Count Data Models

Bruno Crépon

CREST

Plan

Example of Count Data

Poisson Model

Unobserved Heterogeneity
- Negative Binomial Model
- Simulation and finite mixture

Excess Zero

Example of Count Data

Many examples

- number of doctor visit
- number of hospitalization days
- number of car accident
- number of patents

Plan

Example of Count Data

Poisson Model

Unobserved Heterogeneity
- Negative Binomial Model
- Simulation and finite mixture

Excess Zero

Poisson Model : Parametrization

Consider $y \rightsquigarrow \mathcal{P}(\lambda)$. Probabilities are defined as

$$P(y = n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

Poisson model has the property that mean and variance are equal

$$E(y) = \lambda \ V(y) = \lambda$$

One way to model the distribution of y conditional on a set of variable x is to consider

$$y|x \rightsquigarrow \mathcal{P}(exp(xb))$$

which leads to

$$P(y = n|x) = exp(-exp(xb)) \frac{exp(nxb)}{n!}$$

Notice that here also we have an index model : explanatory variables enter the distribution of y|x under the index

$$b_0 + b_1 x_1 + \ldots + x_K b_K$$

Poisson Model : Simple Property

Even if $y|x \rightsquigarrow \mathcal{P}(exp(xb))$ the unconditional distribution exhibit overdispersion

$$E(y) = E_x(E(y|x)) = E_x(exp(xb))$$

and

$$V(y) = E_x(V(y|x)) + V_x(E(y|x)) = E_x(exp(xb)) + V_x(exp(xb))$$

That is

$$V(y) = E_x(exp(xb))(1 + V_x(exp(xb))/E_x(exp(xb))$$

Thus

$$V(y)/E(y) = (1 + V_x(exp(xb))/E_x(exp(xb)) > 1$$

Poisson Model : Marginal Effect

In the conditional Poisson Model

$$E(y|x) = exp(xb)$$

Thus

$$\frac{\partial E(y|x)}{\partial x_k} = exp(xb)b_k$$

That is

$$\frac{\partial lnE(y|x)}{\partial x_k} = b_k$$

Elasticities are constant. Parameters receive therefore a simple interpretation

Poisson Model : Likelihood and estimation

As we have seen the distribution has the following expression

$$l(y|x,b) = exp(-exp(xb))\frac{exp(yxb)}{y_i!}$$

Thus the log likelihood writes as

$$\sum_i y_i x_i b - exp(x_i b) - ln(y!)$$

And first order condition writes as

$$\sum_i (y_i - exp(x_i \widehat{b}))x_i = 0$$

Simple and nice interpretation that the residual $y_i - exp(x_i \widehat{b})$ is orthogonal to the set of explanatory variables

Poisson Model : Estimation

Variance of the ML estimator is given as usual by

$$J^{-1} I J^{-1}$$

with

$$J = -E \frac{\partial^2 lnl(y|x,b)}{\partial b \partial b'} = E\left[exp(xb)xx'\right]$$

and

$$I = E \frac{\partial lnl(y|x,b)}{\partial b} \frac{\partial lnl(y|x,b)}{\partial b}' = E\left[(y - exp(xb))^2 xx'\right]$$

If we believe in the model we have $I = E[(y - E(y|x))^2 xx'] = E[V(y|x)xx']$
As $v(y|x) = exp(xb)$, we have $I = E(exp(xb)xx')$ and thus $I = J$
However we may consider also that we don't want to assume $V(y|x) = exp(xb)$ because we have not total confidence in the Poisson assumption
We will see example like that
In this case we prefer the robust formula with J and I two different matrices

Poisson Model : Stata Estimation

poisson y $x_1 \cdots x_K$

compute the ML estimate of the Poisson regression
predict command provide different predictions

predict *name*, *n*	$exp(x_i * \widehat{b})$	default
predict *name*, $pr(n)$	$p(y = n)$	n can be a scalar given or a variable One interesting case is $pr(y)$ where y is the variable to explain
predict *name*, $pr(a, b)$	$p(a \leq y \leq b)$	a and b scalars
predict *name*, xb	$x_i * \widehat{b}$	

If you want to compute the robust variance $J^{-1}IJ^{-1}$ simply enter the command

poisson y $x_1 \cdots x_K$, r

Same estimation but a different computation of stds

Count Fit

Unconditional distribution $f(y)$ can be directly estimated on the sample

$$\widehat{f_1}(y) = \overline{1(y_i = y)}$$

We have also the formula

$$f(y) = \int f(y|x, \theta) f(x) dx$$

Another way to estimate the distribution, based on the estimated model and its distributional assumption is

$$\widehat{f_2}(y) = \overline{f(y|x_i, \widehat{\theta})}$$

Computing the two estimates and comparing them provide an idea of the goodness of fit of the model

Example

Count	Actual	Predicted
0	0.109	0.007
1	0.085	0.030
2	0.097	0.063
3	0.091	0.095
4	0.092	0.116
5	0.072	0.121
Sum	0.547	0.432

can be directly implemented in stata using "countfit" *countfit* y $x_1 \cdots x_K$,

prm(for Poisson) nbreg(for Negative Binomial) maxcount(#)

Can be used with other count data model like zero inflated (see below) in this case, instead of prm enter zip or zinb instead of nbreg
Can be used to examine the fit of several models simultaneously

Poisson Model : NLSQ Estimation

We have seen that in the Poisson model we have

$$E(y|x) = exp(xb)$$

We may want to apply NLSQ based on the objective function

$$\sum_i \omega_i (y_i - exp(x_i b))^2$$

The first order condition in this case is

$$\sum_i \omega_i (y_i - exp(x_i b)) exp(x_i b) x_i = 0$$

Poisson Model : NLSQ Estimation

The variance is also of the form $J^{-1}IJ^{-1}$ with

$$J = E(\omega exp(2xb)xx')$$

and

$$I = E(\omega^2(y - exp(xb))^2 exp(2xb)xx')$$

Taking $\omega = 1/exp(xb)$ produce an estimator which is similar to the Poisson ML with the same variance
Estimation can be performed in several steps to obtain a first consistent estimation
Then compute weights and reestimate

Plan

Example of Count Data

Poisson Model

Unobserved Heterogeneity
- Negative Binomial Model
- Simulation and finite mixture

Excess Zero

Unobserved heterogeneity

One reason for bad fit of the probabilities is that true distribution is not Poisson
One possibility is unobserved heterogeneity :
Let ν be a positive variable independent from the x with $E(\nu) = 1$ and $V(\nu) = \alpha$
Consider

$$y|x, \nu \rightsquigarrow \mathcal{P}(exp(xb)\nu)$$

y is distributed as a Poisson variable, but conditional on x and ν wich is not observed
One possible explanation for this is the fact that the true model involves additional variables z

$$y|x, z \rightsquigarrow \mathcal{P}(exp(xb + zc))$$

and assume z is not observed, then $\nu = exp(zc)$ is the unobserved heterogeneity term

Effect of unobserved heterogeneity

The question is now how to deal with this unobserved heterogeneity term
The true distribution is simply

$$P(y = n|x, \nu) = exp(-exp(xb)\nu)(exp(xb)\nu)^y/y!$$

But we just want the distribution of $y|x$.
The first thing to note is that if

$$y|x, \nu \rightsquigarrow \mathcal{P}(exp(xb)\nu)$$

then fitting a Poisson Model assuming

$$y \rightsquigarrow \mathcal{P}(exp(xb))$$

is not apropriate
There may likely be some bias in the estimated parameters or in the standard errors

Three estimation methods

There are different possibilities to deal with unobserved heterogenity

- One first possibility is to specify the distribution of ν, say $g(\nu)$
  The distribution of $y|x$ is thus

$$P(y = n|x) = \int exp(-exp(xb)\nu)\frac{(exp(xb)\nu)^y}{y!}g(\nu)d\nu$$

  This expression can be computed
  Using this new expression it is possible to estimate using ML
  Interest is accuracy, problem is robustness (what about if the choice of g is wrong ?)

Three estimation methods

- One issue is about how to proceed to the integration in practice
  There are in this case several possibilities also
  - The first one is to have a specific choice of g that allows to do the computation
  - The second one is to specify and proceed to approximative integration by simulation
  - The last one is to use mixture model, assuming the distribution of the unobserved heterogeneity term is discrete
- Do not specify the distribution and estimate using NLSQ or Pseudo Maximum Likelihood. Robust but potential efficiency loss

Specify and integrate

One first possibility to deal with unobserved heterogeneity is to specify g as a Gamma distribution

$$g(\nu) = \nu^{\frac{1-\alpha}{\alpha}} e^{-\frac{\nu}{\alpha}} \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha}} / \Gamma(\frac{1}{\alpha})$$

Such a distribution has the following moments : $E(\nu) = 1$, $V(\nu) = \alpha$
Let us consider the case where $y \rightsquigarrow \mathcal{P}(\lambda\nu)$. It turns out that in this case integration is possible
We obtain the following distribution :

$$f(y, \lambda, \alpha) = \frac{\Gamma(\frac{1}{\alpha} + y)}{\Gamma(\frac{1}{\alpha})\Gamma(y+1)} \left(\frac{\frac{1}{\alpha}}{\frac{1}{\alpha} + \lambda}\right)^{\frac{1}{\alpha}} \left(\frac{\lambda}{\frac{1}{\alpha} + \lambda}\right)^{y}$$

We can compute the moments of this distribution :

$$E_{\lambda,\alpha}(y) = \lambda \text{ and } V_{\lambda,\alpha}(y) = \lambda(1 + \alpha\lambda)$$

As expected, the distribution exhibit overdispersion

Specify and integrate

The previous density is called Negative Binomial

Parametrization can be done in two ways :

$\lambda = exp(xb)$ and α a scalar : this leads to the so called "Negative Binomiale II" which is widely used

We have in this case $E(y|x) = exp(xb)$ and $V(y|x) = exp(xb)(1 + \alpha exp(xb))$

$\lambda = exp(xb)$ and $\lambda\alpha = \delta$ a scalar : this leads to the so called "Negative Binomiale I"

We have in this case $E(y|x) = exp(xb)$ and $V(y|x) = exp(xb)(1 + \delta)$

Specify and integrate

Consider for the previous distribution the log density term involving parameter λ

$$lnf(y,\lambda,\alpha) = ylnl\lambda - (\frac{1}{\alpha} + y)ln(\frac{1}{\alpha} + \lambda) + g(y,\alpha)$$

derivative with respect to λ yield :

$$\frac{y}{\lambda} - \frac{\frac{1}{\alpha} + y}{\frac{1}{\alpha} + \lambda} = \frac{y - \lambda}{\lambda(1 + \alpha\lambda)}$$

Thus in the case of the Negative binomial II model we obtain as first order condition for b

$$\sum_i \frac{y_i - exp(x_i b)}{exp(x_i b)(1 + \alpha exp(x_i b))} x_i = 0$$

We have here also a simple interpretation in term of orthogonality of the explanatory variables to the residual normalized by the variance

Stata Implementation

This is very simple to use with stata

command is simply

$$nbreg\ y\ x_1 \cdots x_K,\ vce(r)$$

Only fits Negative Binomiale II model

Specify and simulate

The previous choice of a gamma function is related to feasibility of integration. We may prefer have a choice independent and then try to manage integration. The quantity we want to compute is

$$\int f(y|x,b,\nu)g(\nu,\sigma)d\nu = E_\nu(f(y|x,b,\nu)|x)$$

One possibility is the following instead of integrating simulate the integral : For each observation do H draws in the chosen distribution and the compute

$$\widehat{f}(y|x,b,\sigma) = \frac{1}{H}\sum_{h=1}^{H} f(y|x,b,\nu_h,\sigma)$$

where σ is a dispersion parameter for the heterogeneity term (same as α)
By the law of large number this quantity will converge to the desired integral
One can show that optimizing the corresponding simulated log likelihood yield consistent estimates

Finite Mixture

Let us consider a slightly different case where ν is entered into the index as an additional variable ϵ

Any model we may want to consider is of the index form

$$f(y|x,\epsilon) = f(y, xb + \epsilon, \alpha)$$

ϵ is the new heterogeneity term coming in place of ν. Assume it can take a finite number of values

$$\epsilon \in \{\eta_1, \ldots, \eta_J\}$$

and that they can each occur with probabilities

$$\pi_j = P(\epsilon = \eta_j)$$

In this case the likelihood is

$$f(y|x) = \sum_{j=1}^{J} \pi_j f(y, xb + \eta_j, \alpha)$$

The objective function is in this case

$$\sum_i ln \left(\sum_{j=1}^{J} \pi_j f(y_i, x_i b + \eta_j, \alpha) \right)$$

Stata Implementation

This is very simple to use with stata

Command is simply

$$fmm\ y\ x_1 \cdots x_K,\ \ vce(r)\ components(number)\ mixtureof(density)$$

density can be

negbin1	Negative Binomial 1
negbin2	Negative Binomial 2
poisson	Poisson

Do not specify and estimate without ML

One idea would be to do as if the distribution was a Poisson, i.e. optimizing the Poisson ML objective

But do not assuming the true distribution is a Poisson. Not assuming this amount to use $J^{-1}IJ^{-1}$ instead of simply J^{-1}.

This is here an example of a more general estimation method is called Pseudo Maximum Likelihood. It amounts here to simply base estimation on the fact that $E(y|x) = exp(xb)$
The problem with this choice is that there are efficiency losses. However the gain is that the method is robust

One other possibility strongly related is to use NLSQ, here simple NLSQ would lead to even stronger efficiency losses

Plan

Example of Count Data

Poisson Model

Unobserved Heterogeneity
- Negative Binomial Model
- Simulation and finite mixture

Excess Zero

Excess zero two models

Frequently count data exhibit a large number of zero and it is difficult to account for them even with the previous modified distribution.

This leads to specific models

There are mainly of two types :

- Zero Inflated models
- Hurdle models

Excess zero : Zero Inflated models

Zero inflated models correspond to the case where you consider first that individual report or not the events and then if they decide to report the events the number

Typical case is patents. Patents are a measure of invention. You may consider that inventions follow a Poisson or Negative Binomial variable

However firms may decide to protect their invention through patents or through other means. Something you do not observe

What you observe is the resulting number of patents only

There are in this case two types of zeros : zeros because of n report and zeros because of no invention

Excess zero : Zero Inflated models

First is the decision to report, unobservable. Modeled using the tules for binary outcomes.
Let the probability to report be $R(zc)$

Second you have the process itself $f(n, xb)$

Thus the probabilities to observe n event is the following :

$$P(y = 0|x, z) = 1 - R(zc) + R(zc) \times f(0, xb)$$

and

$$P(y = n|x, z) = R(zc) \times f(n, xb) \ \ \forall \ n > 0$$

Stata Command

Syntax is as follows

Zero Inflated Poisson
zip y $x_1 \cdots x_K$, *inflate*($z_1 \cdots z_K$) *(add r for robust)*

Zero inflated Negative Binomial
zinb y $x_1 \cdots x_K$, *inflate*($z_1 \cdots z_K$) *(add r for robust)*

Excess zero : Hurdle models

Hurdle model amount to assume tat the process governing the zeros is not the same as the process governing the strictly positive outcomes.

Assume zeros are for a count model $f_0(n, x_0 b_0, \alpha_0)$ and strictly positive outcomes governed by another process $f_1(n, x_1 b_1, \alpha_1)$

Probabilities are given by :

$$P(y = 0|x_0, x_1) = f_0(0, x_0 b_0, \alpha_0)$$

and

$$P(y = n|x_0, x_1) = \frac{1 - f_0(0, x_0 b_0, \alpha_0)}{1 - f_1(0, x_1 b_1, \alpha_1)} f_1(n, x_1 b_1, \alpha_1) \quad \forall n > 0$$

Excess zero : Hurdle models Stata estimation

There are no direct available estimation of hurdle models in Stata
However there are estimation methods available to estimate zero truncated Poisson or Negative Binomial models
Estimation is based on the truncated sample $y > 0$. The distribution is

$$f(y|x, \theta, y > 0) = \frac{f(y|x, \theta)}{1 - f(0|x, \theta)}$$

Commands are the following
ztp $y\ x_1 \cdots x_K$ *if* $y > 0$*(add ","r for robust)*
ztnb $y\ x_1 \cdots x_K$ *if* $y > 0$*(add ",r" for robust)*
Can be combined with a first estimate on the binary variable $1(y > 0)$ using different model (logit, probit, cloglog)

Excess zero

Both type of model amount to change the distribution for the probability to observe 0 event
Changes are just functional form changes
Also includes another index of explanatory variables One other issue is about identification
Estimation mainly related to choices of functional form