Modeling Health Care Costs and Counts

Partha Deb
Hunter College and the Graduate Center, CUNY, and NBER

Willard G. Manning
University of Chicago

Edward C. Norton
University of Michigan and NBER

iHEA, Milan 2015

Overview

Statistical issues - skewness and the zero mass

Studies with skewed outcomes but no zeroes

Studies with skewed outcomes and zeroes

Studies with count data

Finite mixture models

Conclusions

Top Ten Urban Myths of Health Econometrics

Examples and Characteristics

Examples

- Number of visits to the doctor
- Number of ER visits
- Number of cigarettes smoked per day

Like expenditures / costs

- Many zeros
- Very skewed in non-zero range
- Intrinsically heteroskedastic (variance increases with mean)

Differences

- Integer valued
- Concentrated on a few low values (0, 1, 2)
- Prediction of event probabilities often of interest

Overview

Studies with count data

- **Poisson (canonical model)**
  - **Estimation**
  - **Prediction – Mean, Events**
  - **Interpretation – Marginal effects, Incremental Effects**
  - **Goodness of fit**
- **Negative Binomial**
- **Hurdle Models (Two Part Models for Counts)**
- **Zero Inflated Models**

Model Selection - Discriminating among nonnested models

Poisson

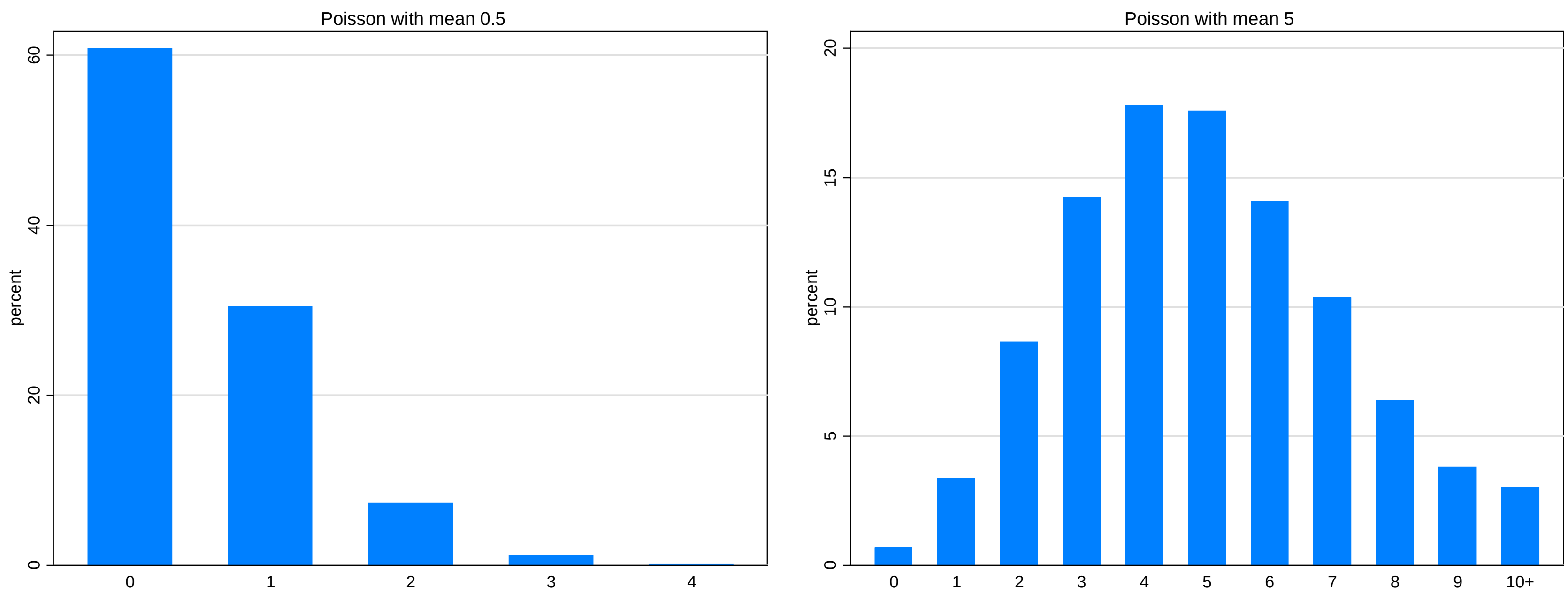
Mean $\mu = \exp(X\beta)$

Variance $\sigma^2 = \exp(X\beta)$

Density

$$\Pr(Y = y|X) = \frac{\exp(-\mu)\,\mu^y}{\Gamma(y+1)}$$

Note that $\Gamma(y+1) = y!$



Estimation

Estimation is usually conducted using Maximum Likelihood

First Order Condition for MLE

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{N} \left(y_i - \mu_i \right) X_i = 0$$

But it is very unlikely that mean = variance property of the Poisson distribution is satisfied for most health count outcomes

The Quasi MLE for a Poisson regression relaxes the mean = variance assumption
But has the same first order condition as the MLE

So

$$\widehat{\beta}_{MLE} = \widehat{\beta}_{QMLE}$$

Estimation

Poisson MLE is robust to misspecification of variance of *y*, i.e.
$$\hat{\beta}_{MLE} = \hat{\beta}_{QMLE}$$

In other words, it is okay to estimate a Poisson regression in terms of point estimates even if the dgp is not Poisson but the weaker QMLE assumptions are satisfied

But standard errors for $\hat{\beta}_{MLE}$ are not correct unless the true dgp is Poisson (mean = variance)

The sandwich form for $Cov(\hat{\beta})$ ("robust") is appropriate because it uses only the QMLE assumptions (mean need not be equal to variance)

Stata command: `poisson use_off age i.female, robust`

Prediction

The typical prediction of interest is the conditional mean.

But, in nonlinear models, predictions of quantities other than the conditional mean are often of interest.

In the context of count data, we might be interested in predictions of the distribution of the count variable

$\Pr(Y = 0 \mid X)$

$\Pr(Y = 12 \mid X)$

We might also be interested in predictions of certain events of interest

$\Pr(Y > 5 \mid X) = 1 - \Pr(Y \leq 5 \mid X)$

Substantively

Probability of exceeding a benefit cap (mental health)

Probability of a “drive-through” delivery

Prediction in Poisson

Conditional Mean: $\hat{\mu} = \exp(X\hat{\beta})$

Stata command: `predict muhat` **(default)**

Distribution and events:

$$\Pr(Y = y | X) = \frac{\exp(-\hat{\mu})\,\hat{\mu}^{y}}{\Gamma(y+1)} \qquad \forall\, y = 0, 1, 2, 3, \ldots$$

Stata commands:

```
predict prhat0, pr(0)
predict prhat12, pr(12)

predict prhat0to5, pr(0,5)
generate prhatgt5 = 1 – prhat0to5
```

Interpretation

Marginal Effects - for continuous variables

$$\frac{\partial E(y_i|X)}{\partial X^k} = \beta^k \ \mu_i$$

Examples: Income, Price, Health status

Incremental Effects - for binary variables

$$E(y_i|X, X^k=1) - E(y_i|X, X^k=0)$$
$$= \left[\mu_i | X^k=0 \right] \left[\exp(\beta^k) - 1 \right]$$

Examples: Treatment/ Control, Insurance, Gender, Race

FYI: Predictions (at specific X), Marginal & Incremental Effects

Approach depends on research question. How one does it can make a big difference

1. **Evaluate for hypothetical individuals**
   a. **Mean (or Median, other quantiles) of X in sample**
   b. **Mean (or Median, other quantiles) of X in sub-sample of interest**
   c. **Hypothetical individual of interest**

2. **Evaluate for each individual**
   a. **Average over sample**
   b. **Average over sub-samples of interest**

3. **For Incremental Effects – (Treatment vs. Control)**
   a. **Switch all individuals from control to treatment**
   b. **Switch controls to treatment**

FYI: Predictions (at specific X), Marginal & Incremental Effects

Stata command for predictions at specific values of X:

```
margins female
margins, at(age=27)
margins female, at(age=32)
```

Stata command for marginal / incremental effects:

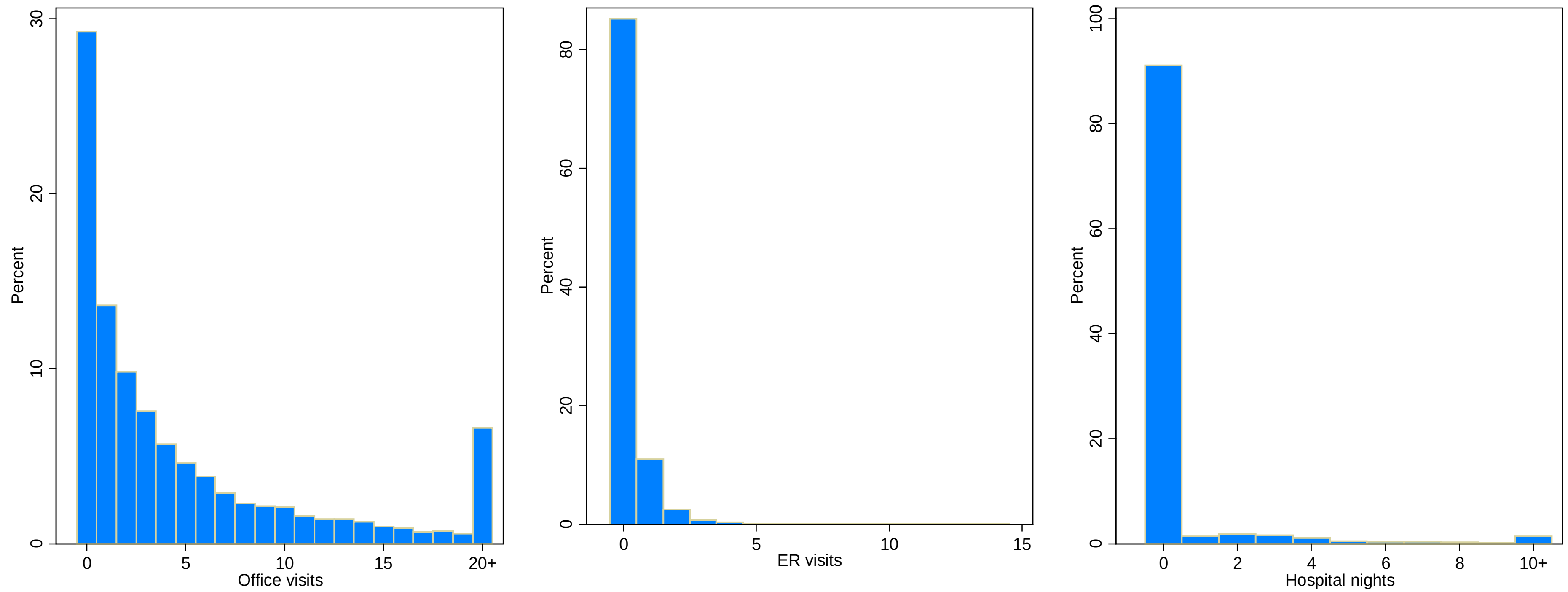
Be sure to code indicator variables using factor notation (`i.female`)

```
margins, dydx(age)
margins, dydx(*)
margins, dydx(*) at(age=27)
margins female, dydx(*) at(age=27)
```

Examples

Data from MEPS

1. Number of office-based visits
2. Number of emergency room visits
3. Number of hospital nights



Poisson Estimates

Poisson Coefficients

	Office visits	ER visits	Hospital nights
Age	0.005**	-0.018**	0.001
	(0.001)	(0.002)	(0.004)
1.female	0.328**	0.171**	-0.044
	(0.027)	(0.044)	(0.138)

* p<0.05; ** p<0.01

Predictive margins from Poisson

```
. margins female

Predictive margins                              Number of obs   =      19386
Model VCE    : Robust

Expression   : Predicted number of events, predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |
          0  |   4.737476   .1074384    44.09   0.000       4.5269    4.948051
          1  |   6.577042   .0963255    68.28   0.000     6.388248    6.765837
------------------------------------------------------------------------------
```

Predictive margins from Poisson

```
. margins, at(age=(30 50 70))

Predictive margins                              Number of obs   =      19386
Model VCE    : Robust

Expression   : Predicted number of events, predict()

1._at        : age             =          30
2._at        : age             =          50
3._at        : age             =          70
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   5.170575   .1441185    35.88   0.000     4.888108    5.453042
          2  |   5.729337   .0719968    79.58   0.000     5.588226    5.870448
          3  |   6.348482   .1421247    44.67   0.000     6.069923    6.627041
------------------------------------------------------------------------------
```

Marginal Effects from Poisson

```
. margins, dydx(age female)

Average marginal effects                          Number of obs   =      19386
Model VCE    : Robust

Expression   : Predicted number of events, predict()
dy/dx w.r.t. : age 1.female

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0297709   .0063503     4.69   0.000     .0173246    .0422171
    1.female |   1.839567   .1465079    12.56   0.000     1.552416    2.126717
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Marginal Effects from Poisson

```
. margins female, dydx(age)

Average marginal effects                          Number of obs   =      19386
Model VCE    : Robust

Expression   : Predicted number of events, predict()
dy/dx w.r.t. : age

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  age        |
      female |
          0  |    .024307   .0052005     4.67   0.000     .0141142    .0344998
          1  |   .0337455    .007205     4.68   0.000      .019624     .047867
------------------------------------------------------------------------------
```

Marginal Effects from Poisson

Poisson Marginal Effects: Office visits

	Average	Mean of X	Median of X
age	0.030**	0.022**	0.021**
	(0.002)	(0.001)	(0.001)
1.female	1.840**	1.402**	1.294**
	(0.035)	(0.027)	(0.025)

Poisson Marginal Effects: ER visits

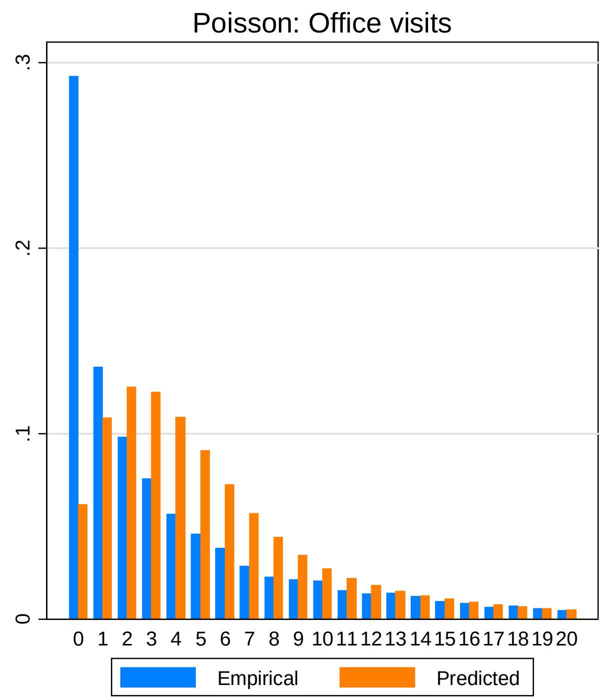
	Average	Mean of X	Median of X
age	-0.004**	-0.003**	-0.003**
	(0.000)	(0.000)	(0.000)
1.female	0.036**	0.029**	0.028**
	(0.007)	(0.005)	(0.005)

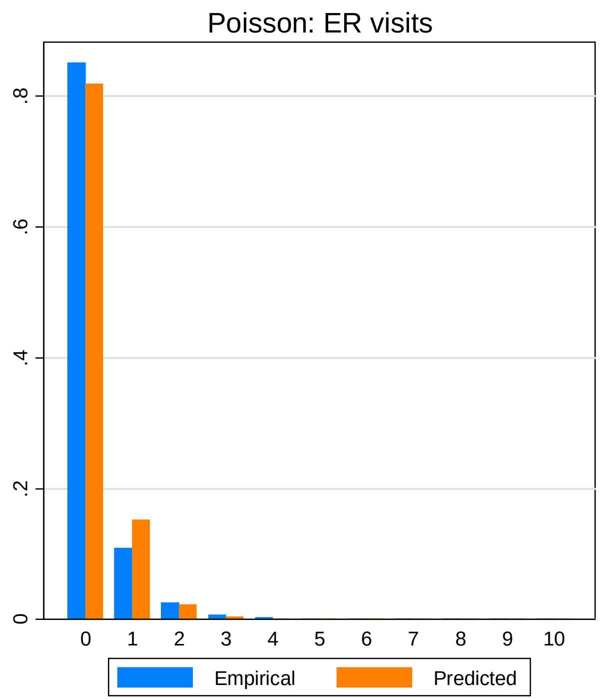
Poisson Marginal Effects: Hospital nights

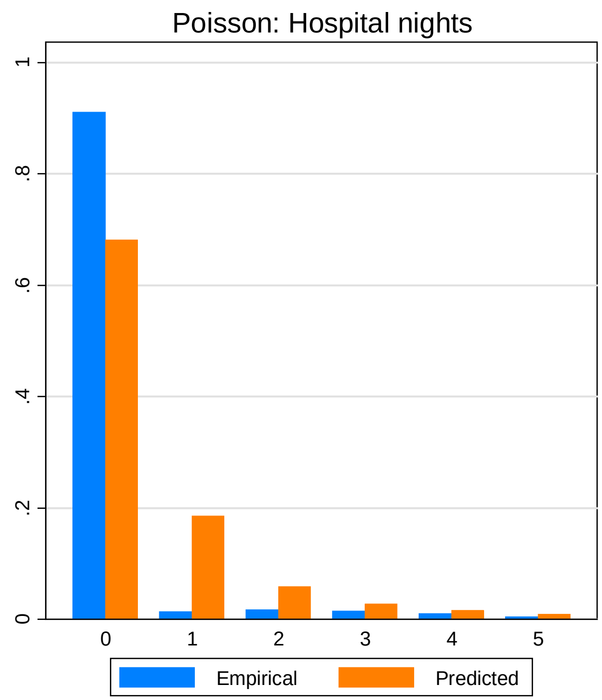
	Average	Mean of X	Median of X
Age	0.001	0.000	0.000
	(0.000)	(0.000)	(0.000)
1.female	-0.028*	-0.013*	-0.013*
	(0.012)	(0.006)	(0.006)

In-sample Goodness of fit

Informal / Graphical - compare empirical distribution of *y* to predicted distribution







In-sample Goodness of Fit

Mean Prediction (of distribution) Error

$$MPE = \frac{1}{J}\sum_{j=0}^{J}(f_j - \hat{f}_j)$$

Mean Square Prediction (of distribution) Error

$$MSPE = \frac{1}{J}\sum_{j=0}^{J}(f_j - \hat{f}_j)^2$$

J should be chosen to cover most of the support (but not all the values of the count variable)

	Office visits (0-20)	ER visits (0-10)	Hospital nights (0-5)
MPE	-0.155	-0.002	-0.129
MSPE	30.615	2.705	139.367

FYI: Stata code for Poisson goodness of fit measures

```
preserve
forvalues j=0/20 {
   gen byte y_`j' = `e(depvar)' == `j'
   predict pr_`j', pr(`j')
}
collapse (mean) y_* pr_*
gen i=_n
reshape long y_ pr_, i(i) j(y)

graph bar (asis) y_ pr_

generate pr_diff = (y_ - pr_)*100
generate pr_diff2 = pr_diff^2

mean pr_diff pr_diff2
restore
```

Poisson - Summary

Advantages

Robust (asymptotic) to misspecification of variance

Easy to compute marginal effects and predictions

Disadvantages

Not robust in finite samples

Possibly sensitive to influential observations and outliers

Not efficient if variance is misspecified

Overview

Studies with count data

- **Poisson (canonical model)**

- **Negative Binomial**
  - **Estimation**
  - **Prediction – Mean, Events**
  - **Interpretation – Marginal effects, Incremental Effects**
  - **Goodness of fit**

- **Hurdle Models (Two Part Models for Counts)**

- **Zero Inflated Models**

Model Selection - Discriminating among nonnested models

Negative Binomial

Canonical model for overdispersed data

Mean $\mu = \exp(X\beta)$

Overdispersion – variance exceeds the mean

$$Var(y|X) = \mu + \alpha\, g(\mu) \;>\; \mu$$

Negative Binomial-1 $Var(y|X) = \mu + \alpha\mu$

Negative Binomial-2 $Var(y|X) = \mu + \alpha\mu^2$

Estimation

Maximum Likelihood

Stata command for NB-2:

```
nbreg use_off age i.female, dispersion(mean)
nbreg use_off age i.female
```

Note: dispersion(mean) is not required – it is the default

Stata command for NB-1:

```
nbreg use_off age i.female, dispersion(constant)
```

Choosing between NB-1 and NB-2

These are non-nested models

Use model selection criteria

FYI: Negative Binomial-2: Estimates

```
. nbreg use_off $X, robust

<snip>

Negative binomial regression                    Number of obs     =      19386
Dispersion           = mean                     Wald chi2(21)     =    4900.92
Log pseudolikelihood = -49111.723               Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |               Robust
     use_off |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0108457   .0010707    10.13   0.000     .0087473    .0129442
    1.female |   .4911042   .0264055    18.60   0.000     .4393504    .5428581
<snip>
-------------+----------------------------------------------------------------
    /lnalpha |   .3581475   .0176417                      .3235703    .3927246
-------------+----------------------------------------------------------------
       alpha |   1.430677   .0252396                      1.382053    1.481011
------------------------------------------------------------------------------

.
. estimates store nb2
```

FYI: Negative Binomial-1:Estimates

```
. nbreg use_off $X, disp(constant) robust

<snip>

Negative binomial regression                       Number of obs   =      19386
Dispersion           = constant                    Wald chi2(21)   =    7663.19
Log pseudolikelihood = -48824.428                  Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |               Robust
     use_off |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0077678   .0006543    11.87   0.000     .0064855    .0090501
    1.female |   .3710594   .0153733    24.14   0.000     .3409282    .4011906
<snip>
-------------+----------------------------------------------------------------
    /lndelta |   2.144767   .0246959                      2.096363     2.19317
-------------+----------------------------------------------------------------
       delta |   8.540047   .2109045                      8.136526     8.96358
------------------------------------------------------------------------------

. estimates store nb1
```

Negative Binomial: Choosing Between NB2 and NB1

```
. estimates stats nb2 nb1
```

Office visits

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
nb2	19386	-52504.62	-49111.72	23	98269.45	98450.51
nb1	19386	-52504.62	-48824.43	23	97694.86	97875.92

Note: N=Obs used in calculating BIC; see [R] BIC note

ER visits

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
nb2	19386	-10671.19	-9995.218	23	20036.44	20217.5
nb1	19386	-10671.19	-10020.41	23	20086.83	20267.89

Hospital nights

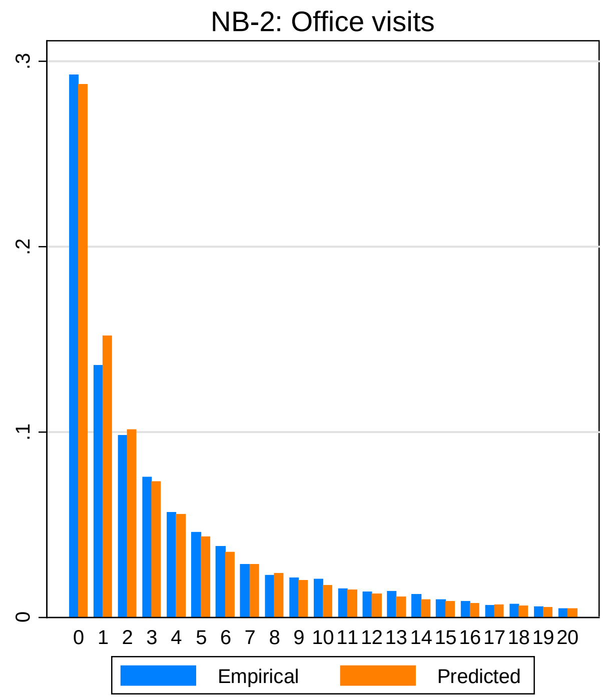
Model	Obs	ll(null)	ll(model)	df	AIC	BIC
nb2	19386	-10635.45	-10033.9	23	20113.8	20294.86
nb1	19386	-10635.45	-9884.171	23	19814.34	19995.41

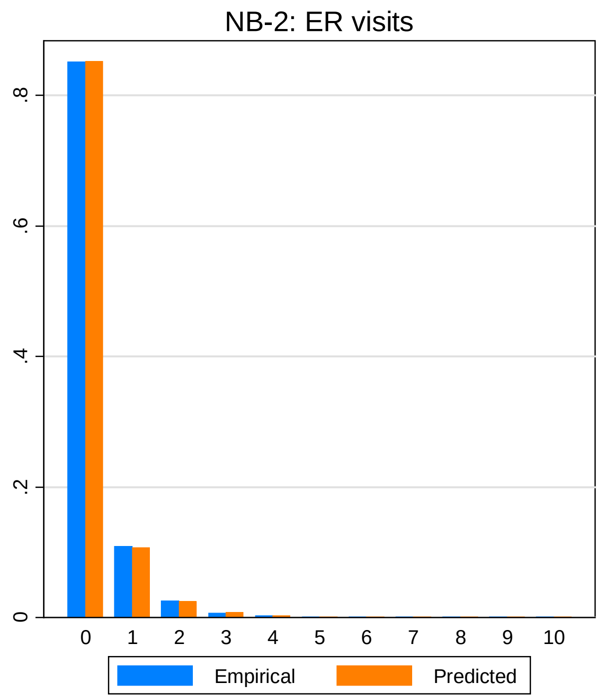
NB Marginal Effects

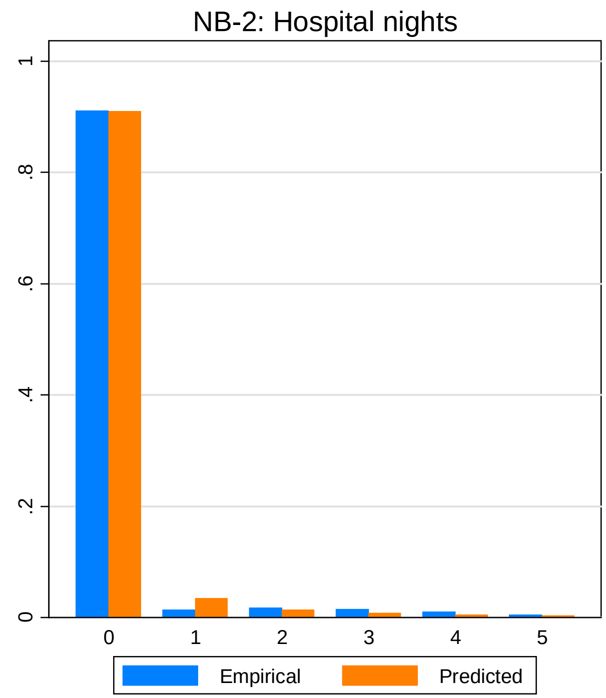
	Office visits		ER visits		Hospital nights	
	NB-2	NB-1	NB-2	NB-1	NB-2	NB-1
age	0.068**	0.045**	-0.004**	-0.003**	-0.002	-0.003*
	(0.007)	(0.004)	(0.000)	(0.000)	(0.004)	(0.001)
1.female	2.909**	2.073**	0.031**	0.033**	0.214**	0.252**
	(0.153)	(0.085)	(0.009)	(0.008)	(0.082)	(0.033)

* $p<0.05$; ** $p<0.01$

In-sample Goodness of Fit







	Office visits	ER visits	Hospital nights
MPE	0.046	-0.001	-0.043
MSPE	0.167	0.005	0.883

Negative Binomial - Summary

Advantages

Much less sensitive to influential observations and outliers

Mean is robust in finite samples

Disadvantages

Distribution is not robust to misspecification of variance

Not efficient if variance is misspecified