nfarr86          nparr86           pcnv
##  Min.   : 0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median : 0.0000   Median :0.0000   Median :0.0000   Median :0.2500
##  Mean   : 0.4044   Mean   :0.2334   Mean   :0.1255   Mean   :0.3578
##  3rd Qu.: 1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.6700
##  Max.   :12.0000   Max.   :6.0000   Max.   :8.0000   Max.   :1.0000
##      avgsen          tottime          ptime86          qemp86
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000   Min.   :0.000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.:1.000
##  Median : 0.0000   Median : 0.0000   Median : 0.0000   Median :3.000
##  Mean   : 0.6323   Mean   : 0.8387   Mean   : 0.3872   Mean   :2.309
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.:4.000
##  Max.   :59.2000   Max.   :63.4000   Max.   :12.0000   Max.   :4.000
##      inc86            durat            black            hispan
##  Min.   :  0.00   Min.   : 0.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:  0.40   1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median : 29.00   Median : 0.000   Median :0.0000   Median :0.0000
##  Mean   : 54.97   Mean   : 2.251   Mean   :0.1611   Mean   :0.2176
##  3rd Qu.: 90.10   3rd Qu.: 2.000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :541.00   Max.   :25.000   Max.   :1.0000   Max.   :1.0000
##      born60          pcnvsq          pt86sq           inc86sq
##  Min.   :0.0000   Min.   :0.0000   Min.   :  0.000   Min.   :     0.00
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:  0.000   1st Qu.:     0.16
##  Median :0.0000   Median :0.0625   Median :  0.000   Median :   841.00
##  Mean   :0.3626   Mean   :0.2841   Mean   :  3.951   Mean   :  7458.93
##  3rd Qu.:1.0000   3rd Qu.:0.4489   3rd Qu.:  0.000   3rd Qu.:  8118.01
##  Max.   :1.0000   Max.   :1.0000   Max.   :144.000   Max.   :292681.00
```

Linear Model

```
lm.res      <-  lm(narr86~pcnv+avgsen+tottime+ptime86+qemp86+inc86+
                   black+hispan+born60, data=crime1)
summary(lm.res)
```

```
##
## Call:
## lm(formula = narr86 ~ pcnv + avgsen + tottime + ptime86 + qemp86 +
##     inc86 + black + hispan + born60, data = crime1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0210 -0.4544 -0.2389  0.2686 11.5207
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.576566   0.037894  15.215  < 2e-16 ***
## pcnv        -0.131886   0.040404  -3.264 0.001111 **
## avgsen      -0.011332   0.012241  -0.926 0.354693
## tottime      0.012069   0.009436   1.279 0.201003
## ptime86     -0.040874   0.008813  -4.638 3.69e-06 ***
## qemp86      -0.051310   0.014486  -3.542 0.000404 ***
## inc86       -0.001462   0.000343  -4.261 2.10e-05 ***
## black        0.327010   0.045426   7.199 7.83e-13 ***
## hispan       0.193809   0.039716   4.880 1.12e-06 ***
## born60      -0.022465   0.033294  -0.675 0.499902
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8287 on 2715 degrees of freedom
## Multiple R-squared:  0.07248,    Adjusted R-squared:  0.0694
## F-statistic: 23.57 on 9 and 2715 DF,  p-value: < 2.2e-16
```

Estimate Poisson model

```
Poisson.res <- glm(narr86~pcnv+avgsen+tottime+ptime86+qemp86+inc86+
                   black+hispan+born60, data=crime1, family=poisson)
summary(Poisson.res)
```

```
##
## Call:
## glm(formula = narr86 ~ pcnv + avgsen + tottime + ptime86 + qemp86 +
##     inc86 + black + hispan + born60, family = poisson, data = crime1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5731  -0.9076  -0.6651   0.2183   7.4609
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.599589   0.067250  -8.916  < 2e-16 ***
## pcnv        -0.401571   0.084971  -4.726 2.29e-06 ***
## avgsen      -0.023772   0.019946  -1.192   0.2333
## tottime      0.024490   0.014750   1.660   0.0969 .
## ptime86     -0.098558   0.020695  -4.763 1.91e-06 ***
## qemp86      -0.038019   0.029024  -1.310   0.1902
## inc86       -0.008081   0.001041  -7.762 8.34e-15 ***
## black        0.660838   0.073834   8.950  < 2e-16 ***
## hispan       0.499813   0.073927   6.761 1.37e-11 ***
## born60      -0.051029   0.064052  -0.797   0.4256
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 3208.5  on 2724  degrees of freedom
## Residual deviance: 2822.2  on 2715  degrees of freedom
## AIC: 4517.5
##
## Number of Fisher Scoring iterations: 6
```

Quasi-Poisson model - quasi-maximum likelihood estimators to adjust s.d -parameter estimates are consistent, but the s.e. and all inference based on them are invalid - similar to the heteroscedasticity-robust inference for OLS - in the results, the standard deviation is different

```
QPoisson.res<- glm(narr86~pcnv+avgsen+tottime+ptime86+qemp86+inc86+
                     black+hispan+born60, data=crime1, family=quasipoisson)
summary(QPoisson.res)
```

```
##
## Call:
## glm(formula = narr86 ~ pcnv + avgsen + tottime + ptime86 + qemp86 +
##     inc86 + black + hispan + born60, family = quasipoisson, data = crime1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5731  -0.9076  -0.6651   0.2183   7.4609
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.599589   0.082824  -7.239 5.84e-13 ***
## pcnv        -0.401571   0.104649  -3.837 0.000127 ***
## avgsen      -0.023772   0.024565  -0.968 0.333268
## tottime      0.024490   0.018166   1.348 0.177732
## ptime86     -0.098558   0.025487  -3.867 0.000113 ***
## qemp86      -0.038019   0.035746  -1.064 0.287609
## inc86       -0.008081   0.001282  -6.303 3.40e-10 ***
## black        0.660838   0.090933   7.267 4.77e-13 ***
## hispan       0.499813   0.091047   5.490 4.40e-08 ***
## born60      -0.051029   0.078885  -0.647 0.517768
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 1.516789)
##
##     Null deviance: 3208.5  on 2724  degrees of freedom
## Residual deviance: 2822.2  on 2715  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

Install stargazer - for output table in a regression stargazer (model1, model2, model3, type="text", keep.stat="n")

```
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
stargazer(lm.res, Poisson.res, QPoisson.res, type ="text", keep.stat="n")
```

```
##
## ================================================
##                         Dependent variable:
##              ----------------------------------
##                               narr86
##                  OLS     Poisson  glm: quasipoisson
##                                          link = log
##                  (1)       (2)           (3)
## ------------------------------------------------
## pcnv          -0.132*** -0.402***     -0.402***
##                (0.040)   (0.085)       (0.105)
##
## avgsen         -0.011    -0.024        -0.024
##                (0.012)   (0.020)       (0.025)
##
## tottime         0.012    0.024*         0.024
##                (0.009)   (0.015)       (0.018)
##
## ptime86       -0.041*** -0.099***     -0.099***
##                (0.009)   (0.021)       (0.025)
##
## qemp86        -0.051***  -0.038        -0.038
##                (0.014)   (0.029)       (0.036)
##
## inc86         -0.001*** -0.008***     -0.008***
##               (0.0003)   (0.001)       (0.001)
##
## black         0.327***  0.661***      0.661***
##                (0.045)   (0.074)       (0.091)
##
## hispan        0.194***  0.500***      0.500***
##                (0.040)   (0.074)       (0.091)
##
## born60         -0.022    -0.051        -0.051
##                (0.033)   (0.064)       (0.079)
##
## Constant      0.577***  -0.600***     -0.600***
##                (0.038)   (0.067)       (0.083)
##
## ------------------------------------------------
## Observations   2,725     2,725         2,725
## ================================================
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

```
stargazer(lm.res, Poisson.res, QPoisson.res, type ="latex", keep.stat="n")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fa
s.harvard.edu
## % Date and time: Sun, Sep 15, 2019 - 1:21:02 AM
## Wbegin{table}[!htbp] Wcentering
##   Wcaption{}
##   Wlabel{}
## Wbegin{tabular}{@{Wextracolsep{5pt}}lccc}
## WW[-1.8ex]Whline
## Whline WW[-1.8ex]
##  & Wmulticolumn{3}{c}{Wtextit{Dependent variable:}} WW
## Wcline{2-4}
## WW[-1.8ex] & Wmulticolumn{3}{c}{narr86} WW
## WW[-1.8ex] & Wtextit{OLS} & Wtextit{Poisson} & Wtextit{glm: quasipoisson} WW
##  & Wtextit{} & Wtextit{} & Wtextit{link = log} WW
## WW[-1.8ex] & (1) & (2) & (3)WW
## Whline WW[-1.8ex]
##  pcnv & $-$0.132$^{***}$ & $-$0.402$^{***}$ & $-$0.402$^{***}$ WW
##   & (0.040) & (0.085) & (0.105) WW
##   & & & WW
##  avgsen & $-$0.011 & $-$0.024 & $-$0.024 WW
##   & (0.012) & (0.020) & (0.025) WW
##   & & & WW
##  tottime & 0.012 & 0.024$^{*}$ & 0.024 WW
##   & (0.009) & (0.015) & (0.018) WW
##   & & & WW
##  ptime86 & $-$0.041$^{***}$ & $-$0.099$^{***}$ & $-$0.099$^{***}$ WW
##   & (0.009) & (0.021) & (0.025) WW
##   & & & WW
##  qemp86 & $-$0.051$^{***}$ & $-$0.038 & $-$0.038 WW
##   & (0.014) & (0.029) & (0.036) WW
##   & & & WW
##  inc86 & $-$0.001$^{***}$ & $-$0.008$^{***}$ & $-$0.008$^{***}$ WW
##   & (0.0003) & (0.001) & (0.001) WW
##   & & & WW
##  black & 0.327$^{***}$ & 0.661$^{***}$ & 0.661$^{***}$ WW
##   & (0.045) & (0.074) & (0.091) WW
##   & & & WW
##  hispan & 0.194$^{***}$ & 0.500$^{***}$ & 0.500$^{***}$ WW
##   & (0.040) & (0.074) & (0.091) WW
##   & & & WW
##  born60 & $-$0.022 & $-$0.051 & $-$0.051 WW
##   & (0.033) & (0.064) & (0.079) WW
##   & & & WW
##  Constant & 0.577$^{***}$ & $-$0.600$^{***}$ & $-$0.600$^{***}$ WW
##   & (0.038) & (0.067) & (0.083) WW
##   & & & WW
## Whline WW[-1.8ex]
## Observations & 2,725 & 2,725 & 2,725 WW
## Whline
## Whline WW[-1.8ex]
## Wtextit{Note:}  & Wmulticolumn{3}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} WW
## Wend{tabular}
## Wend{table}
```

Just quick check for using Python in R market down.

```
import numpy as np
import os
import pandas as pd
test = np.array([1,2,3])
print(test)
```

```
## [1 2 3]
```

```
type(test)
```

```
## <class 'numpy.ndarray'>
```

```
os.getcwd()
```

```
## 'C:WWUsersWWjikhan.jeongWWDocumentsWWRWWEcon_Modelling_R'
```

Data in python is not link to r environment. Therefore, it requires to download in the working directory and r and python interact via woring directory saved file.