Available online at www.sciencedirect.com
ScienceDirect
Economics Letters 99 (2008) 585 – 590
economics letters
www.elsevier.com/locate/econbase

Functional forms for the negative binomial model for count data

William Greene *

Department of Economics, Stern School of Business, New York University, 44 West 4th St., Rm. 7-78, New York, NY 10012, USA

Received 11 June 2007; received in revised form 12 September 2007; accepted 1 October 2007
Available online 11 October 2007

Abstract

This note develops an encompassing model for two well known variants of the negative binomial model (the NB1 and NB2 forms). We conclude with an application of the proposed model using the data employed in a recent health care study.

Keywords: Poisson regression; Negative binomial; Heterogeneity; Encompassing model

JEL classification: C14; C23; C25

1. Introduction

Models for count data have been prominent in many branches of the recent applied literature, for example, in health economics[1], management (e.g., numbers of patents[2]) and industrial organization (e.g., numbers of entrants to markets[3]). The foundational building block in this modeling framework is the Poisson regression model.[4] But, because of its implicit restriction on the distribution of observed counts – in the Poisson model, the variance of the random variable is constrained to equal the mean – researchers generally employ more general specifications such as the negative binomial (NB) model which is the standard choice for a basic count data model.[5] There are two well known, nonnested forms of the negative binomial model, denoted NB1 and NB2 in the literature [see Cameron and Trivedi, (1986)]. Researchers have typically chosen one form or the other (usually NB2), without actually articulating a preference for either. We propose an encompassing model that nests both of them parametrically and allows a statistical test of the two functional forms against a more general alternative.

The study is organized as follows: Section 2 will detail the basic modeling frameworks for count data, the Poisson and NB models. We then develop the NBP model to encompass NB1 and NB2. The model extensions are applied to the Riphahn et al. (RWM) (2003) panel data on health care utilization in Section 3. Some conclusions are drawn in Section 4.

2. Basic functional forms for count data models

The literature abounds with alternative models for counts — see, e.g., CT (1998) and Winkelmann (2003). However, the Poisson and two forms of the negative binomial model overwhelmingly dominate the received applications [see Hilbe (2007).].

2.1. The Poisson regression model

The canonical regression specification for a variable Y that is a count of events is the Poisson regression,

$$\text{Prob}[Y = y_i|\mathbf{x}_i] = \frac{\exp\ (-\lambda_i)\lambda_i^{y_i}}{\Gamma(1+y_i)},$$
$$\lambda_i = \exp\ (\alpha + \mathbf{x}'_i\beta), y_i = 0, 1, \ldots, i = 1, \ldots, N \qquad (2\text{-}1)$$

where $\mathbf{x}_i$ is a vector of covariates and, $i = 1,\ldots, N$, indexes the N observations in a random sample [The regression model is

* Tel.: +1 212 998 0876.
E-mail address: wgreene@stern.nyu.edu.
URL: http://www.stern.nyu.edu/~wgreene.

1 Contoyannis et al. (2004), Munkin and Trivedi (1999), Riphahn et al. (RWN) (2003). See, as well, Cameron and Trivedi (2005).
2 Hausman et al. (1984) and Wang et al. (1998).
3 Asplund and Sandin (1999).
4 HHG (1984), Cameron and Trivedi (1986, 1998), and Winkelmann (2003).
5 The NB model is by far the most common specification. See Hilbe (2007). The latent class (finite mixture) and random parameters forms have also been employed. See, e.g., Wang et al. (1998).

0165-1765/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2007.10.015

developed in detail in a vast number of standard references such as CT (1986, 1998, 2005), Winkelmann (2003) and Greene (2008,in press)]. The signature features of the Poisson model are its loglinear conditional mean function,

$$E[y_i|\mathbf{x}_i] = \lambda_i. \tag{2-2}$$

and its equidispersion,

$$\text{Var}[y_i|\mathbf{x}_i] = \lambda_i. \tag{2-3}$$

Since observed data will almost always display pronounced *over*dispersion, analysts typically seek alternatives to the Poisson model, such as the negative binomial model described below.

2.2. The standard negative binomial model

The negative binomial model is employed as a functional form that relaxes the equidispersion restriction of the Poisson model. A useful way to motivate the model is through the introduction of latent heterogeneity in the conditional mean of the Poisson model.[6] Thus, we write

$$E[y_i|\mathbf{x}_i, \varepsilon_i] = \exp(\alpha + \mathbf{x}_i'\beta + \varepsilon_i) = h_i\lambda_i, \tag{2-4}$$

where $h_i = \exp(\varepsilon_i)$ is assumed to have a one parameter gamma distribution, $G(\theta,\theta)$ with mean 1 and variance $1/\theta = \kappa$;

$$f(h_i) = \frac{\theta^\theta \exp(-\theta h_i) h_i^{\theta-1}}{\Gamma(\theta)}, h_i \geq 0, \theta > 0. \tag{2-5}$$

After integrating h_i out of the joint distribution, we obtain the marginal negative binomial (NB) distribution,

$$\text{Prob}[Y = y_i|\mathbf{x}_i] = \frac{\Gamma(\theta + y_i) r_i^\theta (1 - r_i)^{y_i}}{\Gamma(1 + y_i)\Gamma(\theta)},$$
$$y_i = 0, 1, ..., \theta > 0, \quad r_i = \theta/(\theta + \lambda_i). \tag{2-6}$$

The latent heterogeneity induces overdispersion while preserving the conditional mean;

$$E[y_i|\mathbf{x}_i] = \lambda_i, \tag{2-7}$$

$$\text{Var}[y_i|\mathbf{x}_i] = \lambda_i[1 + (1/\theta)\lambda_i] = \lambda_i[1 + \kappa\lambda_i] \tag{2-8}$$

where $\kappa = \text{Var}[h_i]$.

Maximum likelihood estimation of the parameters of the NB model (α,β,θ) is straightforward, as documented in, e.g., Greene (2007). Inference proceeds along familiar lines. Inference about the specification, specifically the presence of overdispersion, is the subject of a lengthy literature, as documented, e.g., in CT (1990, 1998, 2005) and Hilbe (2007).

2.3. The NB1 and NEGBIN P models

The negative binomial model in Eq. (2-6) was labeled the NEGBIN 2 (NB2) model by CT (1986), in reference to the appearance of the quadratic term for λ_i in the conditional variance function:

$$\text{Var}[y_i|\mathbf{x}_i] = \lambda_i + \kappa\lambda_i^2. \tag{2-11}$$

CT (1986) suggested a reparameterization of the model,

$$\text{Var}[y_i|\mathbf{x}_i] = \lambda_i + \kappa\lambda_i^1 = \lambda_i[1 + \kappa], \tag{2-12}$$

and label the resulting specification NB1. The model is obtained by replacing θ with $\theta\lambda_i$ in Eq. (2-6). After simplification, we obtain the density for NB1,

$$\text{Prob}[Y = y_i|\mathbf{x}_i] = \frac{\Gamma(\theta\lambda_i + y_i) q^{\theta\lambda_i}(1 - q)^{y_i}}{\Gamma(y_i + 1)\Gamma(\theta\lambda_i)},$$
$$y_i = 0, 1, ...; q = 1/1(1 + \theta). \tag{2-13}$$

The authors note in (1998) that other exponents would be possible [see their p. 73 and (3.26)]. By replacing θ with $\theta\lambda_i^{2-P}$, we obtain the NEGBIN P, or NBP model,

$$\text{Prob}[Y = y_i|\mathbf{x}_i] = \frac{\Gamma(\theta\lambda_i^{2-P} + y_i) s_i^{\theta\lambda_i^{2-P}} (1 - s_i)^{y_i}}{\Gamma(y_i + 1)\Gamma(\theta\lambda_i^{2-P})},$$
$$y_i = 0, 1, ...,$$
$$s_i = \frac{\lambda_i}{\lambda_i + \theta\lambda_i^{2-P}}. \tag{2-14}$$

(The log likelihood function and its derivatives are given in the Appendix.) The NB1 and NB2 models are the special cases of $P = 1$ and $P = 2$. The conditional mean in this model is still λ_i, while the conditional variance is

$$\text{Var}[y_i|\mathbf{x}_i] = \lambda_i[1 + (1/\theta)\lambda_i^{P-1}]. \tag{2-15}$$

CT (1998) focus on the $P = 1$ and $P = 2$ forms, but do suggest that the "generalized event count model" (see their Section 4.4.1) does include the NEGBIN P as a special case [CT (1986) also mention the possibly of this extension of the model, but do not develop it at any length.]. The GEC model [Winkelmann and Zimmermann (1991, 1995); King (1989)] which does include NPP is sufficiently cumbersome to have greatly restricted its general use. The NBP model achieves somewhat less of the generality of the GEC model, but is much simpler to implement.[7] Although various authors have suggested the possibility of exponents other than 1 and 2in

[6] This general approach is discussed at length by Gourieroux et al. (1984), CT (1986, 1998), Winkelmann (2003) and HHG (1984).

[7] Winkelmann and Zimmermann (1995) develop a maximum likelihood estimator for the equivalent of the NEGBIN P model, but their formulation adds what appears to be a considerable yet unnecessary layer of difficulty to the derivation. In applications, the direct MLE based on (2-14) appears to be quite well behaved.

Table 1
Variables in German health care data file

Variable	Measurement	Mean	Standard deviation
ID	Household identification, 1,..., 7293		
YEAR	Calendar year of the observation	1987.82	3.17087
YEAR1984	Dummy variable for 1984 observation	.141770	.348820
YEAR1985	Dummy variable for 1984 observation	.138842	.345788
YEAR1986	Dummy variable for 1984 observation	.138769	.345712
YEAR1987	Dummy variable for 1984 observation	.134158	.340828
YEAR1988	Dummy variable for 1984 observation	.164056	.370333
YEAR1991	Dummy variable for 1984 observation	.158823	.365518
YEAR1994	Dummy variable for 1984 observation	.123582	.329110
AGE	Age in years	43.5257	11.3302
AGESQ**	Age squared/1000	2.02286	1.00408
FEMALE	Female=1; male=0	.478775	.499558
MARRIED	Married=1; else=0	.758618	.427929
HHKIDS	Children under age 16 in the household=1; else=0	.402730	.490456
HHNINC***	Household nominal monthly net income, German marks/10,000	.352084	.176908
EDUC	Years of schooling	11.3206	2.32489
WORKING	Employed=1; else=0	.677048	.467613
BLUEC	Blue collar employee=1; else=0	.243761	.429358
WHITEC	White collar employee=1; else=0	.299605	.458093
SELF	Self employed=1; else=0	.0621752	.241478
CIVIL	Civil servant=1; else=0	.0746908	.262897
HAUPTS	Highest schooling degree is Hauptschul=1; else=0	.624277	.484318
REALS	Highest schooling degree is Realschul=1; else=0	.196809	.397594
FACHHS	Highest schooling degree is Polytechnical=1; else=0	.0408402	.197924
ABITUR	Highest schooling degree is Abitur=1; else=0	.117031	.321464
UNIV	Highest schooling degree is university=1; else=0	.0719461	.258403
HSAT	Health satisfaction, 0–10	6.78543	2.29372
NEWHSAT*,**	Health satisfaction, 0–10	6.78566	2.29373
HANDDUM	Handicapped=1; else=0	.214015	.410028
HANDPER	Degree of handicap in pct, 0–100	7.01229	19.2646
DOCVIS	Number of doctor visits in last three months	3.18352	5.68969
DOCTOR**	1 if DOCVIS>0, 0 else	629108	.483052
HOSPVIS	Number of hospital visits in last calendar year	.138257	.884339
HOSPITAL**	1 of HOSPVIS>0, 0 else	.0876455	.282784
PUBLIC	Insured in public health insurance=1; else=0	.885713	.318165
ADDON	Insured by add-on insurance=1; else=0	.0188099	.135856

Data source: http://qed.econ.queensu.ca/jae/2003-v18.4/riphahn-wambach-million/.
From Riphahn, R., A. Wambach and A. Million "Incentive effects in the demand for health care: a bivariate panel count data estimation," *Journal of Applied Econometrics*, 18, 4, 2003, pp. 387–405.
Notes: *NEWHSAT=HSAT; 40 observations on HSAT recorded between 6 and 7 were changed to 7.
**Transformed variable not in raw data file.
***Divided by 1,000 rather than 10,000 by RWM. We used this scale to ease comparison of coefficients.

Eq. (2-14), we have not found an empirical implementation in the received literature. An application appears below.

For choosing statistically between NB1 and NB2, the models are nonnested and essentially equivalently parameterized, so a direct test is precluded. However, one possibility is the Vuong (1989) test based on

$$V = \frac{\sqrt{n}\overline{m}}{s_m}, m_i = \ln L_i(\text{NB2}) - \ln L_i(\text{NB1}). \tag{2-16}$$

When the underlying conditions for its validity are met, the Vuong test statistic has a limiting standard normal distribution. Large positive values would favor NB2. We have found in applications that this statistic is rarely outside the inconclusive region (− 1.96 to + 1.96) for this model. It may be that NB1 and NB2 are not sufficiently different to enable a distinction on this basis. Since the NBP model does nest both of them, it provides a partial solution to the specification problem. For example, in our application below, simple likelihood ratio tests reject both the NB1 and NB2 null hypotheses.

3. Application

In "Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation," Riphahn, Wambach and Million (RWM) (2003) employed a part of the German Socioeconomic Panel (GSOEP) data set to analyze two count variables, DocVis, the number of doctor visits in the last three months and HospVis, the number of hospital visits in the last year. The authors employed a bivariate panel data (random effects) Poisson model to study these two outcome variables. A central focus of the investigation was the role of the choice of private health insurance in the intensity of use of the health care

Table 2
Specifications for the negative binomial model (*t* ratios in parentheses)[a]

Variable	Poisson	NB 1	NB 2	NB P
Constant	2.771 (28.85)	2.7760 (14.06)	3.1488 (13.74)	3.0500 (13.14)
AGE	−0.02387 (−5.44)	−.04768 (−5.61)	−0.03983 (−4.07)	−.04679 (−4.76)
AGESQ	0.3693 (7.45)	.6340 (6.58)	0.5467 (4.77)	0.6373 (5.66)
HSAT	−0.2253 (−104.1)	−.1886 (−44.58)	−0.2392 (−42.44)	−.2279 (−46.17)
HANDDUM	0.06899 (4.09)	.02292 (0.67)	−0.02090 (−0.46)	.01660 (0.41)
HANDPER	0.002858 (10.04)	.004141 (7.33)	0.006614 (8.05)	.005031 (7.30)
MARRIED	0.05831 (3.89)	.1299 (4.50)	0.06582 (2.18)	.1139 (3.55)
EDUC	−0.02348 (−8.43)	−.009550 (−1.80)	−0.02623 (−4.59)	−.01794 (−2.86)
HHNINC	−0.2220 (−5.93)	−.07878 (−1.13)	−0.1917 (−2.48)	−.1462 (−1.78)
HHKIDS	−0.07598 (−5.75)	−.07435 (−2.95)	−0.08440 (−3.32)	−.08672 (−3.10)
SELF	−0.2110 (−8.98)	−.2439 (−5.56)	−0.2179 (−5.02)	−.2628 (−5.42)
CIVIL	0.09144 (3.78)	.02782 (0.60)	0.08411 (1.56)	.05148 (0.93)
BLUEC	0.01779 (1.24)	−.009478 (−0.35)	.03706 (1.20)	.005597 (0.17)
WORKING	−0.05539 (−3.17)	.01258 (0.37)	−0.01545 (−0.38)	−.001046(−0.20)
PUBLIC	0.1001 (4.27)	.06067 (1.38)	.09340 (1.83)	.07823 (1.50)
ADDON	0.06655 (1.63)	.1393 (1.72)	0.05506 (0.50)	.1363 (1.34)
θ[b]	Not estimated	0.2058 (62.08)	0.5707 (59.96)	0.3460 (36.47)
κ[c]	0.0000 (fixed)	4.8598 (62.08)	1.7522 (59.96)	2.8905 (36.47)
P	0.0000 (fixed)	1.0000 (fixed)	2.0000 (fixed)	1.4897 (64.96)[d]
In L	−42774.7	−27410.0	−27480.4	−27306.1

[a] Estimated coefficients for year dummy variables, excluding year 1984, are not reported.
[b] θ=the estimated parameter for the log gamma (NB) model.
[c] $\kappa=1/\theta=\text{Var}[h]$ for log gamma model.
[d] Estimated standard error is 0.02293.

system, i.e., whether the data contain evidence of moral hazard. We will use these data to illustrate the model extensions described above.[8] The authors of this study presented estimates for a Poisson-lognormal model and a bivariate Poisson model.

The RWM data set is an unbalanced panel of 7293 individual families observed from one to seven times. The number of observations varies from one to seven (1525, 1079, 825, 926, 1051, 1000, 887) with a total number of observations of 27,326. The variables in the data file are listed in Table 1 with descriptive statistics for the full sample. RWM estimated separate equations for males and females and did not report any estimates based on the pooled data. In the interest of brevity, we will restrict attention to DocVis, the count of doctor visits, and demonstrate the NBP model with only the subsample of males. Analysis of the count of hospital visits is left for further research. [More extensive analysis of the specification and empirical results appear in Greene (2008a).]

The base case count model used by the authors included the following variables in addition to the constant term:

$$\mathrm{x}_{it} = (\text{Age, Agesq, HSat, Handdum, Handper, Married, Educ, Hhninc, Hhkids, Self, Civil, Bluec, Working, Public, AddOn})$$

and a set of year effects,

$$t = (\text{YEAR1985, YEAR1986, YEAR1987, YEAR1988, YEAR1991, YEAR1994}).$$

The same specification was used for both DocVis and HospVis. We will use their specification in our count models. The estimated year effects are omitted from the reported results in the paper.

Table 2 presents estimates of the parameters of the different specifications of the negative binomial model. The base case Poisson model corresponds to $P=0$ in the encompassing NBP specification. Based on the likelihood ratio tests, any of the alternative specifications in the table, all of which nest the Poisson, will dominate it. As suggested earlier, NB1 and NB2 produce similar results, but nonetheless, are manifestly different specifications. The log likelihood for NB1 is significantly larger than that for NB2. However, as these two models are not nested, the LR test is inappropriate. Using the Vuong statistic in Eq. (2-16), we obtain a value of − 1.63 in favor of NB1. In spite of the log likelihoods, this is in the inconclusive region. As expected, NBP produces a greater likelihood than either NB1 or NB2. Using a likelihood ratio statistic for testing against NB1, we obtain a chi squared of 207.8 with one degree of freedom. Thus, NB1 and, a fortiori, NB2 are rejected in favor of NBP for these data. The estimated standard error for the estimator of P for this model is 0.02293. The t (Wald) test against the null hypothesis that P equals 1.0 gives a statistic of 21.33, which, once again, would decisively reject the NB1 specification.

4. Conclusion

This study has developed the NBP encompassing form for the negative binomial model and applied the techniques in an analysis of a large sample of German households. The NBP

[8] The raw data are published and available for download on the *Journal of Applied Econometrics* data archive website, The URL is given below Table 1.

variant of the negative binomial model is a convenient form that provides a means of formalizing the specification choice. Most received applications of the model have used the NB2 form. In a few other cases, such as HHG (1984), the NB1 model is used. In none of the cases, does the presentation provide a formal means of preferring one or the other. The NBP is an encompassing form that is simple to operationalize. In the application here (and in others we have considered), likelihood ratio tests suggest that the NBP form would be preferred to both NB1 and NB2. The method developed here was applied to the data set used in RWM (2003). Our empirical results were largely similar to theirs. We find that on the question of moral hazard – whether the presence of insurance appears positively to influence demand for health services – the apparent effect that shows up in the simple models (e.g., a pooled Poisson model) almost completely disappears when latent heterogeneity is formally introduced into the model.

Since the NB1 and NB2 models are not nested, there is no simple parametric test that one can employ to choose between them. E.g., CT do not express a preference for either one or the other in (1986) or (1998); they merely note the difference. They do state "[T]he NB2 MLE [is] favored by econometricians and the NB1 GLM [generalized linear model] [is] used extensively by statisticians." This appeal to the estimation algorithm appears to be the closest to a preference for one or the other as appears in the recent literature. On the other hand, the various references to GEC and NBP models do suggest an attraction to a more general specification than NB1 or NB2. Our results suggest that at least in some applications, the NBP model is likely to be preferable yet to either of the more restrictive negative binomial models.

Appendix A. Log likelihood and gradient for NBP model

The NegBin P model is obtained by replacing θ in NB2,

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta)\Gamma(1 + y_i)} r_i^{\theta}(1 - r_i)^{y_i}, r_i = \theta/(\theta + \lambda_i), \quad \text{(A-1)}$$

with $\theta\lambda_i^{P-2}$. For convenience, let $Q = P - 2$. Then, the density is

$$\text{Prob}(Y = y_i|\mathbf{x}_i) = \frac{\Gamma\left(\theta\lambda_i^Q + y_i\right)}{\Gamma\left(\theta\lambda_i^Q\right)\Gamma(1 + y_i)} \left(\frac{\theta\lambda_i^Q}{\theta\lambda_i^Q + \lambda_i}\right)^{\theta\lambda_i^Q} \left(\frac{\lambda}{\theta\lambda_i^Q + \lambda_i}\right)^{y_i}. \quad \text{(A-2)}$$

Derivatives of ln L_i= ln Prob($Y = y_i|\mathbf{x}_i$) for the Negbin P model are straightforward, albeit tedious. We obtain them by writing the density as

$$\ln L_i = \ln \Gamma(y_i + g_i) - \ln \Gamma(g_i) - \ln \Gamma(1 + y_i) + g_i \ln r_i + y_i \ln (1 - r_i)$$
$$g_i = \theta\lambda_i^Q \text{ and } w_i = g_i/(g_i + \lambda_i). \quad \text{(A-3)}$$

Then,

$$\begin{aligned}
\partial \ln L_i/\partial\lambda_i &= [\Psi(y_i + g_i) - \Psi(g_i) + \ln w_i]\partial g_i/\partial\lambda_i \\
&\quad + [g_i/w_i - y_i/(1 - w_i)]\partial w_i/\partial\lambda_i \\
\partial \ln L_i/\partial\theta &= [\Psi(y_i + g_i) - \Psi(g_i) + \ln w_i]\partial g_i/\partial\theta \\
&\quad + [g_i/w_i - y_i/(1 - w_i)]\partial w_i/\partial\theta \\
\partial \ln L_i/\partial Q &= [\Psi(y_i + g_i) - \Psi(g_i) + \ln w_i]\partial g_i/\partial Q \\
&\quad + [g_i/w_i - y_i/(1 - w_i)]\partial w_i/\partial Q.
\end{aligned} \quad \text{(A-4)}$$

where $\Psi(t) = \text{d}\ln\Gamma(t) / \text{d}t$. The inner parts are:

$$\begin{aligned}
\partial g_i/\partial\lambda_i = \theta Q\lambda_i^{Q-1} &= (Q/\lambda_i)g_i \\
\partial g_i/\partial\theta_i = \lambda_i^Q &= (1/\theta)g_i \\
\partial g_i/\partial Q = \theta\lambda_i^Q\log\lambda_i &= \ln \lambda_i g_i \\
\partial w_i/\partial\lambda_i &= [(Q - 1)/\lambda_i]w_i(1 - w_i) \\
\partial w_i/\partial\theta &= (1/\theta)w_i(1 - w_i) \\
\partial w_i/\partial Q &= \log \lambda_i w_i(1 - w_i)
\end{aligned} \quad \text{(A-5)}$$

Collecting terms, now, let

$$\begin{aligned}
A_i &= [\Psi(y_i + g_i) - \Psi(g_i) + \ln w_i], \\
B_i &= [g_i(1 - w_i) - y_i w_i],
\end{aligned} \quad \text{(A-6)}$$

to obtain

$$\partial \ln L_i/\partial \begin{pmatrix} \lambda_i \\ \theta \\ Q \end{pmatrix} = [A_i + B_i] \begin{pmatrix} Q/\lambda_i \\ 1/\theta \\ \log \lambda_i \end{pmatrix} - B_i \begin{pmatrix} 1/\lambda_i \\ 0 \\ 0 \end{pmatrix}. \quad \text{(A-7)}$$

The final element needed is $\partial\ln L_i/\partial\beta = (\partial\ln L_i/\partial\lambda_i)(\partial\lambda_i/\partial\beta)$ where $\partial\ln L_i/\partial\lambda_i$ appears above and $\partial\lambda_i/\partial\beta = \lambda_i\mathbf{x}_i$. We use these and the BHHH estimator to compute the maximum likelihood estimates and their asymptotic standard errors for the NBP model. Good starting values for the NBP iterative estimator are the NB2 estimates of β and θ and $P = 2$ ($Q = 0$).

References

Asplund, M., Sandin, R., 1999. The number of firms and production capacity in relation to market size. Journal of Industrial Economics 47 (1), 69–85. (Mar.).

Cameron, A., Trivedi, P., 1986. Econometric models based on count data: comparisons and applications of some estimators and tests. Journal of Applied Econometrics 1, 29–54.

Cameron, C., Trivedi, P., 1990. Regression-Based Tests for Overdispersion in the Poisson Regression Model. Journal of Econometrics, 46, pp. 347–364.

Cameron, C., Trivedi, P., 1998. Regression Analysis of Count Data. Cambridge University Press, New York.

Cameron, C., Trivedi, P., 2005. Microeconometrics: Methods and Applications. Cambridge University Press, Cambridge.

Contoyannis, C., Jones, A., Rice, N., 2004. The dynamics of health in the British household panel survey. Journal of Applied Econometrics 19 (4), 473–503.

Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: applications to poisson models. Econometrica 52, 701–720.
Greene, W., 2008. Econometric Analysis, 6th ed. Prentice Hall, Englewood Cliffs.
Greene, W., 2007. LIMDEP 9.0 Reference Guide, Econometric Software, Inc. Plainview.
Greene, W., in press (forthcoming). "Functional form and Heterogeneity in Models for Count Data," Foundations and Trends in Econometrics. Working Paper 07-10, Department of Economics, Stern School of Business, New York University, 2007b.
Hausman, J., Hall, B., Griliches, Z., 1984. Economic models for count data with an application to the patents–R&D relationship. Econometrica 52, 909–938.
Hilbe, J., 2007. Negative Binomial Regression. Cambridge University Press, Cambridge, UK.
King, G., 1989. A seemingly unrelated poisson regression model. Sociological Methods and Research 17 (3), 235–255.
Munkin, M., Trivedi, P., 1999. Simulated maximum likelihood estimation of multivariate mixed-poisson regression models, with application. Econometrics Journal 2, 29–49.
Riphahn, R., Wambach, A., Million, A., 2003. Incentive effects in the demand for health care: a bivariate panel count data estimation. Journal of Applied Econometrics 18 (4), 387–405.
Vuong, Q., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–344.
Wang, P., Cockburn, I., Puterman, L., 1998. Analysis of patent data — a mixed-poisson-regression-model approach. Journal of Business and Economic Statistics 16 (1), 27–41.
Winkelmann, R., 2003. Econometric Analysis of Count Data, 4th ed. Springer Verlag, Heidelberg.
Winkelmann, R., Zimmermann, K., 1991. A new approach for modeling economic count data. Economics Letters 37, 139–143.
Winkelmann, R., Zimmermann, K., 1995. Recent developments in count data modeling: theory and application. Journal of Economic Surveys 9 (1), 1–36.