The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$115	1-year subscription	$145
2-year subscription	$210	2-year subscription	$270
3-year subscription	$285	3-year subscription	$375
1-year student subscription	$ 85	1-year student subscription	$115
1-year institutional subscription	$345	1-year institutional subscription	$375
2-year institutional subscription	$625	2-year institutional subscription	$685
3-year institutional subscription	$875	3-year institutional subscription	$965
Electronic only		**Electronic only**	
1-year subscription	$ 85	1-year subscription	$ 85
2-year subscription	$155	2-year subscription	$155
3-year subscription	$215	3-year subscription	$215
1-year student subscription	$ 55	1-year student subscription	$ 55

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, STATA, Stata Press, Mata, MATA, and NetCourse are registered trademarks of StataCorp LP.

The Stata Journal (2015)
15, Number 1, pp. 226–246

Regression models for count data from truncated distributions

James W. Hardin
Institute for Families in Society
Department of Epidemiology and Biostatistics
University of South Carolina
Columbia, SC
jhardin@sc.edu

Joseph M. Hilbe
School of Social and Family Dynamics
Arizona State University
Tempe, AZ
hilbe@asu.edu

Abstract. We present new commands for analyzing count-data regression models for truncated distributions. The `trncregress` command allows specification of a regression model for the mean of the truncated distribution through options. In addition to support for truncated Poisson and negative binomial, `trncregress` fits models based on truncated versions of distributions including generalized Poisson, Poisson-inverse Gaussian, three-parameter negative binomial power, three-parameter Waring negative binomial, and three-parameter Famoye negative binomial.

Keywords: st0378, trncregress, truncation, generalized Poisson, negative binomial, Poisson-inverse Gaussian, Famoye, Waring, PIG, NB-P, NB-F

1 Introduction

Regression modeling of truncated count outcomes is supported by Stata's `tpoisson` and `tnbreg` commands. These commands allow users to fit models for left-truncated $\{y \in (L+1, L+2, \ldots)\}$ distributions. Users may specify either a common truncation value, L, or a variable so that each observation has its own truncation value and thus a uniquely truncated distribution. Though left-truncation is more commonly used in regression models, the commands we introduce here will consider right-truncation $\{y \in (0, 1, \ldots, R-1)\}$ or even truncation on both sides $\{y \in (L+1, L+2, \ldots, R-2, R-1)\}$.

Before Stata offered `tpoisson` and `tnbreg`, support for estimation of truncated regression models was given only for the specific zero-truncated models through commands that are now deprecated. However, Stata still lacks commands that support additional distributions (aside from Poisson and negative binomial) or that support distributions that are right-truncated or truncated on both sides.

In section 2, we present new estimation commands to evaluate count-data regression models for truncated distributions such as Poisson, negative binomial, generalized Poisson, Poisson-inverse Gaussian, negative binomial(P) (NB-P), and negative binomial (Famoye) (NB-F). Hilbe (2011), Hardin and Hilbe (2014), and Harris, Hilbe, and Hardin (2014) discuss the last two distributions and include software for nontruncated regression models.

In section 3, we provide syntax for the new commands, followed by examples in section 4.

2 Extensions of Poisson and negative binomial regression

The Poisson probability mass function is given by

$$f(y;\mu) = \frac{\exp(-\mu)\mu^y}{y!}$$

with mean $E(y) = \mu$ and variance $V(y) = \mu$. Wang and Famoye (1997) introduce a two-parameter distribution that generalizes the distribution. Regression models using the Poisson distribution assume equidispersion; that is, they assume that the mean and variance of the outcome are equal for a given covariate pattern. Most data are characterized as having variance that is larger than the mean. The negative binomial distribution and its generalizations assume different forms of overdispersion. The generalized Poisson can accommodate overdispersion, but its parameterization of the variance also allows underdispersion (a variance less than the mean).

The negative binomial probability mass function is given by

$$f(y;\alpha,\delta) = \frac{\Gamma(y+1/\alpha)}{\Gamma(1/\alpha)\Gamma(y+1)}\left(\frac{1}{1+\delta\alpha}\right)^{1/\alpha}\left(1-\frac{1}{1+\delta\alpha}\right)^{y}$$

with mean $E(y) = \delta$ and variance $V(y) = \delta(1+\delta\alpha)$. Users have access to two parameterizations of the negative binomial distribution. The two results of the parameterizations are referred to as the NB-1 (constant dispersion) and NB-2 (mean dispersion) models. The numerals used in naming these two models correspond to the nature of the variance (as a function of the power of the mean). The NB-1 model results from introducing coefficients via $\alpha = \theta\exp(X\beta) = \theta\mu$. The NB-2 model results from introducing regressors X via $\alpha = \theta$ and $\delta = \exp(X\beta) = \mu$ so that the mean is μ, the variance is $\mu(1+\mu\theta)$, and the dispersion is $1 + \mu\theta$.

Hilbe and Greene (2008) discuss a generalization to the underlying negative binomial probability distribution for which the variance is a function of a parameter power of the mean (also see Greene [2008], Cameron and Trivedi [2013], and Hilbe [2011]). In this NB-P model, regressors X are introduced via $\alpha = \theta\exp(X\beta)^{P-2} = \theta\mu^{P-2}$ and $\delta = \exp(X\beta) = \mu$ so that the mean is μ, the variance is $\mu(1+\mu^{P-1}\theta)$, and the dispersion is $(1+\mu^{P-1}\theta)$. Here we see that the distribution is equal to NB-1 when $P = 1$ and is equal to NB-2 when $P = 2$.

Harris, Hilbe, and Hardin (2014) present two other generalizations to the negative binomial. The authors refer to these generalizations as NB-W for the generalization based on the Waring distribution and as NB-F for the generalization based on the work of Famoye; see also Rodríguez-Avi et al. (2009), Irwin (1968), and Wang and Famoye (1997).

3 Syntax

Software accompanying this article includes the command files as well as supporting files for prediction and help. In the following syntax diagrams, unspecified options include the usual collection of maximization and display options available to all estimation commands.

Equivalent in syntax to the basic count-data commands, the basic syntax for the truncated regression command is

`trncregress` *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, `ltrunc(`*#* | *varname*`)` `rtrunc(`*#* | *varname*`)` `dist(`*distname*`)` `offset(`*varname_o*`)` *display_options* *maximization_options*]

In the commands above, the allowable distribution names are given by `poisson`, `negbin`, `gpoisson`, `pig`, `nbp`, `nbf`, or `nbw`. Help files are included for the estimation and postestimation specifications of these models. The help files include example specifications.

In the output header, we include the summary information for the model. We also include a short description of the support for the outcome by the designated truncated distribution. This description is of the form {$\#_1, \dots, \#_2$}, where $\#_1$ is the minimum and $\#_2$ is the maximum. Thus, for a zero-truncated model, the support is given by $\#_1 = 1$ and $\#_2 =$. (positive infinity).

Model predictions are available through Stata's `predict` command. Specifically, there is support for linear predictions, predictions of the mean, and standard errors of the linear prediction.

4 Examples

Truncated regression models are most commonly used to model zero-truncated count data. Given that the supported count distributions assume the possibility of zero counts, biased results are obtained when zero-truncated count data are modeled using regression methods based on nontruncated distributions. The closer the mean of the response is to zero, the more biased the results. To ameliorate influence on inference from biased results, many analysts prefer standard errors from a sandwich or robust variance adjustment when using nontruncated regression models to model zero-truncated data.

However, zero-truncated data are better modeled using one of the truncated distributions for which we have developed the software accompanying this article. To demonstrate this, we use data from the 1991 Arizona MedPar database, which consist of the inpatient records for Medicare patients. In this study, all patients are over 65 years of age. The diagnostic related group classification is confidential for privacy concerns.

The response variable is the patient length of stay (`los`), which commences with a count of 1. There are no length of stay records of 0, which could indicate that a patient was not admitted to the hospital.

```
. use medpar
. generate byte type = type1 + 2*type2 + 3*type3
. generate offset = uniform()
. generate exposure = ln(offset)
. tabulate los
 Length of |
      Stay |      Freq.     Percent        Cum.
-----------+-----------------------------------
         1 |        126        8.43        8.43
         2 |         71        4.75       13.18
         3 |         75        5.02       18.19
         4 |        104        6.96       25.15
         5 |        123        8.23       33.38
         6 |         97        6.49       39.87
  (output omitted)
        70 |          1        0.07       99.80
        74 |          1        0.07       99.87
        91 |          1        0.07       99.93
       116 |          1        0.07      100.00
-----------+-----------------------------------
     Total |      1,495      100.00
```

The mean of `los` is 9.85. Using a zero-truncated model will make little difference in the estimates. However, if the mean of the response is low (say, under three or four), then there will be a substantial difference in coefficient values. The closer the mean is to zero, the greater the difference in coefficient values. Despite the closeness of coefficients for this example, it is important that we use the appropriate count model for the given data. The explanatory predictors for our example model include an indicator of white race (`white`), an indicator of HMO (`hmo`), an indicator of elective admittance (`type1`, used as the reference group for admittance types), an indicator of urgent admittance (`type2`), and an indicator of emergency admittance (`type3`); all indicators are generated from the classification variable `type`.

We first model the data using a zero-truncated Poisson (ZTP) model. Note that the new truncated regression command included herein supports the `nolog` option to suppress the display of the iteration log, the `eform` option to display model coefficients in exponentiated form, and automatic generation of indicator variables from categorical variable names through the `i.` prefix.

```
. trncregress los white hmo i.type, dist(poisson) ltrunc(0) nolog eform
Truncated Poisson regression                     Number of obs   =       1495
Dist. support on {1, ..., .}                     LR chi2(4)      =     758.68
Log likelihood = -6928.723                       Prob > chi2     =     0.0000

         los      exp(b)   Std. Err.       z   P>|z|     [95% Conf. Interval]

       white    .8573203   .0235048    -5.61   0.000     .8124676    .9046491
         hmo     .930858   .0223067    -2.99   0.003     .8881484    .9756214

        type
           2    1.248297   .0262846    10.53   0.000     1.197829    1.300892
           3    2.033211    .053145    27.15   0.000     1.931672    2.140087

       _cons    10.30738   .2804854    85.73   0.000     9.772044    10.87205

. estat ic
Akaike's information criterion and Bayesian information criterion

       Model      Obs    ll(null)   ll(model)     df          AIC         BIC

           .     1495   -7308.063   -6928.723      5     13867.45       13894

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

We also model the data using standard Poisson regression to determine the dispersion statistic, which indicates the amount of extradispersion in the model. The resulting dispersion value of 6.26 shows that the data are rather markedly overdispersed, which biases the values of the model standard errors. All predictors appear to be significant at the $\alpha = 0.05$ level when, in fact, they may not be. A zero-truncated negative binomial (ZTNB) may account for some of the excess variation.

```
. trncregress los white hmo i.type, dist(negbin) ltrunc(0) nolog eform
Truncated neg. binomial regression               Number of obs   =       1495
Dist. support on {1, ..., .}                     LR chi2(4)      =     106.23
Log likelihood = -4751.396                       Prob > chi2     =     0.0000

         los      exp(b)   Std. Err.       z   P>|z|     [95% Conf. Interval]

       white    .8741019   .0662078    -1.78   0.076     .7535097    1.013994
         hmo     .929911   .0547995    -1.23   0.218     .8284767    1.043764

        type
           2    1.264196   .0706704     4.19   0.000     1.133003     1.41058
           3    2.086729   .1754021     8.75   0.000     1.769773    2.460451

       _cons    9.703802   .7299226    30.21   0.000      8.37364    11.24526

    /lnalpha   -.6007156   .0549884                      -.708491   -.4929402

       alpha     .548419   .0301567                      .4923867    .6108278
```

```
. estat ic
Akaike´s information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1495	-4804.512	-4751.396	6	9514.792	9546.651

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

The Akaike information criterion (AIC) and Bayesian information criterion (BIC) statistics of the ZTNB model are substantially lower than those of the ZTP model, indicating a better fit. Being an HMO member is no longer a significant predictor of length of hospital stay, and white is marginal. By comparing the previous and subsequent outputs, we see that basing standard errors on the robust sandwich variance is not necessary in this case. However, Hilbe (2011) and Cameron and Trivedi (2013) prefer standard errors based on the robust variance estimator, favoring robustness of inference over efficiency.

```
. trncregress los white hmo i.type, dist(negbin) ltrunc(0) nolog eform
> vce(robust)
Truncated neg. binomial regression              Number of obs   =       1495
Dist. support on {1, ..., .}                    LR chi2(4)      =     106.23
Log pseudolikelihood = -4751.396                Prob > chi2     =     0.0000
```

los	exp(b)	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
white	.8741019	.0648392	-1.81	0.070	.7558256	1.010887
hmo	.929911	.0512158	-1.32	0.187	.834758	1.03591
type						
2	1.264196	.0707132	4.19	0.000	1.132928	1.410674
3	2.086729	.248301	6.18	0.000	1.652651	2.63482
_cons	9.703802	.7018808	31.42	0.000	8.421202	11.18175
/lnalpha	-.6007156	.0624481			-.7231116	-.4783196
alpha	.548419	.0342477			.48524	.619824

We then use the `trncregress` command to model the data using a zero-truncated Poisson-inverse Gaussian (PIG), a generalized Poisson, a three-parameter generalized NB-F, and a three-parameter NB-P. The ZINB-P proved to fit the data better than the other zero-truncated models, including the ZTNB.

```
. trncregress los white hmo i.type, dist(nbp) ltrunc(0) nolog eform

Truncated neg. bin(P) regression                  Number of obs   =      1495
Dist. support on {1, ..., .}                      LR chi2(4)      =    128.25
Log likelihood = -4740.387                        Prob > chi2     =    0.0000

------------------------------------------------------------------------------
         los |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       white |   .9392964    .061651    -0.95   0.340     .8259121    1.068247
         hmo |   .9373804   .0452815    -1.34   0.181     .8527022    1.030468
             |
        type |
           2 |   1.225673    .062171     4.01   0.000     1.109681    1.353789
           3 |    2.01843   .2183897     6.49   0.000     1.632735    2.495238
             |
       _cons |   9.177259    .596997    34.08   0.000     8.078688    10.42522
-------------+----------------------------------------------------------------
          /P |   3.177911   .3525741     9.01   0.000     2.486878    3.868943
    /lnalpha |  -3.279836   .7890462                     -4.826338   -1.733334
-------------+----------------------------------------------------------------
       alpha |   .0376344   .0296953                      .0080158    .1766943
------------------------------------------------------------------------------

. estat ic

Akaike´s information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |   1495   -4804.512   -4740.387      7     9494.774    9531.944
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

The AIC statistic is lower by 20 points, and the BIC is lower by 14. Following Hilbe (2009), this classifies as significantly different. Basing standard errors on a robust or sandwich variance estimator produces the following result:

```
. trncregress los white hmo i.type, dist(nbp) ltrunc(0) nolog eform vce(robust)

Truncated neg. bin(P) regression                  Number of obs   =      1495
Dist. support on {1, ..., .}                      LR chi2(4)      =    128.25
Log pseudolikelihood = -4740.387                  Prob > chi2     =    0.0000

------------------------------------------------------------------------------
             |               Robust
         los |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       white |   .9392964   .0568864    -1.03   0.301     .8341642    1.057679
         hmo |   .9373804   .0440311    -1.38   0.169     .8549344    1.027777
             |
        type |
           2 |   1.225673   .0629182     3.96   0.000     1.108356    1.355407
           3 |    2.01843   .2276542     6.23   0.000     1.618112    2.517786
             |
       _cons |   9.177259   .5383163    37.79   0.000     8.180569    10.29538
-------------+----------------------------------------------------------------
          /P |   3.177911   .3517989     9.03   0.000     2.488398    3.867424
    /lnalpha |  -3.279836   .7941904                     -4.836421   -1.723252
-------------+----------------------------------------------------------------
       alpha |   .0376344   .0298889                      .0079354    .1784848
------------------------------------------------------------------------------
```

Neither `white` nor `hmo` is significant at the 0.05 level. The NB-P scale parameter is 3.18. The dispersion parameter is 0.038. The dispersion is parameterized such that it has a direct relationship with the mean, μ. The equation for the variance of the model is given by

$$\mu + \alpha\mu^p = \mu + 0.0376\mu^{3.178}$$

Given the high mean value of `los` (9.85), we expect that the estimates and the adjusted standard errors will be close in values. Though we do not include the output here, we used the command by Hardin and Hilbe (2012) for the PIG model to investigate the similarity of output between the nonzero-truncated and the zero-truncated PIG distributions. However, note that the AIC and BIC statistics are substantially lower in the zero-truncated model, which may be the result of the absence of zero counts in the data. The `trncregress` command adjusts for their absence; `nbregp` does not.

```
. nbregp los white hmo i.type, nolog eform vce(robust)
Negative binomial-P regression                    Number of obs   =       1495
                                                  Wald chi2(4)    =      57.30
Log pseudolikelihood = -4782.519                  Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |               Robust
         los |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       white |   .9320299   .0563996    -1.16   0.245      .8277923    1.049393
         hmo |   .9353262   .0451328    -1.39   0.166      .8509218    1.028103
             |
        type |
          2  |   1.236604   .0630604     4.16   0.000      1.118984    1.366588
          3  |   2.070074    .231388     6.51   0.000      1.662802    2.577099
             |
       _cons |   9.552943   .5622072    38.35   0.000      8.512214    10.72092
-------------+----------------------------------------------------------------
          /P |   3.047995   .2006046    15.19   0.000      2.654817    3.441173
   /lntheta  |  -3.228758   .4663185                      -4.142725    -2.31479
-------------+----------------------------------------------------------------
       theta |   .0396067   .0184693                       .0158795    .0987869
------------------------------------------------------------------------------
Likelihood-ratio test of P=1:     chi2 =     98.47 Prob > chi2     =     0.0000
Likelihood-ratio test of P=2:     chi2 =     29.92 Prob > chi2     =     0.0000
. estat ic
Akaike´s information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |   1495           .   -4782.519      7     9579.037    9616.206
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

The likelihood-ratio test statistics indicate that the data are better modeled by NB-P than by either NB-1 or NB-2. However, not adjusting for the missing-zero counts causes a standard PIG model to not fit as well as any of the `trncregress` options for zero-truncated data except the Poisson.

For another example, we use the German health reform data to model the number of visits to the physician made by patients during the calendar year 1984; these data are used in Hardin and Hilbe (2012). Predictors include age, employment status, and sex. Specifically, `docvis` records the number of physician visits, `age` is the patient's age in years, `outwork` is an indicator that the person is out of work, and `female` is an indicator that the person is female.

```
. use rwm1984, clear
(German health data for 1984; Hardin & Hilbe, GLM and Extensions, 3rd ed)
. summarize age

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         age |       3874    43.99587     11.2401         25         64
. generate cage = (age-r(mean))
. tabulate docvis

          MD |
 visits/year |      Freq.     Percent        Cum.
-------------+-----------------------------------
           0 |      1,611       41.58       41.58
           1 |        448       11.56       53.15
           2 |        440       11.36       64.51
           3 |        353        9.11       73.62
           4 |        213        5.50       79.12
  (output omitted)
          70 |          1        0.03       99.90
          71 |          1        0.03       99.92
          72 |          1        0.03       99.95
          80 |          1        0.03       99.97
         121 |          1        0.03      100.00
-------------+-----------------------------------
       Total |      3,874      100.00
```

The mean of the response, `docvis`, is 3.16. Because 41.5% of the patients did not visit a physician, we also calculate the mean of the visits without zero count. Here we want to model the number of visits made to physicians, excluding those patients who never entered that pool. The mean of the zero-excluded response is 5.41. There will likely be a noticeable difference in the zero-truncated model results and standard results. However, we want to find the best-fitting zero-truncated count model for the given data.

We first model the data using a Poisson regression by simply excluding the zero counts. Given the values of the predictor age, we center it on its mean value (mean-centered ages are in the `cage` variable).

```
. glm docvis outwork female cage if docvis>0, family(poisson) nolog eform
Generalized linear models                         No. of obs      =       2263
Optimization     : ML                             Residual df     =       2259
                                                  Scale parameter =          1
Deviance         =  12162.17413                   (1/df) Deviance =   5.383875
Pearson          =  21997.94599                   (1/df) Pearson  =   9.737913
Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = ln(u)                   [Log]
                                                  AIC             =   8.507555
Log likelihood   = -9622.298504                   BIC             =  -5287.351

------------------------------------------------------------------------------
             |                 OIM
      docvis |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     outwork |   1.178181   .0248475     7.77   0.000     1.130473    1.227901
      female |   1.101225    .022611     4.70   0.000     1.057788    1.146445
        cage |   1.011738   .0008477    13.93   0.000     1.010078    1.013401
       _cons |   4.612541   .0689904   102.21   0.000     4.479285    4.749762
------------------------------------------------------------------------------

. estat ic
Akaike's information criterion and Bayesian information criterion
-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |   2263           .   -9622.299      4      19252.6    19275.49
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

The dispersion statistic is high (9.74), and the AIC value is 19,252. Modeling using a truncated Poisson distribution adjusts the underlying probability density function for the missing zeros.

```
. trncregress docvis outwork female cage if docvis>0, dist(poisson) ltrunc(0)
> nolog eform
Truncated Poisson regression                     Number of obs   =       2263
Dist. support on {1, ..., .}                     LR chi2(3)      =     466.47
Log likelihood = -9605.928                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
      docvis |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     outwork |   1.182679   .0253025     7.84   0.000     1.134112    1.233325
      female |   1.105015   .0230725     4.78   0.000     1.060706    1.151174
        cage |   1.012106   .0008641    14.09   0.000     1.010414    1.013801
       _cons |   4.559868   .0698738    99.02   0.000     4.424954    4.698895
------------------------------------------------------------------------------

. estat ic
Akaike's information criterion and Bayesian information criterion
-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |   2263   -9839.164   -9605.928      4     19219.86    19242.75
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

Note that the AIC and BIC statistics are significantly lower when excluding zero visits. Because of the high dispersion statistic (9.74) and relatively low response mean, we use sandwich or robust standard-error adjustments to model the standard errors.

```
. trncregress docvis outwork female cage if docvis>0, dist(poisson) ltrunc(0)
> nolog eform vce(robust)
Truncated Poisson regression                        Number of obs   =      2263
Dist. support on {1, ..., .}                        LR chi2(3)      =    466.47
Log pseudolikelihood = -9605.928                    Prob > chi2     =    0.0000

------------------------------------------------------------------------------
             |               Robust
      docvis |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     outwork |   1.182679   .1005398     1.97   0.048     1.001166    1.397101
      female |   1.105015   .0882721     1.25   0.211     .9448683    1.292304
        cage |   1.012106    .002791     4.36   0.000      1.00665    1.017591
       _cons |   4.559868   .2141993    32.30   0.000     4.158792    4.999624
------------------------------------------------------------------------------
```

The adjustment causes females to be shown as not contributing to the model and outwork to be shown as only marginally contributing. The centered age (`cage`) is still a significant predictor. However, given the variability in the data, we model the data using a ZTNB model.

```
. trncregress docvis outwork female cage if docvis>0, dist(negbin) ltrunc(0)
> nolog eform vce(robust)
Truncated neg. binomial regression                  Number of obs   =      2263
Dist. support on {1, ..., .}                        LR chi2(3)      =     75.93
Log pseudolikelihood = -5757.054                    Prob > chi2     =    0.0000

------------------------------------------------------------------------------
             |               Robust
      docvis |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     outwork |   1.262669   .1250538     2.35   0.019      1.03989    1.533176
      female |   1.153206    .105388     1.56   0.119     .9640913    1.379417
        cage |   1.016271   .0033972     4.83   0.000     1.009634    1.022951
       _cons |   2.703222    .171369    15.69   0.000     2.387374    3.060858
-------------+----------------------------------------------------------------
    /lnalpha |    .744524   .1218212                      .5057587    .9832892
-------------+----------------------------------------------------------------
       alpha |   2.105439   .2564872                      1.658243    2.673235
------------------------------------------------------------------------------

. estat ic
Akaike´s information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |   2263   -5795.017   -5757.054      5     11524.11    11552.73
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

The AIC statistic drops from 19,220 to 11,524. A standard NB-2 model has an AIC value of 12,371, indicating that the ZTNB is the preferred model. The BIC is similarly reduced.

We then fit zero-truncated generalized Poisson (ZTGP), NB-P, and PIG models. All three fit the data better than the ZTNB, with the ZTGP having the best fit.

```
. trncregress docvis outwork female cage if docvis>0, dist(gpoisson) ltrunc(0)
> nolog eform vce(robust)
Truncated gen. Poisson regression                    Number of obs   =       2263
Dist. support on {1, ..., .}                         LR chi2(3)      =      67.76
Log pseudolikelihood = -5723.069                     Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |               Robust
      docvis |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     outwork |    1.23783   .0949403     2.78   0.005     1.065062    1.438624
      female |   1.200285   .0897923     2.44   0.015     1.036589    1.389831
        cage |   1.014942   .0028927     5.20   0.000     1.009288    1.020627
       _cons |   3.215915   .1696478    22.14   0.000     2.900024    3.566216
-------------+----------------------------------------------------------------
 /atanhdelta |   .7716666   .0234126                      .7257786    .8175545
-------------+----------------------------------------------------------------
       delta |   .6478975   .0135847                       .620476    .6737367
------------------------------------------------------------------------------

. estat ic
Akaike´s information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |   2263   -5756.951   -5723.069      5     11456.14    11484.76
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

Note that all 3 predictors now significantly contribute to the model, and the AIC statistic is 11,456 compared with 11,524, a 68-point drop in value; the BIC similarly reduced from 11,553 to 11,485.

```
. trncregress docvis outwork female cage if docvis>0, dist(pig) ltrunc(0) nolog
> eform vce(robust)
Truncated Poisson IG regression                      Number of obs   =       2263
Dist. support on {1, ..., .}                         LR chi2(3)      =      80.15
Log pseudolikelihood = -5694.797                     Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |               Robust
      docvis |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     outwork |   1.253586   .0960068     2.95   0.003     1.078858    1.456612
      female |   1.173082   .0841799     2.22   0.026      1.01917    1.350237
        cage |   1.015232   .0027737     5.53   0.000      1.00981    1.020683
       _cons |   3.597949   .1730368    26.62   0.000     3.274297    3.953594
-------------+----------------------------------------------------------------
    /lnalpha |   .4397922   .0757686                      .2912885    .5882959
-------------+----------------------------------------------------------------
       alpha |   1.552385    .117622                      1.338151    1.800917
------------------------------------------------------------------------------
```

```
. estat ic
Akaike´s information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	2263	-5734.872	-5694.797	5	11399.59	11428.22

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

Because there are very few observations for which people visited their physician more than 18 times, here we model data only within 1 and 18 visits by using a generalized Poisson distribution truncated on each side. This is referred to as interval truncation.

```
. trncregress docvis outwork female cage if docvis>0 & docvis<19, dist(gpoisson)
> ltrunc(0) rtrunc(19) nolog eform vce(robust)
Truncated gen. Poisson regression                 Number of obs   =       2172
Dist. support on {1, ..., 18}                     LR chi2(3)      =      63.29
Log pseudolikelihood = -4982.805                  Prob > chi2     =     0.0000
```

docvis	exp(b)	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
outwork	1.21593	.0778329	3.05	0.002	1.072562	1.378462
female	1.166139	.0738958	2.43	0.015	1.029939	1.320351
cage	1.009884	.0025166	3.95	0.000	1.004964	1.014829
_cons	3.001998	.1370437	24.08	0.000	2.745063	3.282982
/atanhdelta	.5804847	.0187828			.543671	.6172984
delta	.5230177	.0136448			.4957618	.5492442

```
. estat ic
Akaike´s information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	2172	-5014.451	-4982.805	5	9975.609	10004.03

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

A ZTGP model with a right-truncation point of 19 has an AIC of 9,976, whereas the ZTGP model had an AIC above 11,456. This is a 1,480-point drop in AIC, which is mostly due to fitting the model on a subset of the data.

Finally, we can combine these truncated models with other models to construct hurdle models. For example, we can combine a logistic regression model of the likelihood of a zero outcome with a zero-truncated model. In this example, we also create an interaction term (`femage`) associating centered age (`cage`) and sex (`female`).

```
. generate zerovis = docvis==0
. replace zerovis = . if docvis==.
(0 real changes made)
. generate femcage = female*cage
```

```
. logistic zerovis outwork female cage femcage, nolog vce(robust)
Logistic regression                               Number of obs   =       3874
                                                  Wald chi2(4)    =     202.46
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -2523.1663                 Pseudo R2       =     0.0407

------------------------------------------------------------------------------
             |               Robust
     zerovis | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     outwork |   .7653926   .0623506    -3.28   0.001     .6524444    .8978939
      female |   .5967247   .0450747    -6.84   0.000     .5146084    .6919443
        cage |   .9696342   .0040672    -7.35   0.000     .9616954    .9776387
     femcage |   1.008207   .0061143     1.35   0.178     .9962943    1.020263
       _cons |   .9821952   .0461209    -0.38   0.702     .8958348    1.076881
------------------------------------------------------------------------------

. trncregress docvis outwork female cage if docvis>0, dist(gpoisson) ltrunc(0)
> nolog eform vce(robust)
Truncated gen. Poisson regression                 Number of obs   =       2263
Dist. support on {1, ..., .}                      LR chi2(3)      =      67.76
Log pseudolikelihood = -5723.069                  Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |               Robust
      docvis |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     outwork |    1.23783   .0949403     2.78   0.005     1.065062    1.438624
      female |   1.200285   .0897923     2.44   0.015     1.036589    1.389831
        cage |   1.014942   .0028927     5.20   0.000     1.009288    1.020627
       _cons |   3.215915   .1696478    22.14   0.000     2.900024    3.566216
-------------+----------------------------------------------------------------
 /atanhdelta |   .7716666   .0234126                      .7257786    .8175545
-------------+----------------------------------------------------------------
       delta |   .6478975   .0135847                       .620476    .6737367
------------------------------------------------------------------------------
```

Aging and being female and out of work are all associated with being less likely to never visit the doctor. Similarly, these three characteristics are associated with higher rates of doctor visits.

As a final example, we investigate surgical data from the 1999 Arizona Medicare database. Medicare is a federal health insurance program for U.S. citizens age 65 and over or for those with disability. The exact procedures are withheld from the data for privacy reasons.

The data are not unusual for many types of nonmajor surgical procedures for which the majority of patients are released soon after surgery. However, for some patients, complications occur that necessitate longer recovery periods. We model length of stay (`los`) given explanatory predictors of age in years (`age`), for which we have removed the mean; sex (`gender` indicates male in these data); the type of admission (`1` = emergency/urgent; `0` = elective); and procedure type (`1` = open; `0` = laparoscopic). Our primary interest is how much longer patients stay in the hospital after open surgery compared with laparoscopic surgery, adjusted for gender and emergency status of admission. The following table shows the length of stay, as well as the mean of `los`, which is 3.3 days. Because the nontruncated distributions used to model the data assume the possibility of zero counts, and given the low mean value of the response term, we expect that a zero-truncated model will be preferred.

```
. use azsurgical, clear
(1999 Arizona Medicare surgical data: J. Hilbe)
. summarize age
  (output omitted)
. generate cage = (age-r(mean))
. tabulate los
```

LOS	Freq.	Percent	Cum.
1	1,929	58.23	58.23
2	471	14.22	72.44
3	125	3.77	76.21
4	61	1.84	78.06
5	68	2.05	80.11
6	83	2.51	82.61
7	78	2.35	84.97
8	79	2.38	87.35
9	73	2.20	89.56
10	66	1.99	91.55
11	60	1.81	93.36
12	43	1.30	94.66
13	32	0.97	95.62
14	29	0.88	96.50
15	23	0.69	97.19
16	21	0.63	97.83
17	18	0.54	98.37
18	14	0.42	98.79
19	11	0.33	99.12
20	9	0.27	99.40
21	2	0.06	99.46
22	3	0.09	99.55
23	4	0.12	99.67
24	4	0.12	99.79
25	4	0.12	99.91
26	2	0.06	99.97
27	1	0.03	100.00
Total	3,313	100.00	

```
. summarize los
```

Variable	Obs	Mean	Std. Dev.	Min	Max
los	3313	3.297314	4.24606	1	27

We model the data using a standard Poisson regression to determine whether the data are extradispersed. Given the shape of the data, we suspect overdispersion, and we use a robust or sandwich adjustment on the standard errors. This does not alter the reported dispersion statistic; it adjusts the reported standard errors for the extradispersion.

```
. glm los cage gender type procedure, nolog family(poisson) vce(robust)
Generalized linear models                          No. of obs      =      3313
Optimization     : ML                              Residual df     =      3308
                                                   Scale parameter =         1
Deviance         =  6545.205905                    (1/df) Deviance =  1.978599
Pearson          =  7356.890182                    (1/df) Pearson  =  2.223969
Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]
                                                   AIC             =  4.587385
Log pseudolikelihood = -7594.003885                BIC             = -20268.15

------------------------------------------------------------------------------
             |               Robust
         los |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cage |    .067838   .0030246    22.43   0.000     .0619098    .0737662
      gender |  -.1706062   .0354332    -4.81   0.000    -.2400541   -.1011584
        type |   .5090647   .0361028    14.10   0.000     .4383045    .5798249
   procedure |   1.295007    .029725    43.57   0.000     1.236747    1.353267
       _cons |   .1166878   .0404083     2.89   0.004      .037489    .1958866
------------------------------------------------------------------------------

. estat ic
Akaike´s information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |   3313           .   -7594.004      5     15198.01    15228.54
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

The data are indeed overdispersed given a dispersion statistic of 2.22. Next, we model the data using a ZTP, noting that the AIC reduces from 15,198 to 13,524. The BIC statistic reduces similarly in value, indicating that a zero-truncated model is preferred.

```
. trncregress los cage gender type procedure, dist(poisson) nolog ltrunc(0)
> vce(robust)
Truncated Poisson regression                      Number of obs    =      3313
Dist. support on {1, ..., .}                      LR chi2(4)       =   7408.73
Log pseudolikelihood = -6757.134                  Prob > chi2      =    0.0000

------------------------------------------------------------------------------
             |               Robust
         los |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cage |   .0822975   .0040432    20.35   0.000      .074373    .0902221
      gender |  -.2075121   .0446871    -4.64   0.000    -.2950972   -.1199269
        type |   .6496304   .0481891    13.48   0.000     .5551815    .7440793
   procedure |   1.791857    .051333    34.91   0.000     1.691246    1.892468
       _cons |  -.5070931   .0662515    -7.65   0.000    -.6369437   -.3772425
------------------------------------------------------------------------------

. predict countpoi
(option n assumed; predicted number of events)
```

```
. estat ic
Akaike´s information criterion and Bayesian information criterion

       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |   3313    -10461.5   -6757.134      5     13524.27     13554.8

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

We next attempted a ZTNB, but the model would not converge. A negative binomial model without adjustment was used in its place.

```
. glm los cage gender type procedure, nolog family(nb ml) vce(robust)
Generalized linear models                          No. of obs      =      3313
Optimization     : ML                              Residual df     =      3308
                                                   Scale parameter =         1
Deviance         =  2566.187186                    (1/df) Deviance =  .7757519
Pearson          =  2991.058004                    (1/df) Pearson  =  .9041892
Variance function: V(u) = u+(.3359)u^2             [Neg. Binomial]
Link function    : g(u) = ln(u)                    [Log]
                                                   AIC             =  4.008383
Log pseudolikelihood = -6634.886058                BIC             = -24247.17

             |               Robust
         los |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cage |   .0697254    .002795    24.95   0.000     .0642473    .0752036
      gender |   -.250894   .0306625    -8.18   0.000    -.3109915   -.1907966
        type |   .5077546   .0312664    16.24   0.000     .4464735    .5690357
   procedure |   1.291291   .0277281    46.57   0.000     1.236945    1.345637
       _cons |   .1669493   .0342711     4.87   0.000     .0997791    .2341194

Note: Negative binomial parameter estimated via ML and treated as fixed once
      estimated.
. predict countnbr
(option mu assumed; predicted mean los)
. estat ic
Akaike´s information criterion and Bayesian information criterion

       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |   3313           .   -6634.886      5     13279.77     13310.3

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

Even without adjustment for the absence of zero counts, the AIC and BIC statistics of the negative binomial model are 250 points below that of the ZTP. We attempted to run ZTGP, NB-P, and negative binomial family models, but they also failed to converge. Only the zero-truncated PIG models converged, producing the lowest AIC and BIC values of the model—estimated to be 2,500 points less than the negative binomial. Given the parameterization of the PIG model such that there is a direct relationship between the mean and dispersion parameter, the model performs best on a distribution of counts that are shaped like the data modeled here. Note that a standard PIG model using the `pigreg` command (Hardin and Hilbe 2012) yields an AIC of 13,210.83, nearly 70 points lower than that of the negative binomial. But it is clear that a zero-truncated PIG fits the data best. See Hilbe (2014) for a detailed discussion of the PIG model.

```
. trncregress los cage gender type procedure, dist(pig) nolog ltrunc(0)
> vce(robust)

Truncated Poisson IG regression                    Number of obs   =       3313
Dist. support on {1, ..., .}                       LR chi2(4)      =    1735.61
Log pseudolikelihood =  -5028.09                   Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |               Robust
         los |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cage |   .1719878   .0072864    23.60   0.000     .1577067    .1862689
      gender |  -.9136862   .0802663   -11.38   0.000    -1.071005   -.7563671
        type |   1.085622   .0805258    13.48   0.000     .9277945     1.24345
   procedure |   3.099462   .0920468    33.67   0.000     2.919054    3.279871
       _cons |  -1.912977   .1201499   -15.92   0.000    -2.148466   -1.677487
-------------+----------------------------------------------------------------
    /lnalpha |   1.023893   .0903574                      .8467954     1.20099
-------------+----------------------------------------------------------------
       alpha |   2.784011   .2515559                      2.332161    3.323405
------------------------------------------------------------------------------

. predict countpig
(option n assumed; predicted number of events)

. generate double pigalpha = [lnalpha]_cons

. estat ic

Akaike´s information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |   3313   -5895.893    -5028.09      6     10068.18    10104.81
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

We can interpret the model coefficients more clearly if we exponentiate them. Because the PIG mean was parameterized in `trncregress` using the log link $\{\eta = \log(\mu)\}$, we can interpret the coefficients as we do the incidence-rate ratios of Poisson and negative binomial models.

```
. trncregress, eform

Truncated Poisson IG regression                   Number of obs   =       3313
Dist. support on {1, ..., .}                      LR chi2(4)      =    1735.61
Log pseudolikelihood =  -5028.09                  Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |               Robust
         los |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cage |   1.187663   .0086538    23.60   0.000     1.170823    1.204746
      gender |   .4010432   .0321903   -11.38   0.000     .3426639    .4693685
        type |   2.961282   .2384596    13.48   0.000     2.528925    3.467555
   procedure |   22.18602   2.042153    33.67   0.000     18.52375    26.57234
       _cons |   .1476403    .017739   -15.92   0.000      .116663    .1868429
-------------+----------------------------------------------------------------
    /lnalpha |   1.023893   .0903574                      .8467954     1.20099
-------------+----------------------------------------------------------------
       alpha |   2.784011   .2515559                      2.332161    3.323405
------------------------------------------------------------------------------
```

Open surgery is indeed a predictor of a greater length of stay, as is emergency admission compared with elective and being female.

Following estimation, predicted statistics can be developed to create graphics that help to assess model fit. Following Cameron and Trivedi (2013), we generated the observed and predicted probabilities of the first 10 outcomes (see figure 1). Because the outcome variable has such a large proportion of outcomes of 1 and a few very large outcomes, the models have the most difficulty fitting the distribution for small values. We can see from the listed probabilities, the comparison of BIC values, and the graph that the zero-truncated PIG model is preferred over the negative binomial and ZTP models.

```
. * NOTE: doit is a program we wrote to list out observed
. *       and predicted probabilities.  It is part of the
. *       downloaded files for those interested readers.
. doit 1 10

Outcome  Obs.      Poisson   Nbreg     PIG
1        0.5823    0.3839    0.1523    0.5720
2        0.1422    0.1971    0.0915    0.0840
3        0.0377    0.1127    0.0633    0.0405
4        0.0184    0.0759    0.0477    0.0257
5        0.0205    0.0559    0.0379    0.0182
6        0.0251    0.0424    0.0311    0.0137
7        0.0235    0.0323    0.0261    0.0107
8        0.0238    0.0246    0.0222    0.0086
9        0.0220    0.0186    0.0192    0.0071
10       0.0199    0.0140    0.0167    0.0059
```

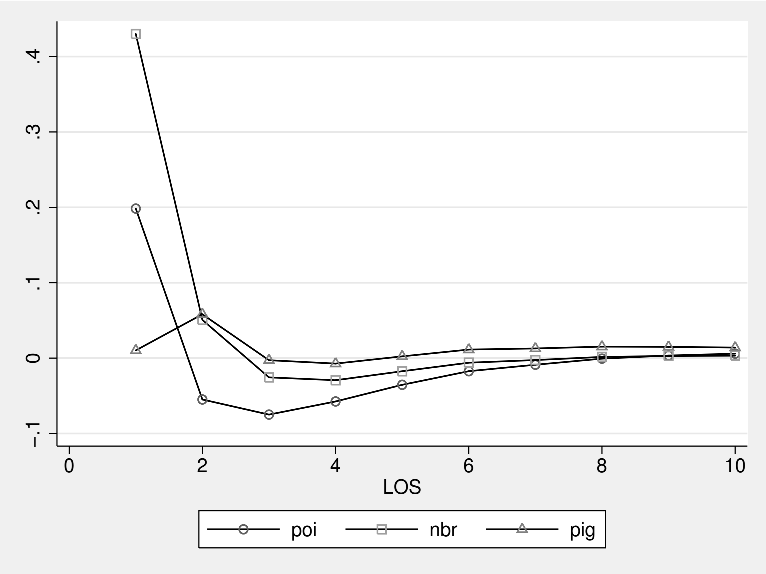

Figure 1. Comparison of differences of observed and predicted probabilities for outcomes

5 References

Cameron, A. C., and P. K. Trivedi. 2013. *Regression Analysis of Count Data*. 2nd ed. Cambridge: Cambridge University Press.

Greene, W. 2008. Functional forms for the negative binomial model for count data. *Economics Letters* 99: 585–590.

Hardin, J. W., and J. M. Hilbe. 2012. *Generalized Linear Models and Extensions*. 3rd ed. College Station, TX: Stata Press.

———. 2014. Regression models for count data based on the negative binomial(p) distribution. *Stata Journal* 14: 280–291.

Harris, T., J. M. Hilbe, and J. W. Hardin. 2014. Modeling count data with generalized distributions. *Stata Journal* 14: 562–579.

Hilbe, J. M. 2009. *Logistic Regression Models*. Boca Raton, FL: Chapman & Hall/CRC.

———. 2011. *Negative Binomial Regression*. 2nd ed. Cambridge: Cambridge University Press.

———. 2014. *Modeling Count Data*. Cambridge: Cambridge University Press.

Hilbe, J. M., and W. H. Greene. 2008. Count response regression models. In *Handbook of Statistics 27: Epidemiology and Medical Statistics*, ed. C. R. Rao, J. P. Miller, and D. C. Rao, 210–252. Amsterdam: Elsevier.

Irwin, J. O. 1968. The generalized Waring distribution applied to accident theory. *Journal of the Royal Statistical Society, Series A* 131: 205–225.

Rodríguez-Avi, J., A. Conde-Sánchez, A. J. Sáez-Castillo, M. J. Olmo-Jiménez, and A. M. Martínez-Rodríguez. 2009. A generalized Waring regression model for count data. *Computational Statistics and Data Analysis* 53: 3717–3725.

Wang, W., and F. Famoye. 1997. Modeling household fertility decisions with generalized Poisson regression. *Journal of Population Economics* 10: 273–283.

About the authors

James W. Hardin is an associate professor in the Department of Epidemiology and Biostatistics and an affiliated faculty member in the Institute for Families in Society at the University of South Carolina in Columbia, SC.

Joseph M. Hilbe is an emeritus professor at the University of Hawaii, an adjunct professor of statistics at Arizona State University in Tempe, AZ, and a Solar System Ambassador with NASA's Jet Propulsion Laboratory in Pasadena, CA.