Communications in Statistics - Simulation and Computation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lssp20

Generalized poisson regression for positive count data

J. M. C. Santos Silva [a]
[a] ISEG/Universidade Técnica de Lisboa R. Miguel , Lisboa, 1200, PortugalLupi 20
Published online: 27 Jun 2007.

COMMUN. STATIST.—SIMULA., 26(3), 1089–1102 (1997)

GENERALIZED POISSON REGRESSION FOR POSITIVE COUNT DATA

J. M. C. Santos Silva

ISEG/Universidade Técnica de Lisboa
R. Miguel Lupi 20, 1200 Lisboa, Portugal

Key Words: parametrization of the variance; specification tests.

ABSTRACT

This paper suggests a flexible parametrization of the generalized Poisson regression, which is likely to be particularly useful when the sample is truncated at zero. Suitable specification tests for this case are also studied. The use of the models and tests suggested is illustrated with an application to the number of recreational fishing trips taken by households in Alaska.

1 INTRODUCTION

The generalized Poisson (GP) distribution, introduced by Consul and Jain (1973), has proved adequate to model many different phenomena [see Consul (1989) and the references therein for a detailed account of empirical results]. Recently, this distribution has also been used in a regression context by Consul and Famoye (1992) and Famoye (1993). In this paper, alternative forms of GP regression are studied, focusing on the case where the sample is truncated at zero.

Researchers often need to model positive count data. This may result either from the fact that only positive counts are observed due to the nature of the sampling scheme [Grogger and Carson (1991) and Creel and Loomis (1990)] or because the researcher feels that the process governing the occurrence of zeros is different from the mechanism generating the other realizations of the dependent variable [Pohlmeier and Ulrich (1995)]. In both cases, consistent estimation of the parameters of interest (generally the population conditional mean) requires the correct specification of the conditional distribution of the truncated variate given the covariates. Therefore, when dealing with truncated samples, special care is needed in choosing a distribution that adequately describes the data.

The applications reported so far in the literature use either the Poisson or Negative Binomial distributions as the basis of their analysis of truncated count data. However, the GP distribution might be preferable, specially when it is unlikely that the occurrence of events being counted is independent of number of occurrences so far registered.

The hypothesis of independent occurrences is certainly adequate when modelling many physical processes but is often inadequate for example in social sciences. The number of visits to the doctor or the number of visits to a recreation site [Grogger and Carson (1991), Creel and Loomis (1990) and Pohlmeier and Ulrich (1995)] are examples of situations in which the independence hypothesis is not very plausible. Because the GP distribution allows for the probability of an event to depend on the number of events already occurred [Consul and Shoukri (1985)], this distribution might be particularly useful in such situations.

The remainder of the paper is organized as follows. Section 2 briefly reviews the application of the GP distribution in a regression context and suggests a flexible parametrization that is likely to be particularly useful for positive count data. Statistical tests to assess the adequacy of the GP model in the context of positive regression are presented in section 3. The application of the main results of the paper are illustrated in section 4 with the study of a well known data set. Finally, section 5 summarises the results and concludes.

2 GENERALIZED POISSON REGRESSION FOR POSITIVE COUNTS

Following Consul (1989) the generalized Poisson distribution, also known as Lagrangian Poisson distribution, can be defined by

$$P(Y=y)=\begin{cases}\dfrac{\theta\,(\theta+\lambda y)^{y-1}\exp\left[-\,(\theta+\lambda y)\right]}{y!} & y=0,1,2,\ldots\\[2ex] 0 & \text{for } y>m,\ \text{when } \lambda<0\end{cases}\quad, \tag{1}$$

where $\theta > 0$, $\max(-1, -\theta/m) < \lambda < 1$, and $m \geq 4$ is the largest positive integer for which $\theta + \lambda m > 0$ when λ is negative. One of the important characteristics of the GP distribution is the fact that it allows for overdispersion ($\lambda > 0$) and underdispersion ($\lambda < 0$) as well as for equidispersion ($\lambda = 0$). It is also well known that [see Consul (1989) for example],

$$E(Y) = \frac{\theta}{(1-\lambda)}, \qquad V(Y) = \frac{\theta}{(1-\lambda)^3}. \tag{2}$$

As pointed out by Nelson (1975), the distribution as originally proposed would lead to negative probabilities for negative values of λ. To avoid this, the distribution is truncated from above when $y > m$ and $\lambda < 0$. Because the distribution is not re-scaled when this form of truncation occurs, the probabilities do not sum to 1 when $\lambda < 0$. Consul and Shoukri (1985) conducted a detailed investigation of this problem and concluded that often the GP can be taken as an adequate model even for negative values of λ. In particular, the fact that this distribution is defective when $\lambda < 0$ has no practical consequences when the number of frequency classes with non-zero probability is 5 or more. This lead Consul and Shoukri (1985) to introduce the restriction $-\theta/m < \lambda$. Otherwise θ and λ are independent [see Consul (1989)].

If besides Y the researcher also observes a set of covariates X, the effect of these on the distribution can be modelled by letting θ and possibly λ depend on X. In count data models it is usual to assume that $\mu(X,\beta)$, the conditional expectation of Y is

$$\mu(X,\beta) = \exp(X\beta),$$

where β is a vector of parameters. This is equivalent to set $\theta = \exp(X\beta)(1-\lambda)$.

In a regression context the fact that the range of λ depends on θ and m may have unpleasant consequences. Because the value of the parameters depends on the covariates, it may happen that for some observations the covariates are such that the lower bound on λ is violated. Even if that does not happen in the sample, the problem may arise when the model is used for out of sample inference. For this reason, in a regression framework, it is perhaps better to take λ as a non-negative parameter. Furthermore, if λ is non-negative, (1) defines a proper distribution that satisfies the standard regularity conditions of likelihood theory. This greatly facilitates inference in these models. The cost of this restriction is that with $\lambda \geq 0$ the GP distribution cannot be used in cases of underdispersion.

For samples of positive counts the GP probability function is defined by

$$P(Y=y|Y>0) = \begin{cases} \dfrac{\theta\,(\theta+\lambda y)^{y-1}\exp\left[-\,(\theta+\lambda y)\right]}{\left[1-\exp(-\theta)\right]y!} & y = 1, 2, \ldots \\ 0 & \text{for } y > m, \text{ when } \lambda < 0 \end{cases},$$

and the conditional expectation of Y is

$$E(Y|Y>0)=\frac{\theta}{(1-\lambda)\left[1-\exp\left(-\theta\right)\right]}.$$

In this case, the conditional expectation of Y depends not only on the population conditional expectation but also on $P(Y=0|X)$. Therefore, with a truncated sample, the correct specification of the first conditional moment depends on the correct specification of the full conditional distribution. As a consequence, it is crucial to use models that are flexible enough to adapt to the characteristics of the data. Namely, it is important to correctly specify the way λ depends on the covariates since, in this case, different parametrizations of λ lead to very different specifications of the conditional expectations of Y in the truncated sample.

2.1 Different parametrizations of λ

In their paper, Consul and Famoye (1992) studied a regression model based on (1) assuming that $\theta=\exp(X\beta)(1-\lambda)$. Following Kumar and Consul (1980), λ and θ can be linked by specifying $\lambda=\alpha\theta$. This leads to the so-called restricted generalized Poisson (RGP) model used in a regression context by Famoye (1993). In this case, it is usual to impose the restriction that $\alpha<\theta^{-1}$ [see for example Consul (1989) and Famoye (1993)]. However, it is easy to show that this restriction is redundant. In fact, with this parametrization, $\theta=E(Y)\left[1+\alpha E(Y)\right]^{-1}$ and therefore, it is obvious that $\alpha<\left[1+\alpha E(Y)\right]/E(Y)$.

Since both the GP and RGP models specify a conditional mean of the form $\exp(X\beta)$, the difference between them is limited to the higher order moments. For example, in the GP model the conditional variance is proportional to the conditional mean while in the case of the RGP, $V(Y|X)=E(Y|X)\left[1+\alpha E(Y|X)\right]^2$. Therefore, the two models imply very different patterns of heteroskedasticity.

When the full range of the count variable is observed, the choice between the two models is dictated by the ability of the models to provide an adequate characterization of the conditional variance and therefore an efficient estimator of β. This is because, in this case, consistent estimates of β can be obtained by pseudo maximum likelihood, provided that the conditional expectation of Y is correctly specified [see Gourieroux, Monfort and Trognon (1984)].

However, if the sample is truncated, different specifications of λ lead to models with potentially very different conditional means. For example, as $E(Y|X)$ goes to zero, $E(Y|X,Y>0)$ converges to $(1-\lambda)^{-1}$ under the standard GP while it passes to 1 if $\lambda=\alpha\theta$, as in the RGP.

Although it has been shown that the parametrization used in the RGP has many interesting properties [see Consul (1989) for a survey], it is not the only way of making λ a function of the covariates. Continuing to specify $\mu(X,\beta) = \exp(X\beta)$, an interesting alternative is to assume that α is itself a function of X and define

$$\lambda = \frac{\alpha(X,\delta)\exp(X\beta)}{1+\alpha(X,\delta)\exp(X\beta)} = \alpha(X,\delta)\theta, \tag{3}$$

where δ is a vector of parameters and $\alpha(X,\delta)$ is function of the covariates such that $\alpha(X,\delta) > \max\left(-0.5/\mu(X,\beta),\ -1/m\right)$.

This parametrization ensures that the bounds on λ are always respected and has several important special cases. When $\alpha(X,\delta)$ is a constant this model coincides with the RGP discussed above and if $\alpha(X,\delta)$ is proportional to $\exp(-X\beta)$ the parametrization of λ defined by (3) reduces to the GP regression model. Naturally, if $\alpha(X,\delta) = 0$, (3) leads to the simple Poisson regression.

Although this parametrization of λ allows both for under and overdispersion, here it is assumed that overdispersion is the case of interest, and therefore $\alpha(X,\delta)$ is taken to be non-negative.

Using (3), the variance of Y given X is $\mu(X,\beta)\left[1+\alpha(X,\delta)\mu(X,\beta)\right]^2$, and the variance to mean ratio is $[1+\alpha(X,\delta)\mu(X,\beta)]^2$. Therefore, (3) allows for a wide range of patterns of heteroskedasticity which can be crucial in providing a good fit with overdispersed data. Moreover, because this parametrization nests both the GP and the RGP regression models, it provides an interesting alternative against which these models can be tested. This point is pursued in section 3.

If λ is specified as in (3), when only positive realizations of y are observed the probability function can be written as

$$P(Y=y|Y>0,X) = \frac{\exp\left[-f(\beta,\delta)g(\delta)\right]\left[f(\beta,\delta)g(\delta)\right]^y}{g(\delta)\left\{1-\exp\left[-f(\beta,\delta)\right]\right\}y!}, \tag{4}$$

where $f(\beta,\delta) = \mu(X,\beta)\left[1+\alpha(X,\delta)\mu(X,\beta)\right]^{-1}$ and $g(\delta) = [1+\alpha(X,\delta)y]$.

Considering this conditional probability model, the contribution of each observation to the log-likelihood function is

$$\begin{aligned}\ell(\beta,\delta) = & -f(\beta,\delta)g(\delta) + y\left\{\ln\left[f(\beta,\delta)\right] + \ln\left[g(\delta)\right]\right\} - \ln\left[g(\delta)\right] \\ & - \ln\left\{1-\exp\left[-f(\beta,\delta)\right]\right\} - \ln(y!)\,.\end{aligned}$$

Since (4) satisfies the standard regularity conditions of likelihood theory, the maximum likelihood estimators of β and δ are defined as the solution of the first order conditions for the maximum of $\ell(\beta,\delta)$. Notice that, in practice, there may be restrictions linking elements of β and δ. In those cases, the appropriate changes to the log-likelihood function have to be made.

For this model, the maximum likelihood estimators are consistent, efficient and asymptotically normally distributed. Because in the presence of truncation any consistent estimator of the parameters of interest requires the correct specification of the conditional distribution of Y, there is little reason to use different estimation methods in this context.

3 SPECIFICATION TESTS

Since the consistency of the maximum likelihood estimator of the parameters of the model for $P(Y=y|Y>0,X)$ depends on the correct specification of the conditional distribution of Y given X, it is crucial to have statistical tests to assess the adequacy of the model. That is, it is not enough that the model provides a "good fit" in some sense, it is necessary to test the assumptions underlaying it. In this section, tests against nested and non-nested alternatives will be considered. The purpose of these tests are very different. Tests against nested alternatives will be used to evaluate the adequacy of the many specifications that are nested within a particular model. However this kind of test is not useful to assess the adequacy of the more general specification. In order to do this, the model can be tested against rival non-nested models using for example the results pioneered by Cox (1961).

3.1 Model reduction tests

As pointed out before, one of the reasons to estimate the model defined by (4) might be to check the adequacy of simpler specifications by testing restrictions on the parameters of $\alpha(X,\delta)$. To this purpose, it is particularly interesting to specify $\alpha(X,\delta)$ as $\delta_0\exp(\delta_1 X\beta)$, where δ_0 and δ_1 are scalars. In this case, a test for $H_0:\delta_1=-1$ is a test for the validity of the hypothesis that λ does not depend on the regressors [as in Consul and Famoye (1992)], while the hypothesis that λ is proportional to θ [as in Famoye (1993)] is equivalent to $H_0:\delta_1=0$.

Likelihood ratios comparing the value of the likelihood functions of the restricted and unrestricted models evaluated at the maximum likelihood estimates can be used to construct such tests. Naturally, asymptotically equivalent score and Wald tests can also be used. Under the null, these test statistics are asymptotically distributed as χ^2 variates with 1 degree of freedom.

The adequacy of the simple Poisson regression model can also be tested. This is a test for $H_0:\delta_0=0$. However this test is not straightforward to perform because δ_1 is a nuisance parameter that is identified only under the alternative. Nevertheless, for a given value of δ_1 such test can be easily performed.

3.2 Tests for non-nested hypotheses

The adequacy of (4) as a model for $P(Y = y|Y > 0, X)$ can be checked comparing the specification of the conditional mean under the null with its specification under alternative assumptions about the conditional distribution of Y. If (4) is true, $E(Y|X, Y > 0)$ can be written as

$$\mu^*(X,\beta,\delta) = \frac{\mu(X,\beta)}{1 - \exp\left\{-\mu(X,\beta)\left[1 + \alpha(X,\delta)\mu(X,\beta)\right]^{-1}\right\}}. \tag{5}$$

Because the form of $\mu^*(X,\beta,\delta)$ depends on the shape of the conditional distribution, this test can be interpreted as a test for the adequacy of the whole conditional distribution specification.

A convenient test for the adequacy of $\mu^*(X,\beta,\delta)$ as a model of the conditional expectation of Y can be constructed using a simple modification of the P test proposed by Davidson and MacKinnon (1981). Letting $\mu^\dagger(X,\phi)$ denote the alternative specification of $E(Y|X, Y > 0)$, the P test of the null against this alternative can be performed by testing $H_0 : \gamma = 0$ in the artificial linear regression model defined by

$$u = \frac{\partial\mu^*(X,\beta,\delta)}{\partial\beta}\eta + \frac{\partial\mu^*(X,\beta,\delta)}{\partial\delta}\nu + \gamma\left[\mu^\dagger(X,\phi) - \mu^*(X,\beta,\delta)\right] + \zeta, \tag{6}$$

where $u = Y - \mu^*(X,\beta,\delta)$, β and δ are evaluated at their maximum likelihood estimates under the null, ϕ is evaluated at its maximum likelihood estimate under the alternative and η, ν and γ are parameters to be estimated. Notice that asymptotically equivalent tests can be constructed evaluating ϕ at any point $\bar{\phi}$ such that $\text{plim}(\bar{\phi}) = \phi$. Although such tests can have better finite sample performance than the test that uses the maximum likelihood estimate of ϕ [see Fisher and McAleer (1981)], their use will not be studied here.

Under the null, $\eta = \nu = \gamma = 0$, and therefore $\zeta = u$. This shows that the errors in this regression are heteroskedastic, and that has to be taken into account when testing $H_0 : \gamma = 0$. A way of tackling this problem is to construct the test statistic using a heteroskedasticity robust covariance matrix of the type described in White (1982) [see also Wooldridge (1990)]. However, noting that under the null the conditional variance of Y is given by

$$V(Y|Y>0,X) = \mu^*(X,\beta,\delta)\left[(1+\alpha(X,\delta)\mu(X,\beta))^2 + \mu(X,\beta) - \mu^*(X,\beta,\delta)\right],$$

the test statistic can also be computed as the t statistic for $H_0 : \gamma = 0$ when (6) is estimated by weighted least squares, using $[V(Y|Y > 0, X)]^{-1/2}$

as weights. Under the null, this statistic is asymptotically distributed as a standard normal variate.

It is worth noting that the power of this test depends critically on the degree of truncation. In fact, if the conditional expectation of Y has the same form both under the null and under the alternative, $\mu^*(X,\beta,\delta)$ will approach $\mu^\dagger(X,\phi)$ as $P(Y=0|X)$ goes to zero. Therefore, in samples for which $P(Y=0|X)$ is on average low, the test will have low power as $\mu^*(X,\beta,\delta)$ and $\mu^\dagger(X,\phi)$ will be practically indistinguishable. This is a natural consequence of the fact that in these circumstances the presence of overdispersion or other forms of misspecification of the higher order moments of the conditional distribution results only in a mild bias of the estimator for the parameters of the conditional mean. This point is illustrated, for example, by the simulation results reported in Grogger and Carson (1991).

4 An Illustrative Application

In an important paper, Grogger and Carson (1991) illustrated several aspects of truncated count data models using a data set on the number of recreational fishing trips for households in Alaska during the fishing season. This data set is from a diary survey of 1063 Alaskan households which took at least one fishing trip.

The following covariates were used by Grogger and Carson (1991): Avlong, the average length of recreational fishing trips taken; Miss, a dummy variable which indicates that the last part of the diary survey was not returned; Crowd, a factor score estimated from attitudinal variables on crowding; Inc, the household income in thousands of dollars; Cperm, the cost per mile of operating the vehicle the household uses for fishing trips; Foff1, a factor score estimated from indicators of pre-season familiarity with different fishing sites; Leisure, a factor score estimated from indicators of the amount of leisure time available and alternative work opportunities; and Trate, a measure of fishing quality at different sites discounted by the household's distance from them.

In their study, Grogger and Carson (1991) estimated several models for this data, including truncated Poisson and NEGBIN2 models [see Cameron and Trivedi (1986)], concluding that there is important overdispersion in the data, even after conditioning on the covariates. In the NEGBIN2 model, overdispersion is usually interpreted as resulting from the researcher inability to condition on all the relevant characteristics of the individuals. However, in this particular example, it may result from the fact that the probability of taking a fishing trip in a given period of time increases with the number of trips taken so far. This occurrence dependence can lead to the negative binomial model [see Gurland (1959)] but it can also lead to the generalized Poisson model [Consul and Jain (1973)], which is the focus of this paper. Therefore, it is interesting to see how generalized Poisson models perform in this case.

To illustrate the application of the models and tests presented in the previous sections, two models were estimated. In both cases the conditional mean was specified as $E(Y|X) = \exp(X\beta)$. In model 1, $\alpha(X,\delta)$ was parametrized as $\delta_0 \exp(\delta_1 X\beta)$. As discussed in section 3.1, this specification is interesting because it nests both the standard GP regression model of Consul and Famoye (1992) and RGP regression model used by Famoye (1993). Furthermore, this is a very parsimonious parametrization which is likely to be adequate for many problems. However, this parametrization is somewhat restrictive since it only allows $\alpha(X,\delta)$ to depend on the covariates through the conditional mean.

For model 2, a different approach was adopted. Although there is very little *a priori* information on the way the covariates influence the conditional variance, an exploratory graphical analysis suggests that Avlong, the average length of the trip, has a negative impact not only on the average number of trips [a result confirmed by the findings of Grogger and Carson (1991)], but also on its dispersion. That is, individuals that prefer longer stays tend to have fewer trips and for large values of Avlong the dispersion around the average number of visits is very small. This suggests that Avlong may influence the conditional variance not only through $\mu(X,\delta)$, but also via $\alpha(X,\delta)$. Since no other variable exhibits such a strong relation with the counts dispersion, in model 2 $\alpha(X,\delta)$ was specified as $\exp(\delta_2 + \delta_3 \text{Avlong})$. This specification, being as parsimonious as the one used in model 1, is also somewhat restrictive in that it excludes all other covariates from the specification of $\alpha(X,\delta)$. Therefore, appropriate test statistics will have to be used to assess the adequacy of these models. A summary of the results obtained with models 1 and 2 is contained in table I.

Although model 2 presents a better fit, as measured by the value of the log-likelihood function at its maximum, both models lead to similar estimates for the parameters of the conditional mean. Furthermore, the magnitude, signs and statistical significance of the estimates are in line with the results reported by Grogger and Carson (1991). As for the parameters of $\alpha(X,\delta)$, the estimate obtained for δ_1 and the associated t statistic suggest that the truncated RGP model is a valid simplification of model 1, while the truncated GP is clearly rejected. The results for model 2 show that, as expected, Avlong has a significant and negative impact on $\alpha(X,\delta)$.

TABLE I
Estimation results for generalized Poisson models

$\alpha(X,\delta)$	Model 1 $\delta_0 \exp(\delta_1 X\beta)$		Model 2 $\exp(\delta_2 + \delta_3 \text{Avlong})$	
Parameters	Estimates	*t ratios*	Estimates	*t ratios*
β_0 (Intercept)	0.0648	0.1640	0.0367	0.1010
β_1 (Avlong)	−0.2518	12.5733	−0.2331	12.7540
β_2 (Miss)	−0.1548	2.4357	−0.1715	2.7072
β_3 (Crowd)	−0.0682	2.3341	−0.0566	1.9978
β_4 (Inc)	−0.0011	1.1482	−0.0009	1.0068
β_5 (Cpmile)	2.4348	4.9273	2.4497	5.3509
β_6 (Foff1)	−0.1669	5.4460	−0.1870	6.5658
β_7 (Leisure)	0.0835	2.8904	0.0659	2.3150
β_8 (Trate)	0.6209	5.3942	0.6165	5.7622
δ_0	0.1298	2.6983	–	–
δ_1	0.2061	1.0688	–	–
δ_2	–	–	−1.1391	8.1873
δ_3	–	–	−0.2746	4.0599
Log-Likelihood	−2837.42		−2828.29	

From the results provided in table I, it is not possible to assess the validity of the hypotheses underlaying the estimated models. In order to gain some insight into their adequacy, models 1 and 2 were tested against a more general specification in which $\alpha(X,\delta) = \exp(X\delta)$ and against a truncated NEGBIN$_k$ model [see Winkelmann and Zimmermann (1991)]. The test against the more general GP regression was performed using a likelihood ratio test, while the test against the NEGBIN$_k$ was performed using the artificial weighted regression described in section 3.2. The results of these specification tests are reported in table II.

The results in table II indicate that model 1 is rejected when compared to a model with a richer parametrization of $\alpha(X,\delta)$. On the other hand, model 2 is not rejected at the conventional 5% level, when tested against this alternative. As for the tests against the NEGBIN$_k$, none of the models can be rejected at any usual significance level. As pointed out before, the power of this test is very low when the probability of occurring a zero is low, as is the case for this example.

TABLE II: Specification tests results

Alternative hypothesis	Asymptotic dist. under the null	Null hypothesis	
		Model 1	Model 2
$\alpha(X,\delta) = \exp(X\delta)$	$\chi^2_{(7)}$	32.02	13.76
NEGBIN_k	$N(0,1)$	-0.815	0.428

In fact, in any of the models considered, the probability of occurring a zero is negligible for most individuals. As a result, different assumptions about the shape of the conditional distribution of Y have little impact on the form of its conditional expectation. Therefore, in this particular example, the specification of the conditional distribution of Y is largely irrelevant if the interest is focused only on its conditional expectation. Indeed, the models considered here and those estimated by Grogger and Carson (1991) lead to remarkably similar estimates for β. However, if other aspects of the conditional distribution are also of interest, its correct specification becomes critical.

5 SUMMARY AND CONCLUSIONS

This paper considers the use of the generalized Poisson distribution to model positive counts in a regression framework. It is argued that when the parameters of the distribution are specified as functions of explanatory variables, the lower bound restriction on λ can lead to awkward situations when this parameter is negative. Although in some applications this problem may not exist, here the analysis was restricted to non-negative values of λ.

Even when limited to the cases of equidispersion and overdispersion, the generalized Poisson distribution is an attractive model due to its simplicity and great flexibility. In fact, it was shown that setting both parameters as functions of the covariates, the generalized Poisson regression can accommodate a great variety of patterns of overdispersion and therefore it can provide an adequate description of many types of overdispersed data. This point is particularly important when dealing with truncated samples because, in this case, consistent estimation of the parameters of interest requires the correct specification of the whole conditional distribution, not only of its first moment.

Besides discussing the estimation of the generalized Poisson regression for positive counts, some specification tests were also suggested. The importance of specification tests for these models is a direct consequence of their lack of robustness to departures from the assumptions concerning the form of the conditional distribution. The tests proposed here are very simple to perform and therefore are adequate to routinely use by applied researchers.

The empirical application considered in section 4 illustrates the ability of the model to adequately describe positive overdispersed counts. Moreover, it emphasizes the usefulness of the specification tests proposed. In fact, the tests not only provide evidence that supports the hypothesis of correct specification of the preferred model, but they also lead to the rejection of alternative specifications.

ACKNOWLEDGEMENTS

This research was partially carried out at University College London with the financial support of the Human Capital and Mobility program of the European Commission, which is gratefully acknowledged. I thank the department of economics at UCL for their hospitality and support. I am grateful to an Associate Editor and Joaquim Ramalho for their most helpful comments. Jeff Grogger kindly provided the data used in section 4. The usual disclaimer applies.

BIBLIOGRAPHY

Cameron, A. C. and Trivedi, P. K. (1986). "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests". *Journal of Applied Econometrics*, 1, 29-53.

Consul, P. C. (1989). *Generalized Poisson Distributions: Properties and Applications*, Marcel Dekker, Inc., New York.

Consul, P. C. and Famoye, F. (1992). "Generalized Poisson Regression Model". *Communications in Statistics - Theory and Methods*, 21(1), 89-109.

Consul, P. C. and Jain, G. C. (1973). "A Generalization of the Poisson Distribution". *Technometrics*, 15, 791-799.

Consul, P. C. and Shoukri, M. M. (1985). "The Generalized Poisson Distribution when the Sample Mean is Larger than the Sample Variance". *Communications in Statistics - Simulations and Computation*, 14(3), 667-681.

Cox, D. R. (1961). "Tests of Separate Families of Hypotheses", in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I, 105-123, University of California Press, Berkeley.

Creel, M. and Loomis, J. (1990). "Theoretical and Empirical Advantages of Truncated Count Data Estimators for the Analysis of Deer Hunting in California". *American Journal of Agricultural Economics*, 72, 681-700.

Davidson, R. and MacKinnon, J. G. (1981). "Several Tests of Model Specification in the Presence of Alternative Hypotheses". *Econometrica*, 49, 781-793.

Famoye, F. (1993). "Restricted Generalized Poisson Regression Model". *Communications in Statistics - Theory and Methods*, 22(5), 1335-1354.

Fisher, G. R. and McAleer, M. (1981). "Alternative Procedures and Associated Tests of Significance for Non-nested Hypotheses". *Journal of Econometrics*, 16, 103-119.

Gourieroux, C., Monfort, A. and Trognon, A. (1984). "Pseudo Maximum Likelihood Methods: Applications to Poisson Models". *Econometrica*, 52, 701-720.

Grogger, J. T. and Carson, R. T. (1991). "Models for Truncated Counts". *Journal of Applied Econometrics*, 6, 225-238.

Gurland, J. (1959). "Some Applications of the Negative Binomial and Other Contagious Distribution". *American Journal of Public Health*, 49, 1388-1399.

Kumar, A. and Consul, P. C. (1980). "Minimum Variance Unbiased Estimation for Modified Power Series Distribution". *Communications in Statistics - Theory and Methods*, 9(12), 1261-1275.

Nelson, D. L. (1975). "Some Remarks on the Generalizations of the Negative Binomial and Poisson Distributions". *Technometrics*, 17, 135-136.

Pohlmeier, W. and Ulrich, V. (1995). "An Econometric Model of the Two-Part Decision Process in the Demand for Health". *The Journal of Human Resources*, 30, 339-361.

White, H. (1982). "Maximum Likelihood Estimation of Misspecified Models". *Econometrica*, 50, 1-25.

Winkelmann, R. and Zimmermann, K. F. (1991). "A New Approach for Modelling Economic Count Data". *Economics Letters*, 37, 139-143.

Wooldridge, J. M. (1990). "A Unified Approach to Robust, Regression-Based Specification Tests". *Econometric Theory*, 6, 17-43.

Received June, 1996; Revised September, 1996.