Première partie

Modèles avec indépendance temporelle

Chapitre 1

Risk Classification for Claim Counts : A Comparative Analysis of Various Zero-Inflated Mixed Poisson and Hurdle Models

Publication prochaine dans *North American Actuarial Journal*
en collaboration avec Michel Denuit et Montserrat Guillén

1. Introduction

In most developed countries, motor third party liability insurance represents a considerable share of the yearly non-life premium collection. Therefore, many attempts have been made in the actuarial literature to find a probabilistic model for the distribution of the annual number of automobile accidents reported to the insurance company. Most of these models are parametric (i.e., an analytical expression is assumed for the probabilities that a policyholder reports k claims during an insurance period, depending on one or several parameters to be estimated from the observations). Let us mention, e.g., the Generalized Geometric and Negative Binomial distributions in Gossiaux and Lemaire (1981), Willmot (1987) and Besson and Partrat (1992), the Poisson-Inverse Gaussian distribution in Willmot (1987), Besson and Partrat (1992) and Tremblay (1992), the Generalized Poisson-Pascal distribution in Consul (1989), the Consul distribution in Islam and Consul (1992) and the Poisson-Goncharov distribution in Denuit (1997). An excellent account of claim frequency distributions can be found in Klugman et al. (2004) (Chapter 4).

In risk classification, a regression component is included in the claim count distribution to take the individual characteristics into account. References for risk classification include, e.g., Dionne and Vanasse (1989) Dionne and Vanasse (1992), Dean et al. (1989), Denuit and Lang (2004), Gourieroux and Jasiak (2004) and Yip and Yau (2005). This paper compares different risk classification models on the basis of a sample of the automobile portfolio of a major company operating in Spain. Specification tests to select the optimal model are proposed. Models are estimated using the SAS system with GENMOD, IML and NLMIXED procedures.

The number of motor liability claims reported to an insurance company exhibits some specific characteristics that must be considered when choosing a distribution that will fit the data. In Section 2, we see that the omission of important classification variables justifies the inclusion of an heterogeneity component in the regression model. The high percentage of zero values motivates zero-inflated models that are presented in Section 3. In Section 4, models for the demand for certain types of health care services (called hurdle distributions) are applied to the number of reported claims. The quality of the fit can be explained by the insured's behavior that is modified when a claim has already been reported in the year. A compound frequency distribution, called the $NegBin_x$ distribution, is explored in Section 5. Section 6 then establishes a comparison between the resulting a priori premiums. The links between models are made apparent in Section 7. Standard specification tests (Wald and Likelihood-Ratio tests) do not keep their simple asymptotic properties when the tested parameter lies on the border of the parameter space. In Section 8, nested models

Variable	Description
v1	equals 1 for women and 0 for men
v2	equals 1 when driving in urban area, 0 otherwise
v3	equals 1 when zone is medium risk (Madrid and Catalonia)
v4	equals 1 when zone is high risk (Northern Spain)
v5	equals 1 if the driving licence is between 4 and 14 years old
v6	equals 1 if the driving licence is 15 or more years old
v7	equals 1 if the client has been in the company between 3 and 5 years
v8	equals 1 if the client has been in the company for more than 5 years
v9	equals 1 if the insured is 30 years old or younger
v10	equals 1 if coverage includes comprehensive except fire
v11	equals 1 if coverage includes comprehensive (material damage and fire)
v12	equals 1 if power is greater or equal to 5500 cc

TAB. 1.1 – Binary variables summarizing the information available about each policyholder.

are compared by means of the Hausman and score tests. Section 9 presents other kinds of tests that can be used to compare non-nested models, more specifically the Vuong test and the selection procedures based on Information Criteria selection. The final Section 10 concludes.

Let us briefly present the data used to illustrate the techniques described in this paper. Here, we work with a sample of the automobile portfolio of a major company operating in Spain. Only private use cars have been considered in this sample. We have 18 exogeneous variables for every policy, as well as the total number of claims at fault that were reported whithin the yearly period. The average annual claim frequency is 6.9% and the maximum number of reported claims recorded is 5.

The exogenous information is coded by means of binary variables, which are described in Table 1. Some of these exogenous variables are included in the model via transformations $h(\beta_0 + \boldsymbol{x}_i'\boldsymbol{\beta})$, where β_0 is the intercept and $\boldsymbol{\beta}' = (\beta_1, ..., \beta_p)$ is a vector of regression parameters for the binary explanatory variables $\boldsymbol{x}_i' = (v_{i,1}, ..., v_{i,p})$ with $p = 12$. More generally, the expression $\boldsymbol{x}'\boldsymbol{\beta} = \beta_0 + \sum_{j=1}^{p} \beta_j v_{i,j}$ is used, where $\boldsymbol{\beta}$ includes an intercept.

Number of reported claims	Observed	Predicted (Poisson)
0	513,814	502,087
1	32,296	44,607
2	2,493	2,068
3	203	67
4+	24	2
Total	548,830	548,830

TAB. 2.1 – Observed claim counts versus Poisson fit.

2. Heterogeneous Models

2.1 Overview

Insurance data often exhibits overdispersion that may be caused by the omission of some important classification variables (swiftness of reflexes, aggressiveness behind the wheel, consumption of drugs, etc.). Table 2 shows the fit of the observed claim frequencies by the Poisson distribution. The large discrepancies between the observed and predicted claim numbers lead to the rejection of the Poisson model. The rejection is interpreted as the sign that the portfolio is heterogeneous : the Poisson frequency may not be the same for all the policyholders.

Equidispersion implied by the Poisson distribution is usually corrected by the introduction of a random heterogeneity term with unit mean and constant variance τ. See, e.g., Boyer et al. (1992) or Lemaire (1995). Specifically, given $\Theta_i = \theta$, the number of claims N_i reported by policyholder i conforms to the Poisson distribution with mean $\lambda_i\theta$, where $\lambda_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$. The variance of N_i is thus equal to $\lambda_i + \tau\lambda_i^2$ and exceeds the mean λ_i. At the portfolio level, the Θ_i's are assumed to be independent and identically distributed.

The distribution to model the heterogeneity component is often taken to be Gamma, Inverse-Gaussian or Log-Normal. As shown by Holgate (1970), all continuous mixtures based on the Poisson distribution have unimodal likelihood functions. This ensures that elementary numerical techniques (Newton-Raphson, for instance) lead to maximum likelihood estimators.

2.1.1 Gamma Heterogeneity

Let us denote as f_{N_i} the discrete probability mass function of N_i, i.e. $f_{N_i}(k) = P[N_i = k]$. If Θ_i is Gamma distributed with unit mean and variance α then policyholder i having reported n_i

claims to the insurer contributes to the likelihood by

$$f_{N_i}(n_i) = \frac{\Gamma(n_i + \alpha^{-1})}{\Gamma(n_i + 1)\Gamma(\alpha^{-1})}\Big(\frac{\lambda_i}{\alpha^{-1} + \lambda_i}\Big)^{n_i}\Big(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i}\Big)^{\alpha^{-1}} \tag{2.1}$$

where $\lambda_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$ and $\Gamma(a) = \int_0^\infty e^{-t}t^{a-1}dt$. We recognize the Negative Binomial distribution. Clearly, $E[N_i] = \lambda_i < Var[N_i] = \lambda_i + \alpha\lambda_i^2$. The Poisson distribution is obtained with $\alpha = 0$. However, the null hypothesis $H_0 : \alpha = 0$ corresponds to the border of the parameter space, and testing for H_0 requires some care. This point, and other specification tests will be analysed in Section 8.

Cameron and Trivedi (1986) considered the NBp distributions having the same mean λ_i as in (2.1), but a variance of the form $\lambda_i + \alpha\lambda_i^p$. This kind of ditribution can be generated with an heterogeneity factor following a Gamma distribution with unit mean and variance $\alpha\lambda_i^{p-2}$. Note that the variance of Θ_i now depends on the individual characteristics of the policyholders. When $p = 2$, the $NB2$ distribution coincides with the Negative Binomial distribution. When p is set to 1, we get the $NB1$ model with probability mass function

$$f_{N_i}(n_i) = \frac{\Gamma(n_i + \alpha^{-1}\lambda_i)}{\Gamma(n_i + 1)\Gamma(\alpha^{-1}\lambda_i)}(1 + \alpha)^{-\lambda_i/\alpha}(1 + \alpha^{-1})^{-n_i}. \tag{2.2}$$

The NB1 model is interesting because the variance $Var[N_i] = \lambda_i + \alpha\lambda_i = \phi\lambda_i$ is the one used in the Poisson GLM approach (such as the one used for the overdispersion correction to the Poisson distribution in the GENMOD procedure of SAS).

Winkelmann and Zimmermann (1991), Winkelmann and Zimmermann (1995) proposed to treat p as an unknown parameter. Putting $p = k+1$, the variance of their model has the form $Var[N_i] = \lambda_i + \sigma^2\lambda_i^{k+1}$. This leads to a distribution called "Generalized Event Count" ($GECk$), which can be expressed using the characterization of the Katz family of distributions. In case of overdispersion, we refer this distribution to $NB_{(k+1)}$. The contribution of policyholder i to the likelihood for the $NB_{(k+1)}$ is

$$f_{N_i}(n_i) = (1 + \sigma^2\lambda_i^k)^{-\lambda_i^{1-k}/\sigma^2}\prod_{j=1}^{n_i}\left(\frac{\lambda_i + \sigma^2(j-1)\lambda_i^k}{(1 + \sigma^2\lambda_i^k)j}\right). \tag{2.3}$$

It is possible to estimate all the parameters by maximum likelihood as shown by Winkelmann and Zimmermann (1991). Note that the GECk distribution is a member of the Panjer family since

$$\frac{\Pr[N_i = n_i]}{\Pr[N_i = n_i - 1]} = a + \frac{b}{n_i},\ n_i \leq 1, \tag{2.4}$$

with $a = \lambda_i^k\sigma^2/(1 + \sigma^2\lambda_i^k)$ and $b = (\lambda_i - \alpha\lambda_i^k)/(1 + \alpha\lambda_i^k)$.

2.1.2 Inverse Gaussian Heterogeneity

If Θ_i follows the Inverse Gaussian distribution with unit mean and variance τ, we get the Poisson Inverse-Gaussian (PIG) distribution with probability mass function

$$f_{N_i}(n_i) = \frac{\lambda_i^{n_i}}{n_i!}\left(\frac{2}{\pi\tau}\right)^{0.5} e^{1/\tau}(1+2\tau\lambda_i)^{-s_i/2}K_{s_i}(z_i) \tag{2.5}$$

where $s_i = n_i - 0.5$ and $z_i = (1+2\tau\lambda_i)^{0.5}/\tau$ and $K_j(.)$ is the modified Bessel fuction of the second kind satisfying

$$\begin{aligned} K_{-1/2}(a) &= \left(\frac{\pi}{2a}\right)^{0.5} e^{-a} \\ K_{1/2}(a) &= K_{-1/2}(a) \\ K_{s+1}(a) &= K_{s-1}(a) + \frac{2s}{a}K_s(a). \end{aligned}$$

Additional properties of the modified Bessel function of the second kind may be found in Shoukri et al. (2004).

As above, the PIG model can be extended to cover many forms of variance. If Θ_i has variance $\tau\lambda_i^{k-1}$, we get the $PIG_{(k+1)}$ distribution with probability mass function

$$f_{N_i}(n_i) = \frac{\lambda_i^{n_i}}{n_i!}\Big(\frac{2}{\pi\tau\lambda_i^{k-1}}\Big)^{0.5} e^{1/\tau\lambda_i^{k-1}}(1+2\tau\lambda_i^{k})^{-s_i/2}K_{s_i}(z_i) \tag{2.6}$$

where $s_i = n_i - 0.5$ and $z_i = (1+2\tau\lambda_i^k)^{0.5}/\tau\lambda_i^{k-1}$. In this case, $Var[N_i] = \lambda_i + \tau\lambda_i^{k+1}$. The parameters of the model can be estimated by maximum likelihood.

2.1.3 Log-Normal Heterogeneity

If Θ_i conforms to the LogNormal distribution with parameters $\mu = -\sigma^2/2$ and σ^2, we get the Poisson LogNormal (PLN) distribution. The contribution of policyholder i to the likelihood is then written

$$f_{N_i}(n_i) = \int_{-\infty}^{\infty} \frac{\exp(-\gamma_i)\gamma_i^{n_i}}{n_i!}\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{1}{2}\left(\frac{\epsilon_i}{\sigma}\right)^2)d\epsilon_i \tag{2.7}$$

where $\gamma_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i)$. This probability mass function cannot be expressed in closed form. Statistical packages such as SAS (NLMIXED procedure) can be used to evaluate 2.7. Here also, other forms of variance can be envisaged, having the same form as the other heterogenous models since $Var[N_i] = \lambda_i + \sigma^2\lambda_i^{k+1}$.

2.2 Insurance Application

The identifiability of the nature of occurrence dependence, through observed contagion has been adressed for a long time by the statistical literature devoted to count data models (see, e.g., Pinquet (2000)). Real positive contagion implies that the occurrence of an event modifies the probability of the next occurrence of the event. By opposition, apparent positive contagion arises from the recognition of the accident proneness of an individual.

Bates and Neyman (1951) stated that in a cross section of count, it is impossible to distinghish between true and apparent contagion. For the number or reported claims, as discussed in more details in Pinquet (2000), the effect of experience rating and the modification in the risk perception should imply negative true contagion, but positive contagion is observed. This indicates that although past events do not truly influence the probability to report a claim, they provide some information about the true nature of the driver. Then, the heterogeneity term of the models can be updated according to accident history of the insured.

The application of these models to the Spanish data set leads to the results displayed in Table 3. Comparing the log-likelihoods, we see that introduction of an heterogeneity term improves the fit compared to the Poisson distribution. The estimates of the parameters are approximately the same for all models. This is expected when the sufficient conditions for consistency are satisfied (see Gourieroux et al. (1984b), Gourieroux et al. (1984a)). Nevertheless, other elements can be analysed, such as the standard error values, which is important in parameter selections, in some forms of premium calculations or in solvency analysis. Moreover, note that the p-values for the dispersion parameters indicate the significance of the heterogeneity component, even if a more formal comparison between these models is made in section 8.

3. Zero-Inflated Models

3.1 Overview

As it can be seen from Table 2, the number of observed zeroes is much larger than under the Poisson assumption. This motivates the use of a mixture of two distributions : a degenerate distribution for the zero case and a standard count distribution. Specifically, the probability mass function is given by

$$f_{N_i}(n_i) = \begin{cases} \phi_i + (1-\phi_i)g_i(0) & \text{for } n_i = 0 \\ (1-\phi_i)g_i(n_i) & \text{for } n_i = 1, 2, ... \end{cases} \tag{3.1}$$

parameter	Poisson		$NB2$		$NB1$	
b0	−2.2625	(0.0337)	−2.2629	(0.0359)	−2.2604	(0.0349)
b1	0.0362	(0.0141)	0.0386	(0.0148)	0.0379	(0.0146)
b2	−0.0471	(0.0109)	−0.0475	(0.0114)	−0.0467	(0.0113)
b3	−0.0515	(0.0129)	−0.0514	(0.0135)	−0.0474	(0.0134)
b4	0.1833	(0.0127)	0.1832	(0.0133)	0.1815	(0.0131)
b5	−0.3672	(0.0267)	−0.3678	(0.0286)	−0.3641	(0.0277)
b6	−0.4238	(0.0290)	−0.4243	(0.0310)	−0.4250	(0.0301)
b7	−0.1364	(0.0169)	−0.1365	(0.0179)	−0.1431	(0.0175)
b8	−0.2164	(0.0157)	−0.2168	(0.0166)	−0.2215	(0.0163)
b9	0.1006	(0.0170)	0.1008	(0.0179)	0.0985	(0.0176)
b10	0.1803	(0.0143)	0.1802	(0.0151)	0.1828	(0.0148)
b11	0.0838	(0.0117)	0.0837	(0.0123)	0.0866	(0.0121)
b12	0.1048	(0.0129)	0.1062	(0.0134)	0.1051	(0.0133)
α	.	.	1.3765	(0.0441)	0.0989	(0.0032)
k	.	.	.	.	.	.
LogLike.	-146,037.9		-145,129.6		-145,105.4	

Tab. 2.2 – Mixed Poisson fits to the Spanish data set (1)

parameter	$NB_{(k+1)}$	$PIG2$	$PIG1$
b0	−2.2626 (0.0347)	−2.2614 (0.0359)	−2.2598 (0.0349)
b1	0.0380 (0.0145)	0.0387 (0.0148)	0.0379 (0.0146)
b2	−0.0464 (0.0112)	−0.0475 (0.0115)	−0.0468 (0.0113)
b3	−0.0460 (0.0133)	−0.0514 (0.0135)	−0.0475 (0.0134)
b4	0.1797 (0.0131)	0.1834 (0.0133)	0.1814 (0.0131)
b5	−0.3599 (0.0276)	−0.3685 (0.0286)	−0.3641 (0.0276)
b6	−0.4213 (0.0299)	−0.4254 (0.0310)	−0.4251 (0.0300)
b7	−0.1446 (0.0173)	−0.1377 (0.0179)	−0.1437 (0.0175)
b8	−0.2216 (0.0161)	−0.2180 (0.0166)	−0.2220 (0.0162)
b9	0.0977 (0.0175)	0.1011 (0.0180)	0.0984 (0.0176)
b10	0.1817 (0.0148)	0.1811 (0.0151)	0.1827 (0.0148)
b11	0.0865 (0.0121)	0.0842 (0.0123)	0.0868 (0.0121)
b12	0.1040 (0.0132)	0.1063 (0.0134)	0.1051 (0.0133)
α or τ	0.0517 (0.0287)	1.4086 (0.0475)	0.1007 (0.0034)
k	−0.2437 (0.2076)	. .	. .
LogLike.	-145,104.6	-145,130.19	-145,107.6

TAB. 2.3 – Mixed Poisson fits to the Spanish data set (2)

parameter	PIG_{k+1}	$PLN2$	$PLN1$	$PLN_{(k+1)}$
b0	−2.2615 (0.0347)	−2.2655 (0.0355)	−2.2585 (0.0349)	−2.2663 (0.0347)
b1	0.0378 (0.0145)	0.0381 (0.0147)	0.0377 (0.0146)	0.0379 (0.0145)
b2	−0.0463 (0.0112)	−0.0473 (0.0114)	−0.0468 (0.0113)	−0.0465 (0.0112)
b3	−0.0461 (0.0133)	−0.0512 (0.0134)	−0.0481 (0.0134)	−0.0472 (0.0133)
b4	0.1794 (0.0131)	0.1825 (0.0132)	0.1814 (0.0131)	0.1808 (0.0131)
b5	−0.3605 (0.0275)	−0.3680 (0.0283)	−0.3643 (0.0276)	−0.3611 (0.0275)
b6	−0.4220 (0.0299)	−0.4252 (0.0306)	−0.4256 (0.0300)	−0.4227 (0.0299)
b7	−0.1448 (0.0173)	−0.1376 (0.0177)	−0.1440 (0.0175)	−0.1449 (0.0173)
b8	−0.2221 (0.0161)	−0.2176 (0.0165)	−0.2226 (0.0162)	−0.2230 (0.0161)
b9	0.0971 (0.0175)	0.1000 (0.0178)	0.0982 (0.0176)	0.0983 (0.0175)
b10	0.1815 (0.0147)	0.1803 (0.0150)	0.1826 (0.0148)	0.1826 (0.0148)
b11	0.0868 (0.0120)	0.0840 (0.0122)	0.0869 (0.0121)	0.0873 (0.0121)
b12	0.1040 (0.0132)	0.1056 (0.0133)	0.1050 (0.0133)	0.1049 (0.0132)
τ or σ^2	0.0478 (0.0251)	0.7932 (0.0185)	0.0972 (0.0034)	0.0398 0.0218
k	−0.2794 (0.1961)	. .	. .	−0.2819 (0.2078)
LogLike.	-145,106.6	-145,188.7	-145,117.1	-145,116.3

TAB. 2.4 – Mixed Poisson fits to the Spanish data set (3)

where

$$\phi_i = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\gamma})}{1+\exp(\boldsymbol{x}_i'\boldsymbol{\gamma})}, \tag{3.2}$$

$\boldsymbol{\gamma}$ is a vector of regression coefficients and g_i is the probability mass function corresponding to the standard distribution to be modified. See, e.g., Lambert (1992). For instance, in the Zero-Inflated Poisson (ZIP) distribution, g_i corresponds to the Poisson distribution with mean λ_i, that is, $g_i(n_i) = e^{-\lambda_i}\frac{\lambda_i^{n_i}}{n_i!}$ for $n_i = 0, 1, \ldots$. The two first moments of the ZIP distribution are $E[N_i] = (1-\phi_i)\lambda_i$ and $Var[N_i] = E[N_i] + E[N_i](\lambda_i - E[N_i])$. ZIP models thus account for overdispersion. Note that the ZIP model is a special case of a mixed Poisson distribution obtained with Θ_i equal to 0 or λ_i (with respective probabilities ϕ_i and $1-\phi_i$).

In some situations, even when the zero-count data are fitted adequately, overdispersion of the underlying distribution (for non-zero count) may be still present. As pointed out by Grogger and Carson (1991), failing to take this overdispersion into account may lead to bad estimates in the context of zero-truncated regression models. This result extends to zero-inflated models. Methods for testing for overdispersion in the zero-inflated models are presented in more details in Section 4.

If there is some overdispersion for the non-zero count data, all the distributions seen in Section 2.1 can be used since an heterogeneity term may be incorporated to the model. For instance, the ZI-$NB_{(k+1)}$ model (which cover all the ZI-Negative Binomial cases) has probability mass function

$$f_{N_i}(n_i) = \begin{cases} \phi_i + (1-\phi_i)(1+\sigma^2\lambda_i^k)^{-\lambda_i^{1-k}/\sigma^2} & \text{for } n_i = 0 \\ (1-\phi_i)(1+\sigma^2\lambda_i^k)^{-\lambda_i^{1-k}/\sigma^2}\prod_{j=1}^{n_i}\left(\frac{\lambda_i+\sigma^2(j-1)\lambda_i^k}{(1+\sigma^2\lambda_i^k)j}\right) & \text{for } n_i = 1, 2, \ldots \end{cases} \tag{3.3}$$

Recently, Yip and Yau (2005) applied this kind of models to insurance data. Note that ZI models with overdispersion for the non-zero counts can be seen as particular cases of Poisson mixtures, obtained when Θ_i has a mixed distribution with an atom ϕ_i at the origin.

3.2 Insurance Application

Most insurance companies, especially in Europe, have implemented experience rating mechanisms, called bonus-malus schemes. In application of this mechanism, a reported claim implies an increase in the premium of the next years. This induces a "hunger for bonus" (Lemaire (1995)) : there is an incentive to not report all incurred claims since the increase of the future premiums can be higher than the insurance benefit. Specifically, it is optimal for the policyholder to retain all the claims with an amount less than some threshold δ depending on the level occupied in the bonus-malus scale. This creates an inflated probability mass at the origin for the observed number

parameter	ZI-Poisson with constant ϕ		ZI-Poisson with ϕ_i given by (3.2) β		γ	
b0	−1.4391	(0.0394)	−0.7308	(0.1036)	−1.8349	(0.0198)
b1	0.0381	(0.0148)	−0.3620	(0.0912)	−0.1513	(0.0460)
b2	−0.0474	(0.0114)	0.0826	(0.0197)	.	.
b3	−0.0510	(0.0135)	.	.	−0.0539	(0.0135)
b4	0.1830	(0.0133)	.	.	0.1817	(0.0133)
b5	−0.3647	(0.0284)	0.8783	(0.0865)	.	.
b6	−0.4204	(0.0308)	0.9865	(0.0900)	.	.
b7	−0.1340	(0.0178)	0.2890	(0.0343)	.	.
b8	−0.2143	(0.0166)	0.4238	(0.0325)	.	.
b9	0.1003	(0.0179)	.	.	0.0943	(0.0172)
b10	0.1787	(0.0150)	−0.3335	(0.0277)	.	.
b11	0.0830	(0.0123)	−0.1524	(0.0210)	.	.
b12	0.1058	(0.0134)	−0.1914	(0.0221)	.	.
ϕ	0.5634	(0.0075)	.	.	.	.
LogLike.	-145,163.6		-145,116.0			

Tab. 3.1 – Zero-Inflated Poisson fits to the Spanish data set

of claims. Consequently, the zero-inflated model can be used to describe this kind of behavior.

Results of the application of these models are shown in Table 3.2. Letting the zero-inflated term ϕ_i depend on observed covariates brings interesting results since it modifies the significance of the parameters of the Poisson distribution. Except for the sex of the insured, all other covariates are meaningful either for the Poisson distribution or the zero-inflated term. Such a distinction between the parameters allows us to interpret differently the insurance data. Indeed, generally speaking, covariates in the zero-inflated term modify the left tail of the distribution while parameters in the Poisson distribution affect the right tail.

4. Hurdle Models

4.1 Overview

A quick view of the pattern of reported claims depicted in Table 2 shows that the vast majority (more than 99.5%) of the insureds reports less than 2 claims per year. Consequently, a classification of the insured based on two processes would be interesting. A dichotomic variable first differentiates insureds with and without claim. In the former case, another process then generates the number of reported claims. The most popular distribution implying the assumption that the data come from two separate processes is the hurdle count model. The simplest hurdle model is the one which sets the hurdle at zero. Formally, given two probability mass functions f_1 and f_2, the hurdle-at-zero model has probability mass function

$$f_{N_i}(n_i) = \begin{cases} f_1(0) & \text{for } n_i = 0 \\ \frac{1-f_1(0)}{1-f_2(0)} f_2(n_i) = \Phi f_2(n_i) & \text{for } n_i = 1, 2, ... \end{cases} \tag{4.1}$$

where $\Phi = (1 - f_1(0))/(1 - f_2(0))$ can be interpreted as the probability of crossing the hurdle (or more precisely in case of insurance, the probability to report at least one claim). Clearly, the model collapses to f if $f_1 = f_2 = f$.

The mean and variance corresponding to (4.1) are given by

$$E[N_i] = \Phi\mu_2 \tag{4.2}$$

$$Var[N_i] = P(N_i > 0)Var[N_i|N_i > 0] + P(N_i = 0)E[N_i|N_i > 0] \tag{4.3}$$

where μ_2 is the expected value associated with the probability mass function f_2. Consequently, the model can be over or underdispersed, depending on the values of the parent processes f_1 and f_2. Many possibilites exist for f_1 and f_2. Nested models where f_1 and f_2 come from the same distribution, such as the Poisson or the Negative Binomial distributions, are often used. However, non-nested models have been used, e.g., by Grootendorst (1995) and Gurmu (1998).

The log-likelihood function of a hurdle model can be expressed as

$$\ell = \sum_{i=1}^{n} I_{(n_i=0)} \log(f_1(0;\theta_1)) + I_{(n_i>0)} \log(1-f_1(0;\theta_1)) + \sum_{i=1}^{n} I_{(n_i>0)} \log(f_2(n_i;\theta_1)/(1-f_2(0;\theta_i)) \tag{4.4}$$

The log-likelihood is then separable and maximisation can be done separately for each part (zero case and positive values).

4.2 Insurance Application

The hurdle models are widely used in connection with health care demands. An application to credit scoring is proposed in Dionne et al. (1996). With health care demand, it is generally accepted that the demand for certain types of health care services depend on two processes : the decisions of the individual and the one of the health care provider. See, e.g. Pohlmeier and Ulrich (1995) or Santos Silva and Windmeijer (2001). The hurdle model also possesses a natural interpretation for the number of reported claims. A reason for the good fitting of the zero-inflated models is certainly the reluctance of some insureds to report their accident (since they would then be penalized by some bonus-malus scheme implemented by the insurer). It is reasonable to believe that the behavior of the insureds is not same when they already have reported a claim. This suggests that two processes govern the total number of claims.

As mentioned earlier, the hurdle models are interesting because the estimators can be found in a two-steps evaluation : for the zero elements and the positive elements of the data. The fit of the zero-part is given in Table 5, whereas Table 6 describes the fit of the positive part.

The complete model is specified in choosing the best combination between the two parts. The behavior of the insureds seems to be different once a claim has been reported in a year because of the bonus-malus scheme. Indeed, direct comparison of the parameters of the Poisson distribution can be done with the parameters of the second process of the hurdle model. We see that the hurdle model has a bigger truncated expected value, implying a worst claim experience once a claim is reported.

Because only 0.5% of the portfolio can be used to model the positive part of the model, only 5 significant covariates remain : variables explaining the geographical zone of the insured and the driving experience. Results show that new insureds exhibit a worst claim experience than older ones when they already have reported a claim.

5. Compound frequency models

5.1 Overview

Compound distributions (or stopped-sum distributions) correspond to counting variables of the form

$$Z = \sum_{i=1}^{N} X_i \tag{5.1}$$

where the X_i's are integer-valued, independent and identically distributed, and where N and the X_i's are independant. Klugman et al. (2004) describe many examples of compound frequency

parameter	Poisson		NB2		PIG2	
b0	−2.3059	(0.0353)	−2.2690	(0.2475)	−2.2435	(0.1472)
b1	0.0400	(0.0147)	0.0417	(0.0185)	0.0426	(0.0164)
b2	−0.0469	(0.0114)	−0.0482	(0.0143)	−0.0490	(0.0127)
b3	−0.0450	(0.0134)	−0.0461	(0.0156)	−0.0468	(0.0146)
b4	0.1795	(0.0132)	0.1841	(0.0333)	0.1872	(0.0221)
b5	−0.3601	(0.0280)	−0.3720	(0.0845)	−0.3796	(0.0526)
b6	−0.4236	(0.0304)	−0.4369	(0.0941)	−0.4455	(0.0585)
b7	−0.1497	(0.0176)	−0.1539	(0.0337)	−0.1568	(0.0243)
b8	−0.2264	(0.0164)	−0.2326	(0.0444)	−0.2367	(0.0288)
b9	0.0976	(0.0177)	0.1001	(0.0251)	0.1018	(0.0209)
b10	0.1843	(0.0150)	0.1890	(0.0346)	0.1921	(0.0235)
b11	0.0883	(0.0122)	0.0903	(0.0188)	0.0918	(0.0150)
b12	0.1047	(0.0134)	0.1075	(0.0231)	0.1094	(0.0174)
α, τ	.	.	0.7018	(4.6252)	1.2355	(2.9186)
LogLike.	-134,299.0		-134,298.6		-134,298.4	

Tab. 4.1 – Hurdle fit to the Spanish data set - Zero parts

parameter	Poisson		NB2		NB1	
b0	−1.5714	(0.0773)	−2.5301	(0.2788)	−2.2527	(0.2287)
b3	−0.0972	(0.0460)	−0.1010	(0.0491)	−0.2756	(0.1635)
b4	0.1218	(0.0429)	0.1258	(0.0461)	0.3122	(0.1300)
b5	−0.2663	(0.0812)	−0.2767	(0.0883)	−0.6131	(0.2189)
b6	−0.2395	(0.0787)	−0.2495	(0.0856)	−0.5389	(0.1954)
α, τ	.	.	1.7257	(0.7435)	0.1076	(0.0171)
LogLike.	-10,829.7		-10,800.4		-10,800.2	

Tab. 4.2 – Hurdle fit to the Spanish data set - Positive parts (1)

parameter	PIG2		PIG1	
b0	−2.1199	(0.1246)	−1.9896	(0.1382)
b3	−0.1013	(0.0491)	−0.1869	(0.0929)
b4	0.1259	(0.0461)	0.2133	(0.0773)
b5	−0.2768	(0.0882)	−0.4357	(0.1370)
b6	−0.2500	(0.0856)	−0.3999	(0.1308)
α, τ	0.8100	(0.1795)	0.0773	(0.0099)
LogLike.	-10,800.4		-10,800.0	

TAB. 4.3 – Hurdle fit to the Spanish data set - Positive parts (2)

distributions. When N is Poisson with mean λ and X_i is Logarithmic with parameter θ, it can be proved that Z is Negative Binomial. Santos Silva and Windmeijer (2001) defined the $NegBin_x$ regression model as follows : the parameter θ_i of the logarithmic distribution is expressed in terms of the available covariates as

$$\exp(\boldsymbol{x}_i'\boldsymbol{\gamma}) = \frac{\theta_i}{1-\theta_i}$$

and the Poisson parameter is taken to be $\lambda_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$. Consequently, Z is Negative Binomial with parameter $\lambda_i / \log(1+\exp(\boldsymbol{x}_i'\boldsymbol{\gamma}))$ and $\exp(\boldsymbol{x}_i'\boldsymbol{\gamma})$. After some simplifications, the probability mass function is given by

$$f_{N_i}(n_i) = \frac{\Gamma\left(n_i + \frac{\lambda_i}{\log(1+\exp(\boldsymbol{x}_i'\boldsymbol{\gamma}))}\right)\exp(-\lambda_i)}{\Gamma(n_i+1)\Gamma\left(\frac{\lambda_i}{\log(1+\exp(\boldsymbol{x}_i'\boldsymbol{\gamma}))}\right)(1+\exp(-\boldsymbol{x}_i'\boldsymbol{\gamma}))^{n_i}}. \tag{5.2}$$

The first two moments are

$$E[N_i] = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta}+\boldsymbol{x}_i'\boldsymbol{\gamma})}{\log(1+\exp(\boldsymbol{x}_i'\boldsymbol{\gamma}))} \tag{5.3}$$

$$Var[N_i] = (1+\exp(\boldsymbol{x}_i'\boldsymbol{\gamma}))E[N_i]. \tag{5.4}$$

The variance is of the NB1 type, with overdispersion parameter depending on the covariates.

When no covariates are statistically significant for the logistic part of the model, the $NegBin_x$ distribution collapses to the NB1 distribution. Note, however, that when there are significant covariates for the logistic part of the model, the $NegBin_x$ model is close (but distinct) to the NB1 distribution with a dispersion parameter that depends on covariates (called dispersion models after Jorgensen (1997)).

5.2 Insurance Application

Santos Silva and Windmeijer (2001) have used the $NegBin_x$ distribution to model the number of visits to a doctor where N is the number of spells of illness and X is the number of visits to the doctor for a given spell. In actuarial science, this model can thus be used for modelling the number of injured persons of the number of third parties involved in a given accident.

The fit of the $NegBin_x$ model and of the NB1 distribution to the Spanish data set is described in Table 7. The reason for the quality of the fit obtained with the $NegBin_x$ distribution can be explained as follows. When several accidents occur within a short period of time, the policyholder has a tendency to report only one and claim for all the damages at the same time. The reason for this is to avoid penalties. This is known to be a source of fraudulent behaviour in companies having a bonus-malus system. So, one may interpret that the reported claims is a sum of preexisting accidents. Another interpretation would be that if an accident occurs, the driver may be careless about his/her car, knowing that the insurance company will repair the vehicle as if it all happened in the first accident.

6. Comparison between a priori claim frequencies

Difference between models can be analysed through the mean and the variance of the annual number of claims for some insured profiles. Several profiles have been selected and are described in Table 8. The first profile is classified as a good driver, while the last one usually exhibits bad loss experience. Other profiles are medium risk. The results are given in Table 9. This table shows that the biggest differences lie in the variance values.

An important comment has to be made here. The maximum likelihood estimators of the Poisson distribution, which is part of the exponential family, have the property of being consistent as long as the mean function is correctly specified. If the underlying data come from a zero-inflated or a two-part distributions, this robustness property does not hold since the mean function cannot be correctly specified.

7. Link between Models

The models considered in this paper are related as described in Figure 1. Indeed, we showed that heterogeneous models having their dispersion parameter (α, τ or σ^2) going to zero come back to a Poisson distribution. Similarly, when the zero-inflated component of the zero-inflated model goes to zero, the model only represents the underlying distribution, which can be Poisson or some

parameter	NB1 (Dispersion Model) β		γ		$NegBin_x$ β		α	
b0	−2.2662	(0.0350)	−2.4625	(0.1164)	−2.3091	(0.0350)	−2.4587	(0.1164)
b1	0.0381	(0.0146)	.	.	0.0381	(0.0146)	.	.
b2	−0.0466	(0.0113)	.	.	−0.0466	(0.0113)	.	.
b3	−0.0526	(0.0135)	−0.2065	(0.0824)	−0.0437	(0.0134)	−0.1998	(0.0821)
b4	0.1815	(0.0131)	.	.	0.1812	(0.0131)	.	.
b5	−0.3648	(0.0277)	−0.1563	(0.0725)	−0.3566	(0.0277)	−0.1654	(0.0727)
b6	−0.4212	(0.0301)	.	.	−0.4206	(0.0301)	.	.
b7	−0.1364	(0.0176)	0.3249	(0.1260)	−0.1497	(0.0176)	0.3262	(0.1259)
b8	−0.2165	(0.0164)	0.2529	(0.1192)	−0.2264	(0.0164)	0.2507	(0.1191)
b9	0.1000	(0.0176)	.	.	0.0992	(0.0176)	.	.
b10	0.1823	(0.0148)	.	.	0.1825	(0.0148)	.	.
b11	0.0866	(0.0121)	.	.	0.0868	(0.0121)	.	.
b12	0.1050	(0.0133)	.	.	0.1048	(0.0133)	.	.
LogLike.	-145,096.1				-145,096.0			

TAB. 5.1 – $NegBin_x$ fit to the Spanish data set.

Profile Number	Kind of Profile	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12
1	Good	0	0	1	0	0	1	0	1	0	0	0	0
2	Average	0	0	0	0	0	0	0	0	0	0	0	0
3	Average	1	1	0	0	1	0	1	0	1	0	1	1
4	Bad	1	0	0	1	0	0	0	0	1	1	0	1

TAB. 6.1 – The four types of policyholders to be compared.

Models	1^{st} Profile		2^{nd} Profile		3^{rd} Profile		4^{th} Profile	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Poisson	0.0521	0.0521	0.1041	0.1041	0.0789	0.0789	0.1906	0.1906
NB2	0.0521	0.0558	0.1040	0.1189	0.0791	0.0877	0.1913	0.2416
NB1	0.0521	0.0573	0.1043	0.1146	0.0794	0.0872	0.1912	0.2101
ZI−Poisson	0.0509	0.0560	0.1077	0.1133	0.1063	0.1101	0.1509	0.1554
Hurdle−Poisson	0.0522	0.0572	0.1051	0.1159	0.0786	0.0838	0.1874	0.1962
Hurdle−NB2	0.0507	0.0559	0.1022	0.1140	0.0772	0.0827	0.1833	0.1948
NegBin X	0.0543	0.0591	0.1081	0.1173	0.0823	0.0891	0.1983	0.2152

Tab. 6.2 – Comparison of a priori claim frequencies.

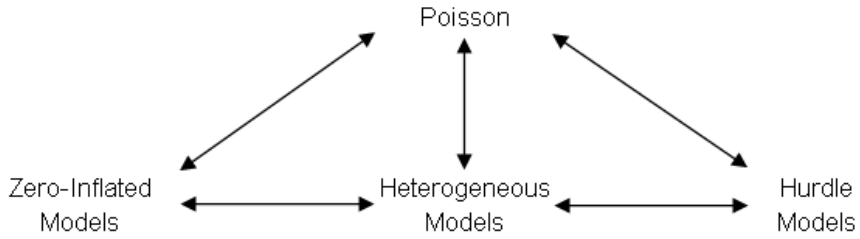


Fig. 7.1 – Links between the models.

heterogeneous models. For the hurdle distributions, we explained that if both process were equal, that is to say $f_1 = f_2 = f$, the model is then equal to f, which can also represent a Poisson distribution or a Poisson with heterogeneity. Even if it is not shown in the graph, we also stated that for specific parameters restrictions, the $NegBin_x$ distribution is nested to the standard NB1 distribution.

On the other hand, when models are not linked with an arrow, we can considered it as non-nested to each other. This can be illustrated with combination of hurdle models where f_1 and f_2 are not parent processes, meaning that they cannot be equal with parameters contrainsts.

8. Nested Models

Classical hypothesis tests can be made to accept or reject some models. The three standard tests are the log-likelihood ratio (LR), the Wald and the Score (or Lagrange Multiplier - LM) tests. Asymptotically, all three tests are equivalent. Some models are equivalent for some simple parameters restrictions. In some cases, the parameter restriction corresponds to the boundary of the parameter space. This particular situation is examined in this section.

8.1 Specification Tests

One problem with standard specification tests (Wald or Likelihood ratio tests) happens when the null hypothesis is on the boundary of the parameter space. When a parameter is bounded by the H_0 hypothesis, the estimate is also bounded and the asymptotic normality of the MLE no longer holds under H_0. Consequently, a correction must be done. Results form Chernoff (1954) for the likelihood ratio statistic and Moran (1971) for the Wald Test, reviewed by Lawless (1987) in the Negative Binomial case, showed that under the null hypothesis, the distribution of the LR statistic is a mixture of a probability mass of $\frac{1}{2}$ on the boundary and $\frac{1}{2}$ $\chi^2_{1-2\partial}$ (rather than $\chi^2_{1-\partial}$). Consequently, in this situation, one-sided test must be used. Analogous result shows that for the Wald test, there is a mass of one half at zero and a Normal distribution for the positive values. In this case, as mentioned by Cameron and Trivedi (1998), one continues to use the usual one-sided test critical value of $z_{1-\partial}$.

Nevertheless, when the hypothesized parameter lies on the boundary of the parameter space, other tests can be used without changing their properties. The asymptotic properties of the Score test, as shown by Moran (1971) and Chant (1974), are not altered when testing on the boundary

of the parameter space. Additionnaly, in some situations [1], an Hausman test can also be used since it is based on the $\boldsymbol{\beta}$ parameters and thus circumvent the boundary problem. Indeed, the Hausman test is used to compare the $\boldsymbol{\beta}$ estimates of two models, to see if the extra parameter has a significant impact on these parameters. For more details about this test, or other ones, we refer the reader, e.g., to Cameron and Trivedi (1998) or Winkelmann (2003a) in the context of count data, as well as to Gourieroux and Monfort (1995) for a general point of view.

8.2 Poisson against Heterogeneous models

The Poisson distribution is the limiting case of the heterogeneous models when the variance parameter of the heterogeneity distribution goes to zero. Another way of seing it is by the variance fonction of the heterogeneous model :

$$Var[n_i|x_i] = \lambda_i + \alpha g(\lambda_i) \tag{8.1}$$

With the function $g(\lambda_i)$ to be replaced by the form of variance to be tested. Then, we have to test the null hypothesis $H_0 : \alpha = 0$ against $H_a : \alpha > 0$.

Cameron and Trivedi (1986) suggested to use the following statistics to test for the Poisson distribution against heterogeneous models having a variance function of the form (8.1).

$$T^1_{LM} = \Big[\sum_{i=1}^{n} \frac{1}{2}\hat{\lambda}_i^{-2} g^2(\hat{\lambda}_i)\Big]^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{1}{2}\hat{\lambda}_i^{-2} g(\hat{\lambda}_i)\big((n_i - \hat{\lambda}_i)^2 - n_i\big) \tag{8.2}$$

$$T^2_{LM} = \Big[\sum_{i=1}^{n} \big(\frac{1}{2}\big)^2 \hat{\lambda}_i^{-4} g^2(\hat{\lambda}_i)\big((n_i - \hat{\lambda}_i)^2 - n_i\big)^2\Big]^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{1}{2}\hat{\lambda}_i^{-2} g(\hat{\lambda}_i)\big((n_i - \hat{\lambda}_i)^2 - n_i\big) \tag{8.3}$$

$$T^3_{LM} = \Big[\frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\hat{\lambda}_i^{-2}\big((n_i - \hat{\lambda}_i)^2 - n_i\big)^2\Big]^{-\frac{1}{2}} \Big[\sum_{i=1}^{n} \frac{1}{2}\hat{\lambda}_i^{-2} g^2(\hat{\lambda}_i)\Big]^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{1}{2}\hat{\lambda}_i^{-2} g(\hat{\lambda}_i)\big((n_i - \hat{\lambda}_i)^2 - n_i\big) \tag{8.4}$$

All of them are normally distributed with mean 0 and variance 1. The function $g(\hat{\lambda}_i)$ has to be replaced by the form of variance to be tested. For example, if $g(\hat{\lambda}_i) = \hat{\lambda}_i$, we have the NB1 or the PIG1 distribution. The test is based on the variance form of the alternative distribution, so all

[1] The test is designed for situations in which at least one of the estimator is inconsistent under the alternative. The Hausman test cannot be used to test Negative Binomial against Poisson distributions since the Poisson model implies consistent maximum likelihood estimators under the alternative.

these tests can be used against any heterogenous models. See also Dean (1992) for a discussion of tests against arbitrary Poisson mixture models.

8.3 Testing the Zero-Inflated Models

The zero-inflated models must be tested carefully since the collapsing happens when the extra parameter is set to zero (thus on the boundary of its parameter space). Even if many zero-inflated models have been fitted, only the simplest one (zero-inflated Poisson with parameter ϕ) is tested : the null hypothesis is $H_0 : \phi = 0$ against $H_a : \phi > 0$. If the model is not rejected, other tests to determine the better construction of the zero-inflated model wil be done.

The test of Poisson against all zero-inflated models is based on

$$\frac{\partial \log f(n_i)}{\partial \boldsymbol{\beta}}\Big|_{\phi=0} = (n_i - \hat{\lambda}_i)\boldsymbol{x}_i \tag{8.5}$$

$$\frac{\partial \log f(n_i)}{\partial \phi}\Big|_{\phi=0} = I_{(n_i=0)}(e^{\hat{\lambda}_i} - 1) - I_{(n_i>0)} \tag{8.6}$$

The expected Fischer information matrix is equal to

$$\begin{bmatrix} J_{\beta\beta} & J_{\phi\beta} \\ J_{\phi\beta}^T & J_{\phi\phi} \end{bmatrix}$$

As $\phi \to 0$, the elements of this matrice are equal to

$$J_{\beta\beta} = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \hat{\lambda}_i \tag{8.7}$$

$$J_{\phi\beta} = \sum_{i=1}^{n} \boldsymbol{x}_i'(n_i - \hat{\lambda}_i)\left[I_{(n_i=0)}(e^{\hat{\lambda}_i} - 1) - I_{(n_i>0)}\right] \tag{8.8}$$

$$J_{\phi\phi} = \sum_{i=1}^{n} \left(I_{(n_i=0)}(e^{\hat{\lambda}_i} - 1) - I_{(n_i>0)}\right)^2. \tag{8.9}$$

The score statistic for testing H_0 is then

$$T = \sqrt{J^{\phi\phi}}\left(I_{(n_i=0)}(e^{\hat{\lambda}_i} - 1) - I_{(n_i>0)}\right) \tag{8.10}$$

where $J^{\phi\phi}$ is the element located at the bootom left of the inverse information matrix evaluated at the maximum likelihood estimates under H_0.

Additionnaly, note that Broeck (1995) shows that a similar LM statistic can be expressed as

$$LM = \frac{\left(\sum_{i=1}^{n}(I_{n_i=0}) - e^{-\hat{\lambda}_i})/e^{-\hat{\lambda}_i}\right)}{\sqrt{\left(\sum_{i=1}^{n}(1 - e^{-\hat{\lambda}_i})/e^{-\hat{\lambda}_i}\right) - \sum_{i=1}^{n} n_i}}. \tag{8.11}$$

Under the null hypothesis, this statistic will have an asymptotic Normal(0,1) distribution. Construction of LM test for heterogeneous models against their zero-inflated modification can be done in the same way.

8.4 Heterogeneity in the Zero-Inflated models

The ZIP tests of overdispersion of Hall and Berenhaut (2002) against heterogeneous models can be used in arbitrary cases of random effects. The only assumption is on the two first moments of the heterogeneity distribution. Formally, we can express the null hypothesis as $H_0 : \alpha = 0$ against $H_a : \alpha > 0$. Based on the test of Ridout et al. (2001), which coincides with the overdispersion test of Hall and Berenhaut (2002) for log-linear Poisson regression models in which the linear predictor includes an intercept, the LM test statistic is given by

$$S = \frac{1}{2}\sum_{i=1}^{n} \hat{\lambda}_i^{k-1}\left[\left((n_i - \hat{\lambda}_i)^2 - n_i\right) - I_{(n_i=0)}\frac{\hat{\lambda}_i^2\phi}{p_{0,i}}\right] \tag{8.12}$$

with $p_{0,i}$ denoting $P(N_i = 0)$ under the null hypothese, which can be expressed as $\phi + (1 - \phi)\exp(-\hat{\lambda}_i)$.

The expected Fischer information matrix is equal to :

$$\begin{bmatrix} J_{\alpha\alpha} & J_{\alpha\beta} & J_{\alpha\gamma} \\ J'_{\alpha\beta} & J_{\beta\beta} & J_{\beta\gamma} \\ J'_{\alpha\gamma} & J^T_{\beta\gamma} & J_{\gamma\gamma} \end{bmatrix}$$

As $\alpha \to 0$, defining $\kappa_i = \hat{\lambda}_i\phi(1 - \frac{\phi}{p_{0,i}})$, the elements of this matrice are equal to :

$$J_{\alpha\alpha} = \frac{1}{4}\sum_{i=1}^{n} \hat{\lambda}_i^{2k}\left(2(1 - \phi) - \hat{\lambda}_i\kappa_i\right) \tag{8.13}$$

$$J_{\alpha\beta} = \frac{1}{2}\sum_{i=1}^{n} \hat{\lambda}_i^{k+1}\kappa_i\boldsymbol{x}'_i \tag{8.14}$$

$$J_{\alpha\gamma} = \frac{1}{2}\sum_{i=1}^{n} \hat{\lambda}_i^{k}\kappa_i \tag{8.15}$$

$$J_{\beta\beta} = \sum_{i=1}^{n} \hat{\lambda}_i \Big((1-\phi) - \kappa_i \Big) \boldsymbol{x}_i \boldsymbol{x}_i' \tag{8.16}$$

$$J_{\beta\gamma} = -\sum_{i=1}^{n} \kappa_i \boldsymbol{x}_i' \tag{8.17}$$

$$J_{\gamma\gamma} = \sum_{i=1}^{n} \frac{\phi^2 (1-p_{0,i})}{p_{0,i}}. \tag{8.18}$$

The score statistic for testing H_0 is then :

$$T = S\sqrt{J^{\alpha\alpha}} \tag{8.19}$$

where $J^{\alpha\alpha}$ is the upper left-hand element of the inverse information matrix evaluated at the maximum likelihood estimates under H_0. This test statistic is normally distributed. Once against, by the right choice of k, we recall that this test covers all the heterogeneous models seen.

8.5 Hurdle Models

8.5.1 Same Distributions

The nested hurdle models (where f_1 and f_2 come from the same distribution) collapse to their parent distribution in case of equality between the regression coefficients, i.e. when $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. The null hypothesis of equality between these two parameters can be tested with a standard Wald test, where the parameters estimates of the zero and the positive parts are independent.

8.5.2 Different Processes

Besides testing if the hurdle model is different from a standard distribution, each part of the model can be tested for overdispersion. Indeed, the positive part of the hurdle model is a truncated distribution (right truncated at one) while the other part is a dichotomous model for the count being zero or positive.

As for the heterogeneous models, the truncated Poisson is a special case of the truncated NB ot PIG distributions for the dispersion parameter going to zero. However, as opposed to the untruncated Poisson distribution, the estimates of the truncated Poisson are not consistent if the heterogeneity is misspecified.

Gurmu and Trivedi (1992) developed score tests of overdispersion for truncated Poisson distribution against truncated Negative Binomial. Following their results, the score statistic for overdis-

persion of the the positive part is given by

$$\tau = \frac{\sum_i \hat{\lambda}_i^{k-1}(\hat{\epsilon}_i^2 - n_i + (\hat{\epsilon}_i + n_i)\hat{\partial}_i)}{2\left[I_{\alpha\alpha}(\hat{\gamma}) - I_{\alpha\beta}(\hat{\gamma})I_{\beta\beta}(\hat{\gamma})I_{\beta\alpha}(\hat{\gamma})\right]^{1/2}} \tag{8.20}$$

where $\lambda_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, $\hat{\epsilon}_i = n_i - \hat{\mu}_i$, $\hat{\mu}_i = \frac{\hat{\lambda}_i}{1-\exp(-\hat{\lambda}_i)}$, $\hat{\partial}_i = \hat{\lambda}_i \frac{\exp(-\hat{\lambda}_i)}{1-\exp(-\hat{\lambda}_i)}$ and the variable k is set to 0 and 1 for the NB1 and the NB2 distributions respectively. Under H_0, the statistic τ is asymptotically Normally distributed. The denominator of the test statistic involves the elements of the information matrix evaluated under the restricted maximum likelihood estimator $\hat{\gamma}$ ($\boldsymbol{\beta}$ and α evaluated under H_0) given by

$$I_{\beta\beta}(\hat{\gamma}) = \sum_i [\hat{\lambda}_i - \hat{\partial}(\hat{\mu} - 1)]\boldsymbol{x}_i\boldsymbol{x}_i' \tag{8.21}$$

$$I_{\beta\alpha}(\hat{\gamma}) = \frac{1}{2}\sum_i \hat{\lambda}_i^k \hat{\partial}_i \hat{\mu}_i \boldsymbol{x}_i \tag{8.22}$$

$$I_{\alpha\alpha}(\hat{\gamma}) = \frac{1}{2}\sum_i \hat{\lambda}_i^{2k-2}\left(\hat{\mu}_i\hat{\partial}_i(1 - \frac{1}{2}\hat{\lambda}_i\hat{\partial}_i)\right). \tag{8.23}$$

For the truncated Poisson-Inverse Gaussian alternative, following Pohlmeier and Ulrich (1995), an Hausman test can be used for the truncated Poisson hypothesis against the truncated PIG. Additionnaly, the dichotomous process of the zero-part of the hurdle model cannot be considered as a truncated distribution. Consequently, direct application of the score tests of Gurmu and Trivedi (1992) cannot be done and a Hausman test should be used to test the significance of the additional parameter.

8.6 Numerical Application

Direct application of the score tests to our insurance data set leads to the results displayed in Table 10. Recall that the Wald and the Score statistics are Normally distributed, while the log-likelihood ratio and the Hausman statistics follow a Chi-Square distribution. These tests are asymptotically equivalent and all conclude that the Poisson distribution should be clearly rejected against all alternatives. The NB1 distribution is also rejected against the more general $NegBin_x$ distribution.

Null Hypothesis	Alternative Hypothesis	Kind of Tests	Results Value	Sig. Level	Decision
Poisson	NB2 form	Wald	29.66	0.00%	Reject
		Likelihood Ratio	1816.48	0.00%	Reject
		Score	28.72	0.00%	Reject
Poisson	NB1 form	Wald	28.24	0.00%	Reject
		Likelihood Ratio	1864.89	0.00%	Reject
		Score	27.99	0.00%	Reject
Poisson	ZI−Poisson	Wald	75.51	0.00%	Reject
		Likelihood Ratio	1748.58	0.00%	Reject
		Score	26.90	0.00%	Reject
NB1	$NegBin_x$	Wald	12.79	0.51%	Reject
		Likelihood Ratio	14.26	0.26%	Reject

Tab. 8.1 – Comparison of models for the Spanish data set.

8.6.1 Hurdle Models

Table 11 shows that estimates of $\boldsymbol{\beta}_1$ are significantly different from $\boldsymbol{\beta}_2$, which shows that all nested hurdle models (Poisson, NB2 and PIG2) are statistically different from their parent distributions.

Remember that the Wald statistic follows asymptotically a Chi-Square distribution with 13 degrees of freedom (the dimension of $\boldsymbol{\beta}$). Additionnaly, results show the rejection of the truncated Poisson in favor of the truncated versions of PIG and NB models. However, an additionnal parameter is not needed for the zero-part of the hurdle model since the Hausman test does not exhibit statistical difference between the $\boldsymbol{\beta}$ parameters of the Poisson and the overdispersed models.

9. Non-nested Models Tests

9.1 Artificial Nesting Test

A convenient method used to discriminate between non-nested models is the construction of a hypermodel, where an additional parameter is added to the tested models. Under restriction of this parameter, the hypermodel reduces to the tested distributions. We have already built this hypermodel when the $NB_{(k+1)}$, $PIG_{(k+1)}$ and the $PLN_{(k+1)}$ were created. Table 12 shows that the

Null Hypothesis	Alternative Hypothesis	Kind of Tests	Results Value	Sig. Level	Decision
Poisson	Hurdle−Poisson	Wald	2375.75	0.00%	Reject
	Nested		.	.	
NB2	Hurdle−NB2	Wald	31.60	0.00%	Reject
	Nested		.	.	
PIG2	Hurdle−PIG2	Wald	70.04	0.00%	Reject
	Nested		.	.	
Hurdle−Poisson	Hurdle−NB2	Wald	2.32	1.01%	Reject
Positive−Part	Positive−Part	Likelihood Ratio	58.54%	0.00%	Reject
		Score	8.63	0.00%	Reject
		Hausman	13.15	2.20%	Reject
Hurdle−Poisson	Hurdle−NB1 form	Wald	6.28	0.00%	Reject
Positive−Part	Positive−Part	Likelihood Ratio	58.94	0.00%	Reject
		Score	8.68	0.00%	Reject
		Hausman	16.17	0.64%	Reject
Hurdle−Poisson	Hurdle−PIG2	Wald	4.51	0.00%	Reject
Positive−Part	Positive−Part	Likelihood Ratio	58.68	0.00%	Reject
		Hausman	35.49	0.00%	Reject
Hurdle−Poisson	Hurdle−PIG1	Wald	7.79	0.00%	Reject
Positive−Part	Positive−Part	Likelihood Ratio	59.48	0.00%	Reject
		Hausman	36.03	0.00%	Reject
Hurdle−Poisson	Hurdle−NB2	Wald	0.15	44.0%	No Reject
Zero−Part	Zero−Part	Likelihood Ratio	0.86	41.8%	No Reject
		Hausman	0.03	100%	No Reject
Hurdle−Poisson	Hurdle−PIG2	Wald	0.42	33.6%	No Reject
Zero−Part	Zero−Part	Likelihood Ratio	1.31	36.3%	No Reject
		Hausman	0.21	100%	No Reject

Tab. 8.2 – Comparison of models for the Spanish data set - Hurdle Models

Models	Value of the Extra Parameter	Standard Error	Confidence Interval (95%) Lower	Upper
Negative Binomial	−0.244	0.208	−0.651	0.163
Poisson−Inverse Gaussian	−0.279	0.196	−0.664	0.105
Poisson−Log Normal	−0.282	0.208	−0.689	0.125

TAB. 9.1 – Comparison of models for the Spanish data set - Artificial Nesting Test

NB2 variance form is rejected. Indeed, for each model, confidence interval of the variable k include the value 0, but not the value 1, that represents the $NB1$ and the $NB2$ variance form respectively. Table 12 also shows the confidence interval of the extra parameter.

9.2 Information Criteria

A standard method for comparing non-nested models refers to information criteria, based on the fitted log-likelihood function. Since the likelihood increases with the addition of parameters, the criteria used to distinguish models must penalize models with a large number of parameters. The classical criteria include Akaike Information Criteria ($AIC = -2\log(L) + 2p$, where p is the number of parameters of the model) and the Bayesian Information Criteria ($BIC = -2\log(L) + \log(n)p$, where p and n represent respectively the number of parameters of the model and the number of observations). Many other penalized criteria have been proposed in the statistical literature (see Kuha (2004) for an overview). However, the AIC and BIC criteria are the most often used in practice.

Application of these criteria on the non-rejected models seen in the preceding section leads to the results displayed in Table 13. Since the models do not contain the same number of parameters, analysis of each information criteria can result in different conclusions. Indeed, even if the $NegBin_x$ and the hurdle models seem to be a lot better than the other remaining distributions, BIC favors other models. Nevethless, since we have rejected the NB1 distribution against its generalisation $NegBin_x$, it does not seem intuitive to choose a similar distribution like the PIG1 or the PLN1 models.

Despite their apparent simplicity, the information criteria are based on explicit theorical considerations, and AIC and BIC do not have the same foundations. As shown by Kuha (2004), the aims of the AIC and BIC criteria are not the same. BIC purposes to identify the model with the highest probability to be true, giving that one model under investigation is true. On the other side,

Models	Number of Parameters	Log-likelihood	AIC	BIC
NegBin X	17	−145,098.30	290,230.59	290,421.81
Hurdle : Poisson−PIG1	19	−145,099.00	290,235.99	290,449.70
Hurdle : Poisson−NB1	19	−145,099.27	290,236.54	290,450.25
Hurdle : Poisson−PIG2	19	−145,099.40	290,236.79	290,450.50
Hurdle : Poisson−NB2	19	−145,099.47	290,236.93	290,450.65
PIG1	14	−145,107.58	290,243.15	290,400.62
Zero−Inflated Poisson	15	−145,116.02	290,262.05	290,430.77
PLN1	13	−145,117.12	290,260.24	290,406.46

TAB. 9.2 – Comparison of models for the Spanish data set - Information Criteria

AIC denies the existence of an identifiable true model and, for example, minimizes the distance or discrepancy between densities. Moreover, in model selection, it has been argued that BIC penalizes large models too heavily. Consequently, at this stage, based on this interpretation and on previous results, the $NegBin_x$ and the Hurdle models are preferred.

9.3 Vuong Test

Since the hypotheses of equality of overlapping models have been rejected by the score-tests seen in the beginning of this section, the test proposed by Vuong (1989) can be applied directly. This test is based on the difference of the log-likelihood models, with a correction that corresponds to the standard deviation of this difference. The test statistic is

$$T_{LR,NN} = \frac{(\ell_f(\hat{\boldsymbol{\beta}}_1) - \ell_g(\hat{\boldsymbol{\beta}}_2))}{\sqrt{n}\omega} \tag{9.1}$$

where

$$\omega^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{f(\hat{\boldsymbol{\beta}}_1)}{g(\hat{\boldsymbol{\beta}}_2)}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\log\frac{f(\hat{\boldsymbol{\beta}}_1)}{g(\hat{\boldsymbol{\beta}}_2)}\right)^2 \tag{9.2}$$

is an estimate of the variance of the log-likelihood difference. None of the two models has to be true. The null hypothesis of the test is that the two model are equivalent, expressed as H_0 : $E[\ell_f(\hat{\boldsymbol{\beta}}_1) - \ell_g(\hat{\boldsymbol{\beta}}_2)] = 0$. Under the null hypothesis, the test statistic is asymptotically Normally distributed. Rejection of the test in favor of f happens when $T_{LR,NN} > c$, or in favor of g if $T_{LR,NN} < c$, where c represents the critical value for some significance level.

Modification of this test is needed if the compared models do not have the same number of parameters. As proposed by Vuong (1989), we may consider the following adjusted statistic :

$$\hat{LR}(\beta_1, \beta_2) = LR(\beta_1, \beta_2) + K(f, g)$$

where $K(f, g)$ is a correction factor based on the difference between the information criteria (seen in the preceeding subsection) of models f and g. Results are displayed in Table 14. Once again, the hurdle and the $NegBin_x$ distributions are the distributions that provide the best results, even if only the PLN1 distribution is rejected against some models.

10. Conclusion

The behavior of the policyholders subject to bonus-malus schemes seems to influence their probability to report a claim. Models such as zero-inflated, $NegBin_x$ or hurdle use this feature to describe the number of reported claims and provide a good fit to the Spanish data set. The choice of the best distribution describing our data has been supported by specification tests for nested or non-nested models. Beside the interpretation of the models, we have seen that the process of classifying policyholders into classes needs adjustments since some classes exhibit much more variability than others.

We also conclude that risk classification based on data that were generated under experience rating schemes must take into account that discount and penalty mechanisms have an infuence on the claiming behavior of the insureds. This is to our knowledge the first time that empirical evidence is shown from a wide range of models.

There is also a general feeling in the insurance industry that drivers have either a good claiming behavior (never claim) or a bad claiming behavior (claim a lot). Indeed many marketing strategies are designed to capture good customers and let the bad drivers go to the competitor. This paper shows that indeed we find evidence that the same customer may indeed have two latent processes and switch his/her probability to report another accident once the first claim has already occured.

Model #1	Model #2	Vuong Test Log-likelihood	AIC	BIC
$NegBin_x$	Hurdle : Poisson−PIG1	0.424	0.229	0.000
	Hurdle : Poisson−NB1	0.396	0.211	0.000
	Hurdle : Poisson−PIG2	0.384	0.202	0.000
	Hurdle : Poisson−NB2	0.376	0.196	0.000
	PIG1	0.025	0.093	0.987
	Zero−Inflated Poisson	0.044	0.065	0.333
	PLN1	0.002	0.010	0.887
Hurdle : Poisson−PIG1	Hurdle : Poisson−NB1	0.251	0.251	0.251
	Hurdle : Poisson−PIG2	0.304	0.304	0.304
	Hurdle : Poisson−NB2	0.273	0.273	0.273
	PIG1	0.042	0.235	1.000
	Zero−Inflated Poisson	0.052	0.107	0.816
	PLN1	0.002	0.028	1.000
Hurdle : Poisson−NB1	Hurdle : Poisson−PIG2	0.436	0.436	0.436
	Hurdle : Poisson−NB2	0.394	0.394	0.394
	PIG1	0.049	0.255	1.000
	Zero−Inflated Poisson	0.054	0.110	0.826
	PLN1	0.003	0.033	1.000
Hurdle : Poisson−PIG2	Hurdle : Poisson−NB2	0.375	0.375	0.375
	PIG1	0.051	0.263	1.000
	Zero−Inflated Poisson	0.057	0.115	0.826
	PLN1	0.003	0.033	1.000
Hurdle : Poisson−NB2	PIG1	0.052	0.267	1.000
	Zero−Inflated Poisson	0.058	0.116	0.828
	PLN1	0.003	0.034	1.000
PIG1	Zero−Inflated Poisson	0.254	0.230	0.119
	PLN1	0.000	0.000	0.100
Zero−Inflated Poisson	PLN1	0.470	0.525	0.801

Tab. 9.3 – Comparison of models for the Spanish data set - Vuong Test

Chapitre 2

Semiparametric Models for Claim Counts : Various Analysis of the Gamma Heterogeneity

Soumis pour publication à ASTIN Bulletin
en collaboration avec Montserrat Guillén

1. Introduction

Risk classification techniques for claim counts have been the topic of many papers appeared in the actuarial literature. Usually, the starting point for modeling the number of reported claims is the Poisson distribution. However, this distribution has some severe drawbacks that limit its use. To provide a better fit to the data, several papers propose to add an heterogeneity term into the mean parameter of the Poisson distribution. A gamma distributed heterogeneity is often chosen, which leads to the well-known negative binomial distribution. However, beside its computational properties, the choice of this specific heterogeneity distribution may be difficult to justify.

Let us assume that N_i is the random variable indicating the number of reported claims to the insurer by the i^{th} policyholder during a fixed period of time. N_i takes values in $\{0, 1, 2, ...\}$. For each policy i, we observe n_i and $x_{i,p}$, respectively, the number of claims and the risk characteristics. such as the age of the driver or the power of the vehicle. Concretely, the equidispersion implied by the Poisson distribution is usually corrected by introducing a random heterogeneity term θ_i with mean equal to 1 and variance τ (e.g. Boyer et al. (1992), and Lemaire (1995)), such as :

$$P[N_i = n_i] = \Pr[n_i] = \int_0^\infty \frac{(\lambda_i\theta_i)^{n_i}\exp(-\lambda_i\theta_i)}{n_i!} g(\theta_i)d\theta_i, \quad i = 1, ..., N \tag{1.1}$$

where the mean parameter $\lambda_i = \exp(x_i'\beta)$ is expressed with covariates (Dionne and Vanasse (1989, 1992)). The introduction of a heterogeneity term brings the possibility to have models with overdispersion since (conditional on the covariates) the variance of N_i is equal to $\lambda_i + \lambda_i^2 Var[\theta_i]$. Additionally, the mixture of the Poisson distribution has an important property referred to as the Two Crossings Theorem by Shaked (1980). This theorem states that the mixed distribution has heavier tails than the original distribution. This result is important since the lack of fit of the Poisson distribution for the number of claims is partially due to the tails of the distribution.

In a parametric setting, heterogeneity distributions other than the gamma can also be used. Examples are the inverse Gaussian distribution (Willmot (1987), Besson and Partrat (1992), and Tremblay (1992)) or the lognormal distribution (Hinde (1982)). In this paper, we also consider flexible, or nonparametric estimation for the mixing distribution $g(\theta_i)$, leading to semiparametric models for the number of reported claims. All models are analyzed by cross-sectional data analysis, where each observation, here each annually contract, is considered to be mutually independent. This kind of models is opposed to panel or longitudinal data analysis where a correlation between

each annually contract of the same insured is allowed.

In section 2, usual parametric models such as the negative binomial, the Poisson-inverse Gaussian and the Poisson-lognormal are reviewed. Three semiparametric methods are analyzed in section 3 : we review the finite mixture models and introduce two new models in actuarial sciences : series expansions for unobserved heterogeneity and a model using weighted least-squares evaluated by a nonparametric estimation of variance. In section 4, all models are illustrated using insurance data from a Spanish insurance company, which allow us to verify if the gamma heterogeneity provides an acceptable estimation of the β parameter, besides approximation to the *true* heterogeneity distribution or link between the mean and the variance of the count model. Additionally, these nonparametric heterogeneity approaches allow us to give new interpretations of the insurance portfolio.

2. Parametric Approaches

There are many possible distributions that can be used to introduce overdispersion in the Poisson model, which we call heterogeneity. Gamma and inverse Gaussian heterogeneities lead to a closed-form distribution. An other important heterogeneity distribution is the lognormal distributions, but closed-form representation is impossible and numerical computations are needed. Other Poisson mixture models are possible (see chapter 8 of Johnson et al. (1992)), but numerically techniques are also needed. As showed by Holgate (1970), all continuous mixtures based on the Poisson distribution have unimodal likelihood functions. This ensures us that many simple numerical techniques (Newton-Raphson,etc.) for maximum likelihood estimation can be used.

2.1 Gamma Heterogeneity

An easy way of handling the heterogeneity is to suppose that the heterogeneity follows a Gamma distribution. It is well-known that this mixed model will result in a negative binomial distribution (NB) (Greenwood and Yule (1920) or Lawless (1987) and Dionne and Vanasse (1989) for an application with insurance data) :

$$\Pr[n_i] = \frac{\Gamma(n_i + \alpha^{-1})}{\Gamma(n_i + 1)\Gamma(\alpha^{-1})}\Big(\frac{\lambda_i}{\alpha^{-1} + \lambda_i}\Big)^{n_i}\Big(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i}\Big)^{\alpha^{-1}} \tag{2.1}$$

where $\lambda_i = \exp(x_i'\beta)$ and Γ is the gamma function, defined by $\Gamma(a) = \int_0^\infty e^{-t}t^{a-1}dt$. To ensure that the heterogeneity mean is equal to 1, both parameters of the Gamma distribution are chosen

to be identical and equal to $1/\alpha$. Winkelmann and Zimmermann (1991) and Winkelmann and Zimmermann (1995), based on a generalization of Cameron and Trivedi (1986), considered a more general class of negative binomial models and proposed to use a Gamma distributed heterogeneity with both parameters equal to λ_i^{1-k}/α that can be expressed as :

$$\Pr[n_i] = (1+\alpha\lambda_i^k)^{-\lambda_i^{1-k}/\alpha} \prod_{j=1}^{N_i} \left[\frac{\lambda_i + \alpha(j-1)\lambda_i^k}{(1+\alpha\lambda_i^k)j} \right].$$

When k is set to 0, the previous specification leads to the $NB1$ model. This model is interesting because its variance is equal to $\lambda_i + \alpha\lambda_i = \phi\lambda_i$, the same used in the Poisson GLM approach (such as the one used for the overdispersion correction of the Poisson distribution in the GENMOD procedure of SAS).

2.2 Inverse Gaussian Heterogeneity

Because it is a longer-tailed distribution, the inverse Gaussian distribution is often chosen instead of the Gamma distribution to model the heterogeneity parameter (Holla (1966) or Willmot (1987), Dean et al. (1989) and Tremblay (1992) for an application with insurance data). It can be proved (Shoukri et al. (2004) or Boucher et al. (2006b)) that the Poisson with inverse Gaussian heterogeneity (PIG) with mean and variance equal to 1 and $\tau\lambda_i^{k-1}$ respectively, has the following density :

$$\Pr[n_i] = \frac{\lambda_i^{n_i}}{n_i!} \Big(\frac{2}{\pi\tau\lambda_i^{k-1}} \Big)^{0.5} e^{1/\tau\lambda_i^{k-1}} (1+2\tau\lambda_i^k)^{-s_i/2} K_{s_i}(z_i), \tag{2.2}$$

where $s_i = n_i - 0.5$ and $z_i = (1+2\tau\lambda_i^k)^{0.5}/\tau\lambda_i^{k-1}$ and $K_j(.)$ is the modified Bessel function of the second kind.

This function has the useful following properties :

$$K_{-1/2}(a) = \Big(\frac{\pi}{2a}\Big)^{0.5} e^{-a} \ , \ K_{1/2}(a) = K_{-1/2}(a) \ , \ K_{s+1}(a) = K_{s-1}(a) + \frac{2s}{a} K_s(a).$$

Additional properties of the modified Bessel function of the second kind may be found in Shrouki et al. Shoukri et al. (2004). If the parameter k is set to 1 on the $PIGk$ density, it leads to the $PIG2$ having a $NB2$ variance form (i.e. $\lambda_i + \alpha\lambda_i^2$), while $k = 0$ results in a $PIG1$ model having the same variance form as the $NB1$ distribution.

2.3 Lognormal Heterogeneity

The Poisson-lognormal (PLN) model (Hinde (1982)) can also be interesting because it has a natural interpretation (Winkelmann (2003a)). The heterogeneity factor used in the Poisson model has a multiplicative form with the mean parameter but the error term can also be expressed as $\theta_i = \exp(\epsilon_i)$. Therefore, the mean parameter can be expressed as $\gamma_i = \exp(x_i'\beta + \epsilon_i)$, with ϵ_i following a Gaussian distribution with mean $\mu = -\sigma^2/2$ (a simplifying normalization) and variance σ^2. Consequently, the error term ϵ_i, having a linear link with the x_i, is often considered as a factor that captures the effects of hidden exogeneous variables. If there are many hidden variables, and if these variables are independent, then central limit theorem can be invoked in order to establish normality of ϵ_i. The distribution can be expressed as :

$$\Pr[n_i] = \int_{-\infty}^{\infty} \frac{\exp(-\gamma_i)\gamma_i^{n_i}}{n_i!} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}\left(\frac{\epsilon_i}{\sigma}\right)^2) d\epsilon_i. \tag{2.3}$$

This distribution cannot be expressed in closed form and numerical computations are needed to evaluate the distribution. Once again, modification of the heterogeneity distribution can be introduce in order to have other forms of variance.

3. Semiparametric Approach

The choice of a specific heterogeneity distribution is difficult to justify in count data models because the variance exceeds the mean only given individual characteristics. When many covariates are used, there is not enough data to determine the mean-variance relationship conditional on the values of the regressors. That is the reason why we consider nonparametric estimation for the mixing distribution $g(\theta_i)$ in this section. Not only this kind of methods allow us several new interpretations of the data, but they can serve to compare the estimates of the β parameters of parametric models (2.1) through (2.3), that are not consistent if the distribution of the heterogeneity term θ_i is misspecified (Gourieroux et al. (1984b)), as opposed to the estimates obtained from exponential family distributions, such as the Poisson distribution.

3.1 Finite Mixture Count Data Models

Mixture models presented in the previous section can be seen as a convolution or continuous mixture models since they can be defined as $f(n_i) = \int_{-\infty}^{\infty} f(n_i|\theta)g(\theta)d\theta$. Nevertheless, it is also possible to consider discrete mixture of different densities, called finite mixture models. Formally, consider various proper distribution functions F^j representing different random variables, $j =$

$1, 2, ...$ and constants a_j with $a_j > 0$ and $\sum_j a_j = 1$, then

$$\Pr[n_i] = \sum_{j=1}^{s} a_j F^j(n_i) \tag{3.1}$$

is a proper distribution function and is called a finite mixture distribution. Walhin and Paris (1999), following Simar (1976), analyzed this kind of semi-parametric models for bonus-malus rating, and showed that is not suitable for ratemaking because it is purely discrete. Nevertheless, this modeling can be useful to interpret and understand insurance data.

As stated by Deb and Trivedi (2002), finite mixture models have some advantages over continuous mixtures. First of all, no requirement is necessary for the distributional assumptions of the mixing variable, which can be seen as a semiparametric model. Second, it can have some natural interpretations, where each latent class may be regarded as "type" or "group". Third, the results of Laird (1978) and Heckman and Singer (1984) suggest that estimates of finite mixture models provide good numerical approximations even if the mixing distribution is continuous. Finally, the choice of a continuous mixing density is sometimes restrictive because the marginal density may not have an analytical solution.

Different types of finite mixture models exist. Brännas and Rosenqvist (1994) proposed Poisson models with intercept heterogeneity, where the F^j functions only differ by their intercept since $\lambda^j = \exp(x_i'\beta + \delta_j)$. Estimators of β can be found using maximum likelihood theory since Simar (1976) showed consistency of the estimator of β.

Finite mixture models can also be extended by allowing different values of β, instead of using only intercepts. This model has been proposed by Wedel et al. (1993) and can have an interesting interpretation. For each observation, the distribution F^j is given according to a multinomial distribution with probability p_j. Consequently, this model allows to separate the portfolio into some classes. This can be seen as an individual heterogeneity, such as models NB1, NBk, PIG1, PIGk, PLN1 or PLNk. In this situation, equation (3.1) can be calculated using distributions F^j having a mean parameter equal to $\lambda^j = \exp(x_i'^j)$, where x_i includes an intercept.

3.2 Series Expansion of the Heterogeneity

Another nonparametric estimation of the density $g(\theta_i)$ of equation (1.1) involves series expansion, as proposed by Gurmu et al. (1999). They based their model on the work of Gallant and Nychka (1987) who approximate an unknown density by a squared K^{th}-order polynomial expansion of the form :

$$g(\theta_i) = \frac{1}{\phi} w(\theta_i)[P_K(\theta_i)]^2 \tag{3.2}$$

where $w(\theta_i)$ denotes the hypothesized base-line density, $P_k(\theta_i)$ the associated polynomial of degree K and $\phi = \int w(\theta_i)[P_K(\theta_i)]^2 d\theta_i$ is the constant of proportionality that makes the density integrate to one. Using the equation (1.1) that can be expressed as :

$$\begin{aligned} \Pr[n_i] &= \frac{\lambda_i^{n_i}}{n_i!} \int_0^\infty \exp(-\lambda_i\theta_i)\theta_i^{n_i} g(\theta_i) d\theta_i & (3.3) \\ &= \frac{\lambda_i^{n_i}}{n_i!} M_{\theta_i}^{(n_i)}(-\lambda_i) & (3.4) \end{aligned}$$

with :

$$\begin{aligned} M_{\theta_i}^{(n_i)}(-\lambda_i) &= (1+\alpha\lambda_i)^{-1/\alpha} \frac{\Gamma(1/\alpha)}{(\lambda_i + 1/\alpha)^{1/\alpha}} \left(\sum_{k=0}^{K} a_k^2\right)^{-1} \sum_{k=0}^{K}\sum_{l=0}^{K} a_k a_l (\eta_k \eta_l)^{1/2} & (3.5) \\ &\times \sum_{r=0}^{k}\sum_{s=0}^{l} \binom{k}{r}\binom{l}{s} \frac{\Gamma(1/\alpha + r + s + n_i)}{\Gamma(1/\alpha + r)\Gamma(s + n_i)} (-1-\alpha\lambda_i)^{-(r+s)} & (3.6) \end{aligned}$$

where :

$$\eta_k = \frac{\Gamma(k + 1/\alpha)}{\Gamma(1/\alpha)\Gamma(k+1)} \tag{3.7}$$

The $M_{\theta_i}^{(n_i)}(.)$ is the n_ith-order derivative of $M_{\theta_i}(.)$ The development of the equation is shown, for example, in Gurmu et al. (1999). The derivatives of $M_{\theta_i}(.)$ are based on the approximation of the unknown density by equation (3.2). The orthonormal polynomial series expansion used in this approximation are the Laguerre polynomials where the leading term in the expansion is the gamma density. This generalisation of the NB2 model is interesting since it allows a more general form of the heterogeneity distribution. Moreover, it also allows to verify the gamma hypothesis because of

the nesting.

Gurmu et al. (1999) showed that the approximation of the mixing distribution $g(\theta_i)$ can be expressed as :

$$g^*(\theta_i) = \frac{(1/\alpha)^{1/\alpha}}{\Gamma(1/\alpha)}\theta_i^{1/\alpha-1}e^{-\theta_i/\alpha}\left(\sum_{k=0}^{K}a_k\eta_k^{-1/2}\sum_{l=0}^{k}\binom{k}{l}\frac{\Gamma(k+\alpha)}{\Gamma(l+\alpha)\Gamma(k+1)}\alpha^l(-\theta_i)^l\right)^2\left(\sum_{k=0}^{K}a_k^2\right)^{-1} \quad (3.8)$$

Identification requires that $E[\theta_i] = 1$, for which $a_0 = 1$. Other parameters to be estimated are $\beta, \alpha, a_1, a_2, ..., a_K$, where K must be selected at the beginning of the evaluation. For example, the NB2 distribution is obtained for $K = 0$. Multiple local maximums exist which implies that careful attention must be taken when evaluating the parameters. Generalizations into polynomial expansion of the NB1 or NBk distributions can be done quite directly by changing the $1/\alpha$ into λ_i^{1-k}/α in the previous equations.

3.3 Unknown Variance Form by Nearest Neighbours

An important property of Poisson and Poisson-Mixture models is that the variance depends on the mean. Based on the work of Robinson (1987), Delgado and Kniesner (1997) use an adaptative method of the general least-squares for count data. They do not suppose a specific form for the conditional variance σ_i^2 of the i^{th} contract, and estimate it using a consistent estimator of β, $\hat{\beta}$, such as the one obtained by a Poisson regression :

$$\hat{\sigma}_i^2 = \sum_{j=1,j\neq i}^{n}[n_i - \hat{\lambda}_j]^2 w_{i,j} \quad (3.9)$$

where $w_{i,j}$ can be seen as weights applied to the j^{th} observation to compute the variance of the i^{th} insured with, obviously, $\sum_{j=1,j\neq i}^{n} w_{i,j} = 1$. Using $w_{i,j} = 1/(n-1)$ leads to an estimated variance almost equal to the empirical variance.

The semiparametric GLS estimator then solves the following first-order vector condition :

$$\sum_{i=1}^{n}(n_i - \exp(x_i\beta))\exp(x_i\beta)x_i\hat{\sigma}_i^{-2} = \mathbf{0} \quad (3.10)$$

Under regularity conditions, the parameter $\hat{\beta}$ is root-n consistent and asymptotically normal with variance equal to :

$$Var[\hat{\beta}] = (\sum x_i' x_i \exp(2x_i'\hat{\beta})\hat{\sigma}^{-2})^{-1} \tag{3.11}$$

A robust sandwich estimator can also be used to estimate the variance of the estimator.

Delgado and Kniesner (1997) evaluated the $w_{i,j}$ using nonparametric k nearest neighbour probabilistic weights. Their method estimates the variance of insured i using its k nearest neighbours, evaluated by its proximity in its normalized covariates. The k nearest neighbours were selected using a neighbour specification k that is proportional to the number of observations, such as $n^{1/2}$ or $n^{3/5}$. Many kinds of weight $w_{i,j}$ can be used, such as uniform, triangular or quadratic weights. However, for ease of computation, only uniform weight $1/k$ was used in Delgado and Kniesner (1997). A tie-breaking rule was established in case where the same normalized proximity was calculated for more than one observation (Robinson (1987)). In such situation, the weight was distributed equally amongst ties.

3.4 Unknown Variance Form by Kernel Estimation

The method of Delgado and Kniesner (1997) can be modified. The use of nearest neighbours estimators to approximate the variance of each n_i has some disadvantages. Indeed, the choice of the k influent observation is based on a selection of each *normalized* covariates $X_{i,m}$. Since almost all covariates used in insurance are categorical, a selection based on the score $\hat{\lambda}_i = \exp(x_i'\hat{\beta})$ seems to be more realistic. Additionally, instead of using the nearest neighbours estimation that gives the same weight on all k observations (which can lead to situations where the $\hat{\sigma}_i^2$ would be estimated using only insureds having the same profiles, while other $\hat{\sigma}_i^2$ would be estimated using other insured's profiles), we can rather use kernel estimation that seems to offer a better smooth.

The kernel estimator was introduced by Rosenblatt (1956) and can be seen as a density estimator like the nearest neighbours technique. Under this construction, the weight given at each j to estimate the variance of the observation i will be based on :

$$\eta_{i,j} = \frac{1}{h} K\left(\frac{\hat{\lambda}_i - \hat{\lambda}_j}{h}\right)$$

from which :

$$w_{i,j} = \frac{\eta_{i,j}}{\sum_{j=1}^{n} \eta_{i,j}}$$

where function $K(\cdot)$ is the Kernel function and h is the bandwidth. Even if many Kernel functions can be chosen to model the weight of each n_j, it is well known that the choice of the bandwidth h is as important as the choice of the Kernel function. We selected the Epanechnikov kernel because since it has good properties and it is widely used. Our selected bandwidth h was chosen using Silverman's plug-in estimate :

$$h^* = 1.3643\delta n^{-0.2} \min\{s, iqr/1.349\}$$

where $\delta = 1.7188$ for the Epanechnikov Kernel, s is the sample standard deviation of $\exp(x_i'\hat{\beta}), i = 1, ..., N$ and iqr is the sample interquartile range. The calculation of each $w_{i,j}$ for insureds i and j can be computationally tedious since it can demand $n \times (n-1)$ computations.

4. Insurance Application

We applied the presented models to the number of reported claim in third-liability for a sample of a portfolio of automobile insurance from a major company operating in Spain. It is well known that the absence of some important classification variables (swiftness of reflexes, aggressiveness behind the wheel, consumption of drugs, etc.) can be used to justify the heterogeneity on the count distribution.

4.1 Data used

In this paper, only private use cars have been considered in the sample. The data contains information from 1991 until 1998. Our sample contains 15,179 policyholders that stay in the company for seven complete periods, resulting in 106,253 insurance contracts that are assumed to be independent. We have 5 exogeneous variables (see table 4.1) that are kept plus the yearly number of accidents. For every policy we have the initial information at the beginning of the period and the total number of claims at fault that took place within this yearly period. The average claim frequency of the portfolio is 6.8412%.

4.2 Parameters Estimates Comparison

The application of the parametric models to the Spanish data set leads to the results displayed in tables 4.2 to 4.6. Comparing the log-likelihoods, we see that introduction of a heterogeneity

Variable	Description
v1	equals 1 for women and 0 for men
v2	equals 1 if the client has been in the company between 3 and 5 years
v3	equals 1 if the client has been in the company for more than 5 years
v4	equals 1 if the insured is 30 years old or younger
v5	equals 1 if power is larger or equal to 5500 cc

TAB. 4.1 – Exogenous variables

parameter	Poisson		NB2		NB1		NBk	
b0	−2.6039	(0.0410)	−2.6046	(0.0434)	−2.6059	(0.0426)	−2.6102	(0.0400)
b1	0.1047	(0.0337)	0.1059	(0.0356)	0.1167	(0.0348)	0.1219	(0.0321)
b2	−0.2106	(0.0390)	−0.2108	(0.0415)	−0.2134	(0.0404)	−0.2004	(0.0388)
b3	−0.2721	(0.0361)	−0.2726	(0.0383)	−0.2803	(0.0373)	−0.2670	(0.0365)
b4	0.1543	(0.0336)	0.1548	(0.0356)	0.1573	(0.0348)	0.1479	(0.0332)
b5	0.1373	(0.0291)	0.1383	(0.0306)	0.1441	(0.0303)	0.1363	(0.0284)
α	.	.	1.5744	(0.1095)	0.1102	(0.0077)	0.0012	(0.0025)
k	.	.	.	.	.	.	−1.6804	(0.7635)
Log-Lik.	-27,133.8		-26,933.0		-26,926.4		-26,922.8	

TAB. 4.2 – Estimates (1)

term improves the fit compared to the Poisson distribution. Only the finite mixture model with 2 intercepts (FMI-2) converge. The model seems to have a good fit since its log-likelihood is close from the best ones obtained with parametric models. To ensure that we avoid local maximums, we repeat estimations with many starting values. Similarly, only the series expansion model with 2 terms converge. For the unknown variance form with Kernel estimator, we also checked $h = 2h^*$ and $h = 0.5h^*$ to be sure that the bandwidth selection is correctly specified.

The estimations of the parameters are approximately the same for all models and by a quick look on the results, we can observe that the fitting is quite the same for each model. A formal approach to compare the fit of the models was also done in Boucher et al. (2006c), where a Vuong

parameter	P-Inverse Gaussian 2		P-Inverse Gaussian 1		P-Inverse Gaussian k	
b0	−2.6040	(0.0434)	−2.6054	(0.0426)	−2.6093	(0.0399)
b1	0.1073	(0.0357)	0.1166	(0.0348)	0.1209	(0.0321)
b2	−0.2122	(0.0415)	−0.2144	(0.0404)	−0.2019	(0.0385)
b3	−0.2746	(0.0384)	−0.2810	(0.0373)	−0.2677	(0.0360)
b4	0.1557	(0.0356)	0.1574	(0.0347)	0.1478	(0.0329)
b5	0.1393	(0.0306)	0.1444	(0.0302)	0.1364	(0.0280)
α	1.6055	(0.1188)	0.1119	(0.0083)	0.0009	(0.0015)
k	.	.	.	.	−1.8019	(0.6426)
Log-Lik.	-26,935.40		-26,929.26		-26,925.61	

Tab. 4.3 – Estimates (2)

parameter	P-LogNormal 2		P-LogNormal 1		P-LogNormal k	
b0	−2.6021	(0.0435)	−2.6044	(0.0425)	−2.6082	(0.0398)
b1	0.1085	(0.0357)	0.1157	(0.0348)	0.1190	(0.0320)
b2	−0.2140	(0.0416)	−0.2152	(0.0403)	−0.2019	(0.0390)
b3	−0.2770	(0.0384)	−0.2812	(0.0373)	−0.2666	(0.0373)
b4	0.1567	(0.0357)	0.1574	(0.0347)	0.1469	(0.0334)
b5	0.1404	(0.0306)	0.1442	(0.0302)	0.1363	(0.0287)
α	0.9819	(0.0531)	0.1072	(0.0085)	0.0004	(0.0014)
k	.	.	.	.	−2.0724	(1.2056)
Log-Lik.	-26,938.3		-26,933.6		-26,929.2	

Tab. 4.4 – Estimates (3)

parameter	Finite Mixture Coeff. - 2		Finite Mixture Intercept - 2		Series Expansion	
b0	−3.1689	(0.1385)	−3.4833	(0.3348)	−2.7165	(0.3784)
b1	0.2165	(0.0659)	0.1051	(0.0356)	0.1051	(0.0356)
b2	−0.3589	(0.0789)	−0.2094	(0.0415)	−0.2096	(0.0415)
b3	−0.4912	(0.0865)	−0.2710	(0.0383)	−0.2712	(0.0383)
b4	0.2687	(0.0672)	0.1541	(0.0356)	0.1542	(0.0356)
b5	0.2645	(0.0689)	0.1376	(0.0305)	0.1377	(0.0305)
c0	−1.1681	(0.1332)	−1.3123	(0.1930)	1.5626	(0.5667)
ϕ-α	0.8880	(0.0314)	0.8190	(0.0728)	0.5634	(0.1877)
Log-Lik.	-26,926.0		-26,932.4		-26,932.5	

Tab. 4.5 – Estimates (4)

	Unkown Variance - Bandwidth (h)					
parameter	$0.5h^*$		h^*		$2h^*$	
b0	−2.6046	(0.0436)	−2.6048	(0.0435)	−2.6019	(0.0434)
b1	0.1032	(0.0353)	0.1028	(0.0353)	0.1021	(0.0352)
b2	−0.2099	(0.0409)	−0.2098	(0.0408)	−0.2108	(0.0407)
b3	−0.2701	(0.0378)	−0.2709	(0.0378)	−0.2729	(0.0376)
b4	0.1579	(0.0354)	0.1559	(0.0353)	0.1561	(0.0352)
b5	0.1369	(0.0309)	0.1381	(0.0309)	0.1366	(0.0308)

Tab. 4.6 – Estimates (5)

test that uses the distribution of each individual's log-likelihood was used.

Estimates of β are quite the same between models. Using models with no link between the variance and the mean lead to estimates of parameters being very similar to other ones. To see more formally if there is a significant difference between the β of the NB2 distribution and the other ones, we propose a Wald test, closely related to the Hausman (1978) test. The idea is to test whether the β^{NB2} are correctly estimated under the hypothesis that the data are distributed under the model $*$. Hypothesis can be expressed as :

$$H_0 : \beta^{NB2} - \beta^* = 0$$
$$H_a : \beta^{NB2} - \beta^* \neq 0$$

where β^* represents the estimate of a specific model while β^{NB2} are the estimates of the NB2 parameter. The β are organized to have the same dimension, more precisely $\beta_1, ..., \beta_5$ since the intercept is excluded from the analysis. The test is easy to calculate, but some difficulties arise because one needs to calculate the variance matrix of $(\beta^{NB2} - \beta^*)$. However, we proceeded with a development similar to the one used for the Hausman (1978) test. Indeed, under the null hypothesis, the model $*$ is asymptotically efficient and the variance matrix of $(\beta^{NB2} - \beta^*)$ is equal to $Var(\beta^{NB2}) - Var(\beta^*)$. As stated in Hausman (1978), the intuitive reasoning behind this result rests on the fact that the efficient estimator must have zero asymptotic covariance with $\beta^{NB2} - \beta^*$ under the null hypothesis. If this were not case, a linear combination of β^* and $\beta^* - \beta^{NB2}$ could result in an estimator which would have smaller asymptotic variance than the efficient estimator which is assumed to attain the Cramer-Rao lower bound. Additionnaly, since β^* attains the asymptotic Cramer-Rao lower bound, it is clear that $Var(\beta^{NB2}) - Var(\beta^*)$ is nonnegative definite. Formal proofs of both results can be found in Hausman (1978). The case where the estimated variance of $(\beta^{NB2} - \beta^*)$ is negative means that the model $*$ does not have efficient estimators. Consequently, in this situation, there is no purpose in looking the closeness of β^{NB2} to β^*. Our proposed test is based on the distance :

$$WT = (\hat{\beta}^{NB2} - \hat{\beta}^*)'[Var(\hat{\beta}^{NB2}) - Var(\hat{\beta}^*)]^{-1}(\hat{\beta}^{NB2} - \hat{\beta}^*). \tag{4.1}$$

The resulting statistic test has a Chi-Squared distribution with r degrees of freedom, representing the dimension of $\hat{\beta}^*$. For a more complete review of the Hausman test, the reader is referred to Gourieroux and Monfort (1995) or to Cameron and Trivedi (1998) and Winkelmann (2003a) in the context of count data analysis. As for the Hausman test, our proposed test is designed for situations in which the estimator of the model $*$ is inconsistent under the alternative. Consequently, it cannot

be used to test NB2 against Poisson distributions since the Poisson model implies consistent MLE under the alternative (see Gourieroux et al. (1984b), Gourieroux et al. (1984a)). As expected, no β^* shows significant difference with the NB2 model (the minimum *p-value* was equal to 0.479), meaning that using a NB2 distribution instead of the more flexible heterogeneity models presented in this paper does not put additional errors on the β estimates.

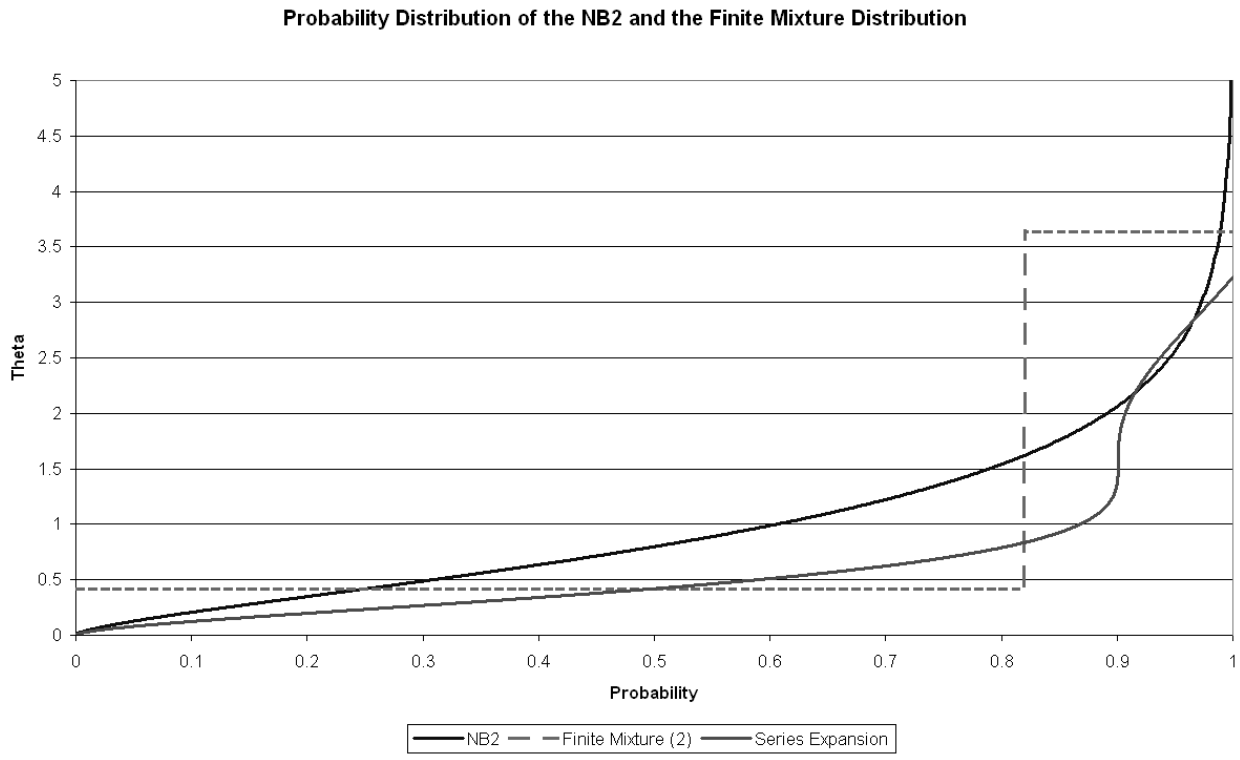
The finite mixture model with coefficient heterogeneity (Finite Mixture Coeff.-2) was not included in the parameters analysis since the β cannot be compared directly. However, this model is interesting since it helps to understand the composition of the insurance porfolio. Only a two-part finite mixture model is presented since additional mixture models did not converge. It shows that the portfolio can be split into two kinds of insureds :

- those who report a claim occasionally : 89% of the insureds having a frequency of report of approximately equal to 4% ;
- those who undergo a lot worst claimming behaviour : 11% of the portfolio. These insureds have a expected frequency of report of more than 31%.

The second portion of this mixture model that shows the part of the portfolio having bad claim experience does not have significant covariates. Based on a comparison of the log-likelihoods, this last model shows a better fit that the other finite mixture model. As noted in Boucher et al. (2006c) for example, it allows us to conclude that our insurance data has a significant portion of bad drivers. However, since these insureds cannot be distinguished by covariates, an insurance company must continue to perform investigations on its portfolio to individually identify these drivers to improve its claim experience.

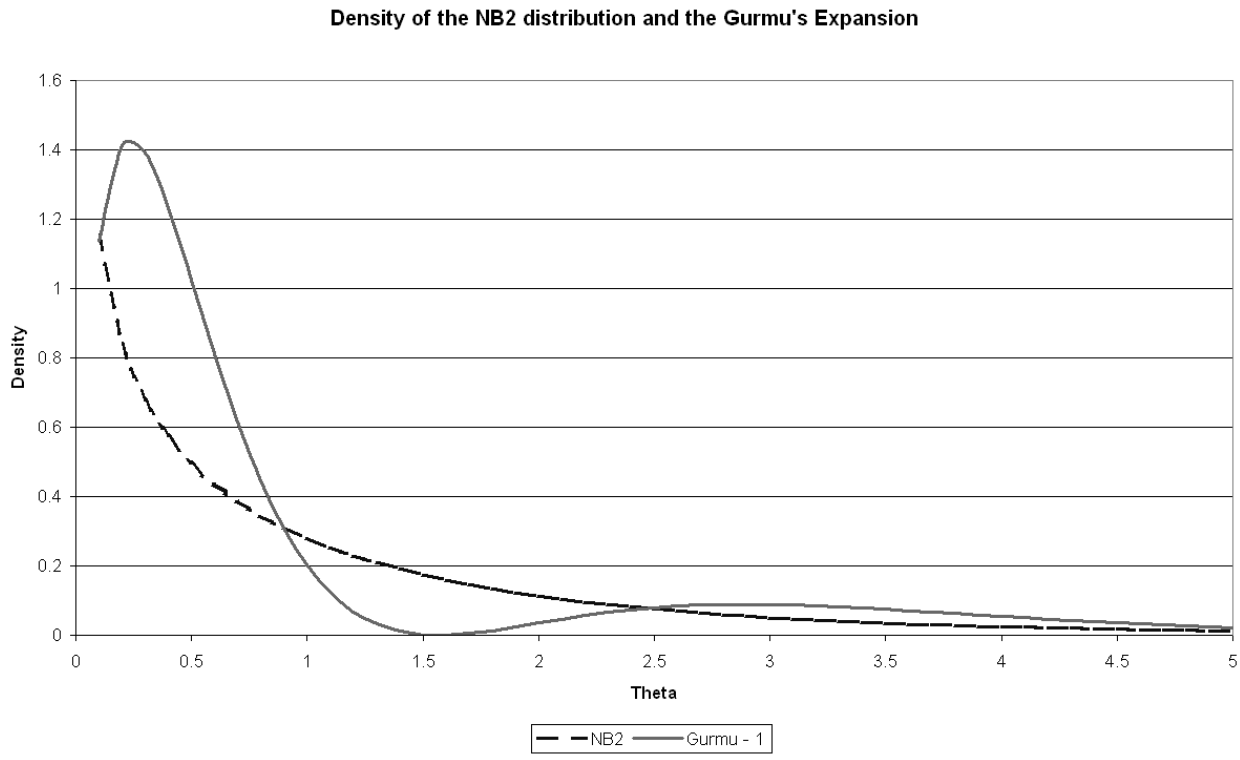
4.3 Comparison of the Heterogeneity Distributions

The following graph presents a comparison between the probability distribution of the gamma heterogeneity (from NB2 distribution) versus the finite mixture heterogeneity is approximated by $\exp(\beta_0)/\exp(\beta_0^*)$, where β^* is the parameters of the Poisson distribution :



We do not include the inverse Gaussian or the Log-Normal distributions since it is almost equal to the gamma distribution. It is interesting to realize that the gamma distribution and the series exapnsion model can be seen as a continuous version of the finite mixture model. The series expansion model seems to be a good alternative to the finite mixture for small θ, but cannot succeed to mimic completely the discrete model for higher values. Even if the discrete expression of the heterogeneity seems better than the continous one, note that Walhin and Paris (1999) showed that the finite mixture distribution is not suitable for ratemaking, because it is purely discrete which can generate difficulties in computing a bonus-malus ratemaking, for example. Several attempts have been done to smooth the finite mixture models, such as Denuit and Lambert (2001), for example.

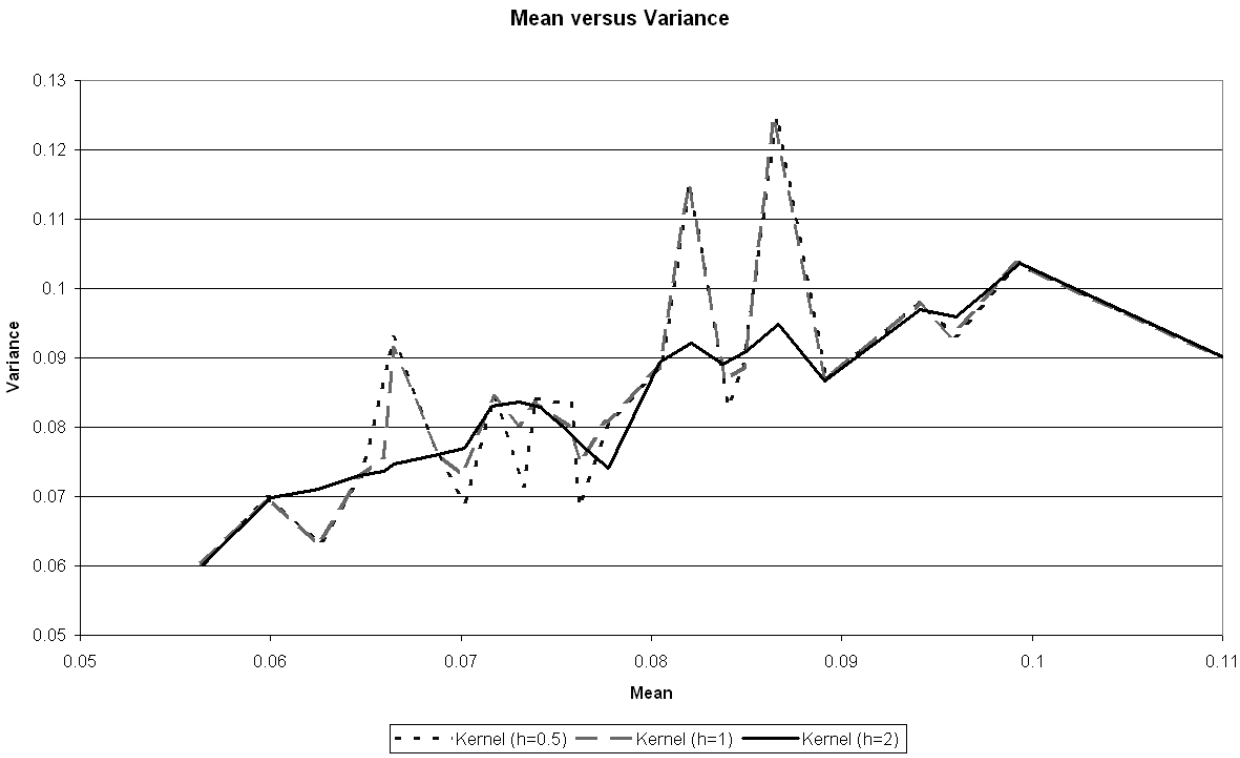
Graphically, we also compare the series expansion model of Gurmu et al. (1999) with the gamma distribution :



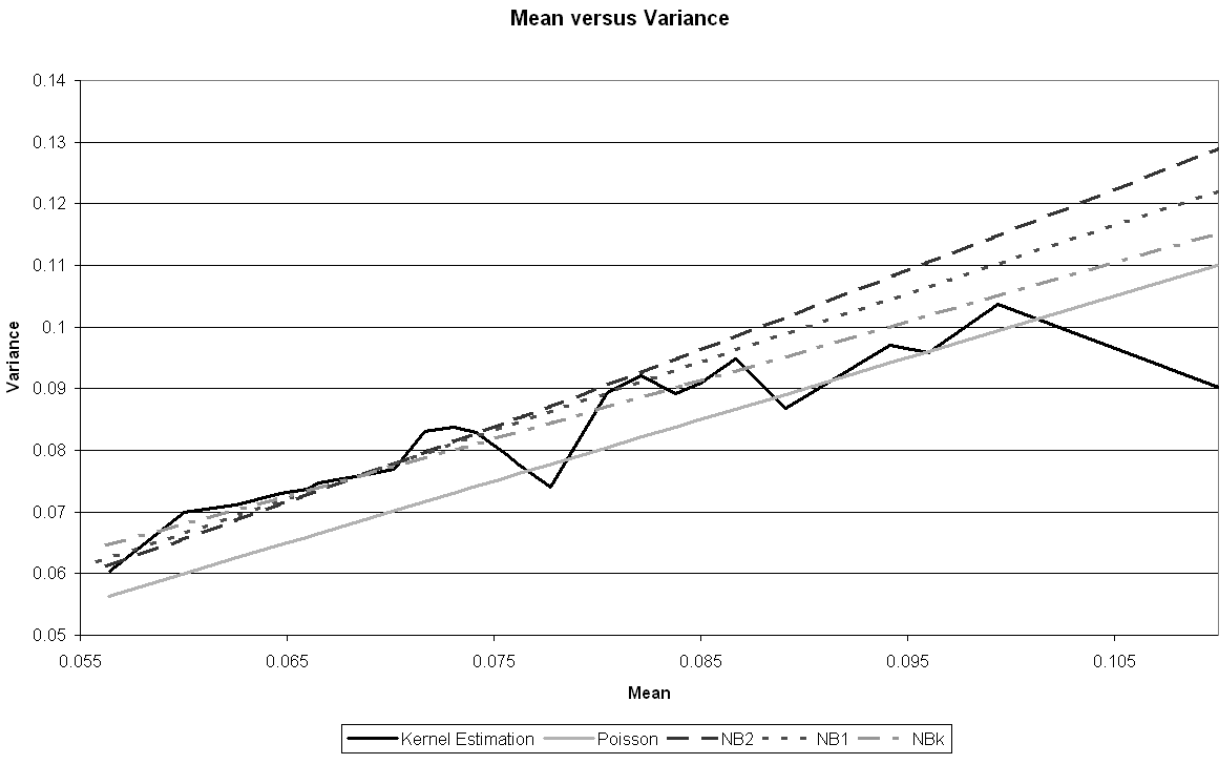
Once again, we can clearly see from the heterogeneity of the model of Gurmu et al. (1999) that the portfolio can be analyze in two parts : drivers have either a good claiming behaviour (left part of the graph) or a bad claiming behaviour (right part of the graph). The gamma heterogeneity tries to smooth that difference. It could be interesting to see what kind of credibility estimates can be computed from this polynomial expression of the heterogeneity.

4.4 Mean-Variance Comparison

For the last analysis of the gamma heterogeneity, we will use the unknown variance model. Indeed, this model allows us to analyze the data differently since the variance of each observation was computed independently of the mean. Depending of the bandwidth values, the variance of each insured was computed using the kernel estimator :



The previous graph exposes the variance of each observation approximated by profiles in function of their mean. Depending on the bandwidth, the variance estimates are smoothed or not. It is clear that some profiles exhibit higher variance than the others. This can be used to identify the kind of insureds that needed to be corrected to obtain a more accurate model (by additional covariates or by a more specific underwriting process, for example). This expression of the variance also allows us to verify if the link between the variance and the mean is correctly defined in the parametric models :



Even if we think that the model with $h = h^*$ offers the best approximation of the variance, the smoothest kernel estimate ($h = 2h^*$) was used for illustration. We can see that the Poisson distribution underestimates the variance. By opposition, the gamma heterogeneities provide a good approximation of the variance for lower risk profiles, while it overestimates it for insureds with high expected values. Other analysis based on this value might be used to construct count models that will imply more precise variance forms, which is useful in the computation of credibility premiums.

5. Conclusion

This paper shows us some of the defaults of the classic gamma distribution to model the heterogeneity of the Poisson distribution for claim count classification. We saw that the estimated parameters $\hat{\beta}$ of the NB2 distribution is quite the same as other flexible models. We see that other parametric distributions such as various inverse Gaussian or Log-Normal do not improve significantly the fit since the densities are approximately the same. Even if the finite mixture models cannot be used directly to compute the insurance premiums, it allows us to see the insurance port-

folio can be split into good and bad drivers and that continuous approximation done by the gamma distribution and the series expansion model seems to be not so bad. By a closer analysis of the series expansion model of Gurmu et al. (1999), we also conclude that the heterogeneity distribution shows a split between good and bad insureds.

The model of Delgado and Kniesner (1997), where the variance was estimated independently from the mean, allows us to verify the link that it exists between the variance and the mean of the count distribution. We saw that the fit of the variance is quite good for low risk profiles, while other insureds do not seem to be modeled correctly with standard heterogeneity distributions.

Even if the *a priori* premiums will not be affected too much by the choice of the heterogeneity distribution, the gamma assumption can generate errors in the analysis of predictive premiums since the form of the heterogeneity is not exactly gamma. We think that more sophisticated parametric heterogeneity distributions might be an interesting future area of research since classic parametric approach miss important properties of the distribution. Nevertheless, another alternative can be the use of more complex count distributions, such as the zero-inflated or the hurdle distributions (Boucher et al. (2006c), Boucher et al. (2006a) or Boucher et al. (2006d)).