

NEGATIVE BINOMIAL MODEL AS AN ALTERNATIVE TO POISSON MODEL FOR POSITIVE CONTAGION-LIKE DATA

Adebowale Olusola Adejumo
Department of Statistics,
University of Ilorin,
Ilorin, Nigeria.
ao123adejumo@yahoo.co.uk

Abstract

Positive contagion in the context of this paper refers to a situation when one event increases the likelihood of another, consequently increases the variance of the observed counts and leads to over dispersion. Positive contagion has two major adverse effects, first the summary statistics have a larger variance than anticipated under the simple model, and secondly, there is a possible loss of efficiency in using statistics appropriate for the assumed distribution. To take care of these effects we propose Negative binomial distribution which provides an example of a variance function containing an unknown parameter that is not a dispersion parameter. Negative binomial is used to measure positive contagion which leads to over dispersion. In comparison with Poisson model, we observed that Negative binomial has a better fit for data with positive contagion.

Keywords: *Positive Contagion, over dispersion, Negative Binomial.*

1. Introduction

In econometrics, positive contagion is broadly defined as the transmission of shocks (or crises) across markets (or countries). These crises may be transmitted in two qualitative different ways: either through stable cross-country linkages (or channels) or through sudden changes (or shifts) in these linkages of varying persistence (Forbes and Rigobon, 2001, King and Wadhwani, 1990). Also Ciccarelli and Rebucci (2003) proposed a way of modeling this sudden change using a Bayesian time-varying coefficient model estimated with Markov Chain Monte Carlo methods (MCMC). This proposed measure worked in the joint presence of heteroscedasticity and omitted variables, which does not require the knowledge of the crises timing.

In this paper, we define positive contagion as a situation when one event increases the likelihood of another. Positive contagion is the one that increases the variance of the observed counts which leads to over dispersion. Over dispersion means the data show evidence that the variance of the response y_i is greater than it should be under the given model. Under

dispersion is also theoretically possible, but rare in practice. McCullagh and Nelder (1989) say that over dispersion is the rule rather than the exception. Deviance and Pearson Chi-square divided by the degrees of freedom are used to detect over dispersion or under dispersion in the Poisson regression. Values greater than 1 indicate over dispersion, that is, the true variance is bigger than the mean, values smaller than 1 indicate under dispersion where the true variance is smaller than the mean. Evidence of under dispersion or over dispersion indicates inadequate fit of the Poisson model. We can test for over dispersion with a likelihood ratio test based on Poisson and Negative binomial distribution. This tests equality of the mean and the variance imposed by the Poisson distribution against the alternative that the variance exceeds the mean. For the negative binomial distribution, the variance equals the mean plus k multiplied by mean square. Plackett (1981). This k is not a dispersion parameter. Estimates of k for single samples and for several samples have been discussed by Anscombe (1949). Many authors like Cox (1983), Breslow (1984), Firth (1987), Cameron and Trivedi (1990), and others have discussed over dispersion. In samples of a Poisson or multinomial distribution, it may occur that the elements show a larger variance than that given by the distribution. This may be due to a violation of the assumption of independence, as, for example, a positive correlation in the sample elements. A frequent cause for this is the cluster structure of the sample.

In herd animals, where you see one, you will likely see more. This fact suggests that counts of antelope will probably violate the independence assumption. One event increases the likelihood of another. This is an example of positive contagion. The effect of this positive contagion is that we will have greater numbers of higher and low counts. Fitting a Poisson model that has equal mean and variance denoted as λ to such data implies that, we will effectively be requiring the variance to be less than it should really be. Consequently the true variability in the data will be underestimated. Lastly, this will lead us to underestimating the standard errors, and so to overestimating the degree of precision in the coefficients. Conversely, when observing animals like foxes, when one sights a fox, it is unlikely that you will see another fox immediately. This phenomenon is called negative contagion which eventually degenerated to under dispersion.

In this paper, we focus on data with positive contagion and fit negative binomial and Poisson models. In sections 2 and 3, we have the models with their properties and methods of fittings as well as estimation of parameters respectively. In section 4, we have some empirical examples while the discussion of results and possible conclusions are given in section 5.

2. Models and their Properties

The Poisson model is used for counts of events that occur randomly over time when outcomes in disjoint periods are independent. The Poisson distribution may be interpreted as the limit distribution of the binomial distribution $B(n;p)$ if $\lambda=np$ is fixed for increasing n . If y is a random variable with a Poisson distribution then

$$f(y) = \frac{\lambda^y e^{-\lambda}}{y!} \quad y = 0, 1, \dots \quad (2.1)$$

$$E(Y) = V(Y) = \lambda.$$

Suppose that $y \sim \text{Poisson}(\lambda)$, but λ itself is a random variable with a gamma distribution.

That is, $y/\lambda \sim \text{Poisson}(\lambda)$ and $\lambda \sim \text{Gamma}(\alpha, \beta)$,

$$P(\lambda) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\lambda/\beta), \quad \lambda > 0 \quad (2.2)$$

It can be shown that the unconditional distribution of y is negative binomial i.e.

$$f(y) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha) y!} \left(\frac{\beta}{1 + \beta} \right)^y \left(\frac{1}{1 + \beta} \right)^\alpha, \quad y = 0, 1, \dots \quad (2.3)$$

This distribution has mean

$$E(Y) = E[E(Y/\lambda)] = E(\lambda) = \alpha\beta$$

and variance

$$\text{Var}(Y) = E[\text{var}(Y/\lambda)] + \text{var}[E(Y/\lambda)]$$

$$\text{Var}(Y) = \text{var}(\lambda) + E(\lambda)$$

$$\text{Var}(Y) = \alpha\beta^2 + \alpha\beta.$$

For building a regression model, it is natural to express the negative binomial distribution in terms of the parameters $\mu = \alpha\beta$ and $\alpha = 1/k$, so that

$$E(Y) = \mu \text{ and}$$

$$\text{Var}(Y) = \mu + k\mu^2 \quad y = 0, 1, 2, \dots \quad (2.4)$$

Notice that the variance function is quadratic. Therefore the distribution of y is then

$$f(y) = \frac{\Gamma(k^{-1} + y)}{\Gamma(k^{-1}) y!} \left(\frac{k\mu}{1 + k\mu} \right)^y \left(\frac{1}{1 + k\mu} \right)^{1/k}, \quad y = 0, 1, 2 \quad (2.5)$$

which approaches $\text{Poisson}(\mu)$ as $k \rightarrow 0$. k is not a dispersion parameter, large value of k corresponds to greater amounts of over dispersion. Anscombe (1949) had worked on estimation of parameter k for single samples and for several samples. See also Santos-Silva

(1997), Habb and McConnell (1996) and Creel and Loomis (1992) and Adejumo (2002) for more on negative binomial models.

3. Estimation of parameters and model fitting.

Based on Generalized Linear model (GLM) approach, let us assume that

$$g(\mu_i) = \eta_i = x_i^T \beta, \quad (3.1)$$

where g is a link function, x_i contains covariates. The most common choice for g is the log link, see McCullagh and Nelder (1989) and Fahrmeir and Tutz (2001).

The log likelihood function is

$$l(\theta, \phi) = \sum_{i=1}^n \left(\frac{(y_i \theta_i - b(\theta_i))}{\phi} + c(y_i, \phi) \right) \quad (3.2a)$$

where θ subsumes all of the θ_i . It could also be written as a function of β and ϕ because (given the x_i), β determines all the θ_i . The main way of maximizing β is by maximizing (3.2). The fact that $G(\mu_i) = x_i \beta$ suggests a crude estimate: regress $G(y_i)$ on x_i , perhaps modifying y_i in order to avoid violating range restrictions (such as taking $\log(0)$), and accounting for the differing variances of the observations.

Fisher scoring iteration is the widely used technique for maximizing the GLM likelihood over β .

Poisson Models

Ignoring constants terms, the log likelihood function for Poisson (λ) is

$$l = \sum_{i=1}^n (y_i \log(\lambda_i) - \lambda_i) \quad (3.2b)$$

And the basic step of Fisher scoring iteration is

$$\beta^{(k+1)} = \beta^k - \left(E \left(\frac{\partial^2 l}{\partial \beta \partial \beta'} \right) \right)^{-1} \frac{\partial l}{\partial \beta} \quad (3.3)$$

which can also be written as,

$$\beta^{(k+1)} = \beta^k + \left(-E \left(\psi'(\beta^{(k)}) \right) \right)^{-1} l'(\beta^{(k)}) \quad (3.4)$$

where l is the loglikelihood function for the entire sample $y_1, \dots, y_N$ and the expectations are taken with $\beta = \beta^{(k)}$. This algorithm is the same as Newton-Raphson.

Fisher scoring simplifies to

$$\beta^{(k+1)} = (XWX)^{-1} XWu \quad (3.5)$$

where W is a diagonal matrix with

$$W = \text{Diag} \left[\text{var}(y_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^2 \right]^{-1}$$

and

$$u = \eta + \left(\frac{\partial \eta}{\partial \mu} \right) (y - \mu).$$

is the working variate.

Because $\eta_i = \log(\mu_i)$, it follows that

$$\frac{\partial \eta_i}{\partial \mu_i} = \frac{1}{\mu_i}, \text{ so that } W_i = \text{Diag}(\mu_i), u_i = \eta_i + \frac{(y_i - \mu_i)}{\mu_i}$$

Fisher scoring may also be written as

$$\beta^{(k+1)} = \beta^{(k)} + (X^T W X)^{-1} X^T W u_i^*$$

where.

$$u_i^* = \eta_i + \frac{(y_i - \mu_i)}{\mu_i}$$

We then regress u on X using the u_i 's as weights to get the new estimate of β and continue this processes until the value of β converges.

Negative Binomial Model

The log likelihood contribution of a single observation is

$$l = \log \Gamma(k^{-1} + y) - \log \Gamma(k^{-1}) + y \log \left(\frac{k\mu}{1 + k\mu} \right) + k^{-1} \log \left(\frac{1}{1 + k\mu} \right) - \log(y!) \quad (3.6)$$

For a fixed value of k , this has the form of a GLM. Therefore we can fit it by Fishers scoring, iteratively re-weight least square (IRWLS) method. One iteration is

$$\beta = (X^T W X)^{-1} X^T W u,$$

where the weights are

$$W = \text{Diag} \left[\text{var}(y_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^2 \right]^{-1}$$

and

$$u = \eta + \left(\frac{\partial \eta}{\partial \mu} \right) (y - \mu).$$

is the working variate. Recall from equation 2.4 that, $\text{Var}(y_i) = \mu_i + k\mu_i^2$, so under log link

$$\frac{\partial \eta}{\partial \mu} = \frac{1}{\mu}$$

the weights become

$$w_i = \frac{\mu_i^2}{\mu + k\mu_i^2},$$

and the working variate is

$$u_i = X_i' \beta + \frac{y_i - \mu_i}{\mu_i}$$

In most applications k will be unknown. We can find the ML estimate for k by a grid search. For each value of k along a grid, we can find the maximized value of the log likelihood. Plotting the log likelihood versus k will allow us to identify the ML estimate $\hat{k}$. The estimated $\hat{\beta}$ at $k = \hat{k}$ is the ML estimate for β .

The ML estimates for k and β are asymptotically uncorrelated. This means that uncertainty in the estimate of k does little to affect the precision of $\hat{\beta}$. If we were to use expectation rather than observed information, it would make no difference whether we regard k as fixed at $\hat{k}$ (Anscombe (1949), Planket (1981)).

4. Examples

Examples with generated values

In order to illustrate these processes, we generate N (which is the number of cells) observations (y_i) which are assumed to be distributed as a negative binomial with an assumed k value and μ_i such that,

$$y_i \sim \text{Negbin}(\mu_i, k)$$

and

$$\log(\mu_i) = \beta_0 + \beta_1 x_i$$

where β_0 and β_1 are fixed a priori and x_i are values that are equally spaced along a grid line. We then fit both Poisson models and negative binomial models to the resulted observations obtained with different values of β_0 , β_1 , k , N and ranges of x_i along the grid line.

Example 1. Using $k=0.85$, $\beta_0=1.5$ and $\beta_1=2.5$, $N=50$, x_i is the value equally spaced along the grid from 0 to 5.

Table 1: Generated data with $k=0.85$, $\beta_0=1.5$ and $\beta_1=2.5$. $N=50$.

6	1	4	11	5
52	15	11	324	460
241	427	1318	707	319
563	13884	21261	22730	15097
14893	3230	164212	94943	724
1182	1812	694	5005	13253
1065	2051	114359	191070	10258
153321	53565	759786	206660	1190886
707036	1702096	16	1	27
95	57	39	146	4

Fitting a Poisson model to these generated observations we obtain the following results:

Coefficients:	Value	Std. Error	t-value	P-value
intercept	-0.7501363	0.006055576	-123.88	< 0.001
X	2.9855719	0.001281526	2329.70	< 0.001

$$\text{Pearson's } X^2 = 1689309.27 \quad \text{Deviance } G^2 = 1826991.29 \quad \text{df} = 48$$

Fitting a negative binomial model to these generated observations we have

Coefficients:	Value	Std. Error	t-value	P-value
intercept	1.199	0.26810152	4.47	< 0.001
X	2.522	0.09122524	27.65	< 0.001

$$\text{Pearson's } X^2 = 37.91 \quad \text{Deviance } G^2 = 47.49 \quad \text{df} = 48$$

Example 2. using $k=0.75$, $\beta_0=0.3$ and $\beta_1=2.0$, $N=100$ is the value equally spaced along the grid from 0 to 5.

Table 2: Generated data $k=0.75$, $\beta_0=0.3$ and $\beta_1=2.0$, $N=100$.

3	0	0	3	4	1	2	5	2	17
1	1	10	0	3	4	14	10	10	8
4	0	9	16	4	4	12	7	63	31
13	5	89	25	33	45	25	60	57	82
39	48	47	211	105	81	49	152	49	66
376	364	342	27	105	230	712	108	703	1261
885	175	296	1239	562	582	422	46	757	263
2142	1642	604	297	4168	3486	590	3291	1372	2995
2343	6811	760	10016	3282	11132	17398	438	14276	26005
1710	38197	10060	84046	17378	24221	58928	208	3445	37456

Fitting a Poisson model to these generated observations we have

Coefficients:	Value	Std. Error	t-value	P-value
intercept	-1.670175	0.0182669	-91.43	< 0.001
X	2.486264	0.0039368	631.53	< 0.001
Pearson's $X^2 = 244867.76$ Deviance $G^2 = 258198.35$ $df = 98$				

Fitting a negative binomial model to these generated observations we have

Coefficients:	Value	Std. Error	t-value	P-value
intercept	0.2130112	0.19332059	1.10	0.271
X	2.0242476	0.06445316	31.41	< 0.001
Pearson's $X^2 = 93.19$ Deviance $G^2 = 110.01$ $df = 98$				

Examples with real data

Example 3. Consider the following data on mean Whitefly (insects) counts across cassava genotypes for four different locations in Nigeria (Ibadan=1, Mokwa=2, Onne=3 and Zaria=4) over six different times of inspection within a year, which was conducted by Agronomic Virology section of IITA Ibadan in 2002.

Table 3: Whitefly counts across cassava genotypes in different locations.

	Location			
Time	1	2	3	4
1	31	1	7	1
2	34	1	11	0
3	18	2	5	0
4	9	2	5	2
5	4	2	7	0
6	3	2	6	0

Using the Poisson distribution we obtain the following results:

Coefficients:	Value	Std. Error	t-value	P-value
intercept	3.523	0.139	25.41	< 0.001
time	-0.440	0.070	-6.58	< 0.001
locat	-1.168	0.150	-7.77	< 0.001
time.locat	0.199	0.053	3.73	< 0.001
Log likelihood= 201.584, Pearson's $X^2 = 74.709$, Deviance $G^2 = 79.610$, $df = 20$.				

For the Negative binomial,

Coefficients:	Value	Std. Error	t value	P-value
intercept	3.483	0.487	7.16	<0.001
time	-0.441	0.170	-2.60	0.009
locat	-1.056	0.300	-3.53	<0.001
time.locat	0.172	0.100	1.71	0.087
Log likelihood= 214.229, Pearson's $X^2 = 20.563$, Deviance $G^2 = 26.694$, $df = 20$				

Example 4. Consider the data on the distribution of reported deaths under one month of age at death in days for 20 years preceding the survey, in Nigeria Demographic and Health Survey, NPC, Nigeria (2000).

Table 4: Total reported death in days under one month.

Age at death (in days)	0	1	2	3	4	5	6	7	8	9
Number of deaths	142	124	34	54	33	42	29	59	29	24

Age at death (in days)	10	11	12	13	14	15	16	17	18	19
Number of deaths	17	5	4	3	27	7	4	1	3	3

Age at death (in days)	20	21	22	23	24	25	27	28	29	30	31
Number of deaths	9	10	1	1	4	2	1	3	1	4	10

Using the Poisson model the results are:

Coefficients:	Value	Std. Error	t-value	P-value
intercept	4.626	0.053	86.81	< 0.001
age	-0.159	0.007	-23.78	< 0.001

Loglikelihood = 1890.241, Pearson's $X^2 = 270.830$, Deviance $G^2 = 204.837$, $df = 29$

For the Negative Binomial the results are:

Coefficients:	Value	Std. Error	t-value	P-value
intercept	4.249	0.335	12.69	< 0.001
age	-0.120	0.020	-5.88	< 0.001

Loglikelihood = 1946.214, Pearson's $X^2 = 25.008$, Deviance $G^2 = 21.052$, $df = 29$

Example 5. Consider the data on total number of facilities that use various HIV test kits ("type") (Capillus=1, Bionor=2, Determine=3, Serocard=4, Hema-strip=5, Multispot/Genne=6, Unigold=7, others=8) for ("principal") (Primary=1, Secondary=2 or confirmatory=3) HIV testing in Uganda, Uganda Health Facilities Survey UHFS (2003).

Table 5: Total principal use of HIV test kits

	Principal		
HIV test types	1	2	3
1	20	0	0
2	5	0	0
3	2	2	2
4	1	7	8
5	0	1	0
6	0	0	8
7	0	1	1
8	0	2	3

Under Poisson model

Coefficients:	Value	Std. Error	t-value	P-value
intercept	2.738	0.220	12.44	< 0.001
type	-0.815	0.133	-6.11	< 0.001
principal	-1.440	0.264	-5.45	< 0.001
type.principal	0.544	0.082	6.59	< 0.001

Loglikelihood= 30.154, Pearson's $X^2 = 66.251$, Deviance $G^2 = 61.083$, $df= 20$

while for Negative Binomial we have

Coefficients:	Value	Std. Error	t-value	P-value
intercept	2.505	0.539	4.65	< 0.001
type	-0.762	0.199	-3.82	< 0.001
principal	-1.404	0.464	-3.02	0.002
type.principal	0.556	0.132	4.20	< 0.001

Loglikelihood= 38.152, Pearson's $X^2 = 29.519$, Deviance $G^2 = 30.750$, $df= 20$

The results of the working examples are summarized for the two models in Table 6.

Table 6: Summary of results for generated data.

Example	Mean	Variance	Poisson Model		Negative Binomial		df	$\hat{k}$
			Pearson's (X^2)	L'hood Ratio (G^2)	Pearson's (X^2)	L'hood Ratio (G^2)		
1	109398.4	101034819784	1689309.27	1826991.30	37.91	47.49	48	0.85
2	3995.81	144878779	244867.77	258198.35	93.19	110.01	98	0.75
3	6.375	82.592	74.71	79.61	20.56	26.69	20	0.55
4	5.607	46.004	270.83	204.83	25.01	21.05	29	0.85
5	2.625	20.245	66.25	61.08	29.52	30.74	20	0.5

5. Conclusion

From Tables 1 to 5, the estimates of variances are greater than their respective means indicating that each of the data used possess the element of positive contagion which eventually leads to over dispersion. From the results of the generated data, the estimates for β s are still in the neighborhood of their respective pre-assigned values of β s used to generate the data. Based on the estimates of the Pearson's chi-squares X^2 and likelihood ratios G^2 statistics for Poisson which are far higher than that obtained under negative binomial models with the same number of degrees of freedom for each of the example (Table 6) it is clear that negative binomial model is better than the Poisson model.

Examination of plots of fitted counts against absolute residuals as well as the plots of least squares fits of fitted counts with absolute residuals also showed superiority of the

Negative Binomial model. . We may therefore deduce that the Negative Binomial model is better than the Poisson model under positive contagion.

References

- Adejumo, A. O. (2002). On the Estimation of the Relative Incident Parameter of Negative Multinomial Distribution. *Journal of The Nigerian Statistical Association*, 15, 36-43.
- Agresti, A. (1990). *Categorical Data Analysis*. Wiley, New York.
- Anscombe, F. J. (1949). The statistical analysis of insect counts based on the negative binomial distribution. *Biometrics*, 15, 165-173.
- Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Appli. Statist.*, 33, 38-44.
- Cameron, A. C and Trivedi, P. K. (1990). regression based tests for overdispersion in the Poisson model. *J. Econometrics*, 46, 347-364.
- Consul, P. C and Famoye F. (1992). Generalized Poisson regression model. *Comm. in Statist., Theory and Methods*, 21, 89-109.
- Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika*, 70, 269-274.
- Ciccarelli, M. and Rebucci, A. (2003). Measuring Contagion with a Bayesian, time-varying coefficient model. *European Central Bank, Working paper series*, 263.
- Creel, M. E. and Loomis, J. B. (1993). Modelling Hunting Demand in the presence of Bag limit, with tests of alternative specifications. *J. Environmental and Management*, 22, 93-113.
- Fahrmeir, L., and Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models*. Springer.
- Famoye, F. (1993). Restricted generalized Poisson regression model. *Comm. in Statist., Theory and Methods*, 22, 1335-1354.
- Firth, D. (1987). On the efficiency of Quasi-likelihood estimation. *Biometrika*, 74, 233-245.
- Forbes K. and Rigobon, R. (2001). Contagion in Latin American: Definitions, Measuring, and Policy Implications. *Enonomia.*, 1(2), 1-46.
- Habb, T. C. and McConnell, K. E. (1996). Count data models and the problem of zeros in recreation demand analysis *Ameri. J. Agricultural Econ.*, 78, 89-103.
- King, M. A. and Wadhvani, S. (1990). Transmission of volatility between stock markets. *Review of Finan. Studies*, 3(1), 5-33.
- McCullagh, P., and Nelder, J. A. (1989). *Generalized linear models*, 2nd Edit. Chapman and Hall.

- Ministry of Health (MoH)[Uganda]. (2003). *Uganda Health Facilities Survey 2002*.
Calverton, Maryland and Arlington, Virginia, USA: ORC Macro and John Snow,
Inc./Deliver.
- National Population Commission [Nigeria]. (2000). *Nigeria Demographic and Health Survey
1999*. Calverton, Maryland: National Population Commission and ORC/Macro.
- Plackett, R. L. (1981). *The analysis of categorical data*. (2nd edition), London: Griffin.
- Santos.Silva J. M. C. (1997). Generalized Poisson regression for positive count data. *Comm.
in Statist., Simulation Methods*, 26, 1089-1102.