Advanced Data Analysis

Poisson and Negative Binomial Regression

Shenyang Guo, PhD

1. The count data
2. The Poisson distribution
3. The Poisson regression model
4. The negative binomial regression model
5. Comparisons between Poisson & negbin
6. Interpretations of negbin

The count data (1)

- In this type of GLMs, we consider an outcome variable that counts the number of times an event has happened.
- Examples of count variables:

  Number of times using a service (visiting a doctor, hospitalization), daily homicides, number of nations at war, derogatory reports in an individual’s credit history, consumption of beverages, industrial injuries, the emergence of new companies, police arrests…
- The primary feature of a count variable: it has a skewed distribution.

The count data (2)

- The OLS regression model assumes a normally distributed dependent variable.
- Applying OLS regression to a count dependent variable results in inefficient, inconsistent, and biased estimates.
- There are two basic models for the count outcomes:
  - The Poisson regression model: with this model the probability of a count is determined by a Poisson distribution; the model has a defining characteristic that the conditional mean of the outcome is equal to the conditional variance;
  - The negative binomial regression model: this model is needed when the conditional variance exceeds the conditional mean;
  - Other models: zero-truncated models, the hurdle regression models, and zero-inflated models.

The Poisson distribution (1)

- Let y be a random count variable, y has a Poisson distribution with parameter $\mu > 0$ if

$$\Pr(y \mid \mu) = \frac{\exp(-\mu)\mu^{y}}{y!}$$

for y = 0, 1, 2, …

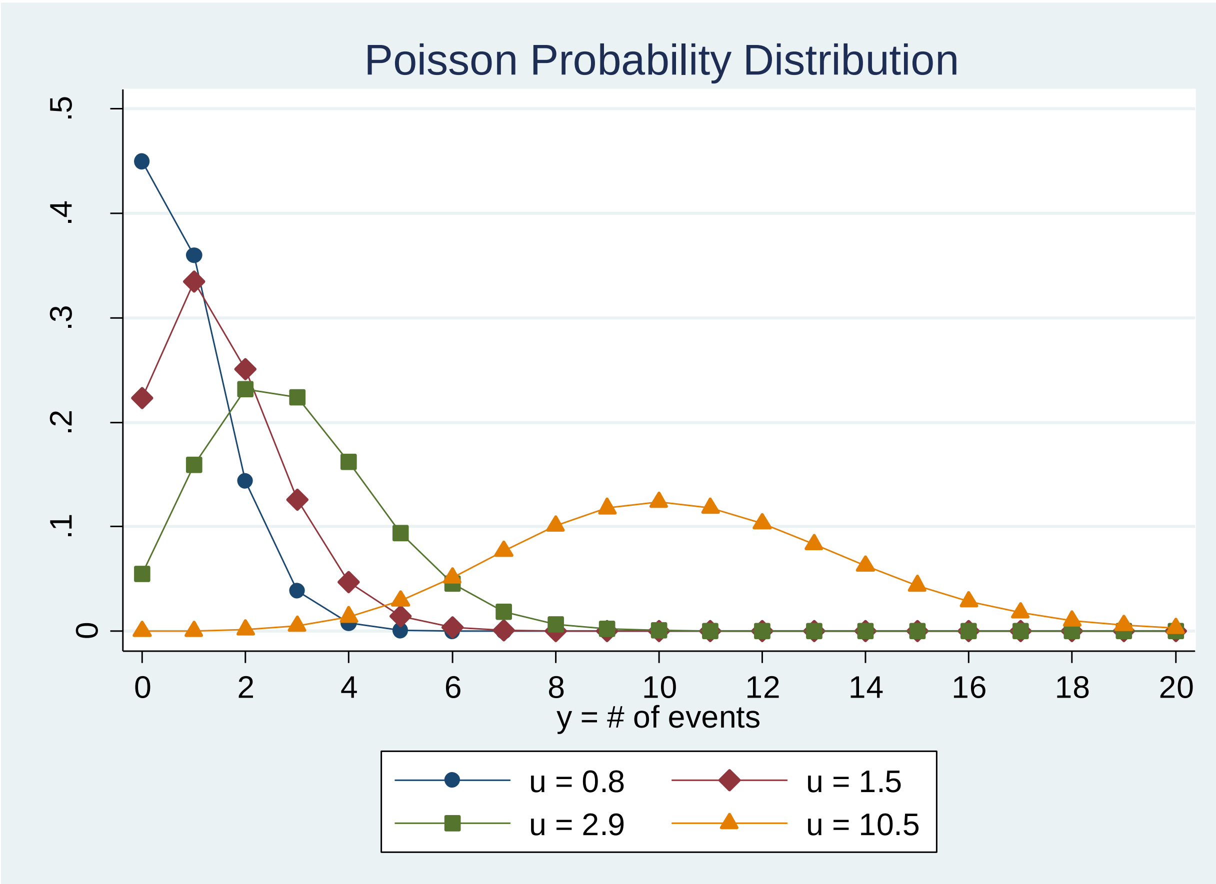
- The probability of y is a function of μ:

$\Pr(y=0 \mid \mu) = \exp(-\mu)$

$\Pr(y=1 \mid \mu) = \exp(-\mu)\,\mu$

$\Pr(y=3 \mid \mu) = \exp(-\mu)\,\mu^{3}/6$

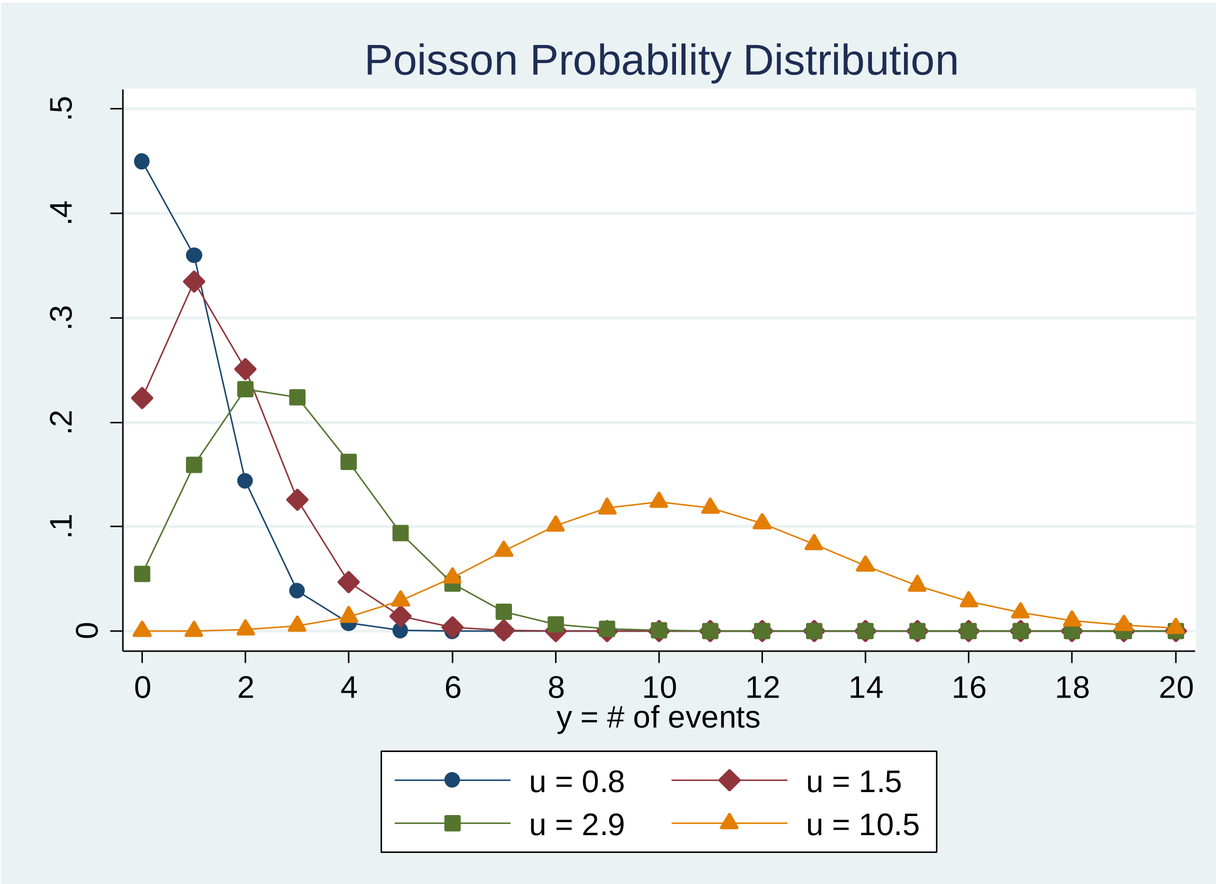
The Poisson distribution (2)



- This plot shows Pr(y) when μ=.8, 1.5, 2.9, and 10.5.
- The plot illustrates important properties of the Poisson distribution.

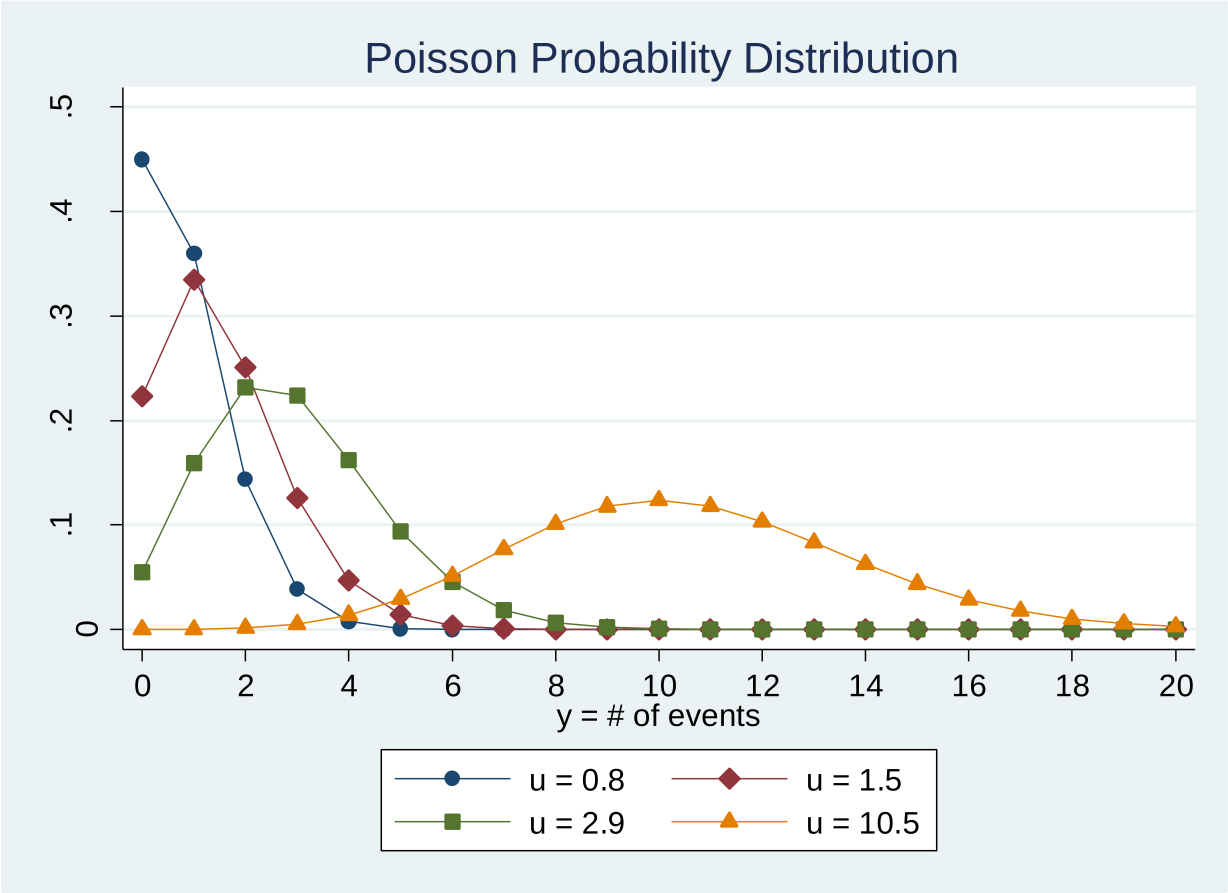
See Long & Freese (2014, p.483) for the Stata syntax that generates this figure.

The Poisson distribution (3)



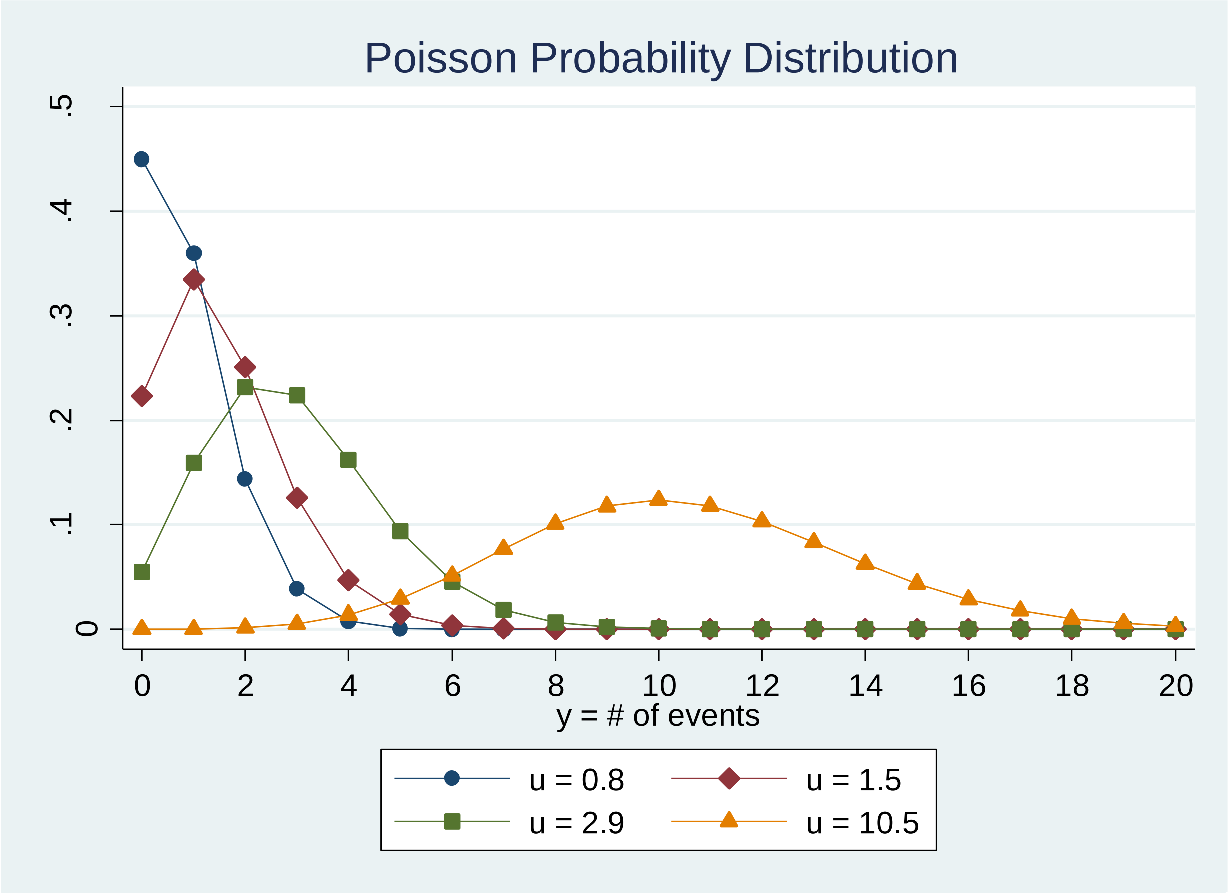
- As μ increases, the mass of the distribution shifts to the right. μ is known as the rate: it is the expected number of times that an event has occurred per unit of time. μ is also the mean or expected count.

The Poisson distribution (4)



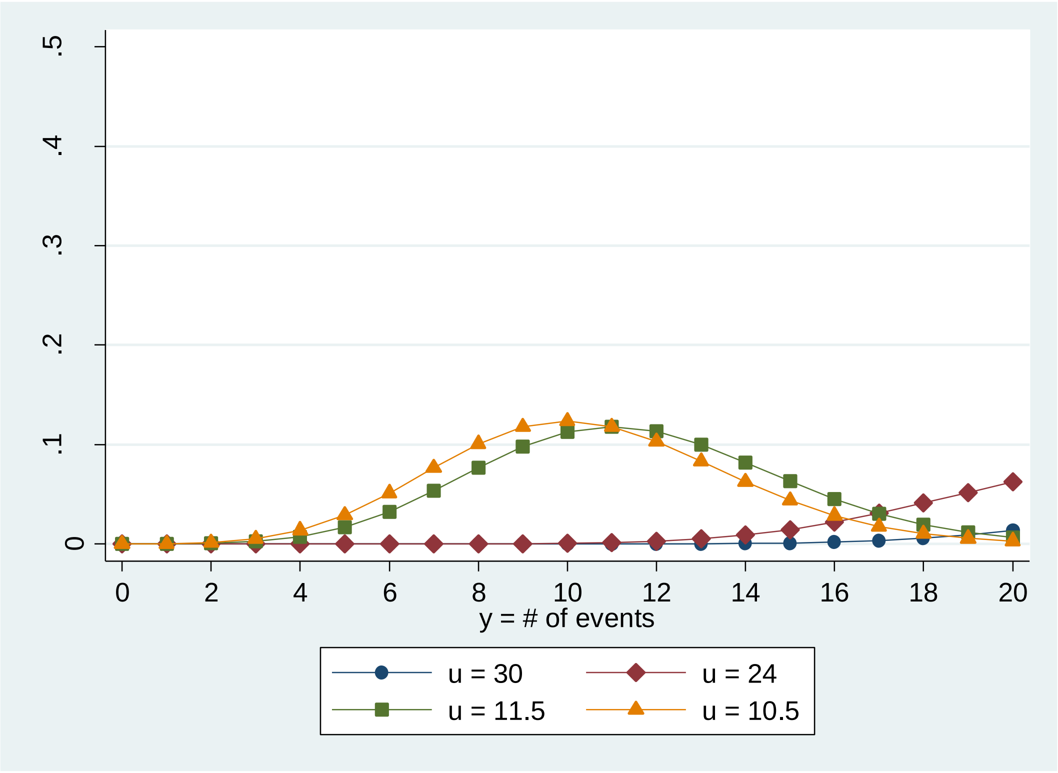
- The variance equals the mean: *Var(y)=E(y)=μ*
- This equality is known as ***equidispersion***. When the variance is greater than the mean, we say that the count variable *y* has an ***overdispersion***.
- When overdispersion occurs, you should use negative binomial model.

The Poisson distribution (5)



- As μ increases, the probability of 0's Pr(y=0) decreases.
- As μ increases, the Poisson distribution approximates a normal distribution. When E(μ)=Var(μ)=10.5, a normal distribution superimposes on the Poisson distribution.

The Poisson distribution (6)



- This figure shows how a negatively skewed Poisson distribution looks like.
- We know for this dataset Poisson approximates to a normal distribution at μ=10.5. So any μ > 10.5 creates a negatively skewed Poisson. When μ is close to 10.5 such as μ=11.5, the distribution approaches normal.

Heterogeneity

- The key idea of statistical modeling is to use independent variables to model heterogeneity. When enough observed heterogeneities are accounted for, the model would fit to the study data to an desirable degree.
- In Poisson regression, this task is known as finding important predictors of the rate of change μ (or the expected count).
- In a univariate Poisson model (i.e., use no predictors), the model often fails to account for heterogeneity, and the model-predicted outcomes are largely different from the observed outcomes.
- Failure to account for heterogeneity also leads to overdispersion. When important predictors are used but one still encounters overdispersion, he or she needs to employ negative binomial regression.

The Poisson regression model (1)

- ***The sampling model:*** The sampling of a count variable *y* follows the Poisson distribution with parameter *μ*, where

$$\mu = E(y) = Var(y)$$

- ***The link function:*** take a logarithm transformation of the parameter *μ* so that:

$$\eta_i = \ln(\mu_i) = x_i\beta = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}$$

- ***The structural model:***

$$\mu_i = E(y_i \mid x_i) = \exp(x_i\beta) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi})$$

Or $\Pr(y_i \mid x_i) = \dfrac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!}$ where $\mu_i = \exp(x_i\beta)$

The Poisson regression model (2)

- Use either one of the two equations from the structural model to express your Poisson regression when describing your analytic model.

- The second equation is also the formula for obtaining the predicted probability

$$\hat{\Pr}(y = m \mid x) = \frac{\exp(-\hat{\mu})\hat{\mu}^{m}}{m!} \quad \text{where} \quad \hat{\mu} = \exp(x\hat{\beta})$$

The Poisson regression model (3) Estimation

- The likelihood function for the Poisson regression is

$$L(\beta \mid y, x) = \prod_{i=1}^{N} \Pr(y_i \mid \mu_i) = \prod_{i=1}^{N} \frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!}$$

where $\mu = \exp(x\beta)$

- After taking logarithm, one can use Newton-Raphson method to maximize the log-likelihood function. Maddala (1983) shows the gradients and Hession for this log-likelihood function.

The Poisson regression model (4)

- Use either exp(B) or model-predicted probabilities to interpret findings of the Poisson regression.
- The exp(B) in the Poisson regression has a friendly meaning. It's not called "odds ratio", it's called "incidence-rate ratio", or IRR.
- Using the first equation of the structural model, we can derive the IRR for a change of δ in x_k: $\dfrac{E(y \mid x, x_k + \delta)}{E(y \mid x, x_k)}$

$$= \frac{\exp(\beta_0)\exp(\beta_1 x_1)\ldots\exp(\beta_k x_k)\exp(\beta_k \delta)\ldots\exp(\beta_K x_K)}{\exp(\beta_0)\exp(\beta_1 x_1)\ldots\exp(\beta_k x_k)\ldots\exp(\beta_K x_K)}$$

$$= \exp(\beta_k \delta).$$

The Poisson regression model (5)

- Interpretations:
  - For a unit change in x_k, the expected count changes by a factor of $\exp(\beta_k)$, other things being equal;
  - For every one standard-deviation unit change, approximately δ, the expected count changes by a factor of $\exp(\beta_k * \delta)$, other things being equal.
- You can also interpret the change in terms of percentage change.

The Poisson regression model (6)

- You can always use predicted probabilities to present and interpret your findings.

- Three types of predicted probabilities: average marginal effects (AMEs), marginal effects at means (MEMs), and marginal effects at representative values (MERs).

The negative binomial regression model (1)

- The Poisson regression assumes equidispersion, that is, the conditional variance equals to the conditional mean. This assumption rarely holds in practice. Therefore, to handle the problem of overdispersion, statisticians and econometricians developed a new model called negative binomial regression model or ***Negbin*** model.
- The derivation of this model dates to the work of Greenwood and Yule in 1920.
- Because the variance in this model equals to the mean of the Poisson distribution plus an overdispersion parameter multiplies by a quadratic term of the mean, Cameron and Trivedi called this model as ***Negbin 2*** model.

The negative binomial regression model (2)

- In a Poisson regression, the mean and variance of the dependent variable y are both a function of the x's and the β's:

$$\mu = E(y \mid x) = \exp(x\beta), \ Var(y \mid x) = \exp(x\beta)$$

- To address the problem of overdispersion, the negbin model adds a parameter (so called overdispersion parameter a) that allows the conditional variance of y to exceed the conditional mean.
- The motivation of adding this parameter is to model unobserved heterogeneity. In the negbin model, the mean μ is replaced with the random variable $\tilde{\mu}$:

$$\tilde{\mu}_i = \exp(x_i \beta + \varepsilon_i)$$

The negative binomial regression model (3)

- As the formula shows, the primary feature of this new model is to add a random error ε that is assumed to be uncorrelated with ***x*** (i.e., it follows the exogenous assumption).
- Doing so, the relationship between $\tilde{\mu}$ and μ is

$$\tilde{\mu} = \exp(x_i \beta)\exp(\varepsilon_i) = \mu_i \exp(\varepsilon_i) = \mu_i \delta_i$$

- When the error term ε is absent, or $\varepsilon = 0$, the expectation of δ_i becomes $E(\delta_i) = \exp(0) = 1$, so that $\tilde{\mu} = \mu_i \delta_i = \mu_i$, the negbin becomes a Poisson reg.
- You can assume different parametric distributions for the error term ε, but the most common approach is to assume a gamma distribution for the $\exp(\varepsilon_i)$ *or* δ_i with parameter v_i : $g(\delta_i) = \dfrac{v_i^{v_i}}{\Gamma(v_i)} \delta_i^{v_i - 1} \exp(-\delta_i v_i) \quad for\ v_i > 0$

The negative binomial regression model (4)

- To identify the parameters, we further assume that ν is the same for all individuals in our sample:

$$v_i = \alpha^{-1} \qquad for\ \alpha > 0.$$

- With the above constraints, we now have the expected value of *y* and its variance for the negative binominal distribution as:

$$E(y_i \mid x_i) = \exp(x_i \beta) = \mu_i,$$

$$Var(y_i \mid x) = \mu_i \left(1 + \frac{\mu_i}{\alpha^{-1}}\right) = \mu_i + \alpha \mu_i^2$$

- Note again that when $\alpha = 0$, $E(y_i \mid x_i) = Var(y_i \mid x) = \mu_i$, the negbin becomes Poisson reg.
- The variance equation $Var(y_i \mid x) = \mu_i + \alpha \mu_i^2$ leads Cameron & Trivedi (1986) call the model ***Negbin 2 model***.

The negative binomial regression model (5)

We now have a formal expression of Negbin:

- ***The sampling model:*** The sampling of a count variable *y* follows the negative binomial distribution with conditional expectation and variance:

$$E(y_i \mid x_i) = \mu_i, \quad and \quad Var(y_i \mid x) = \mu_i + \alpha\mu_i^2$$

- ***The link function:*** take a logarithm transformation of the expected count or the rate of change μ so that:

$$\eta_i = \ln(\mu_i) = x_i\beta = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_i$$

The negative binomial regression model (6)

- ***The structural model:***

$$\mu_i = E(y_i \mid x_i) = \exp(x_i \beta) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_i)$$

or

$$\hat{\Pr}(y \mid x) = \frac{\Gamma(y + \hat{\alpha}^{-1})}{y!\Gamma(\hat{\alpha}^{-1})} \left(\frac{\hat{\alpha}^{-1}}{\hat{\alpha}^{-1} + \hat{\mu}} \right)^{\hat{\alpha}^{-1}} \left(\frac{\hat{\mu}}{\hat{\alpha}^{-1} + \hat{\mu}} \right)^{y}$$

$$where \quad \hat{\mu} = \exp(x\hat{\beta})$$

The negative binomial regression model (7)

- Use either one of the two equations from the structural model to express your negative binomial regression model when describing your analytic method.

- The second equation is also the formula that can be used to obtain the model-predicted probability:

$$\hat{\Pr}(y \mid x) = \frac{\Gamma(y+\hat{\alpha}^{-1})}{y!\Gamma(\hat{\alpha}^{-1})}\left(\frac{\hat{\alpha}^{-1}}{\hat{\alpha}^{-1}+\hat{\mu}}\right)^{\hat{\alpha}^{-1}}\left(\frac{\hat{\mu}}{\hat{\alpha}^{-1}+\hat{\mu}}\right)^{y}$$

$$\textit{where} \quad \hat{\mu} = \exp(x\hat{\beta})$$

The negative binomial regression model (8)

- Because overdispersion, the standard error in both Poisson regression and negbin regression are downwardly biased. As a convention, we always use robust standard errors. That is, in Stata, use ***vce(robust)***.

The negative binomial regression model (9)

Estimation

- The likelihood function for the negbin regression is

$$L(\beta \mid y, x) = \prod_{i=1}^{N} \hat{\Pr}(y_i \mid x_i) = \prod_{i=1}^{N} \frac{\Gamma(y_i + \hat{\alpha}^{-1})}{y_i!\Gamma(\hat{\alpha}^{-1})} \left(\frac{\hat{\alpha}^{-1}}{\hat{\alpha}^{-1} + \hat{\mu}_i} \right)^{\hat{\alpha}^{-1}} \left(\frac{\hat{\mu}_i}{\hat{\alpha}^{-1} + \hat{\mu}_i} \right)^{y_i}$$

$$where \quad \hat{\mu} = \exp(x\hat{\beta})$$

- After taking logarithm, one can use Newton-Raphson method to maximize the log-likelihood function. Lawless (1987) shows the gradients and Hession for this log-likelihood function.

Comparisons between Poisson & negbin

- Below is a summary of the differences between the two models:
  - The overdispersion parameter $\alpha = 0$ for Poisson reg, but $\alpha > 0$ for negbin.
  - The Poisson regression underestimates the probability of (y=0); that is, Pr(y=0) from Poisson generally is smaller than Pr(y=0) from negbin.
  - Estimated standard errors from Poisson tends to be smaller than those from negbin, so Poisson estimates are inefficient.
- Determining which model should be used in a real data analysis is important. Use likelihood ratio test to address this question:

  $$G^2 = 2(\ln L_{\text{negbin}} - \ln L_{\text{possion}})$$

  Stata ***nbreg*** automatically reports this!

Interpretations of negbin (1)

- Use either exp(B) or model-predicted probabilities to interpret findings of the negbin regression.

- The exp(B) in the negbin regression has exactly the same meaning as that from a Poisson regression. It's the "incidence-rate ratio", or IRR.

Interpretations of negabin (2)

- Interpretations:
  - For a unit change in x_k , the expected count changes by a factor of exp(β_k), other things being equal;
  - For every one standard-deviation unit change, approximately δ , the expected count changes by a factor of exp($\beta_k * \delta$), other things being equal.
- You can also interpret the change in terms of percentage change.

Interpretations of negbin (3)

- You can always use predicted probabilities to present and interpret your findings.
- Three types of predicted probabilities: average marginal effects (AMEs), marginal effects at means (MEMs), and marginal effects at representative values (MERs).
- Use a line chart to present the impact of a continuous IV on the expected event counts, or effects of interactions.