An introduction to count data models

Xiang Ao

March 4, 2009

1 Poisson Model

If the interested variable is a count data variable, it is natural to model it as a Poisson process. The number of occurance follows a Poisson process:

$$\Pr[Y = y] = \frac{e^{-\mu}\mu^y}{y!} \tag{1}$$

One of Poisson's properties is that it has equal variance as its mean:

$$\mathrm{E}[Y] = \mathrm{Var}[Y] = \mu \tag{2}$$

In a count data model such as Poisson regression model, we are modeling log of the expected counts:

$$\log(\mathrm{E}(Y)) = \mathbf{X_i'}\beta \tag{3}$$

The log-likelihood function is

$$\log L(\beta) = \sum_{i=1}^{N} -\exp(X_i'\beta) + y_i X_i'\beta - \ln y_i! \tag{4}$$

To maximize it, the first order condition is

$$\sum_{i=1}^{N} (y_i - \exp(X_i'\beta))X_i' = 0 \tag{5}$$

We can use Newton-Raphson or other optimization algorithms to find the solution for the first order condition.

2 QMLE Poisson

2.1 Model

Note that in equation 5, if X_i' includes a constant, then $(y_i - \exp(X_i'\beta))$ sums to zero. If $\mathrm{E}[y_i|X_i] = \exp(X_i'\beta)$, then the summation on the left-hand side has expectation of zero. Hence the only specification needed to apply equation 5 is the conditional expectation of Y given X. Even the data is not Poisson-distributed, the estimator by equation 5 is still consistent. Therefore, this estimator is called quasi_ML (QML) Poisson estimator.

Although the QML Poisson estimator is consistent under relatively weak condition, the regular variance estimator for the coefficients are not valid anymore.

If a stronger assumption is made that the data follows a Poisson distribution, then the error term has a variance which equals to the mean of Y. The estimator $\hat{\beta}_P$ (ML estimator) follows a normal distribution asymptotically with a variance matrix

$$\mathrm{Var}_{ML}[\hat{\beta}_P] = (\sum_{i=1}^{N} \mu_i X_i X_i')^{-1} \tag{6}$$

On the other hand, the variance matrix for the QML Poisson estimator $\hat{\beta}_{QML}$ is

$$\mathrm{Var}_{QML}[\hat{\beta}_P] = (\sum_{i=1}^{N} \mu_i X_i X_i')^{-1} (\sum_{i=1}^{N} \omega_i X_i X_i')^{-1} (\sum_{i=1}^{N} \mu_i X_i X_i')^{-1} \tag{7}$$

where $\omega_i = \mathrm{Var}[y_i|X_i]$ is the conditional variance of y_i.

2.2 Implementation

Poisson estimation is implemented in almost every statistical package. However, some of them may not work if you have a continuous dependent variable.

Stata's implementation of Poisson model: “poisson” and “xtpoisson” do take continuous dependent variable. Note that standard errors reported without any specification in “vce” will be likely downward biased. Therefore, always use “robust” option at least. Bootstrapped standard errors are also preferred. However, if you intend to use it as QMLE-Poisson, standard errors need to be adjusted. Those two procedures do not adjust for standard errors. A user-written program called xtpqml calls for xtpoisson and it calculates robust standard error which is suggested by Wooldridge (1999).

3 Fixed-effect Poisson model

We specify a panel data count model as:

$$E[y_{it}|\alpha_i, X_{it}] = \alpha_i \exp(X_{it}'\beta) = \exp(\gamma_i + X_{it}'\beta). \tag{8}$$

It turns out that for fixed-effect Poisson model, just like in linear regression case, there is NO incidental parameter problem. Poisson MLE is the same for conditional and unconditional likelihood.(See Cameron and Trivedi, 2005, Page 805).

Therefore, the coefficient estimates are the same for conditional or unconditional fixed-effect Poisson model. That is, in Stata, “xtpoisson, fe” will return the same results as “xi: poisson i.group”, is the same as “xtpqml, fe”. The only difference is that “xtpqml, fe” returns “correct” standard errors for QMLE Poisson. You can also calculate standard errors with the other commands, using clustered standard errors, or bootstrapped standard errors.

Conditional fixed effect negative binomial model is also possible to be consistent for some specifications. But Poission fixed effect is more popular since it is QMLE.

4 Negative Binomial model

The Poisson model has a restrictive property that the conditional variance equals the conditional mean. Often we'll see “overdispersion”, that is, the variance exceeds the mean, in real data. The negative binomial model is a generalization of poisson model which allows for overdispersion by introducing an unobserved heterogeneity term. In a negative binomial model,

$$\mathrm{E}[Y] = \mu\tau = e^{X'\beta} \cdot \tau \tag{9}$$

This extra τ term follows a $Gamma(\theta, \theta)$ distribution, with $E(\tau) = 1$ and $Var(\tau) = 1/\theta$.

Conditional on X and τ, the distribution of Y is still poisson:

$$\Pr[Y = y|X, \tau] = \frac{e^{-\mu\tau}(\mu\tau)^y}{y!} \tag{10}$$

Conditional on X only, the distribution of Y is negative binomial:

$$\Pr[Y = y|X] = \frac{\theta^\theta \mu^y \Gamma(\theta + y)}{\Gamma(y+1)\Gamma(\theta)(\mu + \theta)^{\theta + y}} \tag{11}$$

This distribution still has conditional mean of μ, but variance is $\mu(1 + (1/\theta)\mu)$.

When $\alpha = 1/\theta$ approaches zero, the negative binomial model converges to poisson model.

5 Models for truncated counts

In some situations, all we observe are non-zero counts, because of the way data was collected. For example, we only observe people's visits to a park, only if they visit. To model how many times they visit, based on this kind of data set, we need to use zero-truncated count models.

$$\Pr[Y = y|y > 0, X] = \frac{\Pr(y|X)}{\Pr(y > 0|X)} = \frac{\Pr(y|X)}{1 - \exp(-\mu)} \tag{12}$$

Thus we are modeling the counts given that we have a positive outcome. Both zero-truncated poisson ("ztp" in Stata) and zero-truncated negative binomial ("ztnb" in Stata) are estimated by Maximum-likelihood method.

We need to pay extra attention to the "over-dispersion" problem when dealing with zero-truncated data. When data is not truncated, coefficients estimated by poisson model is consistent, even if data is "overdispersed". However, if we have zero-truncated sample, then "over-dispersion" results in inconsistent coefficient estimation (Grogger and Carson, 1991). Therefore, "over-dispersion" needs to be tested before using truncated poisson model or truncated negative binomial model.

6 The hurdle regression model

Although negative binomial model relaxes the assumption of equal mean and variance, some researchers prefer modeling the process of generating zeros differently from the process of generating other values. Suppose zero counts are generated by a binary process:

$$\Pr[y = 0|X] = \frac{\exp(X\gamma)}{1 + \exp(X\gamma)} = \pi \tag{13}$$

Positive counts are generated by a truncated count process, such as in equation 12.

In this model, zero is a "hurdle" to get past before reaching positive counts. The hurdle model is estimated by two seperate equations. It is easy to estimate. However, the marginal effect is tricky to calcuate, since and independent variable, if apprearing in both equations, will have effect through both equations.

$$\mathrm{E}(y|X) = [\pi \times 0] + (1 - \pi) \times \mathrm{E}(y|y > 0, X) = (1 - \pi) \times \mathrm{E}(y|y > 0, X) \tag{14}$$

In stata, there are commands such as "hplogit", which is a hurdle model with first equation logit, and second equation truncated poisson. It is the same as two seperate estimations; namely "logit" first, then "ztp" on positive counts.

The difference between hurdle model and a Heckman selection model is that in a Heckman model, two equations need to be jointly estimated. There must be exlusion condition to make the selection equation identifiable. That is, the second stage equation need to take account of the selection bias; therefore, some instrument for selection is needed for the first stage equation. On the other hand, a hurdle model assumes that the two processes are seperate; there is no selection.

7 Zero-inflated count models

A zero-inflated count model also models two different processes. One is a binary process: generating zero or non-zeros. The second one is a count model. Note there is a different between zero-inflated and hurdle model: In hurdle model, the second process is a truncated count model (positive counts). In a zero-inflated model, a zero can come from either the binary process, or from the count process.

A second difference is that the estimation of zero-inflated model is a joint estimation of the two processes.

$$y_i \sim \begin{cases} 0 \text{ with probability } \psi \\ g(y_i|X_i)\text{with probability } 1-\psi \end{cases} \tag{15}$$

Then

$$P(Y_i = y_i|X_i, Z_i) = \begin{cases} \psi(\gamma' Z_i) + (1-\psi(\gamma' Z_i))g(0|X_i) \text{ if } y_i = 0 \\ (1-\psi(\gamma' Z_i))g(y_i|X_i) \text{ if } y_i > 0 \end{cases} \tag{16}$$

This says that every non-zero observation follows a count process $g(y_i|X_i)$. Every zero observation has two possible sources: from a binary process (with probability $\psi(\gamma' Z_i)$) and from a count process (with probability $1 - \psi(\gamma' Z_i)$ into the count process and then with probability of $g(0|X_i)$ to be a zero). The count process can be a Poisson process (“zip” in Stata) or a Negative Binomial process (“zinb” in Stata).