The Mixture Poisson Exponential–Inverse Gaussian Regression Model: An application in Health Services

Emilio Gómez–Déniz[1], Enrique Calderín–Ojeda[2]

Abstract

In this paper a mixed Poisson regression model for count data is introduced. This model is derived by mixing the Poisson distribution with the one–parameter continuous exponential–inverse Gaussian distribution. The obtained probability mass function is over–dispersed and unimodal with modal value located at zero. Estimation is performed by maximum likelihood. As an application, the demand for health services among people 65 and over is examined using this regression model since empirical evidence has suggested that the over–dispersion and a large portion of non–users are common features of medical care utilization data.

1 Introduction

Counting data are common in many social and biomedical studies to explain differences among cases that generate small counts of events. The Poisson distribution plays an important role in the modeling of count data. In this regard, Poisson regression models have been traditionally used to analyze data with a nonnegative integer response variable in a wide range of different applied areas, for example, biostatistics, epidemiology, accident analysis and prevention, insurance and criminology among other fields. Nevertheless, the rigidity of the Poisson mean–variance relationship makes the Poisson regression models exposed to over–dispersion (i.e. the empirical variance is larger than the empirical mean). This is a crucial modeling issue for count data since inadequate confidence interval coverage is produced when over–dispersed count data are considered. The Poisson model does not allow for heterogeneity among individuals. Often there is additional heterogeneity between individuals that is not accounted for by the predictors in the model which results in over–dispersion. To overcome this difficulty, practitioners usually use more general specifications, e.g. negative binomial regression model (Hilbe (2007) and Greene (2009)). The latter model is an example of mixed Poisson regression model. Mixed Poisson regression models are natural extensions of the Poisson regression model allowing for

[1]Department of Quantitative Methods and TiDES Institute, University of Las Palmas de Gran Canaria, Gran Canaria, Spain; emilio.gomez-deniz@ulpgc.es

[2] Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Australia; enrique.calderin@unimelb.edu.au

over–dispersion. This feature can be included in the model by assuming that the parameter of the Poisson distribution is not fixed due to the heterogeneity of the population, being likewise considered a random variable. For instance, for over–dispersed count–panel data the negative binomial and Poisson–Inverse Gaussian regression models are well–known in the statistical literature. In this regard, by using a gamma distribution for the unknown parameter θ, the former model is obtained. The latter model was proposed by Dean et al. (1989), in this case an inverse Gaussian distribution is used to describe the parameter of the Poisson distribution. These models account for over–dispersion by assuming that there will be unexplained variability among individuals who have the same predicted value. It leads to larger variance in the overall outcome distribution but has no effect on the mean.

Regrettably, other mixed Poisson regression models have not been used since they involve special functions and appropriate numerical methods are required. Nevertheless, due to the fast improvement of mathematical software these models can be handled relatively easily. In this article a new mixed Poisson regression model is proposed. As mixing distribution, a particular case of the continuous Exponential–Inverse Gaussian distribution in Bhattacharya and Kumar (1986) when one of the parameter tends to infinity is considered. Furthermore, as it arises from a mixed Poisson distribution, many of its properties can be derived from the ones of the mixing distribution. In this sense, it displays interesting features such as over–dispersion, unimodality, closed–form expressions for factorial moments of any order among other nice properties. The mixed Poisson regression model introduced in this paper does not belong to the linear exponential family of distributions. However, as Wedderburn (1974) showed, the parameter estimation and inference theory developed for the exponential family (i.e. generalized linear models), can be extended to models where a relation between the mean and variance of the response variable can be specified, even though they were not associated with a known likelihood. In this sense, the unconditional distribution obtained in the Poisson–Inverse Gaussian regression model (Dean et al. (1989)) is not part of the exponential family of distributions.

In this manuscript, the demand for health services among people 65 and over is analyzed by using this new mixed Poisson regression model. In particular, the number of hospital stays among the elderly population is considered as response variable. Moreover, as it will be shown later, the data include two important features a high proportion of zeros and over–dispersion. The use of regression model to explain the demand for health services has been studied by Gurmu and Elder (2000) where bivariate regression model for count data was used and also by Lahiri and Xing (2004) by using two–parts model based on Poisson selection model.

The remainder of the paper is structured as follows. Section 2 introduces the new Poisson distribution together with some properties; additionally parameter estimation is discussed; section 3 describes the mixed Poisson regression model derived from this distribution. Estimation is performed by maximum likelihood. Next, a numerical application to analyze factors explaining medical care of people 65 and over is examined in section 4. Finally, some conclusions are drawn in section 5.

2 The discrete model

The continuous Exponential–Inverse Gaussian distribution in Bhattacharya and Kumar (1986) can be simplified by letting one of its parameters tends to infinity. Then a more simple probability density function (pdf) is obtained. Then, the pdf of a random variable Θ following an Exponential–Inverse Gaussian distribution with a single scale parameter ϕ (henceforward $\mathcal{EIG}(\phi)$) is given by

$$f(\theta|\phi) = \sqrt{\frac{\phi}{2\,\theta}} \exp\left(-\sqrt{2\,\phi\,\theta}\right), \text{ with } \theta > 0 \text{ and } \phi > 0. \tag{2.1}$$

Let us now consider the Poisson distribution (henceforward $\mathcal{P}(\theta)$) whose probability mass function is given by

$$\Pr\{Y = y\} = e^{-\theta}\frac{\theta^y}{y!}, \ \ y = 0, 1, \dots, \ \theta > 0. \tag{2.2}$$

Definition 1. *We say that a random variable Y has a Poisson–Exponential–Inverse Gaussian distribution if it admits the stochastic representation:*

$$\begin{aligned} Y|\theta &\sim \mathcal{P}(\theta) & (2.3) \\ \theta &\sim \mathcal{EIG}(\phi), & (2.4) \end{aligned}$$

with $\phi > 0$. We will denote this distribution by $Y \sim \mathcal{PEIG}(\phi)$.

Then, the unconditional probability mass function (pmf) of Y is given by

$$p_y = \frac{\sqrt{2\phi}\,\Gamma(2y+1)}{2^{2y+1}\,y!}\,\mathcal{U}\left(\frac{1}{2}+y, \frac{1}{2}, \frac{\phi}{2}\right), \ \ y = 0, 1, \dots, \tag{2.5}$$

where $\mathcal{U}(a, b, z)$ represents the Tricomi confluent hypergeometric function given by ($a, z > 0$):

$$\mathcal{U}(a, b, z) = \frac{1}{\Gamma(a)}\int_0^\infty e^{-zs}s^{a-1}(1+s)^{b-a-1}ds \tag{2.6}$$

(see Gradshteyn and Ryzhik (1994), page 1085, formula 9211–4).

The probability generating function is given by

$$G_Y(s) = \sqrt{\frac{\phi\pi}{1-s}}\exp\left\{\frac{\phi}{2(1-s)}\right\}\left[1 - \mathrm{erf}\left(\sqrt{\frac{\phi}{2(1-s)}}\right)\right], \tag{2.7}$$

where $\mathrm{erf}(z)$ is the error function given by

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt = \frac{2z}{\sqrt{\pi}}\,{}_1F_1(1/2, 3/2, -z^2),$$

being ${}_1F_1(\cdot, \cdot, \cdot)$ the confluent hypergeometric function.

The factorial moments of order k can be obtained from (2.5). They are provided by

$$\mu_{[k]}(Y) = E[Y(Y-1)\cdots(Y-k+1)] = \frac{2k\,\Gamma(2k)}{(2\phi)^k}, \tag{2.8}$$

with $k = 1, 2, \ldots$

From the latter expression it can be seen that (2.5) is over–dispersed, since

$$\frac{var(Y)}{E(Y)} = \frac{5}{\phi} + 1 > 1.$$

Additionally, as (2.1) has an asymptotic mode at 0, the discrete model (2.5) is uni-modal with mode at 0 (see Holgate (1970)). Besides, as (2.1) is log–convex, then (2.5) is infinitely divisible and therefore, it is a compound Poisson distribution (see Propositions 8 and 9 in Karlis and Xekalaki, 2005).

Let us now suppose that $\mathbf{Y} = (Y_1, ..., Y_n)$ is a random sample of size n from the $\mathcal{PEIG}$ distribution with pmf (2.5). The log–likelihood function is proportional to

$$\ell(\phi; \mathbf{Y}) \propto \frac{n}{2} \log \phi + \sum_{i=1}^{n} \log \mathcal{U}\left(\frac{1}{2} + Y_i, \frac{1}{2}, \frac{\phi}{2}\right). \tag{2.9}$$

Having into account that

$$\frac{\partial}{\partial z} \mathcal{U}(a, b, z) = -a\, \mathcal{U}(a+1, b+1, z),$$

the maximum likelihood estimate of the parameter ϕ can be simply obtained by solving this normal equation

$$\frac{\partial \ell(\phi; \mathbf{Y})}{\partial \phi} = \frac{n}{\phi} - \sum_{i=1}^{n} \frac{\left(\frac{1}{2} + Y_i\right) \mathcal{U}\left(\frac{3}{2} + Y_i, \frac{3}{2}, \frac{\phi}{2}\right)}{\mathcal{U}\left(\frac{1}{2} + Y_i, \frac{1}{2}, \frac{\phi}{2}\right)} = 0. \tag{2.10}$$

The Fisher's information matrix can be approximated from

$$\frac{\partial^2 \ell(\phi; \mathbf{Y})}{\partial \phi^2} = -\frac{n}{\phi^2} - \sum_{i=1}^{n} \frac{\left(Y_i - \frac{1}{2}\right) \left\{\mathcal{M}_1(Y_i, \phi) + [\mathcal{M}_2(Y_i, \phi)]^2\right\}}{[\mathcal{M}_3(Y_i, \phi)]^2}, \tag{2.11}$$

where

$$\mathcal{M}_1(Y_i, \phi) = -\left(\frac{3}{2} + Y_i\right) \mathcal{U}\left(\frac{5}{2} + Y_i, \frac{5}{2}, \frac{\phi}{2}\right) \mathcal{U}\left(\frac{1}{2} + Y_i, \frac{1}{2}, \frac{\phi}{2}\right),$$

$$\mathcal{M}_2(Y_i, \phi) = \left(\frac{1}{2} + Y_i\right) \left[\mathcal{U}\left(Y_i + \frac{3}{2}, \frac{3}{2}, \frac{\phi}{2}\right)\right]^2,$$

$$\mathcal{M}_3(Y_i, \phi) = \mathcal{U}\left(\frac{1}{2} + Y_i, \frac{1}{2}, \frac{\phi}{2}\right).$$

This maximum likelihood estimate can also be calculated by using the EM algorithm. This method is a powerful technique that provides an iterative procedure to compute maximum likelihood estimation when data contain missing information. This methodology is suitable for distributions arising as mixtures since the mixing operation produces missing data. One of the main advantages of the EM algorithm is its numerical stability, increasing the likelihood of the observed data in each iteration. It does not guarantee convergence to the global maximum. It can be usually reached by starting the parameters at the moment estimates. The EM algorithm maximizes $\ell(\phi; \mathbf{Y})$ by iteratively

maximizing $E(\ell(\phi; \mathbf{Y}, \mathbf{Z}))$ where $\mathbf{Y} = (Y_1, ..., Y_n)$ denotes the sample observations and $\mathbf{Z} = (\theta_1, ..., \theta_n)$ denotes the missing observations and $\ell(\phi; \mathbf{Y}, \mathbf{Z})$ is the complete log–likelihood function.

The EM algorithm is based on two steps, the E–step, or expectation, fills in the missing data. Once the missing data are built–in, the parameters are estimated in the M–step (maximization step).

At the E–step of the *(j+1)*-th iteration the expected log–likelihood of the complete data model is computed by

$$E(\ell(\phi; \mathbf{Y}, \mathbf{Z}) \mid \mathbf{Y}, \hat{\phi}^{(\mathbf{j})}). \tag{2.12}$$

In the M–step, the updated parameter estimate is computed from maximizing the quantity (2.12) with respect to ϕ. Then, if some terminating condition is satisfied we stop iterating, otherwise move back to E–step for more iterations.

In mixed Poisson distributions (Karlis, 2005) the unobserved quantities are the realizations of θ_i of the unobserved mixing parameter for each data point Y_i, $i = 1 \ldots n$. Additionally, we assume that the distribution of $Y_i \mid \theta_i$ is Poisson with θ_i following (2.1). On the other hand, when the complete model is from the exponential family then the E–step computes the conditional expectations of its sufficient statistics. As it can be seen below, the continuous distribution given in (2.1) is a member of the exponential family of probability distributions since it can be written as

$$f(\theta|\phi) = h(\theta) \exp\left(A(\phi)\,T(\theta) - B(\phi)\right) \quad \text{where}$$

$h(\theta) = \dfrac{1}{\sqrt{2\theta}}$, $A(\phi) = -\sqrt{2\phi}$, $T(\theta) = \sqrt{\theta}$ and $B(\phi) = -\log\sqrt{\phi}$. Then, $T(\theta)$ is a sufficient statistic of this distribution.

The EM type algorithm for this model can be described as follows. From the current estimates $\phi^{(j)}$

- **E–step:** Calculate the pseudo–values

$$t_i = E(\sqrt{\theta_i} \mid Y_i, \hat{\phi}^{(j)})$$

 for $i = 1, \ldots, n$.

- **M–step:** Find the new estimates $\hat{\phi}^{(j+1)}$

$$\hat{\phi}^{(j+1)} = \frac{1}{2}\left(\frac{n}{\sum_{i=1}^{n} t_i}\right)^2.$$

- If some convergence condition is satisfied then stop iterating, otherwise move back to the E–step for another iteration.

3   The regression model

Let us now consider a random variable Y_i denoting event counts and a vector of covariates or explanatory variables $\mathbf{x_i} = (x_{i1}, \ldots, x_{ip})^t$, including an intercept, related to the *i*-th

observation that denotes a weight of observable features. In this model with fixed effects, it is assumed that

$$
\begin{aligned}
Y_i|\theta_i &\sim \mathcal{P}(\theta_i \mu_i) \\
\theta_i &\sim \mathcal{EIG}(\phi) \\
\mu_i &= \exp(\mathbf{x_i}^t \beta),
\end{aligned}
\tag{3.1}
$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^t$ a vector of regression coefficients.

The $\mathcal{PEIG}$ distribution has mean $\mu = 1/\phi$ and variance $1/\phi + 5/\phi^2$. If we parameterize $\mu_i = 1/\phi = \exp(\mathbf{x_i}^t \beta)$, the marginal mean and the marginal variance of the response distribution distribution are given by

$$
\begin{aligned}
E(Y_i|x_i) &= \exp(\mathbf{x_i}^t \beta) \text{ and} \\
var(Y_i|x_i) &= E(Y_i|x_i) + 5E(Y_i|x_i)^2,
\end{aligned}
$$

respectively.

Likewise the conditional mean of the response variable is related to the explanatory variables through a link function, $g(E(Y_i|x_i)) = \mathbf{x_i}^t \beta$, where $g(\cdot)$ is a monotonic function. The link function determines the function of the conditional mean that is predicted by $\mathbf{x_i}^t \beta$. As the mean of (2.5) is non-negative, the log–link is the usual choice for $\mathcal{PEIG}$ regression model since it guarantees a non-negative value for the conditional mean. Additionally, as $var(Y_i|x_i) > E(Y_i|x_i)$, this mixed Poisson regression model is over–dispersed. In addition to this, as the variance is determined by the mean, no additional variance estimate is required. Besides, this model does not nest the Poisson regression model. Maximum likelihood estimation for this fixed effect regression model involves setting the partial derivatives of the log–likelihood function with respect to regression coefficients β_j with $j = 1, \ldots, p$ equal to zero.

Let us now suppose that $(y_i, \mathbf{x_i})$, $i = 1, \ldots, n$ are n independent realizations of the regression model given in (3.1) where y_i is the response variable and $\mathbf{x_i}$ a vector of explanatory variables. Then, the log–likelihood function can be expressed as

$$
\begin{aligned}
\ell(\beta_1, \ldots, \beta_p) &= \sum_{i=1}^{n} \ell_i(\mu_i; \beta_1, \ldots, \beta_p) \\
&= -\frac{n}{2} \log \mu_i + \sum_{i=1}^{n} \log \Gamma(2y_i + 1) - \left(2 \sum_{i=1}^{n} y_i + \frac{n}{2}\right) \log 2 \\
&\quad - \sum_{i=1}^{n} \log y_i! + \sum_{i=1}^{n} \log \mathcal{U}\left(\frac{1}{2} + y_i, \frac{1}{2}, \frac{1}{2\mu_i}\right).
\end{aligned}
\tag{3.2}
$$

Then, the normal equations to obtain the maximum likelihood estimates are given by

$$
\frac{\partial \ell}{\partial \beta_s} = \frac{n}{2} \sum_{i=1}^{n} x_{is} + \sum_{i=1}^{n} \left(\frac{1}{2} + y_i\right) \frac{x_{is}}{2} \frac{1}{\mu_i} \frac{\mathcal{U}\left(\frac{3}{2} + y_i, \frac{3}{2}, \frac{1}{2\mu_i}\right)}{\mathcal{U}\left(\frac{1}{2} + y_i, \frac{1}{2}, \frac{1}{2\mu_i}\right)},
$$

with $s = 1, 2, \ldots, p$.

Furthermore, the required expressions to approximate the Fisher's information matrix associated with maximum–likelihood estimates are provided by

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \beta_s \partial \beta_k} &= -\sum_{t=1}^{n} \left(\frac{1}{2} + y_i \right) \frac{x_{is} x_{ik}}{2} \left(\frac{1}{2} + y_i \right) \frac{1}{\mu_i} \\
&\times \frac{\left[\mathcal{U}\left(\frac{3}{2} + y_i, \frac{3}{2}, \frac{1}{2\mu_i} \right) - \left(\frac{3}{2} + y_i \right) \mathcal{U}\left(\frac{5}{2} + y_i, \frac{5}{2}, \frac{1}{2\mu_i} \right) \right]}{\mathcal{U}\left(\frac{1}{2} + y_i, \frac{1}{2}, \frac{1}{2\mu_i} \right)} \\
&+ \left[\left(\frac{1}{2} + y_i \right) \frac{x_{is} x_{ik}}{2} \left(\frac{1}{2} + y_i \right) \frac{1}{\mu_i} \frac{\mathcal{U}\left(\frac{3}{2} + y_i, \frac{3}{2}, \frac{1}{2\mu_i} \right)}{\mathcal{U}\left(\frac{1}{2} + y_i, \frac{1}{2}, \frac{1}{2\mu_i} \right)} \right]^2,
\end{aligned}
$$

for $s = 1, 2, \ldots, p$ and $k = 1, 2, \ldots, p$.

4 Application to health service data

4.1 Estimation of parameters

In the following, we are going to illustrate the performance of this mixed Poisson regression model. For that reason, let us consider now the number of hospital stays among the elderly population age 65 and over in the U.S. This amount represents a significant portion of the annual expenditures on hospital care since government insurance programs in the U.S. bear the highest financial burden for health care. Moreover, it has been forecasted that the number of elderly will continue to grow in the coming years. This set of data appears originally in Deb and Trivedi (1997) in their analysis of various measures of health–care utilization using a sample of 4406 single–person households in 1987. Data have been obtained from the Journal of Applied Econometrics 1997 Data Archive. Estimation of model and all the data analyses were done using *Mathematica 9.0* software package. All the codes used to obtain reported results and all additional information useful to make research reproducible can be found on the journal's website or it will be made available by the authors on request. Our goal is to model the number of hospital stays (HOSP) as the response variable. This measure includes two interesting features, on the one hand over–dispersion, the mean and variance of the empirical distribution are 0.30 and 0.56 respectively, and, on the other hand, a very high proportion of non–users (80.36%). Since the Poisson regression model is not able to capture the the heterogeneity among individuals found in the data, the $\mathcal{PEIG}$ regression model is used to explain the demand for health services.

Let us firstly considered the model without covariates. Parameter estimates, standard errors (in brackets) and the maximum of the log–likelihood ($\ell_{\max}$) of the distribution of the hospital stays are $\widehat{\theta} = 0.296\,(0.01)$ and $\ell_{max} = -3304.51$ for Poisson model and $\widehat{\mu} = 0.308\,(0.01)$, $\ell_{max} = -3021.92$ for $\mathcal{PEIG}$ model respectively. For the latter model, the estimate can also be obtained by using the EM algorithm after 25 iterations when the relative change of the estimate between two successive iterations is smaller than 1×10^{-10},

after taking initial starting value in the neighborhood of the moment estimate. Therefore, it can be concluded that the $\mathcal{PEIG}$ model provides a better fit to the data than Poisson distribution by considering maximum of the log–likelihood as criterion of comparison. For the standard model given in (2.5) the estimated value of ϕ is 3.24783 with a standard error of 0.138. Since the empirical distribution is over-dispersed the Poisson model seems to be inadequate for estimating these count data. Next, in Figure 1 the histogram of the empirical distribution of the number of hospital stays (Observed), together with fitted distribution, obtained from the Poisson distribution and $\mathcal{PEIG}$ distribution has been plotted. As it can be observed, there is a clear spike of extra zeros representing the non-hospitalization of the elderly population with the best fit to the data obtained with the $\mathcal{PEIG}$ model.
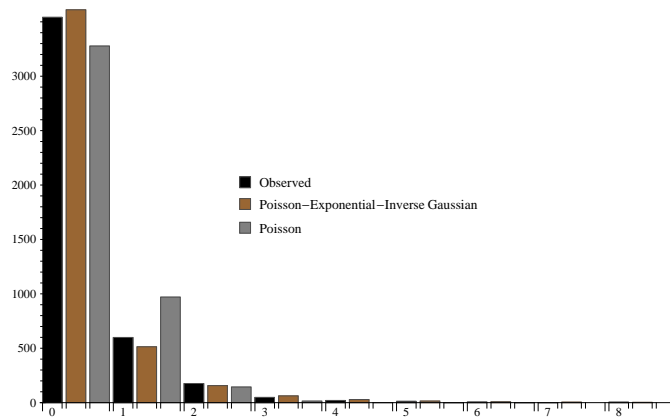
Figure 1: Observed and fitted ($\mathcal{PEIG}$ and Poisson) distribution of the number of hospital stays (HOSP)

Let us now analyze the model with covariates. The explanatory variables are as follows: (1) a dummy variable (EXCLHLTH) which takes the value 1 if self–perceived health is excellent; (2) a dummy variable (POORHLTH) which takes the value 1 if self–perceived health is poor; (3) a count variable (NUMCHRON) giving the number of chronic disease and condition (cancer, heart attack, etc.); (4) age (AGE) divided by 10; (5) a dummy variable (MALE) with value 1 if the patient is male. For the ith patient, the number of hospital stays Y_i follows a $\mathcal{PEIG}$ whose mean depends on a set of covariates trough the log–link function. The goal is to predict the number of hospital stays Y_i (response variable) using a vector of explicative variables (covariates).

At first sight, it seems logical that due to the presence of over–dispersion, a relative large long right tail, and a high proportion of zeros as compared to the proportion of other values, a simple Poisson regression model is not adequate to explain the number of hospital stays since it tends to overestimate the probability of lower values and underestimate the probability of larger values. For that reason, it is expected that a a mixed Poisson regression model will describe in a more accurate way the right tail of empirical data and the high proportion of zeros in the sample. As it can be observed in Table 1, the $\mathcal{PEIG}$ and a Poisson (in brackets) regression model have been fitted to data. From left to right parameter estimates, standard errors, t-Wald and p-values are shown for both models. After

observing the values of the estimated regressors, there exists some differences between estimated effects of both models. In this sense, the $\mathcal{PEIG}$ regression model predicts a higher use of the health service when self–perceived health is poor, the number of chronic disease and condition and age increases and the patient is male. Furthermore, when self–perceived health is excellent then the predicted change in the number of hospital stays decreases at a lower rate than in the Poisson regression model. The intercept coefficient -3.959 is the predicted logarithm of the number of hospital stays when the values of EX-CLHLTH, POORHLTH, NUMCHR, AGE and MALE are equal to 0. Having said that, it can be concluded, from this numerical application, that the $\mathcal{PEIG}$ regression model predicts a higher use of the health service for this set of explanatory variables. All of parameter estimates are significant at the usual nominal level.

Table 1: Parameter estimates, standard errors, t-Wald and p-values for $\mathcal{PEIG}$ and Poisson (in brackets) regression models for the number of hospital stays.

Parameter	Estimate	S.E.	t-Wald	$\Pr > |t|$
INTERCEPT	–3.959(–3.220)	0.52(0.32)	–7.63(–10.19)	0.00(0.00)
EXCLHLTH	–0.688(–0.720)	0.22(0.18)	–3.15(–4.10)	0.00(0.00)
POORHLTH	0.683(0.613)	0.12(0.07)	5.60(9.18)	0.00(0.00)
NUMCHRON	0.326(0.264)	0.03(0.02)	9.72(14.48)	0.00(0.00)
AGE	0.268(0.183)	0.07(0.04)	3.93(4.39)	0.00(0.00)
MALE	0.196(0.109)	0.10(0.06)	2.17(1.94)	0.03(0.05)

Following the work of Wedderburn (1974), we have also estimated the parameters by using a quasi–likelihood model. In this case, we need only to specify the marginal response variance in terms of the marginal mean, i.e. $var(Y_i) = \mu_i + 5\mu_i^2$, $(i = 1, \ldots, n)$. Via quasi–likelihood estimation, the estimates are very close to the ones shown in Table 1. Note that they are given in the same order as in Table 1, that is, –3.92958, –0.679321, 0.605773, 0.307492, 0.262405 and 0.187604. The value of the negative of the maximum of the log–likelihood is 2896.79.

4.2 Model assessment

Several measures of model validation to compare the $\mathcal{PEIG}$ and Poisson regression model are shown in Table 2. Firstly, the value of the negative of the maximum of the log–likelihood (NLL) and Akaike Information Criterion (AIC) are given in the first two rows of this Table; as a lower value of these measures is desirable, the $\mathcal{PEIG}$ regression model is preferable. Bozdogan (1987) proposed a corrected version of AIC, the Consistent Akaike Information Criteria (CAIC), in an attempt to overcome the tendency of the AIC to overestimate the complexity of the underlying model. Bozdogan (1987) also observed that AIC does not directly depend on the sample size and, as a result, it lacks certain properties of asymptotic consistency. See also Anderson et al. (1998). When formulating the CAIC, a correction factor based on the sample size is used to compensate for the overestimating nature of AIC. The CAIC is defined as $\mathrm{CAIC} = 2\,\mathrm{NLL} + (1 + \log n)\,p$,

where p refers to the number of estimated parameters and n is the sample size. Again, a model that minimize the Consistent Akaike Information Criteria is preferable. As it can be observed, the $\mathcal{PEIG}$ regression model also dominates the Poisson regression model in terms of the CAIC.

Table 2: Measures of model selection for the models considered.

	Distribution	
Criterion	Poisson	$\mathcal{PEIG}$
NLL	3047.32	2895.11
AIC	6116.63	5802.22
CAIC	6150.98	5846.57
Pearson statistic, $(\epsilon_i^P)^2$	7071.90	4626.74
Deviance residual/df	–0.30183	–0.33572

Now we perform some diagnostic checks based on analysis of residuals. This is a useful method to detect outliers and check the variance assumption in a more general setting (see Cameron and Trivedi (1986), for details). Perhaps the most common choice is Pearson's residuals. They are used to identify discrepancies between models and data, and they are based upon differences between observed data points and fitted values predicted by the model. The i-th Pearson residual for a given model is provided by

$$\epsilon_i^P = \frac{y_i - \widehat{\mu}_i}{\sqrt{var(\widehat{\mu}_i)}}, \tag{4.1}$$

where $\widehat{\mu}_i$ is the fitted marginal mean and $var(\widehat{\mu}_i)$ is the estimated marginal variance under the discussed model. Hence, if the model is correct, the variability of these residuals should appear to be fairly constant, when they are plotted against fitted values or predictors. The Pearson's residuals are often skewed for non–normal data, and this make the interpretation of the residual plots more difficult to interpret. For that reason, other quantifications of the discrepancy between observed and fitted values have been suggested in the literature. In this regard, another choice in the analysis of residual is the signed square root of the contribution to the deviance goodness–of–fit statistic (i.e. deviance residuals). This is given by $D = \sum_{i=1}^{n} d_i$, where

$$d_i = \text{sgn}(y_i - \widehat{\mu}_i)\sqrt{2(\ell(y_i) - \ell(\widehat{\mu}_i))}, \quad i = 1, 2, \ldots, n,$$

and sgn is the function that returns the sign (plus or minus) of the argument. The $\ell(y_i)$ term is the value of the log likelihood when the mean of the conditional distribution for the i-th individual is the individual's actual score of the response variable. The $\ell(\widehat{\mu}_i)$ is the log–likelihood when the conditional mean is plugged into the log–likelihood. Usually the deviance divided by its degree of freedom is examined by taking into account that a value much greater than one indicates a poorly fitting model. See for example De Jong and Heller (2008).

It is well–known that for the Poisson distribution with parameter θ_i the deviance residuals are given by (see Dunteman and Ho 2006))

$$d_i = \text{sgn}(y_i - \widehat{\theta}_i) \left[2 \left(y_i \log\left(\frac{y_i}{\widehat{\theta}_i}\right) - (y_i - \widehat{\theta}_i) \right) \right]^{1/2}, \quad i = 1, 2, \ldots, n. \qquad (4.2)$$

For the model introduced in this manuscript the deviance residuals are easily obtained by

$$d_i = \text{sgn}(y_i - \widehat{\mu}_i) \left\{ 2 \left[\log\left(\frac{\mathcal{U}(0.5 + y_i, 0.5, (2y_i)^{-1})}{\mathcal{U}(0.5 + \widehat{\mu}_i, 0.5, (2\widehat{\mu}_i)^{-1})} \right) - \frac{1}{2} \log\left(\frac{y_i}{\widehat{\mu}_i}\right) \right] \right\}^{1/2},$$
$$i = 1, 2, \ldots, n.$$

Note that the deviance does not exist whenever there are zero responses in the data. However, it is usually assumed that $d_i = 0$ when $y_i = 0$ (e.g. $y_i \log y_i$ is zero for $y_i = 0$). The Pearson's statistics together with the deviance residual divided by the degree of freedom are shown in Table 2. The $\mathcal{PEIG}$ dominates widely the Poisson distribution in terms of the Pearson's statistics and small differences appear in the value of the deviance residual. Recall that we have taken this value as zero when the observed response variable takes the value zero.

Graphical model diagnostic may also be developed using expression (4.1). In this case, for the Poisson regression model this reduces to $\epsilon_i^P = (y_i - \widehat{\theta}_i)/\sqrt{\widehat{\theta}_i}$, while for the distribution $\mathcal{PEIG}$ regression model, this expression is given by $\epsilon_i^P = (y_i - \widehat{\mu}_i)/\sqrt{\widehat{\mu}_i(1 + 5\widehat{\mu}_i)}$ as it can be easily verified. For this example, not much differences are found between these plots and those ones produced by the raw residuals, $y_i - \widehat{\theta}_i$, which are shown in Figure 2. On the other hand, the Pearson's residuals are usually standardized by divid-
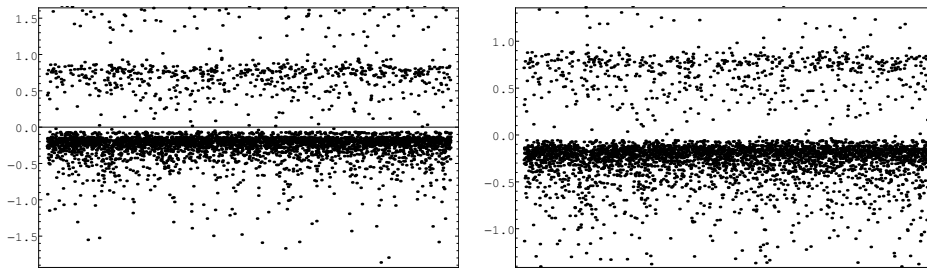
Figure 2: Plots of the raw residuals for the Poisson (left) and the $\mathcal{PEIG}$ (right) regression models.

ing by $\sqrt{1 - h_i}$, where h_i are the leverages obtained from the diagonal of the hat matrix $W^{1/2}X(X'WX)^{-1}X'W^{1/2}$, being W equal to the $n \times n$ diagonal matrix with i–th entry w_i, given by $w_i = (\partial\theta_i/\partial x'\beta))^2 /var(Y_i)$. This results θ_i for the Poisson regression model and $\mu_i/(1 + 5\mu_i^2)$ for the regression based on the new distribution presented here. See Cameron and Trivedi (1986) for details about the construction of the hat matrix. The standardized Pearson's residuals have also been plotted, they are shown in Figure 3. As it can be seen, for the Poisson regression model many of the values of the Pearson's standardized residuals lie outside the range $(-2, 2)$, pointing out a poorer fit to data than the
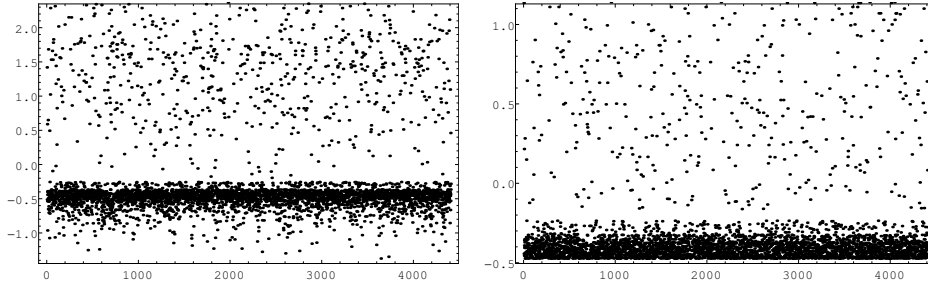
Figure 3: Standardized Pearson's residuals for the Poisson (left) and the $\mathcal{PEIG}$ (right) distributions

one obtained for the $\mathcal{PEIG}$ regression model presented in this work. See Hilbe (2007) for details.

In the following, as the regression model introduced in this paper is not nested in the Poisson regression model, the Vuong's test can be used to compare the estimates of the Poisson regression model and $\mathcal{PEIG}$ regression model. In this regard, one might be interested in testing the null hypothesis that the two models are equally close to the actual model, against the alternative one that one of the model is closer (see Vuong (1989)). The z-statistic is

$$Z = \frac{1}{\omega\sqrt{n}} \left(\ell(\widehat{\mu}) - \ell(\widehat{\theta}) \right),$$

where

$$\omega^2 = \frac{1}{n} \sum_{i=1}^{n} \left[\log\left(\frac{f(\widehat{\mu})}{g(\widehat{\theta})} \right) \right]^2 - \left[\frac{1}{n} \sum_{i=1}^{n} \log\left(\frac{f(\widehat{\mu})}{g(\widehat{\theta})} \right) \right]^2$$

and f and g represent here the $\mathcal{PEIG}$ and Poisson distributions, respectively.

Due to the asymptotic normal behaviour of the Z statistic under the null hypothesis, rejection of null hypothesis in favour of the alternative one that f occurs with significance level α, when $Z > z_{1-\alpha}$ being $z_{1-\alpha}$ the $(1 - \alpha)$ quantile of the standard normal distribution. For the Vuong's test, $Z = 3.95754$, then the $\mathcal{PEIG}$ model is preferred at the usual nominal levels.

4.3   Comparisons with other models

Finally the fit obtained with the $\mathcal{PEIG}$ regression model is compared to two other mixed Posisson regression models traditionally used in the statistical literature, the negative binomial and the Poisson–Inverse Gaussian regression models (see Dean et al. (1989)). Furthermore, when the empirical data includes a high presence of zeros it is usual to consider a reparameterization of the parent distribution to capture all zeros in the sample, the zero–inflated (ZI) model. If the parent distribution is $p(x)$, a ZI distribution is built as follows (see Cohen (1966))

$$p(x) = \begin{cases} (1 - \psi) + \psi\, p(0), & x = 0, \\ \psi\, p(x), & x > 0, \end{cases}$$

where $p(x)$ is the parent distribution and $0 < \psi \leq 1$ is the inflated parameter. The $\mathcal{PEIG}$, negative binomial and Poisson–Inverse Gaussian distributions have been reparameterized to obtain the maximum likelihood estimates under the ZI model and the results, together with the homogeneous models (without inflation), are displayed in Table 3.

Table 3: Maximum of the log–likelihood and Consistent Akaike Information Criteria (CAIC) for different homogeneous and ZI models.

Distribution	Homegeneous		ZI	
	NLL	CAIC	NLL	CAIC
$\mathcal{PEIG}$	2895.11	5846.57	2851.90	5769.74
NB	2857.11	5779.95	2853.37	5781.87
PIG	2877.33	5820.40	2847.69	5770.51

As it can be seen in this Table, the (ZI) $\mathcal{PEIG}$ regression model provides the best fit to data for this particular dataset when the CAIC is used as a criterion of comparison since the other two mixed Poisson regression models include an additional parameter. Since the global maximum of the log–likelihood surface is not guaranteed, different initial values of the parametric space were considered as a seed point. The calculations have been completed by using the `FindMaximum` function of Mathematica software package v.9.0 (Wolfram (2003)) (the derivative of the modified Bessel function of the third kind is available in this package). Additionally, by using other different methods such as `Newton`, `PrincipalAxis` and `QuasiNewton` the same results were obtained.

5   Conclusions

In this paper, a new mixed Poisson regression model to explain the demand for health services among people 65 and over to account for a large portion of non–users has been proposed. This model has been derived by mixing the Poisson distribution with a particular case of the continuous Exponential–Inverse Gaussian distribution when one of its parameter tends to infinity. Additionally, it is over–dispersed and unimodal with modal value located at zero. The model might be considered an alternative to Poisson regression model when the empirical data include a high proportion of zeros. In this regard, several measures of model assessment, including the Vuong's test for non-nested model selection, have been provided to support this goal. Apart from that, due to the high proportion of zeros in the empirical data, a zero–inflated version of this model has also been used to explain the demand for health services of elderly people.

Acknowledgment

References

[1] Anderson, D.R., Burnham, K.P. and White, G.C. (1998). Comparison of Akaike information criterion and consistent Akaike information criterion for model selection and statistical inference from capture–recapture studies. *Journal of Applied Statistics*, **25**, 2, 263–282.

[2] Bhattacharya, S.K., Kumar, S. (1986): E–IG model in life testing. *Calcutta Statistical Association Bulletin*, **35**, 85–90.

[3] Bozdogan, H. (1987): Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, **52**, 3, 345–370.

[4] Cameron, A.C. and Trivedi, P.K. (1986): Econometric models based on counts data: comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, **1**, 1, 29–53.

[5] Cohen, A. C. (1966): A note on certain discrete mixed distributions. Biometrics, **22**, 3, 566–572.

[6] De Jong, P. and Heller, G.H. (2008): *Generalized Linear Models for Insurance Data*. Cambridge University Press.

[7] Dean, C.B., Lawless, J., and Willmot, G.E. (1989): A mixed Poisson–inverse Gaussian regression model. *Canadian Journal of Statistics*, **17**, 2, 171–181.

[8] Deb, P. and Trivedi, P.K. (1997): Demand for Medical Care by the Elderly: A Finite Mixture Approach. *Journal of Applied Econometrics*, **12**, 3, 313–336.

[9] Dunteman, G.H. and Ho, M–H.R. (2006): *An Introduction to Generalized Linear Models*. SAGE Publications.

[10] Gradshteyn, I.S., Ryzhik, I.M. (1994): Table of Integrals, Series, and Products. Alan Jeffrey, Editor. Fifth Edition. Academic Press, Boston.

[11] Greene, W. (2009): Models for count data with endogenous participation. *Empirical Economics*, **36**, 133–173.

[12] Gurmu, S. and Elder, J. (2000): Generalized bivariate count data regression models. *Economics Letters*, **68**, 31–36.

[13] Hilbe, J.M. (2007): *Negative Binomial Regression*. New York: Cambridge University Press.

[14] Holgate, P. (1970): The modality of some compound Poisson distribution. *Biometrika*, **57**, 666–667.

[15] Karlis, D. (2005): EM algorithm for mixed Poisson and other discrete distributions. *Astin Bulletin*, **35**, 3–24.

[16] Karlis, D. and Xekalaki, E. (2005): Mixed Poisson distributions. *International Statistical Review*, **73**, 35–58.

[17] Lahiri, K. and Xing, G. (2004): An econometric analysis of veterans' health care utilization using two–part models. *Empirical Economics*, **29**, 431–449.

[18] Vuong, Q. (1989): Likelihood ratio tests for model selection and non–nested hypotheses. *Econometrica*, **57**, 307–333.

[19] Wedderburn, R.W.M. (1974): Quasi–likelihood functions, generalized linear models and the Gauss–Newton method. *Biometrika*, **61**, 439–447.

[20] Wolfram, S. (2003): *The Mathematica Book*. Wolfram Media, Inc.