Troisième partie

Généralisations des modèles gonflés à zéro et des modèles à barrière

Chapitre 6

Modelling of Insurance Claim Counts with Hurdle Distribution for Panel Data

Soumis pour publication dans un livre suivant l'International Conference on Mathematical and Statistical Modeling in Honor of Enrique Castillo, June 2006
en collaboration avec Michel Denuit et Montserrat Guillén

1. Introduction

The basis for insurance agreements is risk-sharing between individuals, so that each one contributes economically to constitute a fund that is used to re-establish the wealth of the one that suffers a loss. This is the principle of mutualization. In automobile insurance, for example, when an accident occurs, the responsible driver is liable for the damages caused to others, but since third-party liability automobile insurance is compulsory in most countries, the insurance company takes up the economic compensation for losses. If accident occurrence were completely hazardous and the insured risk was the same (i.e. they all share the same characteristics), all insureds would agree to pay the same price (premium) to sign an insurance contract. Risk factors come into scene because there exists a heterogeneity in risk exposure due to the vehicle, the driver, the mileage or even the geographical location.

The aim of statistics in insurance is to predict future claims conditional on risk factors. Claims amounts (or economic compensations to be paid) are more difficult to predict than claiming frequency (the number of accidents for a given period of time). Nowadays, panel data that contain information on the number of claims reported by insureds each year (or month) together with their characteristics, are the main source of information for insurers to derive models that should be used to estimate the expected number of claims conditional on the information on the risk covariates. The number of individuals is usually very large, compared to the time framework for which information is available. Covariates (i.e. the risk characteristics) may change over time and insureds are also free to cancel an insurance contract to go to a competitor.

In this paper we discuss models suitable to this situation. We also show the practical difficulties of applying them in practice and illustrate with real data examples from the Spanish insurance market. We note that the kind of data that we present here, are central in insurance and quantitative risk management, but they are also very frequent in many other microeconomic problems. Some examples from different contexts are : repeated purchase of products made by a consumer, number of visits to doctors, number of patents written by an industry.

We especially want to stress that our contribution is in the way we address the excess-of-zeros phenomenon that is common in this kind of data sets.

1.1 Data used

In this paper, we worked with a sample of the automobile portfolio of a major company operating in Spain. Only private use cars have been considered in this sample. The panel data contains information from 1991 until 1998. Our sample contains 15,179 policyholders that stay in the com-

Variable	Description
v1	equals 1 for women and 0 for men
v2	equals 1 if the client was in the company between 3 and 5 years
v3	equals 1 if the client was 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. 1.1 – Exogenous variables

pany for seven complete periods, resulting in 106,253 insurance contracts. We have 5 exogeneous variables (see table 1.1) that are kept in the panel 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%.

To analyse these data, we make the following assumption : all covariates used are the ones observed at the beginning of year 1. Thus, all regressors can be seen as time independent. Nevertheless, this assumption is not so restricting as it might seem. Indeed, the majority of the time dependent covariates that are used in insurance involves the age of the insured or the lenght of stay in the company. These variables do not evolve randomly in time since the change can already be known in advance. Other covariates can also change in time, such as the city or the type of vehicule of the driver but this kind of major change often involves the creation of an other policy.

In this paper, the exogenous variables, shown in table 1.1, are used to model the parameters of the distributions. The caracteristics of the insureds are expressed throught some fonctions $h(\beta_0 + x_i'\beta)$, where β_0 is the intercept, $\beta' = (\beta_1, ..., \beta_p)$ is a vector of regression parameters for explanatory variables $x_i = (x_{i,1}, ..., x_{i,p})$.

2. Cross Section versus Panel Data

Usually, data are analysed by cross-sectional analysis, where each observation, here each annually contract, is considered to be mutually independent. Thus, in this situation, we worked with $N \times T$ independent observations. However, it is clear that insurance data is a repetition of one-year contract for an insured, where some dependence over contracts of the same insured exists. Longitudinal data (or panel data) consists of repeated observations of individual units that are observed over time. Each individual is assumed to be independent but correlation between observation of

the same individual is permitted. This fact must be regarded as an advantage and must be used in the construction of frequency models.

2.1 Modelling

There exists many models that imply time dependence but the most popular way of dealing with these data is the use of a common individual term (Hausman et al. (1984)) that affects all contracts of the same insured. To illustrate, this random effects model can represent individual specificities not captured by the covariates, such as swiftness of reflexes, aggressiveness behind the wheel, consumption of drugs, etc. Given the insured-specific random effect term θ_i, the annual claim numbers $N_{i,1}, N_{i,2}, \ldots, N_{i,T}$ are independent. The joint probability function of $N_{i,1}, ..., N_{i,T}$ is thus given by

$$\begin{aligned} &\Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] && (2.1)\\ &= \int_0^\infty \Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}|\theta_i] g(\theta_i) d\theta_i \\ &= \int_0^\infty \left(\prod_{t=1}^{T} \Pr[N_{i,t} = n_{i,t}|\theta_i] \right) g(\theta_i) d\theta_i. && (2.2) \end{aligned}$$

Many conditional distributions for the random variables $N_{i,t}$ can be chosen as well as a distribution for the random effect θ_i, as we will see in the following sections.

2.1.1 Endogeneous regressors and fixed effects

In linear regression, correlation between the regressors and the random effect term leads to inconsistency of the estimated parameters (Mundlak (1978), or for a general overview Hsiao (2003)). The same problem exists for the count data regression when $E[\theta|x_i] \neq E[\theta]$ (Mullahy (1997)) and it leads to biased estimates of parameters. In insurance, correlation between regressors and error term is often present (Boucher and Denuit (2006)) and may be caused by omitted variables that are correlated with the included ones.

As noted in Winkelmann (2003a), consistent estimates may be found if corrections are made to the standard estimation procedures. However, as shown by Boucher and Denuit (2006), for insurance applications, standard methods of estimation can still be used. Indeed, the resulting estimates of parameters, while being biased, represent the apparent effect on the frequency of claim, which is the real interest when the correlated omitted variables cannot be used in classification.

3. Poisson Distribution

3.1 Overview

Commonly, the starting point for the modelling of the number of reported claims is the Poisson distribution :

$$f_{N_{i,t}}(n_{i,t}) = \frac{\lambda_i^{n_{i,t}} e^{-\lambda_i}}{n_{i,t}!} \tag{3.1}$$

where covariates are included in the model by the parameter $\lambda_i = \exp(x_i'\beta)$ Dionne and Vanasse (1989). The Poisson distribution is equidispersed since its mean and variance are both equal to λ_i. Because the Poisson distribution has some severe drawbacks that limit its use, other distributions can be used, such as zero-inflated or hurdle models Boucher et al. (2006c) which is analyzed in the next section.

3.2 Panel data

To generalize the Poisson distribution for panel data, an individual random effect term θ_i is added to its mean parameter. Formally, we can express the classic Poisson random effects model as :

$$N_{i,t}|\theta_i \sim Poisson(\theta_i\lambda_i), \quad i = 1, ..., N \quad t = 1, ..., T$$

where i represents an insured and t his covered period. As for the heterogeneity models of the cross-sectionnal model Boucher et al. (2006c), many possible distribution for the random effects can be chosen. The use of a Gamma heterogeneity is a natural possibility because it can express the joint distribution in a closed form since the gamma is conjugated to the Poisson distribution. Formally, the joint distribution of the number of claims, when the heterogeneity term follows a Gamma distribution of mean 1 and variance α, is equal to Hausman et al. (1984) :

$$\begin{aligned} Pr(n_{i,1}, ..., n_{i,T}) &= \left[\prod_{t=1}^{T} \frac{(\lambda_i)^{n_{i,t}}}{n_{i,t}!}\right] \frac{\Gamma(\sum_{i=1}^{T} n_{i,t} + 1/\alpha)}{\Gamma(1/\alpha)} \left(\frac{1/\alpha}{T\lambda_i + 1/\alpha}\right)^{1/\alpha} \\ &\quad \times (T\lambda_i + 1/\alpha)^{-\sum_{i=1}^{T} n_{i,t}} \end{aligned} \tag{3.2}$$

This distribution is known as a Multinomial Negative Binomial or negative multinomial, which has been often applied (see chapter 36 of Johnson et al. (1996)). Note that this distribution can also be seen as the generalization of the bivariate Negative Binomial of Marshall and Olkin (1990). For this distribution, $E[N_{i,t}] = \lambda_i$ and $Var[N_{i,t}] = \lambda_i + \alpha\lambda_i^2$, so overdispersion can be accounted. Maximum likelihood estimations of parameters and variances of these estimates are straightforward.

The generalization of the Negative Multinomial to other kinds of Multivariate Negative Binomial distributions can also be done. If we suppose that the random effects are following a Gamma distribution with both parameters equal to λ_i^{1-k}/α, the joint distribution can be expressed as :

$$\begin{aligned} Pr(n_{i,1},...,n_{i,T}) &= \left[\prod_{t=1}^{T}\frac{(\lambda_i)^{n_{i,t}}}{n_{i,t}!}\right]\frac{\Gamma(\sum_{i=1}^{T}n_{i,t}+\lambda_i^{1-k}/\alpha)}{\Gamma(\lambda_i^{1-k}/\alpha)} \\ &\quad\times\left(\frac{\lambda_i^{1-k}/\alpha}{T\lambda_i+\lambda_i^{1-k}/\alpha}\right)^{\lambda_i^{1-k}/\alpha}(T\lambda_i+\lambda_i^{1-k}/\alpha)^{-\sum_{i=1}^{T}n_{i,t}}, \end{aligned} \tag{3.3}$$

where $k=1$ is the standard negative multivariate distribution (MVNB2), $k=0$ is the multivariate NB1 equivalence (MVNB1). The variable k can also be numerically estimated, which can be used to construct a method of discriminating between the MVNB2 and the MVNB1 models. This method, used by Boucher et al. (2006c) is based on the creation of an hypermodel, where the additional parameter, here the random variable k, is used to test whether MVNB2 or MVNB1 is statistically the better by doing a simple confidence interval. Expressed in its general form, the MVNBk distributions has the following moments :

$$\begin{aligned} E[N_{i,t}] &= E[E[N_{i,t}|\theta_i]] \\ &= \lambda_i \end{aligned} \tag{3.4}$$

$$\begin{aligned} Var[N_{i,t}] &= E[Var[N_{i,t}|\theta_i]]+Var[E[N_{i,t}|\theta_i]] \\ &= \lambda_i+\alpha\lambda_i^{k+1} \end{aligned} \tag{3.5}$$

$$\begin{aligned} Cov[N_{i,t},N_{i,t+j}] &= Cov[E[N_{i,t}|\theta_i],E[N_{i,t+j}|\theta_i]]+E[Cov[N_{i,t},N_{i,t+j}|\theta_i]] \\ &= Cov[\lambda_i\theta_i,\lambda_i\theta_i]+0 \\ &= \alpha\lambda_i^{k+1}, \quad j>0 \end{aligned} \tag{3.6}$$

As for the cross-sectionnal data, other distributions can be chosen to model the random effects, such as the Inverse Gaussian or the LogNormal distributions, which result in distributions having the same form for the two first moments, distinctions found using higher moments. Note that the MVNB2 and MVNB1 models are nested to the Poisson distribution in the case when $\alpha \to 0$.

4. Hurdle Models

4.1 Overview

Because the vast majority of the insureds reports less than 2 claims per year, Boucher et al. (2006c) proposed to model the number of reported claims by two different processes. Firstly, a dichotomic distribution to differentiate insureds with and without claim. Secondly, an other process that precises, conditionally on having reported at least one time, the number of reported claims. The distribution that uses two processes to determine the number of reported claims can be driven by the same explanatory variables, that may be interpreted differently depending on the process involved. This kind of modelling looks like the double modelling of the amount of claims, where the limited costs and the excess-of-loss costs (costs above a certain value) are modelised separately.

The most popular distribution implying the assumption that the data come from two separate processes is the Hurdle count models that were introduced by Mullahy (1986). The hurdle model is characterized by the process below the hurdle and the one above. Obviously, the most widely used hurdle model is the one which sets the hurdle at zero. Formally, the hurdle-at-zero model is expressed as :

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

The variable Φ 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 $f_1 = f_2 = f$. Expected values and variance of this Hurdle model are expressed as :

$$E[N_i] = \frac{1 - f_1(0)}{1 - f_2(0)} \sum_{k=0}^{\infty} k f_2(k) \tag{4.2}$$

$$Var[N_i] = \Phi \sum_{k=1}^{\infty} k^2 f_2(k) - [\Phi \sum_{k=1}^{\infty} k^2 f_2(k)]^2, \tag{4.3}$$

where $\Phi = (1-f_1(0))/(1-f_2(0))$. Consequently, the model can be over or underdispersed, depending on the values of the parent processes. Many possibilites exist for the choice of the processes f_1 and f_2. Nested models where f_1 and f_2 come from the same distribution, such as the Poisson distribution Mullahy (1986) or the Negative Binomial Pohlmeier and Ulrich (1995), which is by far the most popular hurdle model Winkelmann (2000), are possible. However, non-nested models Grootendorst (1995), Gurmu (1998), or Winkelmann (2003a) can also be used. These models do not nest with a standard count distributions such as the Poisson or the NB types, but are overlapping Vuong (1989) since models can be equivalent for certain parameter restrictions.

Estimation of the parameters by maximum likelihood is easy. The log-likelihood function of a hurdle model can be expressed as :

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

The hurdle model is interesting because it allows to estimate the parameters by two separate steps. Indeed, the zero-part parameters can be estimated using MLE on the first part of the loglikelihood equation while the other parameters only use the second part, only composed with non-zero elements. This charateristic of the model is very useful to save computer time in the estimation.

4.1.1 Unobserved heterogeneity

To add some uncertainty due to the absence of important classification variables, an heterogeneity term can be added to the distribution. The standard approach is to integrate the heterogeneity prior to the conversion into a hurdle model :

$$\hat{f}_1(n_i) = \int f_1(n_i|\theta) g_1(\theta) d\theta \tag{4.5}$$

$$\hat{f}_2(n_i) = \int f_2(n_i|\theta) g_2(\theta) d\theta \tag{4.6}$$

$$P(N_i = n_i) = \begin{cases} \hat{f}_1(0) & \text{for } n_i = 0 \\ \frac{1-\hat{f}_1(0)}{1-\hat{f}_2(0)} \hat{f}_2(n_i) & \text{for } n_i = 1, 2, ... \end{cases} \tag{4.7}$$

However, as noted by Santos Silva (2003), it seems more natural to estimate the heterogeneity on the positive part and on the zero-part of the model. Formally,

$$P(N_i = n_i) = \begin{cases} \int f_1(0|\theta) g_1(\theta) d\theta & \text{for } n_i = 0 \\ \left(\int (1 - f_1(0|\theta)) g_1(\theta) d\theta\right) \left(\int \frac{f_2(n_i|\theta)}{1-f_2(0|\theta)} g_2(\theta) d\theta\right) & \text{for } n_i = 1, 2, ... \end{cases} \tag{4.8}$$

Conceptually, models (4.7) and (4.8) differ in interpretation. As noted in Santos Silva (2003) and reviewed in Winkelmann (2003a), the zero and positive processes of the hurdle model are evaluated separately and possesses their own heterogeneity. Consequently, model (4.8) seems to be more intuitive since we avoid the step of needing to compute $\int f_2(0|\theta) g_2(\theta) d\theta$ althought the zeros are supposed to be generated by an other process.

When an heterogeneity term is chosen for model (4.8), closed expression can be found using a transformation. Indeed, a new random variable $n_i* = n_i - 1$ can be used to model the second part of the hurdle model with standard count models, such as a Poisson distribution or other variations.

4.1.2 Model interpretation

The Hurdle models are quite popular in the modelling in health care demands. Indeed, 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 (Stoddart and Barer (1981), Pohlmeier and Ulrich (1995), Mullahy (1998), Santos Silva and Windmeijer (2001). Consequently, conditionnaly on certain assumptions, the use of a hurdle model is intuitive and the parameters can have a structural interpretation.

Many direct applications of this model can be used for insurance data, such as the number of coverage implied for a claim, the number of injured people or the number of implicated third parties. Such modellings using separate processes are appropriate since these processes are not the same as the one that drives the number of accident.

However, beside these specific examples, such a modelling can also have a natural interpretation for the number of reported claims. Indeed, since it can exist a reluctance from some insureds to report their accident since they would lose their good bonus-malus rating, we can suppose that the behavior of the insureds is not same when they already have reported a claim.

4.2 Panel data

The hurdle model can be directly generalized to panel data using the transformation $n_i^* = n_i - 1$:

$$\begin{aligned} P(N_{i,1},...,N_{i,t}) &= \int \prod_t f_1(0|\theta_{i,1})^{I_{(n_{i,t}=0)}} \left(1 - f_1(0|\theta_{i,1})\right)^{1-I_{(n_{i,t}=0)}} g_1(\theta_{i,1}) d\theta_{i,1} \\ &\times \int \prod_t f_2^*(n_{i,t}^*|\theta_{i,2})^{1-I_{(n_{i,t}=0)}} g_2(\theta_{i,2}) d\theta_{i,2}. \end{aligned} \tag{4.9}$$

As for the cross-sectionnal data, the hurdle model is interesting because it allows to estimate the parameters on two separate steps. The transformation of the positive part allows to fit the data using standard Poisson random effects models seen in section 3, where the mean variable can be expressed as $\gamma_i = \exp(x_i\delta)$. For the zero-part of the model, instead of using the classic Poisson parametrization approach that results in a complicated form for the joint distribution expression, a more intuitive model could be the use of a Bernouilli distribution where the parameter is beta distributed to account for the invidual specificities.

Using a Bernouilli-Beta combination for the zero part, the joint distribution of the variable $Z_{i,t}$,

that can take values of 0 or 1, is well known and can be expressed as :

$$\begin{aligned} P(Z_{i,1},...,Z_{i,t}) &= \int \prod_t \theta_i^{z_{i,t}} (1-\theta_i)^{1-z_{i,t}} \frac{\Gamma(a_i+b)}{\Gamma(a_i)\Gamma(b)} \theta_i^{a_i-1} (1-\theta_i)^{b-1} d\theta_i \\ &= \frac{\Gamma(a_i+b)}{\Gamma(a_i)\Gamma(b)} \frac{\Gamma(\sum_t z_{i,t}+a_i)\Gamma(T-\sum_t z_{i,t}+b)}{\Gamma(T+a_i+b)} \end{aligned} \tag{4.10}$$

The covariates can be included in the model as $a_i = \exp(x_i\beta)$. Having the upper part modeled with the transform random variable n_i^* and basing the computations on equations (4.2) and (4.3), the complete hurdle model involves the following moments :

$$\begin{aligned} E[N_{i,t}] &= \frac{a_i}{a_i+b} \sum_{j=0}^{\infty} (1+j) f_{n_i^*}(j) \\ &= \frac{a_i}{a_i+b}(1+\gamma_i) \end{aligned} \tag{4.11}$$

$$\begin{aligned} E[N_{i,t}^2] &= \frac{b}{a_i+b} \sum_{j=0}^{\infty} (1+j)^2 f_{n_i^*}(j) \\ &= \frac{a_i}{a_i+b}\left[1+3\gamma_i+\alpha\gamma_i^{k+1}-\gamma_i^2\right] \end{aligned} \tag{4.12}$$

$$Var[N_{i,t}] = E[N_{i,t}^2] - E[N_{i,t}]^2 \tag{4.13}$$

$$\begin{aligned} Cov[N_{i,t}, N_{i,t+j}] &= (1+\gamma_i)^2 \frac{a_i b}{(a_i+b+1)(a_i+b)^2} \\ &+ \frac{\gamma_i^{k+1}}{\alpha}\left(\frac{a_i b}{(a_i+b+1)(a_i+b)^2} + \left(\frac{a_i}{a_i+b}\right)^2\right), \quad j>0. \end{aligned} \tag{4.14}$$

5. Predictive Distribution

For the Poisson and the hurdle models for panel data, at each insured period, the random effects θ_i and p_i can be updated for past claims experience, revealing some individual informations. Formally, the predictive distribution can be found using the following development :

$$\begin{aligned} \Pr[n_{i,T+1}|n_{i,1},...,n_{i,T}] &= \frac{\Pr(n_{i,1},...,n_{i,T+1})}{\Pr(n_{i,1},...,n_{i,T})} = \frac{\int \Pr(n_{i,1},...,n_{i,T+1},\theta_i)d\theta_i}{\int \Pr(n_{i,1},...,n_{i,T},\theta_i)d\theta_i} \\ &= \int \Pr(n_{i,T+1}|\theta_i)\left(\frac{\Pr(n_{i,1},...,n_{i,T}|\theta_i)g(\theta_i)}{\int \Pr(n_{i,1},...,n_{i,T}|\theta_i)g(\theta_i)d\theta_i}\right)d\theta_i \\ &= \int \Pr(n_{i,T+1}|\theta_i)\left(\frac{\left[\prod_t \Pr(n_{i,t}|\theta_i)\right]g(\theta_i)}{\int \left[\prod_t \Pr(n_{i,t}|\theta_i)\right]g(\theta_i)d\theta_i}\right)d\theta_i \\ &= \int \Pr(n_{i,T+1}|\theta_i) g(\theta_i|n_{i,1},...,n_{i,T}) d\theta_i, \end{aligned} \tag{5.1}$$

where $g(\theta_i|n_{i,1},...,n_{i,T})$ is the a posteriori distribution of the random effects θ_i, reflecting the past claims experience of insured i. If this a posteriori distribution can be expressed in closed form, moments of the predictive distribution can be found easily by conditioning on the random effects θ_i.

5.1 Poisson

As it is well known, we found that the a posteriori distribution of the random effect term is also a Gamma distribution having parameters equal to $T\lambda_i + \lambda_i^{(1-k)}/\alpha$ and $\sum_t^T n_{i,t} + \lambda_i^{(1-k)}/\alpha$. The two first moments of the predictive distributions are equaled to :

$$E[N_{i,t+1}|N_{i,1},...,N_{i,t}] = \lambda_i \frac{\sum_t^T n_{i,t} + \lambda_i^{(1-k)}/\alpha}{T\lambda_i + \lambda_i^{(1-k)}/\alpha} \tag{5.2}$$

$$Var[N_{i,t+1}|N_{i,1},...,N_{i,t}] = \lambda_i \frac{\sum_t^T n_{i,t} + \lambda_i^{(1-k)}/\alpha}{T\lambda_i + \lambda_i^{(1-k)}/\alpha} + \lambda_i^2 \frac{\sum_t^T n_{i,t} + \lambda_i^{(1-k)}/\alpha}{(T\lambda_i + \lambda_i^{(1-k)}/\alpha)^2}. \tag{5.3}$$

We see that the future premium only depends on the sum of the number of reported claims. Additionnally, we observe that the premium of insured i goes to its average number of reported claim if the number of insured periods goes large. For models where the variable k is different than 1, the variance of the heterogeneity distribution depends on characteristics of the insureds. Then, the impact on premium of having a claim is more important for low variance heterogeneity and thus, premium modifications for negative composants of λ_i are more severe.

5.2 Hurdle

The same development can be done for the two processes of the hurdle distribution, since the panel hurdle model can be expressed as :

$$\begin{aligned} P(N_{i,1},...,N_{i,t}) &= \int \prod_t f_1(0|p_i)^{z_{i,t}} \left(1 - f_1(0|p_i)\right)^{1-z_{i,t}} g(p_i) dp_i \\ &\times \int \prod_t f_2^*(n_{i,t}^*|\theta_i)^{1-I_{(n_{i,t}=0)}} h(\theta_i) d\theta_i. \end{aligned} \tag{5.4}$$

Because the two processes can be analyzed separately, the a posteriori distribution of the random effect term of the first process can be found as follows :

$$\begin{aligned} g[p_i|z_{i,1},...,z_{i,T}] &\propto p_i^{\sum_t^T z_{i,t}} (1-p_i)^{T-\sum_t^T z_{i,t}} \frac{\Gamma(a_i+b)}{\Gamma(a_i)\Gamma(b)} p_i^{a-1} (1-p_i)^{b-1} dp_i \\ &\propto p_i^{\sum_t^T z_{i,t}+a_i-1} (1-p_i)^{T-\sum_t^T z_{i,t}+b-1} dp_i. \end{aligned}$$

From which, we can see that it is Beta distributed with parameters $\sum_t^T z_{i,t}+a_i$ and $T-\sum_t^T z_{i,t}+b$. For the first part of the distribution, the expected value of the predicted distribution is then calculated as :

$$E[z_{i,T+1}|z_{i,1},...,z_{i,T}] = \frac{\sum_t^T z_{i,t} + a_i}{T + b + a_i} \tag{5.5}$$

The second process $N_i^* = N_i - 1$ of the hurdle distribution is following a MVNB distribution. Then, the a posteriori distribution of the random effect term can be found using the same development as section 5.1, and thus the expected value of the second process is equal to :

$$\begin{aligned} E[n_{i,J+1}|n_{i,J+1} > 0, n_{i,1},...,n_{i,J}] &= 1 + E[N^*_{i,J+1} = n^*_{i,J+1}|n_{i,1},...,n_{i,J}] \\ &= 1 + \gamma_i \frac{\sum_t^J n_{i,t} + \gamma_i^{(1-k)}/\alpha}{J\gamma_i + \gamma_i^{(1-k)}/\alpha}, \end{aligned} \tag{5.6}$$

where J is equaled to the number of insured period where the insured has reported at least one claim. Combining equations (5.5) and (5.6) leads to the predictive expected value of the count distribution, while the variance of the model can be computed using equations (4.3) and (4.13) :

$$E[n_{T+1}|n_1,...,n_T] = \frac{\sum_t^T z_{i,t} + a_i}{T + b + a_i} \left(1 + \gamma_i \frac{\sum_t^J n_{i,t} + \gamma_i^{(1-k)}/\alpha}{J\gamma_i + \gamma_i^{(1-k)}/\alpha}\right) \tag{5.7}$$

$$\begin{aligned} E[n^2_{T+1}|n_1,...,n_T] &= \frac{\sum_t^T z_{i,t} + a_i}{T + a_i + b} \left[1 + 3\gamma_i \frac{\sum_t^J n_{i,t} + \gamma_i^{(1-k)}/\alpha}{J\gamma_i + \gamma_i^{(1-k)}/\alpha}\right. \\ &+ \left. \gamma_i^2 \frac{\sum_t^J n_{i,t} + \gamma_i^{(1-k)}/\alpha}{(J\gamma_i + \gamma_i^{(1-k)}/\alpha)^2} - \left(\gamma_i \frac{\sum_t^J n_{i,t} + \gamma_i^{(1-k)}/\alpha}{J\gamma_i + \gamma_i^{(1-k)}/\alpha}\right)^2\right]. \end{aligned} \tag{5.8}$$

As opposed to the Multivariate Negative Binomial models, the future premium not only depends on the sum of reported claims, but also on the number of insured periods without claim. Additionnally, when the insured periods goes to infinite, the premium of the insured does not converge to the average number of reported claims, but on the product of two limited values : the average number of time periods with at least one claim and the average of the number of reported claims greater or equal to one.

5.3 Linear credibility

Predictive and a posteriori distributions used bayesian theory and are strongly related to credibility theory Bühlmann (1967), Bühlmann and Straub (1970), Hachemeister (1975), Jewell (1975).

Linear credibility is a theory used to obtain premium based on a weighted average of past experience and a priori premium, such as :

$$P_{T+1} = Z\bar{n}_{i,t} + (1 - Z) \times P_0, \tag{5.9}$$

where $\bar{n}_{i,t} = \sum_{t=1}^{T} \frac{n_{i,t}}{T}$, P_{T+1} is the predictive premium a time $T + 1$, P_0 is the a priori premium and $n_{i,t}$ is the number of reported claim for insured i at period t. The coefficient Z, that gives weight to the two components, is the value that minimizes the squared error of the predictive value (Bühlmann, 1967) which corresponds to :

$$Z = \frac{Cov(\bar{N}_{i,t}, N_{n+1})}{Var(N_i^*)}.$$

The linear credibility model gives exact results only for conjugated distributions Jewell (1975), such as the Poisson with Gamma random effects or Bernoulli with Beta parameters. Then, using standard results of the credibility theory, it is possible to obtain the same premium as the one obtained in equation (5.7).

6. Insurance Application

The models seen in the preceding sections can be used to calculate the impact of the covariates on the number of reported claims, but can also be used to calculate a priori premiums or predictive premiums of the frequency part of an insurance premium. The amount of claims is the other part of a standard insurance premium. The a priori premium represents the premium that a new insured must paid to be covered. By opposition, the predictive premium is the premium of the other insureds, that depends on the experience of each insured.

6.1 Estimations

Applications of the Poisson and the Hurdle models on our insurance data lead to the results shown in table 6.1 for the models based on Poisson generalization, while tables 6.2 and 6.3 refer to the Hurdle model since it can be decomposed in two parts.

For the Multivariate Negative Binomial models, the estimated parameters are quite the same for all models. Without surprises, we see that women exhibit less claims than men, while new insureds in the company seems to have a worst loss experience than older client. In the presence of other covariates, we can also see that young drivers exhibit a better claim experience, but it is not statistically significant. To finish, we observe that insureds with powerful vehicle suffer more accident than other drivers. A look on the p-values of the parameters implying time dependence

Parameter	MVNB2		MVNB1		MVNBk	
b0	−2.6600	(0.0352)	−2.6635	(0.0341)	−2.6646	(0.0337)
b1	0.1087	(0.0409)	0.1170	(0.0387)	0.1166	(0.0380)
b2	−0.1805	(0.0327)	−0.1731	(0.0314)	−0.1684	(0.0326)
b3	−0.2103	(0.0370)	−0.2068	(0.0358)	−0.2022	(0.0368)
b4	0.0471	(0.0346)	0.0599	(0.0329)	0.0613	(0.0324)
b5	0.0990	(0.0316)	0.0931	(0.0305)	0.0904	(0.0305)
α	0.8832	(0.0432)	0.0610	(0.0031)	0.0339	(0.0426)
k	.	.	.	.	−0.2188	(0.4682)
LogLike	26,703.0		26,699.4		26,699.3	

Tab. 6.1 – Multivariate Negative Binomial Models

between contracts of the same insured leads to the conclusion that it improves the model. However, more statistical tests must be done to conclude that the introduction of this term improves the Poisson models since it cannot be negative, and the tested hypothesis is on the boundary of the parameter space. Indeed, when a parameter is bounded by the H_0 hypothesis, the estimate is also bounded and the asymptotic normality of the MLE no longer hold under H_0.

By comparison of the log-likelihoods, we see that the form of the random effects do not really modify the fitting of the data. The confidence interval implies by the k variable of MVNBk model concludes that the MVNB2 and the MVNB1 are both accepted.

For the Hurdle models, we have to notice that the parameters obtained cannot be compared directly to the ones obtained with the multivariate negative binomial distributions since it does not have the same impact on the expected value, see equations (4.11) and (3.4). However, by the sign of the estimated parameters of the zero-part of the hurdle model (table 6.2), we can conclude the same conclusions as the ones done with the MVNB distribution for the probability to report at least one claim in a time period. As opposed to the Poisson generalization models, the time dependence assumption cannot be analysed directly with this parametrization.

The parameters of the table 6.3 allow us to observe what is the kind of insureds that are most likely to report a high number of claim in a single time period. The result is a little bit strange since only the insureds that was in the company between 3 and 5 years at the beginning of the study are statistically better than the others. The α parameter seems to be quite significative and the k variable seems to favorise the MVNB2 distribution, even if its confidence interval accept both

Parameter	Beta	
b0	0.3066	(0.0682)
b1	0.1298	(0.0401)
b2	−0.1726	(0.0325)
b3	−0.2124	(0.0370)
b4	0.0616	(0.0341)
b5	0.0975	(0.0315)
b	19.9640	(1.2279)
Loglike.	24,637.9	

Tab. 6.2 – Hurdle Models : Zero Part

Parameter	MVNB2		MVNB1		MVNBk	
b0	−2.3688	(0.0525)	−2.3695	(0.0524)	−2.3687	(0.0526)
b2	−0.1951	(0.0933)	−0.1922	(0.0933)	−0.1953	(0.0935)
α	0.8122	(0.2070)	0.0718	(0.0183)	0.6955	(4.7872)
k	.	.	.	.	0.9359	(2.8422)
Loglike.	2,050.80		2,050.85		2,050.80	

Tab. 6.3 – Hurdle Models : Positive Part

models.

6.2 Premiums

Difference between models can be analysed throught the mean and the variance of some insured profiles. Several profiles have been selected and are described in Table 6.4. The first profile is classified as a good driver, while the last one usually exhibits bad loss experience. The other profile is medium risk. The results are given in Table 6.5. This table shows that the expected values of all profiles are quite the same for the 4 models studied. The biggest differences lie in the variance values, where the hurdle models exhibit higher values.

Table 6.6 shows the predictive premiums of the 2^{nd} profile for MVNB2 and hurdle-MVNB2 (chosen arbitrary between available models), which depends on the sum of reported claims and on the number of insured periods with at least one reported claim ($t = T - T_0$). To illustrate, we selected a loss experience of 10 years but other situations can be easily illustrated since equations

Profile Number	Kind of Profile	v1	v2	v3	v4	v5
1	Good	0	0	1	0	0
2	Average	1	1	0	0	0
4	Bad	1	0	0	1	1

Tab. 6.4 – Profiles analyzed

	1^{st} Profile		2^{nd} Profile		3^{rd} Profile	
Models	Mean	Variance	Mean	Variance	Mean	Variance
MVNB2	0.0567	0.0595	0.0651	0.0688	0.0902	0.0974
MVNB1	0.0567	0.0601	0.0659	0.0699	0.0911	0.0966
H. Be-MVNB2	0.0570	0.0667	0.0659	0.0753	0.0911	0.1065
H. Be-MVNB1	0.0570	0.0667	0.0659	0.0753	0.0911	0.1065

Tab. 6.5 – A priori Premiums

(5.2) and (5.7) and simple to use.

Interesting conclusions can be done from the analysis of predictive premiums. Indeed, we can see that the number of insured periods with a claim have a greater impact of the next year premium than the total number of reported claims. Indeed, the premium increases only by 16.4% if the insured reports 4 claims instead of one in a single insured period. By opposition, if the insured reports his 4 claims on 4 different time periods instead of only one, his premium increases by more than 95%. For insureds who reported 1 claim or less, the hurdle model offer a discount that is smaller than the MVNB model. For higher claims reporter, we can see that the hurdle model exhibits a wide panel of premium values that go from 0.3 to 1.8 times the MVNB's premiums.

7. Conclusion

The behavior of the insureds toward their bonus-malus scheme seems to influence their probability to report a claim. Model such as the Hurdle distribution use this feature to models the number of reported claims and provides good fitting. The generalization of the hurdle model to panel data can be done directly and shows interesting property, such as the discrimination of the insureds depending of the number of reported claims and the number of insured periods without claim. The choice of the best distribution describing our data must be supported by specification tests for nested or non-nested models and should be the subject of further analysis. There is also a

		Sum of claims						
Models	t	0	1	2	3	4	10	20
PG2	.	0.0413	0.0778	0.1143	0.1509	0.1874	0.4064	0.7715
H. PG2	0	0.0448	.	.	.	.	.	
	1	.	0.0790	0.0833	0.0876	0.0920	0.1180	0.5067
	2	.	.	0.1128	0.1187	0.1246	0.1598	0.5345
	3	.	.	.	0.1465	0.1538	0.1972	0.5623
	4	.	.	.	.	0.1800	0.2309	0.5901
	10	.	.	.	.	.	0.3786	0.7570
	20	.	.	.	.	.	0.7393	1.0351

TAB. 6.6 – Predictive Premiums

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 this classification must go a little beyond where the number of insured periods without claim has some importance.

Chapitre 7

Independent and Correlated Random Effects for Hurdle Models Applied to Panel Count Data

Soumis pour publication à *Econometrics Journal*
en collaboration avec Michel Denuit et Montserrat Guillén

1. Introduction

The literature on regression analysis of count data has grown considerably in recent years. The existing literature has mainly advocated the use of the classical Poisson regression approach, but little has been said about alternative models and model selection. When data exhibit a high number of zero values, interesting alternatives involve the use of zero-inflated or hurdle models which often provide a good fit for cross-section data. It is also well known that panel data models using random effects are well suited to fit the dependence between all the observations of a single unit.

Combining these two ideas, we present a new model that consists of a generalization of the hurdle distribution for panel count data. The hurdle distribution is a model that supposes that the count data are generated by two processes, while longitudinal data (or panel data) consist of repeated observations of individual units over time. These individual units can be individuals or experimental units. Each individual unit is assumed to be independent but a correlation implied by the random effects is allowed between their repeated observations. The hurdle model is generalized to include random effects in each of its process. Independence and dependence between these random effects is supposed. Predictive distribution is derived analytically for independent random effects, while Markov chain Monte Carlo simulations are needed to compute posterior distributions of the random effects for correlated models. A numerical illustration of reported insurance claims supports the discussion, where we see that the dependence between random effects must be considered when computing predictive distribution.

The paper is organized as follows. In section 2, hurdle models are reviewed and presented for cross-section data. The theory for panel data is given in section 3. In section 4, we propose our generalization of the hurdle model for panel data under the assumption that random effects are independent or correlated. In the next section, we show that the predictive distribution can be expressed in a closed form for independent random effects, while Markov chain Monte Carlo simulations are needed to compute the posterior distribution of the correlated random effects models. In section 6, numerical examples of a Spanish insurance portfolio are given. Computation of the predictive premiums, analysis of the dependence of the random effects over time and a comparison of models are all provided.

2. Hurdle Distributions

2.1 Overview

The hurdle distribution was introduced by Mullahy (1986). This model is characterized by the processes below and above the hurdle. Obviously, the most widely used hurdle model is that which sets the hurdle at zero, which leads to a distribution with the following two processes : firstly, a dichotomic distribution that allows the participation of the second process ; secondly, another process that specifies, on the condition that the first process *succeeds*, the count number. The first part of the model is a binary outcome model, and the second part is a truncated count distribution. Formally, this hurdle model is expressed as follows. Let $f_1(\cdot|\theta_1)$ and $f_2(\cdot|\theta_2)$ be two probability mass functions with respective support $\{0,1\}$ and $\{0,1,\ldots\}$ depending on parameter vectors θ_1 and θ_2. The counting random variable N obeys to the hurdle distribution if :

$$P(N=n) = \begin{cases} f_1(0|\theta_1) & \text{for } n=0 \\ \frac{1-f_1(0|\theta_1)}{1-f_2(0|\theta_2)} f_2(n|\theta_2) = \Psi f_2(n|\theta_2) & \text{for } n=1,2,... \end{cases}, \tag{2.1}$$

where $\Psi = \frac{1-f_1(0|\theta_1)}{1-f_2(0|\theta_2)}$. The variable Ψ can be interpreted as the probability of crossing the hurdle, which means the probability of registering a count. Clearly, the model collapses to the distribution f if $f_1 = f_2 = f$. Expected values and variance of this hurdle model are expressed as :

$$E[N_i] = \frac{1-f_1(0|\theta_1)}{1-f_2(0|\theta_2)} \sum_{k=0}^{\infty} k f_2(k|\theta_2), \tag{2.2}$$

$$Var[N_i] = \Psi \sum_{k=1}^{\infty} k^2 f_2(k|\theta_2) - \left(\Psi \sum_{k=1}^{\infty} k^2 f_2(k|\theta_2) \right)^2. \tag{2.3}$$

Consequently, the model can be over- or underdispersed, depending on the parent processes f_1 and f_2. Many possibilities exist for the choice of the parent-processes f_1 and f_2 : nested models in which f_1 and f_2 come from the same distribution, such as the Poisson distribution (Mullahy (1986)), or the negative binomial (NB) (Pohlmeier and Ulrich (1995)), which is by far the most popular hurdle model (Winkelmann (2003b)). However, non-nested models such as those offered by Grootendorst (1995), Gurmu (1998) or Winkelmann (2003a) can also be used. These models do not nest with standard count distributions such as the Poisson or NB types, but are instead overlapping (Vuong (1989)) since models can be equivalent for certain parameter restrictions. Estimation of parameters by maximum likelihood is easy since each process can be estimated separately.

In this paper, instead of using a truncated count distribution for the second process in the model, we propose a specific transformation. By using $n^* = n - 1$, the second part of the hurdle model can be constructed with a standard count distribution, such as the Poisson or the negative binomial distribution. We do not transform the data for the first hurdle process, which means that it can be modeled using a Bernoulli distribution. Specifically, we now use the following variant of the hurdle distribution :

$$P(N = n) = \begin{cases} f_1(0|\theta_1) & \text{for } n = 0 \\ (1 - f_1(0|\theta_1)) f_2(n^*|\theta_2) & \text{for } n = 1, 2, ... \end{cases} . \tag{2.4}$$

Using this transformation, the log-likelihood function of a random sample of observations N_i, $i = 1, \ldots, n$, coming from a hurdle model can be expressed as :

$$\ell = \sum_{i=1}^{n} I_{(n_i=0)} \ln f_1(0|\theta_1) + I_{(n_i>0)} \ln(1 - f_1(0|\theta_1)) + \sum_{i=1}^{n} I_{(n_i>0)} \ln(f_2(n_i^*|\theta_2)). \tag{2.5}$$

Provided the model is separable (that is, θ_1 and θ_2 do not share common components), we can see that the estimation of the parameters can be divided into two independent steps : one for the Bernoulli distribution and another one for the Poisson distribution. Many researchers into count data are sceptical of more sophisticated models than the Poisson distribution because they can lead to inconsistent estimates. Indeed, the parameters of Poisson distribution, which is part of the exponential family, are consistent as long as the mean function is correctly specified (Gourieroux et al. (1984b) and Gourieroux et al. (1984a)). However, if the underlying data come from two-processes distributions like the hurdle model, this robustness property does not hold since the mean function cannot be correctly specified.

Using the n^* transformation, the two processes of the hurdle distribution can be derived from distributions of the exponential family. Indeed, if we choose the Bernoulli and the Poisson distributions for the first and the second processes, the parameters obtained by maximum likelihood estimation are consistent if the mean function is well defined. Consequently, unlike the Poisson distribution, the parameters of the hurdle model under this parametrization will be well evaluated even if the data are taken from a two-process distribution. Additionally, if the data come from a one-process distribution, the parameters will also be consistent since the product of two consistent estimators is also consistent.

2.2 Heterogeneity

To take into account the uncertainty due to the absence of important explanatory variables, a heterogeneity term can be added to the distribution. The standard approach is to integrate the heterogeneity prior to conversion into a hurdle model, as for the NB2 hurdle model, where the gamma heterogeneity has been added to a Poisson distribution. Formally, when the heterogeneities are independent, two different heterogeneity expressions can be used :

$$\hat{f}_1(n_i) = \int f_1(n_i|\theta) g_1(\theta) d\theta, \tag{2.6}$$

$$\hat{f}_2(n_i) = \int f_2(n_i|\theta) g_2(\theta) d\theta, \tag{2.7}$$

$$P(N_i = n_i) = \begin{cases} \hat{f}_1(0) & \text{for } n_i = 0 \\ \frac{1-\hat{f}_1(0)}{1-\hat{f}_2(0)} \hat{f}_2(n_i) & \text{for } n_i = 1, 2, ... \end{cases} \tag{2.8}$$

and

$$P(N_i = n_i) = \begin{cases} \int f_1(0|\theta) g_1(\theta) d\theta & \text{for } n_i = 0 \\ \left(\int (1 - f_1(0|\theta)) g_1(\theta) d\theta\right) \left(\int \frac{f_2(n_i|\theta)}{1-f_2(0|\theta)} g_2(\theta) d\theta\right) & \text{for } n_i = 1, 2, ... \end{cases}. \tag{2.9}$$

It seems more natural to estimate the heterogeneity on the positive part and the zero-part of the model (Santos Silva (2003)). Conceptually, models (2.8) and (2.9) differ in interpretation. As noted in Santos Silva (2003) and reviewed in Winkelmann (2003a), the zero and positive processes of the hurdle model are evaluated separately and possess their own heterogeneity. Consequently, model (2.9) seems to be more intuitive since the need to compute $\int f_2(0|\theta) g_2(\theta) d\theta$ is avoided although the zeros are supposed to be generated by an other process. For example, note that if we choose Poisson distributions with two gamma heterogeneities, model (2.9) is clearly not an NB2 hurdle model.

Using the transformation $n_i^* = n_i - 1$ seen earlier, the second part of the hurdle model can constructed with a Poisson distribution, where many combinations with heterogeneity distributions are well known, such as gamma, inverse Gaussian or the lognormal distribution (see Boucher and Denuit (2006) or Boucher et al. (2006b) for some applications to insurance).

3. Cross Section versus Panel Data

The data used for panel data analysis consist of N individual units, each having T observations. Usually, data are analyzed by cross-sectional analysis, where each observation of an individual unit is considered to be mutually independent. Thus, in this situation, we worked with $N \times T$ independent observations. However, it is common for some data to be a repetition for the same individual. Consequently, a dependence over these observations must be modeled.

Longitudinal data (or panel data) consists of repeated observations of individual units that are observed over time. Each individual is assumed to be independent but correlation between observations of the same individual is allowed. This fact must be regarded as an advantage and must be used in the construction of count models.

3.1 Modeling

There are many models involving time dependence but the most popular way of dealing with these data is the use of a common individual term (Hausman et al. (1984)) that affects all the observations of an individual unit. This random effect represents individual specificities that are constant over time. The hidden individual characteristics that are not captured by the covariates are also captured by this random effect.

Given the insured-specific random effect term θ_i, the annual claim numbers $N_{i,1}, N_{i,2}, \ldots, N_{i,T}$ are independent. Under the assumption that the covariates are independent from the random effects (see Mundlak (1978) or Hsiao (2003) for a general review), the joint probability function of $N_{i,1}, ..., N_{i,T}$ is thus given by :

$$\begin{aligned} P(N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}) &= \int_0^\infty P(N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}|\theta_i) g(\theta_i) d\theta_i \\ &= \int_0^\infty \left(\prod_{t=1}^{T} P(N_{i,t} = n_{i,t}|\theta_i) \right) g(\theta_i) d\theta_i \end{aligned} \tag{3.1}$$

where i represents an individual unit and t is the time of each observation. For example, if we use a Poisson distribution such as

$$N_{i,t}|\theta_i \sim Poisson(\theta_i \lambda_i), \quad i = 1, ..., N \quad t = 1, ..., T,$$

with θ_i a gamma distribution having both parameters equal to $1/\alpha$, it can be shown (Hausman et al. (1984)) that the joint distribution is :

$$P(N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}) = \left(\prod_{t=1}^{T} \frac{\lambda_i^{n_{i,t}}}{n_{i,t}!}\right) \frac{\Gamma(\sum_{i=1}^{T} n_{i,t} + 1/\alpha)}{\Gamma(1/\alpha)} \left(\frac{1/\alpha}{T\lambda_i + 1/\alpha}\right)^{1/\alpha} (T\lambda_i + 1/\alpha)^{-\sum_{i=1}^{T} n_{i,t}} . \tag{3.2}$$

Many conditional distributions for the random variables $N_{i,t}$ can be chosen as well as a distribution for the random effect θ_i. Under conditioning, moments of the joint distribution can be found.

4. Hurdle Model for Panel Data

The modeling of the panel data for the hurdle distribution can be performed as for the Poisson distribution. However, since random effects can be added to the model for both the first and the second process, the joint distribution expressed by equation (3.1) can be generalised for more than one θ_i. Indeed, the model can be expressed as :

$$\begin{aligned} & P(N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}) \\ = & \int\int \prod_{t=1}^{T} f_1(0|\theta_{i,1})^{I_{(n_{i,t}=0)}} (1 - f_1(0|\theta_{i,1}))^{1-I_{(n_{i,t}=0)}} f_2(n^*_{i,t}|\theta_{i,2})^{1-I_{(n_{i,t}=0)}} g(\theta_{i,1}, \theta_{i,2}) d\theta_{i,1} d\theta_{i,2}, \end{aligned} \tag{4.1}$$

where $g(\theta_{i,1}, \theta_{i,2})$ is the joint probability density function of the random effects, which can be expressed by a copula. A special case of bivariate copula is the product copula that consists of two independent random variables, which will be seen in the next section.

Note that the use of a heterogeneity term on the zero part or on the positive part alone cannot be satisfactory for many kinds of data. For insurance data for instance, the exclusive use of a heterogeneity term on the zero-part leads to a situation in which the increase of premiums is the same whether an insured has reported one claim or more in the same one-year period. Similarly, if we choose a model where only the positive part includes random effects, no discount will be given to an insured with no reported claims. Moreover, since the mean of the positive part is necessarily greater than one (except in the case where the maximum number of reported claims is one for

the entire portfolio), an insured who reports one claim will have a premium reduction for the next insured period. Consequently, the hurdle model with panel data must be analyzed with two random effects.

For the zero-part of the model, a Bernoulli distribution is used where the parameter is beta-distributed to account for the individual specificities. The beta probability density function has the following form :

$$h_1(\theta_{i,1}; a_i, b) = \frac{\Gamma(a_i + b)}{\Gamma(a_i)\Gamma(b)} \theta_{i,1}^{a_i - 1} \left(1 - \theta_{i,1}\right)^{b-1}. \tag{4.2}$$

Under the notation that $K_{i,t} = I_{(n_{i,t}=0)}$ and using a Bernoulli-beta combination for the zero part, the joint distribution of the variable $K_{i,t}$, that can take values of 0 or 1, is well known and can be expressed as :

$$\begin{aligned} P(K_{i,1} = k_{i,1}, ..., K_{i,T} = k_{i,T} | a_i, b) &= \int \prod_{t=1}^{T} \theta_{i,1}^{k_{i,t}} \left(1 - \theta_{k,1}\right)^{1-k_{i,t}} \frac{\Gamma(a_i + b)}{\Gamma(a_i)\Gamma(b)} \theta_{i,1}^{a_i - 1} \left(1 - \theta_{i,1}\right)^{b-1} d\theta_{i,1} \\ &= \frac{\Gamma(a_i + b)}{\Gamma(a_i)\Gamma(b)} \frac{\Gamma(\sum_{t=1}^{T} k_{i,t} + a_i)\Gamma(T - \sum_{t=1}^{T} k_{i,t} + b)}{\Gamma(T + a_i + b)}, \end{aligned} \tag{4.3}$$

where the covariates can be included in the model as $a_i = \exp(x_i\beta)$. The transformation of the positive part means the data can be fitted using standard Poisson random effects models, where the mean variable can be expressed as $\gamma_i = \exp(x_i\delta)$. The generalization of the Poisson distribution for panel data can be performed with an individual random effect term $\theta_{i,2}$ that is added to its mean parameter. Formally, we can express the classic random effects Poisson model as :

$$N_{i,t}^{*} | \theta_i \sim Poisson(\gamma_i \theta_{i,2}), \quad i = 1, ..., N \quad t = 1, ..., J.,$$

where J is the total number of time periods during which at least one count was registered. For the heterogeneous models for cross-section data (Boucher et al. (2006c)), many possible distributions for the random effects can be chosen. The use of a gamma heterogeneity is a natural possibility because it can express the joint distribution in a closed form since the gamma is conjugated to the Poisson distribution. The density of the gamma distribution is defined as :

$$h_2(\theta_{i,2} | \beta, r) = \frac{\beta^r}{\Gamma(r)} \theta_{i,2}^{r-1} \exp(-\beta\theta_{i,2}). \tag{4.4}$$

When the heterogeneity term follows a Gamma distribution of mean 1 and variance α (i.e. both parameters are equal to $1/\alpha$) the joint distribution of the number of claims, is equal to (Hausman et al. (1984)) :

$$\begin{aligned} & P(N_{i,1}^* = n_{i,1}^*, ..., N_{i,J}^* = n_{i,J}^*) \\ = & \left(\prod_{t=1}^{J} \frac{(\gamma_i)^{n_{i,t}^*}}{n_{i,t}^*!} \right) \frac{\Gamma(\sum_{t=1}^{J} n_{i,t}^* + 1/\alpha)}{\Gamma(1/\alpha)} \left(\frac{1/\alpha}{J\gamma_i + 1/\alpha} \right)^{1/\alpha} (J\gamma_i + 1/\alpha)^{-\sum_{t=1}^{J} n_{i,t}^*} \end{aligned}$$

. (4.5)

This distribution is known as a multinomial-negative binomial or negative multinomial, which has often been applied. This distribution has the following moment : $E[N_{i,t}^*] = \gamma_i$ and $Var[N_{i,t}^*] = \gamma_i + \alpha\gamma_i^2$, so overdispersion can be accounted. Maximum likelihood estimations of the parameters and variances of these estimates are straightforward.

Note that all covariates included in the models are time independent since $\lambda_{i,t} = \lambda_i$, $\gamma_{i,t} = \gamma_i$, and $a_{i,t} = a_i$, for all $t = 1, ..., T$. This assumption is needed to obtain simple equations forms of the multivariate hurdle models, but generalization to include a time-variant is not difficult since the MVNB can directly support time-variant covariates and a probit transformation with Gaussian disturbance can be supposed for the first process as proposed by (Winkelmann (2003b)) for the correlated hurdle model for cross-section data.

The first moments of the hurdle for panel data, when conditioning on the random effects $\Theta_i = \{\theta_{1,i}, \theta_{2,i}\}$ can be expressed as :

$$\begin{aligned} E[N_{i,t}|\Theta_i] &= \theta_{i,1} \sum_{j=0}^{\infty} (1+j) f_2(j|\theta_{i,2}) \\ &= \theta_{i,1} + \gamma_i \theta_{i,1} \theta_{i,2} \qquad (4.6) \\ E[N_{i,t}^2|\Theta_i] &= \theta_{i,1} \sum_{j=0}^{\infty} (1+j)^2 f_2(j|\theta_{i,2}) \\ &= \theta_{i,1} \left(\gamma_i^2 \theta_{i,2}^2 + 3\gamma_i \theta_{i,2} + 1 \right) \qquad (4.7) \\ Var[N_{i,t}|\Theta_i] &= E[N_{i,t}^2|\Theta_i] - E[N_{i,t}|\Theta_i]^2 \\ &= \gamma_i^2 \theta_{i,1} \theta_{i,2}^2 + 3\gamma_i \theta_{i,1} \theta_{i,2} - 2\gamma_i \theta_{i,1}^2 \theta_{i,2} - \gamma_i^2 \theta_{i,1}^2 \theta_{i,2}^2 + \theta_{i,1} - \theta_{i,1}^2, \qquad (4.8) \end{aligned}$$

which gives in :

$$\begin{aligned} E[N_{i,t}] &= E[E[N_{i,t}|\Theta_i]] \\ &= E[\theta_{i,1}] + \gamma_i E[\theta_{i,1}\theta_{i,2}] \qquad (4.9) \\ Var[N_{i,t}] &= E[Var[N_{i,t}|\Theta_i]] + Var[E[N_{i,t}|\Theta_i]] \\ &= \gamma_i^2 E[\theta_{i,1}\theta_{i,2}^2] + E[\theta_{i,1}\theta_{i,2}]\left[3\gamma_i - 2\gamma_i E[\theta_{i,1}]\right] + E[\theta_{i,1}] - E[\theta_{i,1}]^2 - \gamma_i^2 E[\theta_{i,1}\theta_{i,2}]^2 \qquad (4.10) \end{aligned}$$

$$Cov[N_{i,t}, N_{i,t+j}] = Var[\theta_{i,1}] + 2\gamma_i \left(E[\theta_{i,1}^2\theta_{i,2}] - E[\theta_{i,1}\theta_{i,2}]E[\theta_{i,1}]\right) + \gamma_i^2 \left(E[\theta_{i,1}^2\theta_{i,2}^2] - E[\theta_{i,1}\theta_{i,2}]^2\right). \qquad (4.11)$$

4.1 Independent Copula

The first copula that can be used to model the dependence between the random effects is the product copula. This copula means that each random effect distribution is independent. Using this modeling with the equation (4.1) for the Bernoulli-beta and the Poisson-gamma combination leads to the following joint distribution :

$$\begin{aligned} P(N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}) &= \int \prod_{t=1}^{T} f_1(0|\theta_{i,1})^{k_{i,t}} \left(1 - f_1(0|\theta_{i,1})\right)^{1-k_{i,t}} g_1(\theta_{i,1}) d\theta_{i,1} \\ &\times \int \prod_{t=1}^{T} f_2(n_{i,t}^*|\theta_{i,2})^{1-k_{i,t}} g_2(\theta_{i,2}) d\theta_{i,2}. \qquad (4.12) \end{aligned}$$

The hurdle model for cross-section data, this hurdle model with independent random effects is interesting because it allows the parameters to be estimated in two separate steps. Using equations (4.9) to (4.11), the first moments of the hurdle model with independent random effects can be easily calculated from :

$$\begin{aligned} E[\theta_{i,1}] &= \frac{a_i}{a_i + b} \qquad (4.13) \\ E[\theta_{i,2}] &= 1 \qquad (4.14) \\ E[\theta_{i,1}^x \theta_{i,2}^y] &= E[\theta_{i,1}^x] E[\theta_{i,2}^y]. \qquad (4.15) \end{aligned}$$

Consequently :

$$E[N_{i,t}] = \frac{a_i}{a_i+b}(1+\gamma_i) \tag{4.16}$$

$$Var[N_{i,t}] = \frac{a_i}{a_i+b}\left[1+3\gamma_i+(\alpha+1)\gamma_i^2\right] - \left(\frac{a_i}{a_i+b}\right)^2(1+\gamma_i)^2 \tag{4.17}$$

$$Cov[N_{i,t}, N_{i,t+j}] = (1+2\gamma_i+\gamma_i^2(\alpha+1))\frac{a_i b}{(a_i+b+1)(a_i+b)^2} - \left(\frac{a_i}{a_i+b}\right)^2(1+\gamma_i)^2 \text{ j} > 0 \tag{4.18}$$

4.2 Gaussian Copula

In this subsection, we propose to use an other kind of copula for modeling the joint distribution of the random effects. We chose to model this dependence using a Gaussian copula which leads to the following expression for the joint distribution of the random effects :

$$g(\theta_{i,1}, \theta_{i,2}) = c(u,v) f_1(\theta_{i,1}) f_2(\theta_{i,2}), \tag{4.19}$$

with the Gaussian copula expressed as :

$$c^{Ga}(u,v) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\left(\frac{\rho^2\Phi^{-1}(u)^2 + \rho^2\Phi^{-1}(v)^2 - 2\rho\Phi^{-1}(u)\Phi^{-1}(v)}{1-\rho^2}\right)\right) \tag{4.20}$$

where Φ is the standard Normal distribution function, $u = F_1(\theta_{i,1})$ and $v = F_2(\theta_{i,2})$. The marginal density functions $f_1(\theta_{i,1})$ and $f_2(\theta_{i,2})$ are beta and gamma distributions, having the same parameters as the model with independent random effects.

Obviously, no closed form expression can be generated from this last model using equation (4.1) and an alternative modeling approach must be used, such as numerical integration techniques, Markov chain Monte Carlo methods or others. However, we used an other technique based on the NLMIXED procedure from the SAS System, as described by Nelson et al. (2006).

To approximate densities involving random effects, the NLMIXED procedure uses numerical evaluation techniques such as Gaussian quadrature. This technique is only available for normally distributed random effects. Nevertheless, nonnormal random effects can be accommodated within

the numerical integration techniques via the use of the probability integral transformation. Since the NLMIXED procedure allows for multivariate normal random effects, we can also extend the procedure for random effects linked by Gaussian copula. Formally, if we suppose that ϵ_1 and ϵ_2 are bivariate normal with means 0, unit variance and a correlation parameter ρ, we can state that $u_1 = \Phi(\epsilon_1)$ and $u_2 = \Phi(\epsilon_2)$ are both uniform linked with a Gaussian copula. Application of the probability integral transform to each variable leads to $\theta_1 = F_1^{-1}(u_1)$ which has density $f_1(\theta_1)$ and $\theta_2 = F_2^{-1}(u_2)$ having a density $f_2(\theta_2)$, both random variable θ_1 and θ_2 having a joint distribution modeled by a Gaussian copula. Numerical computations can be helpful in obtaining the mean and the variance of the model.

4.3 Frechet-Hoeffding Copula

Another popular copula that can be useful here is the Frechet-Hoeffding copula. It is the upper bound for all copulas as it represents perfect positive dependence between variates. The NLMIXED procedure mentioned in the previous subsection can also be used to evaluate the parameters of the model. In fact, the estimation of this model is a lot easier that for the previous one, since only one random effect ϵ must be used. Using the same probability integral transformation $u = \Phi(\epsilon)$, we can generate the beta and the gamma values as $\theta_1 = F_1^{-1}(u)$ and $\theta_2 = F_2^{-1}(u)$, where $F_1(u)$ is the Beta(a,b) distribution function and $F_2(u)$ is the Gamma($1/\alpha$,$1/\alpha$) distribution distribution. Moments must also be computed numerically.

5. Predictive Distribution

At each successive time period, the random effects $\theta_{1,i}$ and $\theta_{i,2}$ of the hurdle model for panel data, composed of a Poisson and a Bernoulli distributions, can be updated for past experience, revealing some individual information. Formally, the predictive distribution can be found using the following development :

$$
\begin{aligned}
& P(N_{i,T+1} = n_{i,T+1} | N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] \\
= & \frac{P(N_{i,1} = n_{i,1}, ..., N_{i,T+1} = n_{i,T+1})}{P(N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T})} \\
= & \int\int \Pr(N_{i,T+1} = n_{i,T+1} | \theta_{i,1}, \theta_{i,2}) \\
& \left(\frac{\Pr(N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T} | \theta_{i,1}, \theta_{i,2}) g(\theta_{i,1}, \theta_{i,2})}{\int\int \Pr(N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T} | \theta_{i,1}, \theta_{i,2}) g(\theta_{i,1}, \theta_{i,2}) d\theta_{i,1} d\theta_{i,2}} \right) d\theta_{i,1} d\theta_{i,2} \\
= & \int\int \Pr(N_{i,T+1} = n_{i,T+1} | \theta_{i,1}, \theta_{i,2}) \left(\frac{[\prod_t \Pr(N_{i,t} = n_{i,t} | \theta_{i,1}, \theta_{i,2})] \, g(\theta_{i,1}, \theta_{i,2}) d\theta_{i,1} d\theta_{i,2}}{\int [\prod_t \Pr(N_{i,t} = n_{i,t} | \theta_{i,1}, \theta_{i,2})] \, g(\theta_{i,1}, \theta_{i,2})} \right) d\theta_{i,1} d\theta_{i,2} \\
= & \int\int \Pr(N_{i,T+1} = n_{i,T+1} | \theta_{i,1}, \theta_{i,2}) g(\theta_{i,1}, \theta_{i,2} | n_{i,1}, ..., n_{i,T}) d\theta_{i,1} d\theta_{i,2}, \qquad (5.1)
\end{aligned}
$$

where $g(\theta_{i,1}, \theta_{i,2} | n_{i,1}, ..., n_{i,T})$ is the joint posterior distribution of the random effects $\theta_{i,1}, \theta_{i,2}$, reflecting the past experience of an individual unit i. If this a posteriori distribution can be expressed in closed form, moments of the predictive distribution can easily be found by conditioning on the random effects $\theta_{i,1}$ and $\theta_{i,2}$.

5.1 Independent Random Effects

As shown by equation (4.12), each process of the hurdle model for panel data with independent random effects can be expressed separately. Consequently, the two processes can also be analyzed separately. Following the developments in the last subsection, the *a posteriori* distribution of the random effect term of the first process can be found as follows :

$$
\begin{aligned}
g(\theta_{i,1} | k_{i,1}, ..., k_{i,T}) & \propto \theta_{i,1}^{\sum_t^T k_{i,t}} (1 - \theta_{i,1})^{T - \sum_t^T k_{i,t}} \frac{\Gamma(a_i + b)}{\Gamma(a_i)\Gamma(b)} \theta_{i,1}^{a-1} (1 - \theta_{i,1})^{b-1} \\
& \propto \theta_{i,1}^{\sum_t^T k_{i,t} + a_i - 1} (1 - \theta_{i,1})^{T - \sum_t^T k_{i,t} + b - 1},
\end{aligned}
$$

from which, we can see that the posterior distribution is beta with parameters $\sum_t^T k_{i,t} + a_i$ and $T - \sum_t^T k_{i,t} + b$. For the first part of the distribution, the expected value of the predicted distribution is then calculated as :

$$
E[\theta_{i,1} | k_{i,1}, ..., k_{i,T}] = \frac{\sum_t^T k_{i,t} + a_i}{T + b + a_i}. \qquad (5.2)
$$

The second process $N_i^* = N_i - 1$ of the hurdle distribution is to follow a Poisson distribution with random effects. When found that (as is well known) the *a posteriori* distribution of the random

effect term is also a gamma distribution having parameters equal to $T\lambda_i + 1/\alpha$ and $\sum_t^T n_{i,t} + 1/\alpha$. The expected value of the posterior distributions is equal to :

$$E[\theta_{i,2}|n_{i,1},...,n_{i,T}] = \frac{\sum_t^J n_{i,t}^* + 1/\alpha}{J\gamma_i + 1/\alpha}, \tag{5.3}$$

where J is equal to the number of periods in which at least one event occurred. Using these last equations and those shown in (4.9) to (4.11) gives the first moments of the predictive count distribution. For example, the expected value of the distribution, having past experience $1, 2, ..., T$ is equal to :

$$\begin{aligned} E[N_{i,T+1}|n_{i,1},...,n_{i,T}] &= \frac{\sum_t^T k_{i,t} + a_i}{T + b + a_i}\left(1 + \gamma_i \frac{\sum_t^J n_{i,t}^* + 1/\alpha}{J\gamma_i + 1/\alpha}\right) \\ &= \left(\frac{J\gamma_i + a_i\gamma_i}{J\gamma_i + 1/\alpha}\right)\left(\frac{\sum_t^T n_{i,t} + (1+\gamma_i)/\alpha}{T + b + a_i}\right). \end{aligned} \tag{5.4}$$

Unlike the standard Poisson-gamma models, expressed in equation (5.3), the predictive mean not only depends on the sum of past counts, but also on the number of periods without a registered count. Moreover, note that when the observed periods become infinite, the expected value of an individual unit converge to the product of two limited values : the average number of time periods with at least one count and the average of the number of counts greater than or equal to one, which is equal to the average number of reported claims, as expressed in equation (5.4).

5.2 Gaussian Copula

Exact predictive and posterior distributions for the random effects can only be expressed in closed form for some specific distributions. For other models, such as the correlated random effects hurdle models, these distributions cannot be evaluated analytically. Consequently, a possible approach for evaluating these predictive distributions is the use of numerical computations or simulations. In our case, we used Markov chain Monte Carlo (MCMC) simulations to compute posterior and predictive distributions.

5.2.1 Overview of Markov Chain Monte Carlo Simulations

The idea of MCMC simulations is to simulate realizations from a Markov chain that converges to the joint distribution of the random effects. The resulting random draws are no longer independent,

but under mild regularity conditions (as described in the Appendix of Smith and Roberts (1993), for example), the value of the draw tends in distribution to that of a random draw from the joint distribution as the number of draws becomes moderately large. See Scollnik (2001) for an introduction to MCMC simulations in actuarial sciences.

5.2.2 MCMC Algorithm

As expressed in the development of equation (5.1), the posterior distribution of the random effects can be expressed as :

$$\begin{aligned} g(\theta_{i,1}, \theta_{i,2}|n_{i,1}, ..., n_{i,T}) &\propto g(n_{i,1}, ..., n_{i,T}|\theta_{i,1}, \theta_{i,2}) g(\theta_{i,1}, \theta_{i,2}) \\ &\propto c^{Ga}(F_1(\theta_{i,1}), F_2(\theta_{i,2})) \times \theta_{i,1}^{A-1}(1-\theta_{i,1})^{B-1}\theta_{i,2}^{C-1}\exp(-\theta_{i,2}D) \end{aligned} \tag{5.5}$$

where :

$$\begin{aligned} A &= \sum_t^T k_{i,t} + a \\ B &= T - \sum_t^T k_{i,t} + b \\ C &= \sum_t^T (n_{i,t} - k_{i,t}) + 1/\alpha \\ D &= \lambda_i \sum_t^T k_{i,t} + 1/\alpha \end{aligned}$$

Our strategy is to simulate values of $\theta_{i,1}$ and $\theta_{i,2}$ that approximate the expression (5.5) using the Metropolis-Hasting algorithm. The most obvious choice for simulated realizations of these two random variables is to simulate independent beta and gamma distributions with the parameters A,B and C,D, respectively.

The (j+1) iteration of the Metropolis-Hasting algorithm is as follows :

1. Given $\theta_{i,1}^{(j)}, \theta_{i,2}^{(j)}$, simulate $\hat{\theta}_{i,1}, \hat{\theta}_{i,2} \sim g(x, y|\theta_{i,1}^{(j)}, \theta_{i,2}^{(j)})$;
2. Specifying the distribution of interest by :

$$\pi(X_1, X_2) = c^{Ga}(F_1(X_1), F_2(X_2)) X_1^{A-1}(1-X_1)^{B-1} X_2^{C-1} \exp(-X_2 D), \tag{5.6}$$

compute the ratio :

$$P = \frac{\pi(\hat{\theta}_{i,1}, \hat{\theta}_{i,2})}{\pi(\theta_{i,1}^{(j)}, \theta_{i,2}^{(j)})} \frac{g(\theta_{i,1}^{(j)}, \theta_{i,2}^{(j)} | \hat{\theta}_{i,1}, \hat{\theta}_{i,2})}{g(\hat{\theta}_{i,1}, \hat{\theta}_{i,2} | \theta_{i,1}^{(j)}, \theta_{i,2}^{(j)})} \tag{5.7}$$

3. Simulate $U \sim U(0,1)$
4. Take :

$$\theta_{i,1}^{(j+1)}, \theta_{i,2}^{(j+1)} = \begin{cases} \hat{\theta}_{i,1}, \hat{\theta}_{i,2} & \text{if } U \leq P \\ \theta_{i,1}^{(j)}, \theta_{i,2}^{(j)} & \text{otherwise} \end{cases} \tag{5.8}$$

This algorithm is then updated a large number of times, and following given minimal regulatory conditions, it will generate a Markov chain with distribution (5.5) as a stationary distribution. Since the first simulated values depend on the first input $\theta_{i,1}^{(1)}, \theta_{i,2}^{(1)}$, it is convenient to *discard* the first simulated values, which is frequently referred to as burn-in. Additionally, to avoid the correlation between consecutive results, it is advisable to *thin* the sequence by keeping every k^{th} simulation draw from each sequence.

5.2.3 Monitoring Convergence

Once the discarding and thinning have been carried out, the convergence of the n simulated values of each m chain can be monitored. Graphical analysis comparing each simulated chains is a common alternative. The expected value, the variance and the median of each simulated chain can also be checked. An other way to monitor the convergence is to compute the following statistics :

$$B = \frac{n}{m-1} \sum_{j=1}^{m} (\bar{\psi}_{.j} - \bar{\psi}_{..})^2, \quad W = \frac{1}{n} \sum_{j=1}^{m} s_j^2 \tag{5.9}$$

where

$$\bar{\psi}_{.j} = \frac{1}{n}\sum_{i=1}^{n}\psi_{ij}, \quad \bar{\psi}_{..} = \frac{1}{m}\sum_{j=1}^{m}\psi_{.j} \text{ and } s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\psi_{ij} - \bar{\psi}_{.j})^2. \tag{5.10}$$

The estimated statistics B and W can be seen as the between and the within-sequence variances. We can estimate the variance of the marginal posterior distribution by a weighted average of B and W, such as :

$$\hat{Var}(\psi) = \frac{n-1}{n}W + \frac{1}{n}B \tag{5.11}$$

The estimated values W and $\hat{Var}(\psi)$ can be seen respectively as the lower and the upper bound of the real variance of ψ. In the limit, both W and $\hat{Var}(\psi)$ approach $Var(\psi)$, but from opposite directions. A ratio between the upper and the lower bound of the variance can be computed, which will estimate the factor by which the upper bound might be reduced. This ratio is called the estimated potential scale reduction :

$$\sqrt{\hat{R}} = \sqrt{\frac{\hat{Var}(\psi)}{W}} \tag{5.12}$$

If $\hat{R}$ is near 1 (less than 1.1 seems to be generally accepted) for all scalar estimands of interest, the convergence is accepted and the chain of simulations can be seen as the required *a posteriori* distribution.

5.3 Frechet-Hoeffding Copula

The posterior distribution of the random effects coming from a Frechet-Hoeffding copula must also be computed using MCMC simulation. Using the same notation as the Gaussian copula, it can be expressed as :

$$g(\theta_i|n_{i,1},...,n_{i,T}) \quad \propto \quad (F_1^{-1}(u))^{A-1}(1-F_1^{-1}(u))^{B-1}(F_2^{-1}(u))^{C-1}\exp(-F_1^{-1}(u)D) \tag{5.13}$$

where u is a uniform distribution on $[0,1]$, $F_1(u)$ is a Beta(a,b) cumulative distribution and $F_2(u)$ is a Gamma($1/\alpha$,$1/\alpha$) cumulative distribution. The MCMC algorithm and the monitoring of the convergence are the same as the ones explained in section 5.2.3.

6. Numerical Application

The hurdle models are relatively popular in the modeling of health care demands. Indeed, it is generally accepted that the demand for certain types of health care services depend on two processes : the decisions of the individual and thoses taken by the health care provider (Stoddart and Barer (1981), Pohlmeier and Ulrich (1995), Mullahy (1998), Santos Silva and Windmeijer (2001)). Consequently, conditionally on certain assumptions, the use of a hurdle model is intuitive and the parameters can have a structural interpretation.

We propose to use the hurdle model for panel data to an insurance portfolio. Using the same notation as in the preceding sections, N represents the number of insureds while T is the total number of insured periods by insured. The hurdle model is suitable to model these kind of data since, as shown in Boucher et al. (2006c), the majority of insureds reports less than 2 claims per year. This kind of modeling appears to be as the double modeling of the amount of claims, where the limited costs and the excess-of-loss costs (costs above a certain value) are modelled separately. Many direct applications of this model can be used for insurance data, such as the number of coverage implied for a claim, the number of injured peoples or the number of third parties involved. Such modeling using separate processes is appropriate since these processes are not the same as the one that drives the number of accident.

However, besides these specific examples, this type of modeling can also have a natural interpretation for the number of reported claims. The reluctance of some insureds to report their accident (since they would lose their good bonus-malus scheme) can supports the use of an hurdle model for the number of reported claims. Indeed, we can suppose that the behaviour of the insureds is not same when they have already reported a claim, which confirms the hypothesis of the two processes that determine the total number of claims.

6.1 Assumptions

As seen in the presentation of the hurdle model for panel data, covariates are time independent, that is to say that they do not change over all the observations of an individual unit. This assumption was also made in by Gourieroux (1999) for insurance data. Even if this assumption causes additional heterogeneity in the data, and then some anti-selection, it is not as restricting as it might seem. Indeed, the majority of the time dependent covariates that are used in insurance involves the age of the insured or the length of time spent in the company. These variables do not

evolve randomly in time since the change can already be known in advance. Consequently, the estimated coefficient of a parameter can still be interpretable since the evolution of the parameter is known. Other covariates can also change in time, such as the city or the type of vehicle belonging to the driver but this kind of major change often involves the creation of another policy.

Another required assumption was linked to the transformation of the cross-section data model into a panel data model. No correlation between the regressors and the random effect term was permitted. In linear regression, this correlation leads to an inconsistency of the estimated parameters (Mundlak (1978) or Hsiao (2003) for a general review). The same problem exists for the count data regression when $E[\theta|x_i] \neq E[\theta]$ (Mullahy (1997)) which leads to biased estimates of parameters. In insurance, a correlation between regressors and the error term is often present (Boucher and Denuit (2006)) and may be caused by omitted variables that are correlated with those included in the model.

As noted in Winkelmann (2003a), consistent estimates may be found if corrections are made to the standard estimation procedures. However, as shown by Boucher and Denuit (2006), for insurance applications, standard methods of estimation can still be used. Indeed, the resulting parameter estimates, althought biased, represent the apparent effect on the frequency of claims, which is the main point of interest when the omitted correlated variables cannot be used in classification.

6.2 Data used

We worked with a sample of the automobile portfolio of a major company operating in Spain. Only cars for private use were considered in this sample. The panel data contains information for the period from 1991 to 1998. Our sample contains 15,179 policyholders that remain in the company for seven complete periods, resulting in 106,253 insurance contracts. We have 5 exogeneous variables that are kept in the panel plus the annual number of accidents. For every policy of a single insured we used the initial information contained in his first contract. The total number of at-fault claims that took place within each year-long period was used for analysis.

The following table contains the observed annual frequency of at-fault claims from the first to the seventh period, together with the maximum number of claims per policyholder. The average claim frequency is 6.8412%.

The exogenous variables are defined in the next table :

Table 1 : Frequency of claims

Period	Frequency	Maximum
1	7.5367%	3
2	6.9569%	3
3	6.4035%	3
4	6.1466%	3
5	6.4695%	4
6	6.8450%	4
7	7.5301%	3
Total	6.8412%	4

Table 2 : Exogenous variables

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

In this paper, these exogenous variables are used to model the distribution parameters. The characteristics of the insureds are expressed through the functions $h(\beta_0 + x_i'\beta)$, where β_0 is the intercept and $\beta' = (\beta_1, ..., \beta_p)$ is a vector of regression parameters for explanatory variables $x_i = (x_{i,1}, ..., x_{i,p})$.

6.3 A Priori Analysis

6.3.1 Estimated Parameters

Estimated parameters for the hurdle model with independent random effects are shown in table 6.1. For ease of comparison, we also included the standard Poisson distribution with gamma random effects which is the most popular model used to describe the number of reported claims. As expected, an analysis of estimations leads to the conclusion that women report less than men, while new insureds in the company seem to have a worse loss experience than older clients. In the presence of other covariates, we can also see that young drivers exhibit a better claim experience, although it is not statistically significant. To finish, we observe that insureds with powerful vehicles show more insured periods with reported accidents than other drivers. It is important to notice that the parameters obtained cannot be compared directly to those obtained with the Poisson distribution since they do not have the same impact on the expected value. However, by a look on the sign of the estimated parameters of the zero-part of the hurdle model and the Poisson distribution, we can draw tha same conclusion for both models.

The parameters from the positive part allow us to observe what is the kind of insureds that are most likely to report a high number of claims in a single time period. The results are a little unusual since only those insureds that had been policyholder with the company for between 3 and 5 years at the beginning of the study are statistically better than the others.

We can see the estimated parameters of the hurdle model with correlated random effects in the table 6.2. Estimates of the parameters are very close to those obtained with the independent random effects model. More variation occur for the second process, where the intercept shows is lesser. Additionally, note that the second process exhibits higher overdispersion since its α value is greater than the one obtained with independent processes. All those differences come from the correlation added to the model. The reason why only the second process seems to be affected by the correlated random effects comes from the composition of the portfolio. Indeed, for all insureds who did not report a claim (approximately 66% of the portfolio), the model does not need to use

Parameter	Poisson-Gamma	hurdle Parts: Zero	hurdle Parts: Positive
b0	−2.6600 (0.0352)	0.3066 (0.0682)	−2.3688 (0.0525)
b1	0.1087 (0.0409)	0.1298 (0.0401)	. .
b2	−0.1805 (0.0327)	−0.1726 (0.0325)	−0.1951 (0.0933)
b3	−0.2103 (0.0370)	−0.2124 (0.0370)	. .
b4	0.0471 (0.0346)	0.0616 (0.0341)	. .
b5	0.0990 (0.0316)	0.0975 (0.0315)	. .
α / b	0.8832 (0.0432)	19.9640 (1.2279)	0.8122 (0.2070)
Loglike.	-26,702.98	-24,637.90	-2,050.80

Tab. 6.1 – Estimated Parameters for the hurdle Model with Independent Random Effects

the second random effects and its associated correlation.

The model shows a significant value of 0.8424 for the parameter ρ associated with the Gaussian copula. It clearly shows that the insureds who report in many insured periods also experience time periods with a large number of reported claims. Nevertheless, it is important to understand what is the meaning of the parameter ρ of a Gaussian copula with nonnormal marginals. As opposed to other tools that measure dependence, the parameter ρ depends on the marginal density. It means that the *real* linear correlation between the Beta and the Gamma distribution is not equal to ρ (in our data, it is equal to 0.8121). However, the Kendall' tau, usually denoted by τ, is a constant of the copula. Consequently, any correlated variates with the same copula will have the τ of that copula. For a Gaussian copula, the Kendall's tau is equal to $2\arcsin(\rho)/\pi = 0.6377$.

The model with perfect correlation between the random effects shows approximately the same parameters values as the one with Gaussian copula. It is not surprising since the Gaussian copula model exhibits a strong dependence between its random effects.

6.3.2 **Premiums**

The differences between models can be analyzed through the mean and the variance of some insured profiles. Several profiles were selected and are described in Table 6.3. The first profile is

Parameter	Gaussian Copula		Frechet-Hoeffding Copula	
	Zero	Positive	Zero	Positive
b0	0.3043 (0.0531)	−2.9100 (0.0882)	0.3104 (0.0429)	−2.9622 (0.0788)
b1	0.1362 (0.0416)	. .	0.1368 (0.0399)	. .
b2	−0.1703 (0.0323)	−0.2278 (0.0939)	−0.1702 (0.0326)	−0.2330 (0.0927)
b3	−0.2075 (0.0368)	. .	−0.2069 (0.0366)	. .
b4	0.0600 (0.0440)	. .	0.0597 (0.0346)	. .
b5	0.0966 (0.0315)	. .	0.0964 (0.0315)	. .
b,α	19.9568 (1.0844)	0.7533 (0.2038)	20.0836 (0.5725)	0.8818 (0.2442)
ρ	0.8424 (0.1165)	. .	. .	. .
Loglike.	-26,662.47		-26,663.28	

Tab. 6.2 – Estimated Parameters for the Hurdle Model with Correlated Random Effects

Profile Number	Kind of Profile	v1	v2	v3	v4	v5
1	Good	0	0	1	0	0
2	Average	1	1	0	0	0
4	Bad	1	0	0	1	1

Tab. 6.3 – Profiles analyzed

classified as a good driver, while the last one usually exhibits bad loss experience. The other profile is medium risk. The results are given in Table 6.4. This table shows that the expected values of all profiles are fairly similar for the four models studied. However, note the independent random effects model seems to underestimate low risk profile when compared to the correlated hurdle models. The greatest differences between models can be found in the variance estimates since the Poisson-gamma model exhibits lower variances than the other models. Additionnaly, for low risk profile, we can note that the independent random effects model shows variance values that are lower than the ones obtained for the correlated random effects models.

Models	1^{st} Profile Mean	1^{st} Profile Variance	2^{nd} Profile Mean	2^{nd} Profile Variance	3^{rd} Profile Mean	3^{rd} Profile Variance
Poisson-Gamma	0.0567	0.0595	0.0651	0.0688	0.0902	0.0974
H - Independent	0.0570	0.0644	0.0659	0.0717	0.0911	0.0997
H - Gaussian	0.0577	0.0657	0.0664	0.0721	0.0911	0.0989
H - Frechet-H.	0.0576	0.0655	0.0664	0.0719	0.0909	0.0985

TAB. 6.4 – A priori Premiums

6.4 Predictive distributions

6.4.1 Markov Chain Monte-Carlo Simulations

Predictive distribution and corresponding moments can be found directly for the independent hurdle model. However, as described before, this distribution for the correlated random effects using a Gaussian copula cannot be found analytically and Markov chain Monte Carlo simulations must be performed. Following the algorithm of section (5.2.2), we were able to find a sample of the posterior distribution of the random effects, from which predictive distribution of the $N_{i,T+j}, j = 1, 2, ...$ can be approximated.

Two jump functions $g(x, y|\theta_{i,1}^{(j)}, \theta_{i,2}^{(j)})$ were tested for our model :

1. An independent Metropolis-Hastings, where $\hat{\theta_{i,1}}$ and $\hat{\theta_{i,1}}$ are distributed as the product of a Beta(A, B) and a Gamma(C, D). In this situation, the values generated by the proposal distribution (jump function) do not depend on the past realizations. Using this jump function, the acceptance probability can be simplified as :

$$P = \frac{c^{Ga}(F_1(\hat{\theta}_{i,1}), F_2(\hat{\theta}_{i,2}))}{c^{Ga}(F_1(\theta_{i,1}^{(j)}), F_2(\theta_{i,2}^{(j)}))} \tag{6.1}$$

   where (as we recall) F_1 is Beta distributed with parameters a and b, while F_2 follows a distribution gamma($1/\alpha$,$1/\alpha$).

2. A second jump function was used where $\hat{\theta_{i,1}}$ and $\hat{\theta_{i,1}}$ are distributed as the product of a beta and a gamma distributions, having respective means equal to $\theta_{i,1}^{(j)}$ and $\theta_{i,2}^{(j)}$. The second terms

of the distributions, needed to compute the second moment, were set to $T - \sum_t^T k_{i,t} + b$ for the beta distribution and to $1/\alpha$ for the gamma random effects.

The same process was used for the hurdle model with a Frechet-Hoeffding copula.

6.4.2 Predictive Premiums

Table 6.5 shows the predictive premiums for a medium risk profile. It depends on the sum of reported claims and on the number of insured periods with at least one reported claim $(T - T_0)$. To illustrate this, we selected a loss experience of 10 years although other situations can easily be illustrated for the independent random effects since the equation (5.4) is simple to use. For the correlated random effects model, the first jump process was used to compute posterior distributions that are not too far from the prior, while the extreme situations (10 reported claims, 10 insured periods with at least one claim, etc.) required the use of the second jump process.

We simulated 5 chains of 500,000, selected a burn-in value of 10,000 draws for our MCMC simulations and a lag of 10 to eliminate the possible correlation between successive draws. Results of these MCMC simulations and results found in analytic computations for the Poisson-gamma and the independent random effects models are described in the following tables :

Interesting conclusions can be drawn from the analysis of predictive premiums. Indeed, for the independent random effects, we can see that the number of insured periods with a claim have a greater impact on the premium for the following year than the total number of reported claims. For insureds who reported one claim or less, the hurdle model offer a discount that is smaller than the MVNB model. For insureds making a higher number fo claimq, we can see that the hurdle model contains a wide panel of premium values that range from 0.3 to 1.8 times the MVNB premiums.

However, we also observe that the dependence between the two random effects has a great impact on hte premium. The correlated random effects hurdle models show premium values that are closer to those of the Poisson-gamma rather than the independent random effects model, since the impact of the pattern of reporting is reduced. However, except for the claim-free insurer, predictive premiums are shown to be lower in the Gaussian model when compared to the MVNB model.

For the Frechet-Hoeffding copula, we can see that the impact of the number of reported claims compared to the number of insured periods with a claim is not the same as in the independent of the Gaussian copula models. Indeed, it is the exact opposite : the model applies harshner penalties

Models	$T - T_0$	A priori	Sum of claims 0	1	2	3	4	10
MVNB	.	0.0651	0.0413	0.0778	0.1143	0.1509	0.1874	0.4064
Independent	0	0.0659	0.0448	.	.	.	.	.
	1	0.0659	.	0.0790	0.0833	0.0876	0.0920	0.1180
	2	0.0659	.	.	0.1128	0.1187	0.1246	0.1598
	3	0.0659	.	.	.	0.1465	0.1538	0.1972
	4	0.0659	.	.	.	.	0.1800	0.2309
	10	0.0659	.	.	.	.	.	0.3786
Gaussian	0	0.0663	0.0441	.	.	.	.	.
	1	0.0663	.	0.0776	0.1063	0.1336	0.1598	0.3217
	2	0.0663	.	.	0.1113	0.1403	0.1687	0.3341
	3	0.0663	.	.	.	0.1444	0.1750	0.3442
	4	0.0663	.	.	.	.	0.1774	0.3529
	10	0.0663	.	.	.	.	.	0.3770
Frechet-Hoef.	0	0.0663	0.0441	.	.	.	.	.
	1	0.0663	.	0.0775	0.1137	0.1500	0.1861	0.4044
	2	0.0663	.	.	0.1106	0.1469	0.1826	0.4006
	3	0.0663	.	.	.	0.1438	0.1795	0.3965
	4	0.0663	.	.	.	.	0.1764	0.3928
	10	0.0663	.	.	.	.	.	0.3704

Tab. 6.5 – Predictive Premiums

to insureds with more reported claims in the same time period.

Our model provides a strong dependence between the random effects (Kendall's Tau of 0.6377), which can explain why the correlated model is so close to the MVNB model. However, for a less correlated model, we can suppose that the expected predicted values of the correlated model can be more associated with those of the independent random effects model.

Analysis of the variance of these predictive premiums has also been carried out. The results are shown in the table below :

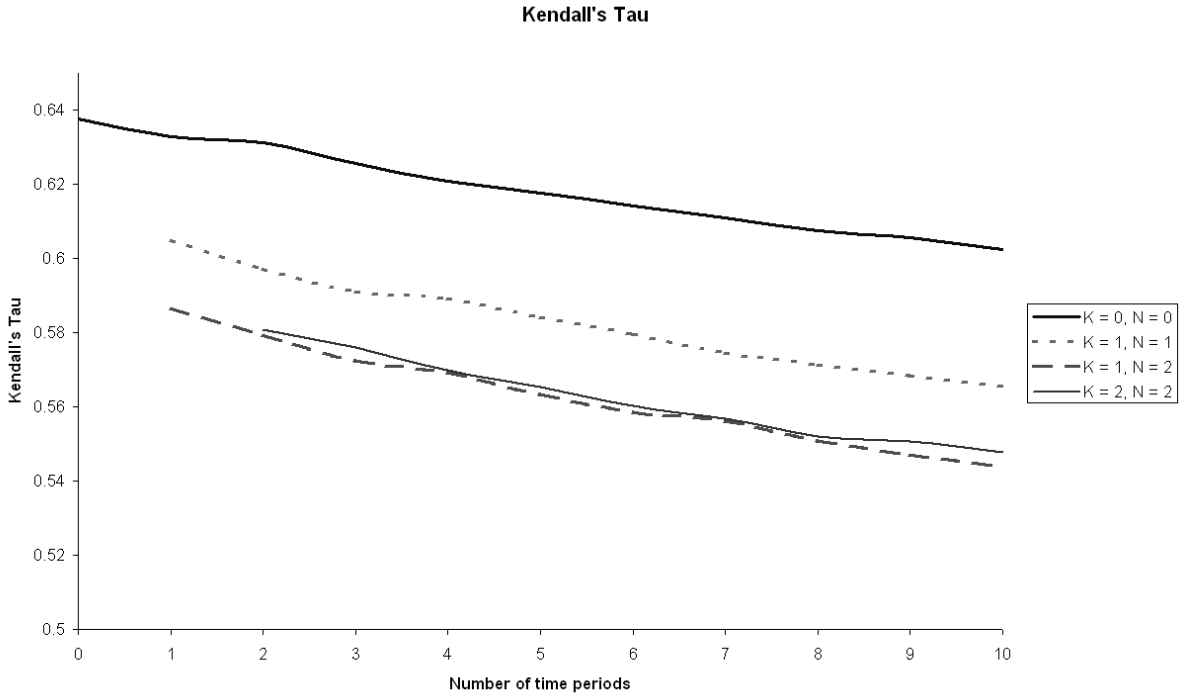
Large differences between the independent random effects model and the correlated models can also be observed. Indeed, in many situations, the independent random effects model significantly underestimates the value of the variance when compared to the other models. However, in contrast to the expected values, the correlated hurdle models do not show close similarities with the MVNB model for variance values. The difference is greatest for insureds who report a lot of claims.

6.4.3 **Dependence between Random Effects**

An interesting analysis can be performed for the correlated random effects model. We compute the Kendall's Tau for a number of selected experiences to check the behaviour of the dependence between θ_1 and θ_2. The following graph illustrates the Kendall's Tau for the medium profiles, depending on the pattern of reporting claims :

Models	$T - T_0$	A priori	Sum of claims 0	1	2	3	4	10
MVNB	.	0.0688	0.0428	0.0807	0.1185	0.1564	0.1942	0.4212
Independent	0	0.0717	0.0497	.	.	.	.	.
	1	0.0717	.	0.0841	0.0975	0.1114	0.1258	0.2220
	2	0.0717	.	.	0.1155	0.1332	0.1515	0.2735
	3	0.0717	.	.	.	0.1439	0.1652	0.3066
	4	0.0717	.	.	.	.	0.1697	0.3256
	10	0.0717	.	.	.	.	.	0.2710
Gaussian	0	0.0720	0.0465	.	.	.	.	.
	1	0.0720	.	0.0818	0.1139	0.1458	0.1771	0.3862
	2	0.0720	.	.	0.1169	0.1498	0.1831	0.3924
	3	0.0720	.	.	.	0.1509	0.1858	0.3953
	4	0.0720	.	.	.	.	0.1840	0.3966
	10	0.0720	.	.	.	.	.	0.3694
Frechet-Hoef.	0	0.0718	0.0463	.	.	.	.	.
	1	0.0718	.	0.0820	0.1213	0.1610	0.2011	0.4469
	2	0.0718	.	.	0.1179	0.1577	0.1973	0.4425
	3	0.0718	.	.	.	0.1543	0.1938	0.4379
	4	0.0718	.	.	.	.	0.1903	0.4338
	10	0.0718	.	.	.	.	.	0.4083

Tab. 6.6 – Variance



where K and N express the number of time periods with at least one claim and the total number of reported claims, respectively. The dependence between the random effects vanishes when the number of insured periods increases. This was expected since the increase of time periods reveals the unobserved characteristics of the risk for each process. It is interesting is to observe that the decrease of the Kendall's Tau seems to be linear in time, for all patterns of reporting. Additionally, we can see that the gradient is also the same since each line seems to be parallel.

The graphic illustrates the effect on the dependence between the random effects when an insured reports a claim. The Kendall's Tau decreases a lot more for the first claim reported than for subsequent claims. The dependence seems to be affected more by the number of reported claims in one time period that the number of insured periods in which a claim is reported.

6.5 Models Comparison

On the models seen above, some are equivalent for certain parameters restrictions. This can be tested using a classical hypothesis with two standard tests : the log-likelihood ratio and the Wald tests. These tests are asymptotically equivalent. In some cases, the parameter restriction corresponds to the boundary of the parameter space. One problem with the Wald or the log-likelihood

ratio tests occurs 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 performed. Results from Chernoff (1954) for the likelihood ratio statistic and Moran (1971) for the Wald Test 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, the usual one-sided test critical value of $z_{1-\partial}$ is still used.

The correlated hurdle model with the Frechet-Hoeffding copula and the independent hurdle model are nested to the correlated hurdle model with the Gaussian copula. Indeed, for $\rho = 0$ we obtain the independent copula while a value $\rho = 1$ gives the Frechet-Hoeffding copula. This can be tested using the Wald or the log-likelihood ratio test, using the precautions noted above for the Frechet-Hoeffding copula. The data indicates that the independent copula assumption is rejected against the Gaussian copula (*p-value* ≤ 0.001), while no statistical difference is shown between the Frechet-Hoeffding copula and the Gaussian copula (*p-value* varying between 32.74% and 8.81%).

Some remaining models could not be compared directly since they would be non-nested models. Two kinds of non-nested models exist : overlapping and strictly non-nested models. Overlapping models are those where neither model can be derived form the other parameter restriction, but in which some joint parameter restrictions result in the same model. A standard method of comparing non-nested models is through the information criteria, such as the Akaike information Criteria $(AIC) = -2\log(L) + 2k$ where k is the number of parameters in the model. A rule of the thumb states that a difference greater than 10 indicates a significant difference between models Burnham and Anderson (2002). In our numerical example, the remaining model is the hurdle model with the Frechet-Hoeffding copula. For the prupose of illustration, this distribution can be tested against the standard MVNB (Poisson-gamma) model with the information criteria (AIC). The MVNB model gives a AIC value of 53,419.96 and the hurdle model has a value of 53,346.44. Because the difference between values is greater than 10, it indicates a significant difference between the models and we favour the hurdle distribution.

Formally, for independent observations, a log-likelihood ratio test for non-nested models, de-

veloped by Vuong (1989) and generalized by Rivers and Vuong (2002) can be used to determine whether the hurdle model with the Frechet-Hoeffding copula model is statistically better than the other models. This test is based on the likelihood ratio approach of Cox (1961), with a correction applied to the standard deviation of this difference. As proposed by Golden (2003), this non-nested models test can be generalized fairly directly to correlated observations. Boucher et al. (2006b) and Boucher et al. (2006d) applied this test directly on panel data models. Because complex intermediary steps are needed when using this statistical test (gradient evaluation, autocorrelation check, etc.), we chose to base our results on AIC analysis only.

7. Conclusion

Even if the *a priori* and the predictive premiums of the correlated hurdle model are fairly close to the those obtained with a standard Poisson with gamma random effects, the estimated variances, the predicted premiums and the predictive variances are not same. This can be very important in some analyses, for example when estimating the minimal capital needed in insurance. We saw that the dependence between the heterogeneities of the process decreases linearly in time, which can be particularly interesting in the analysis of missing informations.

Since we observed that the correlation existing between the two processes of an hurdle count distribution must be considered, disregarding this dependence can lead to serious underestimations of the risk. The same conclusion can be made for the analysis of multivariate count distributions, such as the analysis of two lines of business in insurance. Further analysis of the credibility estimators indicated by the hurdle model would be an interesting area of study.

Chapitre 8

Zero-Inflated Poisson Models for Panel Data

Soumis pour publication à *Journal of Applied Statistics*
en collaboration avec Michel Denuit et Montserrat Guillén

1. Introduction

In various count applications, the data exhibit a high number of zero values. This led the idea that a distribution with excess zeros can provide a good fit, such as the zero-inflated distribution. We propose a new zero-inflated model that generalizes the model for longitudinal data. Longitudinal data (or panel data) consists of repeated observations of individual units over time. Each individual is assumed to be independent but correlation between observation of the same individual is permitted. This fact must be regarded as an advantage and should be used in the construction of count distributions.

An overview of the zero-inflated model and the bases for the construction of panel data distributions are given in the second part of the paper. In the third section, we show various types of parametrizations of zero-inflated models for panel data. Random effects are added to the zero-inflated term and the count distribution. We show that the correlation between these random effects is of no use since the construction of the distribution already assumes an interaction between the distribution processes. Another kind of multivariate distribution, comprising a special degenerated random effects distribution is shown to have interesting properties. In section 4, we show that predictive distributions and their expected predictive value can be expressed in a closed form for all proposed distributions.

Section 5 contains a numerical illustration performed of the reported claims in a sample from a Spanish insurance company to support the discussion. We show that the expected value, the variance and these moments for the predictive distribution can differ significantly according to the model. Statistical tests to compare models are explained and a Vuong-Golden test is used to compare the fitting of non-nested models. Our results show that some of the models presented here have a better fit than the commonly used Poisson distribution with gamma random effects.

2. Overview

2.1 Zero-Inflated Distribution

A high number of zero-values are often observed in the fitting of count data. This leads to the idea that a distribution with excess zeros can provide a good fit. The idea of this model is to use a finite mixture models of two distributions combining an indicator distribution for the zero case and a standard count distribution (Mullahy (1986), Lambert (1992)). As a result, this distribution

will account for the excess zeros of the empirical distribution. When applied to cross-section data, the density of this type of model, with $0 < \phi < 1$, can be expressed as :

$$Pr[N = n] = \begin{cases} \phi + (1 - \phi)Pr[K = 0] & \text{for } n = 0 \\ (1 - \phi)Pr[K = n] & \text{for } n = 1, 2, ... \end{cases}, \tag{2.1}$$

where the random variable K follows the standard distribution to be modified by an additional excess zero function. In the limiting case, where $\phi \to 0$, the zero-inflated model correspond to the distribution of K.

Many distributions may be used with the zero-inflated models. Obviously, the classic distribution to be used is the zero-inflated Poisson (ZIP) distribution. With the use of the equation (2.1), the density of the ZIP model is :

$$Pr[N = n] = \begin{cases} \phi + (1 - \phi)e^{-\lambda} & \text{for } n = 0 \\ (1 - \phi)\frac{e^{-\lambda}\lambda^n}{n!} & \text{for } n_i = 1, 2, ... \end{cases}. \tag{2.2}$$

From this, we can determine the first two moments of the ZIP distribution $E[N] = (1 - \phi)\lambda$ and $Var[N] = E[N] + E[N](\lambda - E[N])$. The ZIP model could be useful for modeling the counts because it accounts for overdispersion. Since the only difference between the ZIP model and the Poisson distribution is found when $N = 0$, it is easy to adjust the MLE equations of the Poisson distribution to find the parameters of the ZIP model. Other count distributions can also be used with the zero-inflated distributions, such as the negative-binomial, the Poisson-inverse Gaussian or the Poisson-lognormal (Boucher et al. (2006c)).

2.2 Cross-Section versus Panel Data

The data used for panel data analysis consist of N individual units, each having T observations. Data are usually analyzed by cross-sectional analysis, where each observation of an individual unit is considered to be mutually independent. Thus, in this situation, we worked with $N \times T$ independent observations. However, it is often the case that some data are a repetition for the same individual. Consequently, it can be useful to model the dependence between these observations.

Longitudinal data (or panel data) consists of repeated observations of individual units that are observed over time. Each individual is assumed to be independent but correlation between

observations of the same individual is allowed. This fact must be regarded as an advantage and should be used in the construction of count models.

2.2.1 Modeling

There are several models that include time dependence but the most popular way of dealing with these data is to use a common individual term (Hausman et al. (1984)) that affects all the observations of an individual unit. This random effect represents individual specificities that are constant over time. The hidden individual characteristics that are not captured by the covariates are also captured by this random effect.

Given the insured-specific random effect term θ_i, the annual claim numbers $N_{i,1}, N_{i,2}, \ldots, N_{i,T}$ are independent. Under the assumption that the covariates are independent from the random effects (see Mundlak (1978) or Hsiao (2003) for a general review), the joint probability function of $N_{i,1}, \ldots, N_{i,T}$ is thus given by :

$$\begin{aligned} Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] &= \int_0^\infty Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}|\theta_i] g(\theta_i) d\theta_i \\ &= \int_0^\infty \left(\prod_{t=1}^{T} Pr[N_{i,t} = n_{i,t}|\theta_i] \right) g(\theta_i) d\theta_i, \end{aligned} \tag{2.3}$$

where i represents an individual unit and t is the t^{th} observation of this individual. Many conditional distributions for the random variables $N_{i,t}$ can be chosen as well as a distribution for the random effect θ_i. Under conditioning, moments of the joint distribution can be found quite easily.

2.2.2 Multivariate Negative Binomial

The simplest random effects model for count data is the Poisson distribution with an individual heterogeneity term that follows a specified distribution. Formally, we can express the classic random effects Poisson model as :

$$N_{i,t}|\theta_i \sim Poisson(\theta_i \lambda_i), \quad i = 1, ..., N \;\; t = 1, ..., T.,$$

where $\lambda_i = exp(x_i'\beta)$, which is time independent, is used to model covariates in the distribution (Dionne and Vanasse (1992)). Many possible distributions for the random effects can be chosen. A gamma distribution can be used to express the joint distribution in closed form. Indeed, the

joint distribution of all the observations of a singe unit, when the random effects follow a gamma distribution of mean 1 and variance α, is equal to (Hausman et al. (1984)) :

$$Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] = \left[\prod_{t=1}^{T} \frac{(\lambda_i)^{n_{i,t}}}{n_{i,t}!}\right] \frac{\Gamma(\sum_{i=1}^{T} n_{i,t} + 1/\alpha)}{\Gamma(1/\alpha)} \left(\frac{1/\alpha}{T\lambda_i + 1/\alpha}\right)^{1/\alpha} (T\lambda_i + 1/\alpha)^{-\sum_{i=1}^{T} n_{i,t}}. \quad (2.4)$$

This distribution, which has often been applied (see chapter 36 of Johnson et al. (1996) for an overview), is known as a multivariate negative binomial (MVNB) or negative multinomial. Note that this distribution can also be seen as the generalization of the bivariate negative binomial of Marshall and Olkin (1990). Using time invariant covariates, a sufficient statistic for this distribution is the sum of counts over time T, as we can see in the equation above. For the MVNB distribution, $E[N_{i,t}] = \lambda_i$ and $Var[N_{i,t}] = \lambda_i + \alpha\lambda_i^2$, so overdispersion can be accounted for. Maximum likelihood and variance estimates of the parameters are straightforward.

The generalization of the negative multinomial to other kinds of multivariate negative binomial distributions can also be done. If we assume that the random effects are following a gamma distribution with both parameters equal to λ_i^{1-k}/α, the joint distribution can be expressed as (Boucher et al. (2006b)) :

$$Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] = \frac{\Gamma(\sum_{i=1}^{T} n_{i,t} + \lambda_i^{1-k}/\alpha)}{\Gamma(\lambda_i^{1-k}/\alpha) \prod_t^T n_{i,t}!} \left(\frac{\lambda_i^{1-k}/\alpha}{T\lambda_i + \lambda_i^{1-k}/\alpha}\right)^{\lambda_i^{1-k}/\alpha} \left(\frac{\lambda_i}{T\lambda_i + \lambda_i^{1-k}/\alpha}\right)^{\sum_{i=1}^{T} n_{i,t}}, \quad (2.5)$$

where $k = 1$ is the standard negative multivariate distribution (MVNB2) and $k = 0$ is the multivariate NB1 equivalence (MVNB1). Note that each time that a gamma distribution is used to model the random effects of a distribution in this paper, this generalization will be possible. The variable k can also be numerically estimated, which can be used to construct a method of distinguish between the MVNB2 and the MVNB1 models. This method, used in Boucher et al. (2006c) is based on the creation of a hypermodel, where the additional parameter, here the random variable k, is used to test whether MVNB2 or MVNB1 is statistically the better by setting a simple confidence interval. Expressed in its general form, the MVNBk distributions have the following moments :

$$\begin{aligned} E[N_{i,t}] &= E[E[N_{i,t}|\theta_i]] \\ &= \lambda_i \qquad (2.6) \\ Var[N_{i,t}] &= E[Var[N_{i,t}|\theta_i]] + Var[E[N_{i,t}|\theta_i]] \\ &= \lambda_i + \alpha\lambda_i^{k+1} \qquad (2.7) \\ Cov[N_{i,t}, N_{i,t+j}] &= Cov[E[N_{i,t}|\theta_i], E[N_{i,t+j}|\theta_i]] + E[Cov[N_{i,t}, N_{i,t+j}|\theta_i]] \\ &= Cov[\lambda_i\theta_i, \lambda_i\theta_i] + 0 \\ &= \alpha\lambda_i^{k+1}, \quad j > 0. \qquad (2.8) \end{aligned}$$

Other mixing distributions can be chosen to model the random effects, such as the inverse Gaussian (Holla (1966), Willmot (1987), Dean et al. (1989) or (Shoukri et al. (2004))) or the lognormal (Hinde (1982)) distributions, which result in distributions with the same form for the two first moments, with distinctions found using higher moments. The MVNB distribution has been often used to model panel count data. Consequently, it can be used to compare with zero-inflated models presented in this paper.

3. Multivariate Zero-Inflated Models

3.1 Generalizations of the Zero-Inflated Poisson Distribution

The zero-inflated Poisson model has been shown to be a useuful alternative to the Poisson distribution for cross-section data. Indeed, it often provides a good fit of the data and can be interpreted quite easily. Obviously, the addition of a random effect to the model can generalize it for panel data. However, if we look at the equation (2.2), we will see that this generalization can be carried out in many ways. Indeed, we can treat the zero-inflated component as an individual parameter, add random effects to the mean parameter of the Poisson distribution or even use them together. By conditioning on these two random effects, the joint distribution can be expressed as :

$$\begin{aligned} Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}|\phi_i, \theta_i] &= \prod_{t=1}^{T_i} \Big(I_{(n_{i,t}=0)}\phi_i + (1-\phi_i) Pr[N_{i,t} = n_{i,t}] \Big) \\ &= \sum_{j=0}^{T_0} \binom{T_0}{j} V_j^{Poi} (\sum_t^T n_{i,t}|\theta_i) \phi_i^{T_0-j} (1-\phi_i)^{(T-T_0)+j}, \end{aligned} \qquad (3.1)$$

where the $V^{Poi}(.)$ function has the following Poisson form :

$$V_j^{Poi}(\sum_t^T n_{i,t}|\theta_i) = \frac{(\lambda_i\theta_i)^{\sum_t^T n_{i,t}} \exp\left(-(T-T_0+j)\lambda_i\theta_i\right)}{\prod_t^T n_{i,t}!}. \quad (3.2)$$

Using one or both random effects, we will show that the joint distribution can be expressed in closed form. Moreover, using the moments described in the zero-inflated model definition, since the conditional moments of this distribution are equal to :

$$\begin{aligned} E[N_{i,t}] &= \lambda_i \left(E[\theta_i] - E[\theta_i\phi_i]\right) \quad (3.3) \\ Var[N_{i,t}] &= \lambda_i \left(E[\theta_i] - E[\theta_i\phi_i]\right) \\ &+ \lambda_i^2\Big(E[\phi_i\theta_i^2] - E[\phi_i^2\theta_i^2] + Var[\theta_i] \\ &\quad +E[\phi_i]^2 Var[\theta_i] + E[\theta_i]^2 Var[\phi_i] + Var[\phi_i]Var[\theta_i] - 2E[\phi_i]Var[\theta_i]\Big), \quad (3.4) \end{aligned}$$

we will show that the moments of the new models we present here can also be defined in closed form.

3.1.1 Zero-Inflated Term

We begin with the case where the individual term ϕ_i is beta distributed with parameters a_i and b. In this situation, the joint distribution that we can call multivariate zero-inflated Poisson beta (MZIP-Beta) can now be expressed in the form :

$$Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] = \sum_{j=0}^{T_0} \binom{T_0}{j} V_j^{Poi}(T) \frac{\beta(a+T_0-j, b+(T-T_0)+j)}{\beta(a,b)}, \quad (3.5)$$

where the function $V^{Poi}(.)$ has the following Poisson form :

$$V_j^{Poi}(\sum_t^T n_{i,t}) = \frac{\lambda_i^{\sum_t^T n_{i,t}} \exp\left(-(T-T_0+j)\lambda_i\right)}{\prod_t^T n_{i,t}!}. \quad (3.6)$$

The parameter a_i can be modeled with covariates such as $\exp(x_i'\gamma)$ while the mean parameter of the Poisson distribution can be set as $\lambda_i = \exp(x_i'\beta)$. Note that the distribution does not depend on the count number by period, but only on the number of periods in which at least one count has

been recorded. Using equations (3.3) and (3.4), the first two moments of the distribution are equal to :

$$E[N_{i,t}] = \lambda_i \left(1 - \frac{a_i}{a_i + b}\right) \tag{3.7}$$

$$Var[N_{i,t}] = \lambda_i \left(1 - \frac{a_i}{a_i + b}\right) + \lambda_i^2 \left(\frac{a_i}{a_i + b} - \left(\frac{a_i}{a_i + b}\right)^2\right) \tag{3.8}$$

$$Cov[N_{i,t}, N_{i,t+j}] = \lambda_i^2 \frac{a_i b}{(a_i + b)^2 (a_i + b + 1)}, \quad j > 0. \tag{3.9}$$

From these last equations, we can see that beta random effects imply overdispersion in the zero-inflated model.

3.1.2 Count Distribution

By contrast, we can suppose that a heterogeneity term θ_i is added to the Poisson distribution and follows a gamma distribution of mean 1 and variance α. In this case, the joint distribution will be named multivariate zero-inflated Poisson gamma (MZIP-Gamma) and will be equal to :

$$\begin{aligned} Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] &= \int_0^\infty \prod_{t=1}^{T_i} \left(I_{(n_{i,t}=0)}\phi_i + (1 - \phi_i) Pr[N_{i,t} = n_{i,t}|\theta_i]\right) h(\theta_i) d\theta_i \\ &= \sum_{j=0}^{T_0} \binom{T_0}{j} V_j^{NB}(\sum_t^T n_{i,t}) \phi_i^{T_0 - j} (1 - \phi_i)^{(T - T_0) + j}, \end{aligned} \tag{3.10}$$

where, in this case, the function $V_j^{NB}(.)$ has the following multivariate negative binomial form :

$$V_j^{NB}(\sum_t^T n_{i,t}) = \frac{\Gamma(\sum_t^T n_{i,t} + 1/\alpha)}{\Gamma(1/\alpha) \prod_t^T n_{i,t}!} \left(\frac{1/\alpha}{(T - T_0 + j)\lambda_i + 1/\alpha}\right)^{1/\alpha} \left(\frac{\lambda_i}{(T - T_0 + j)\lambda_i + 1/\alpha}\right)^{\sum_t^T n_{i,t}} \tag{3.11}$$

As noted earlier, a gamma distribution of mean 1 and variance λ_i^{1-k}/α can also be used, which would lead to a $V_j^{NBk}(.)$ function taking the form of the equation (4.19). For the purposes of illustration, let us also mention that apart from the gamma, the inverse-Gaussian can also be computed in a closed form, which leads to a multivariate zero-inflated Poisson inverse-Gaussian distribution (MZIP-Inverse Gaussian). Using an inverse-Gaussian distribution of unit mean and variance τ, the $V_j^{IG}(.)$ function is equal to :

$$V^{IG}(.) = \frac{\lambda_i^{\sum_t^T n_{i,t}}}{\prod_t^T n_{i,t}!} \left(\frac{2}{\pi\tau}\right)^{0.5} e^{1/\tau}(1 + 2\tau\lambda_i(T - T_0 + j))^{-s_i/2} K_{s_i}(z_i), \tag{3.12}$$

where $s_i = \sum_t^T n_{i,t} - 0.5$, $z_i = (1 + 2\tau\lambda_i(T - T_0 + j))^{0.5}/\tau$ and $K_j(.)$ is the modified Bessel function of the second kind. This Bessel function has the following useful properties :

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

Additional properties of the modified Bessel function of the second kind may be found in Shoukri et al. (2004). A lognormal distribution can also be used but the joint distribution cannot be expressed in closed form. The distribution uses the parameter ϕ_i that can be modeled with covariates such as $\phi_i = \exp(x_i'\gamma)/(1 + \exp(x_i'\gamma))$, to ensure that the parameter is contained in the domain $[0, 1]$. In this case, as opposed to the MVNB and the MZIP-Beta distributions, we can see that the distribution depends on the number of count by period and on the number of periods with at least one registered count.

Whichever of the three distributions defined above is chosen to model the random effects, they all share the same first two moments (where α is changed for τ for the MZIPIG distribution) :

$$E[N_{i,t}] = \lambda_i(1 - \phi_i) \tag{3.13}$$

$$Var[N_{i,t}] = \lambda_i(1 - \phi_i) + \lambda_i^2(1 - \phi_i)(\phi_i + \alpha) \tag{3.14}$$

$$Cov[N_{i,t}, N_{i,t+j}] = (1 - \phi_i)^2\lambda_i^2\alpha, \quad j > 0 \tag{3.15}$$

where, again, we can see that the distribution implies overdispersion.

3.1.3 Zero-Inflated Term and Count Distribution

Obviously, the two random effects seen with the MZIP-Beta and the MZIP-Gamma models can be used simultaneously if we assume the independence between them, as proposed in a GEE approach by Song (2005). This would lead to a multivariate zero-inflated Poisson beta gamma model (MZIP-BetaGamma) that can be expressed as :

$$Pr[N_{i,1}=n_{i,1},...,N_{i,T}=n_{i,T}] = \sum_{j=0}^{T_0}\binom{T_0}{j}V_j^{NB}(\sum_t^T n_{i,t})\frac{\beta(a+T_0-j,b+(T-T_0)+j)}{\beta(a,b)}, \tag{3.16}$$

where the moments can be found using the conditional calculation :

$$E[N_{i,t}] = \lambda_i(1-\frac{a_i}{a_i+b}) \tag{3.17}$$

$$Var[N_{i,t}] = \lambda_i(1-\frac{a_i}{a_i+b})+\lambda_i^2\left(1-\frac{a_i}{a_i+b}\right)\left(\frac{a_i}{a_i+b}+\alpha\right) \tag{3.18}$$

$$Cov[N_{i,t},N_{i,t+j}] = \lambda_i^2\left[\left(1-\frac{a_i}{a_i+b}\right)^2\alpha+\frac{a_ib}{(a_i+b)^2(a_i+b+1)}(1+\alpha)\right], \quad j>0. \tag{3.19}$$

3.1.4 Estimation and Correlation

Parameters can be evaluated using maximum likelihood estimation. Programs such as the NLMIXED procedure of the SAS system allow this type of estimations when the log-likelihood can be expressed in closed form. Instead of using the complex computation leading to equations (3.5) and (3.10), alternative estimation approaches can be used, such as numerical integration techniques, Markov chain Monte Carlo methods or others.

An other interesting technique based on the NLMIXED procedure of the SAS system, as described by Nelson et al. (2006) can also be considered. To approximate densities involving random effects, the NLMIXED procedure uses numerical evaluation techniques such as Gaussian quadrature. This technique is only available for normally distributed random effects. Nevertheless, nonnormal random effects can be accommodated in the numerical integration techniques via the use of the probability integral transform. Since the NLMIXED procedure allows for normal random effects ϵ, application of the probability integral transform to the normal random effect such as $u=\Phi(\epsilon)$ leads to $b=F^{-1}(u)$ which has density $f(b)$.

In the MZIP-BetaGamma model, we assume independence between the processes. However, a correlation between each process, as proposed by Crepon and Duguet (1997) for example, can also be supposed. Indeed, disregarding the correlation between the random effects can lead to bad

estimates of the model (see Boucher et al. (2006a) for an example with the multivariate hurdle Poisson model). This procedure can also be used to estimate models where a correlation between the random effects is assumed. Indeed, as explained in Boucher et al. (2006a), if we suppose that ϵ_1 and ϵ_2 are bivariate normal with means 0, unit variance and a correlation parameter ρ, we can state that $u_1 = \Phi(\epsilon_1)$ and $u_2 = \Phi(\epsilon_2)$ are both uniform linked by a Gaussian copula. Application of the probability integral transform to each variable leads to $\theta_1 = F_1^{-1}(u_1)$ which has density $f_1(\theta_1)$ and $\theta_2 = F_2^{-1}(u_2)$ with a density $f_2(\theta_2)$, both random variable θ_1 and θ_2 have a joint distribution modeled by a Gaussian copula.

Nevertheless, unlike the multivariate hurdle distribution explained in Boucher et al. (2006a), correlation between the random effects does not seem to be an important issue here. Indeed, because the two processes are combined in the construction of the model, an interaction between the processes is already supposed. For example, the introduction of a random effect to the count distribution produces thicker distribution tails that will influence the zero-inflated component. Consequently, the aim of a dependence between these random effects might already be assumed in the model.

3.2 Generalization of the Poisson Distribution

Instead of considering the zero-inflated term and its standard count distribution as two different elements that can possess their own random effects, we propose to consider a Poisson distribution which has a special degenerated random effects distribution.

The motivation for is taken from Boucher and Denuit (2006) who used fixed effects estimation to approximate the random effects of Poisson distribution for panel data. They observed a significant presence of zero-values for each individual since a lot of insureds did not a report a single claim. This situation complicates the regression analysis since a mass point at zero exists. As a result, the authors use a weighted regression for gamma, inverse-Gaussian and lognormal distributions to approximate the random effects distribution.

Instead of using a weighted regression, we propose to use a two-part distribution : the first part determines whether θ_i is different from zero while the second part, conditionally on the success of the first part, determines its value. Formally, the distribution of the random effects could be modeled using the following distribution :

$$g(\theta_i|\mu_i,\tau) = \begin{cases} \phi_i & \text{for } \theta_i = 0 \\ (1-\phi_i)\, f(\theta_i) & \text{for } \theta_i \geq 0 \end{cases}, \tag{3.20}$$

where ϕ_i is as a mass point modeled as $\frac{\exp(x_i'\gamma)}{1+\exp(x_i'\gamma)}$ and $f(.)$ has a standard heterogeneity distribution of mean 1 (without loss of generality) and a dispersion parameter ν, such as gamma, inverse-Gaussian or lognormal. When f follows a gamma distribution, the joint distribution of all claims reported by insured i can be expressed as :

$$\begin{aligned} Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] &= \int_0^\infty \prod_{t=1}^{T} \frac{e^{-\lambda_{i,t}\theta_i}(\lambda_{i,t}\theta_i)^{n_i}}{n_i!} g(\theta_i|\mu_i,\tau)d\theta_i & (3.21) \\ &= I_{(T_0=T)}\phi_i + \\ &\quad (1-\phi_i)\frac{\Gamma(\sum_{i=1}^{T} n_{i,t} + 1/\alpha)}{\Gamma(/\alpha)\prod_t^T n_{i,t}!}\left(\frac{1/\alpha}{T\lambda_i + 1/\alpha}\right)^{1/\alpha}\left(\frac{\lambda_i}{T\lambda_i + 1/\alpha}\right)^{\sum_{i=1}^{T} n_{i,t}}. & (3.22) \end{aligned}$$

This joint distribution can be called zero-inflated multivariate negative binomial (ZI-MVNB) since it has the form of a multivariate negative binomial distribution with a zero-inflated term equal to ϕ_i. For the MZIP-inverse Gaussian distribution, note that we can also use an inverse-Gaussian to model the degenerated random effects distribution, which would also lead to a closed form expression for the joint distribution. The model exhibits the following moments :

$$\begin{aligned} E[N_{i,t}] &= \lambda_i\left(1 - \frac{\mu_i}{1+\mu_i}\right) & (3.23) \\ Var[N_{i,t}] &= \lambda_i\left(1 - \frac{\mu_i}{1+\mu_i}\right) + \lambda_i^2\alpha\left(1 - \frac{\mu_i}{1+\mu_i}\right)^2 & (3.24) \\ Cov[N_{i,t}, N_{i,t+j}] &= \lambda_i^2\alpha\left(1 - \frac{\mu_i}{1+\mu_i}\right)^2. & (3.25) \end{aligned}$$

The model implies overdispersion, as do the other models constructed on a generalization of the zero-inflated Poisson distribution.

4. Predictive Distributions

An interesting characteristic of the random effects models is the presence of Bayesian properties. For each t period, the heterogeneity terms (θ_i or ϕ_i) can be updated for past experience,

revealing some individually informations. Formally, the computation of the predictive distribution is performed using Bayesian analysis :

$$\begin{aligned} & Pr[N_{i,T+1} = n_{i,T+1}|n_{i,1}, ..., n_{i,T}] \\ &= \int Pr[N_{i,T+1} = n_{i,T+1}|\theta_i] \left(\frac{\left[\prod_t^T Pr[N_{i,t} = n_{i,t}|\theta_i]\right] g(\theta_i)}{\int \left[\prod_t^T Pr[N_{i,t} = n_{i,t}|\theta_i]\right] g(\theta_i) d\theta_i} \right) d\theta_i \\ &= \int Pr[N_{i,T+1} = n_{i,T+1}|\theta_i) g(\theta_i|n_{i,1}, ..., n_{i,T}) d\theta_i, \end{aligned} \tag{4.1}$$

where $g(\theta_i|n_{i,1}, ..., n_{i,T})$ is the *a posteriori* distribution of the random effects θ_i, reflecting the claims history of insured i. If this *a posteriori* distribution can be expressed in closed form, moments of the predictive distribution can easily be found by conditioning on the random effects θ_i.

4.1 MVNB distribution

As is well known, the posterior distribution of the random effect term for the Poisson model with gamma random effects is also gamma-distributed with parameters $T\lambda_i + 1/\alpha$ and $\sum_t^T n_{i,t} + 1/\alpha$. Using the posterior distribution, the predictive distribution can be computed directly as in section 2.2.2. For the pruposes of illustration, we can express the expected predictive value of this model as :

$$E[N_{i,T+1}|N_{i,1}, ..., N_{i,t}] \quad = \quad \lambda_i \frac{\sum_t^T n_{i,t} + 1/\alpha}{T\lambda_i + 1/\alpha}. \tag{4.2}$$

This can be used for comparison with the other expected predictive values of the Multivariate Zero-Inflated models. Note that the predictive distribution will only depend on the sum of past counts. Additionally, note that when the observed periods go to infinite, the expected predictive value of an individual unit will converge to its average number of past counts.

4.2 MZIP-Beta Distribution

The predictive distribution of MZIP-Beta model can be computed numerically :

$$\begin{aligned} & Pr[N_{i,T+1} = n_{i,T+1}|n_{i,1}, ..., n_{i,T}] \\ &= \frac{\sum_{j=0}^{T_0^*} \binom{T_0^*}{j} V_j^{Poi}(\sum_t^T n_{i,t} + n_{i,T+1}) \beta(a + T_0^* - j, b + (T + 1 - T_0^*) + j)}{\sum_{k=0}^{T_0} \binom{T_0}{k} V_k^{Poi}(\sum_t^T n_{i,t}) \beta(a + T_0 - k, b + (T - T_0) + k)} \end{aligned} \tag{4.3}$$

where T_0^* is the updated value of T_0, considering $n_{i,T+1}$. Using the following notation :

$$G(j) = \frac{\binom{T_0}{j} e^{-\lambda_i j} \Gamma(a+T_0-j)\Gamma(b+T+1-T_0+j)}{(a_i+b+T)\sum_{k=0}^{T_0}\binom{T_0}{k} e^{-\lambda_i k}\Gamma(a+T_0-k)\Gamma(b+T-T_0+k)} \tag{4.4}$$

the predictive distribution can be expressed as :

$$Pr[N_{i,T+1}=n_{i,T+1}|n_{i,1},...,n_{i,T}] = \begin{cases} 1-\sum_{j=0}^{T_0} G(j)(1-e^{-\lambda_i}) & \text{for } n_{i,T+1}=0 \\ \sum_{j=0}^{T_0} G(j) Pr_{Poi}[N_{i,T+1}=n_{i,T+1};\lambda_i] & \text{for } n_{i,T+1}=1,2,... \end{cases} \tag{4.5}$$

where $Pr_{Poi}[.;\lambda_i]$ is the probability function of a Poisson distribution with a mean parameter of λ_i. Under this notation, the expected predictive value can be computed directly :

$$E[N_{i,T+1}|n_{i,1},...,n_{i,T}] = \lambda_i \sum_{j=0}^{T_0} G(j) \tag{4.6}$$

We can see that the number of past reported claims can be excluded from the predictive distribution and, therefore, from the predictive premium. Consequently, the distribution does not consider the number of reported claims, but only on the number of insured periods with at least one claim. This can be seen as a disadvantage since the future premium of an insured will be the same whatever the number of claims he reports in the same insured period. Obviously, unlike the MVNB distribution, the predictive distribution does not converge to the average number of past count. Indeed, it can be shown that $G(j)$ is bounded by $[0,1]$, where the value 0 occurs in the case when $T_0=0$ and $T\to 0$, while the value $G(j)=1$ occurs when $T_0=T\to 0$. In these situations, the expected premium for the following year is 0 or λ_i, which is the premium for a classic Poisson distribution, meaning that no zero-inflated term is needed.

4.3 MZIP-Gamma Distribution

Using the model with random effects that only affect the count distribution, the predictive distribution is equal to :

$$Pr[N_{i,T+1}=n_{i,T+1}|n_{i,1},...,n_{i,T}] = \frac{\sum_{j=0}^{T_0^*}\binom{T_0^*}{j} V_j^{NB}(\sum_t^T n_{i,t}+n_{i,T+1})\phi_i^{T_0^*-j}(1-\phi_i)^{(T+1-T_0^*)+j}}{\sum_{j=0}^{T_0}\binom{T_0}{j} V_j^{NB}(\sum_t^T n_{i,t})\phi_i^{T_0-j}(1-\phi_i)^{(T-T_0)+j}} \tag{4.7}$$

In contrast to the MZIPB model, here we cannot remove the T_0 variable from the predictive distribution. The future premiums will depend on the number of past claims and on the number of insured periods with at least one reported claim. This is different to the hurdle model for panel data proposed by Boucher et al. (2006a) which showed some similarities with the zero-inflated model for panel data. Indeed, the predictive distribution of the multivariate hurdle model with only one random effect put in the count process depends only on the number of claims.

Using the notation :

$$H(j) = \frac{\binom{T_0}{j}\phi_i^{T_0-j}(1-\phi_i)^{(T+1-T_0)+j}((T-T_0+j)\lambda_i+1/\alpha)^{-(\sum_t^T n_t+1/\alpha)}}{\sum_{k=0}^{T_0}\binom{T_0}{k}\phi_i^{T_0-j}(1-\phi_i)^{(T-T_0)+k}((T-T_0+k)\lambda_i+1/\alpha)^{-(\sum_t^T n_t+1/\alpha)}}, \tag{4.8}$$

the predictive distribution can be expressed as :

$$Pr[N_{i,T+1}=n_{i,T+1}|n_{i,1},...,n_{i,T}] = \begin{cases} 1-\sum_{j=0}^{T_0}H(j)(1-p^r) & \text{for } n_{i,T+1}=0 \\ \sum_{j=0}^{T_0}H(j)Pr_{NB}[N_{i,T+1}=n_{i,T+1};r,p] & \text{for } n_{i,T+1}=1,2,... \end{cases}, \tag{4.9}$$

where :

$$Pr_{NB}[N_{i,T+1}=n_{i,T+1};r,p] = \binom{n_{i,T+1}+r}{r}p^r q^{n_{T+1}} \tag{4.10}$$

is the probability function of a negative binomial distribution with parameters equal to :

$$r = \sum_t^T n_{i,t}+1/\alpha, \quad p=\frac{(T-T_0+j)\lambda_i+1/\alpha}{(T+1-T_0+j)\lambda_i+1/\alpha}. \tag{4.11}$$

The expected predictive value is equal to :

$$E[N_{i,T+1}|n_{i,1},...,n_{i,T}] = \lambda_i\sum_{j=0}^{T_0}\frac{\left(\sum_t^T n_{i,t}+1/\alpha\right)H(j)}{(T+1-T_0+j)\lambda_i+1/\alpha} \tag{4.12}$$

4.4 MZIP-BetaGamma Distribution

Using both individual effects leads to the following predictive distribution :

$$\begin{aligned} &Pr[N_{i,T+1} = n_{i,T+1}|n_{i,1}, ..., n_{i,T}] \\ &= \frac{\sum_{j=0}^{T_0^*} \binom{T_0^*}{j} V_j^{NB} (\sum_t^T n_{i,t} + n_{i,T+1}) \beta(a + T_0^* - j, b + (T + 1 - T_0^*) + j)}{\sum_{j=0}^{T_0} \binom{T_0}{j} V_j^{NB} (\sum_t^T n_{i,t}) \beta(a + T_0 - j, b + (T - T_0) + j)}. \end{aligned} \tag{4.13}$$

As before, using the notation :

$$\begin{aligned} &K(j) = \\ &\frac{\binom{T_0}{j} \Gamma(a + T_0 - j) \Gamma(b + T + 1 - T_0 + j) ((T - T_0 + j)\lambda_i + 1/\alpha)^{-(\sum_t^T n_t + 1/\alpha)}}{(a_i + b + T) \sum_{k=0}^{T_0} \binom{T_0}{k} \Gamma(a + T_0 - k) \Gamma(b + T - T_0 + k) ((T - T_0 + k)\lambda_i + 1/\alpha)^{-(\sum_t^T n_t + 1/\alpha)}}, \end{aligned} \tag{4.14}$$

the predictive distribution can be expressed as :

$$Pr[N_{i,T+1} = n_{i,T+1}|n_{i,1}, ..., n_{i,T}] = \begin{cases} 1 - \sum_{j=0}^{T_0} K(j)(1 - p^r) & \text{for } n_{i,T+1} = 0 \\ \sum_{j=0}^{T_0} K(j) Pr_{NB}[N_{i,T+1} = n_{i,T+1}; r, p] & \text{for } n_{i,T+1} = 1, 2, ... \end{cases} \tag{4.15}$$

where :

$$r = \sum_t^T n_{i,t} + 1/\alpha, \quad p = \frac{(T - T_0 + j)\lambda_i + 1/\alpha}{(T + 1 - T_0 + j)\lambda_i + 1/\alpha} \tag{4.16}$$

and the expected predictive value is the equal to :

$$E[N_{i,T+1}|n_{i,1}, ..., n_{i,T}] = \lambda_i \sum_{j=0}^{T_0} \frac{\left(\sum_t^T n_{i,t} + 1/\alpha\right) K(j)}{(T + 1 - T_0 + j)\lambda_i + 1/\alpha} \tag{4.17}$$

4.5 ZI-MVNB Distribution

The predictive distribution of the degenerated random effects model can also be computed analytically :

$$Pr[N_{i,T+1} = n_{i,T+1}|n_{i,1},...,n_{i,T}] = \frac{I_{(n_{i,T+1}=0)}\phi_i + (1-\phi_i)L(n_{i,1},...,n_{i,T+1};T+1)}{I_{(T=T_0)}\phi_i + (1-\phi_i)L(n_{i,1},...,n_{i,T};T)}, \quad (4.18)$$

with :

$$L(n_{i,1},...,n_{i,T};T) = \frac{\Gamma(\sum_{i=1}^{T} n_{i,t} + 1/\alpha)}{\Gamma(1/\alpha)\prod_t n_{i,t}!}\left(\frac{1/\alpha}{T\lambda_i + 1/\alpha}\right)^{1/\alpha}\left(\frac{\lambda_i}{T\lambda_i + 1/\alpha}\right)^{\sum_{i=1}^{T} n_{i,t}}. \quad (4.19)$$

This model can be analyzed for two kinds of insureds : those who reported at least one claim and the others who did not report at all. For insureds who reported at least once, the predictive distribution returns to a standard multivariate negative binomial distribution since the first parts of the numerator and the denominator of equation (4.18) fall. Indeed, we found that the a posteriori distribution of the random effect term is a gamma distribution with parameters equal to $T\lambda_i + 1/\alpha$ and $\sum_t^T n_{i,t} + 1/\alpha$, as for the standard Poisson-Gamma combination. The first moments can be found easily where, for example, the expected value is equal to :

$$E[N_{i,t+1}|N_{i,1},...,N_{i,t},T_0 \neq T] = \lambda_i \frac{\sum_t^T n_{i,t} + 1/\alpha}{T\lambda_i + 1/\alpha}. \quad (4.20)$$

When the insured is claim-free, the predictive distribution has the following form :

$$Pr[n_{i,T+1}|n_{i,1},...,n_{i,T},T = T_0] = \frac{I_{(n_{i,T+1}=0)}\phi_i + (1-\phi_i)\left(\frac{1/\alpha}{T\lambda_i+1/\alpha}\right)^{1/\alpha} Pr_{NB}[n_{i,T+1}]}{\phi_i + (1-\phi_i)\left(\frac{1/\alpha}{T\lambda_i+1/\alpha}\right)^{1/\alpha}} \quad (4.21)$$

where $Pr_{NB}[n_{i,T+1}]$ has a Negative Binomial form that can be expressed as :

$$Pr_{NB}[x;r,p] = \frac{\Gamma(x+r)}{\Gamma(r)\Gamma(x+1)}(1-p)^x p^r \quad (4.22)$$

$$Pr_{NB}[n_{i,T+1}] = \frac{\Gamma(n_{i,T+1} + 1/\alpha)}{\Gamma(1/\alpha)\Gamma(n_{i,T+1}+1)}\left(\frac{T\lambda_i + 1/\alpha}{(T+1)\lambda_i + 1/\alpha}\right)^{1/\alpha}\left(\frac{\lambda_i}{(T+1)\lambda_i + 1/\alpha}\right)^{n_{i,T+1}}. \quad (4.23)$$

Since the moments of this negative binomial are easily found, we can obtain the predictive moments quite directly. For example, the first moment of this predictive distribution can be expressed as :

$$E[N_{i,t+1}|N_{i,1},...,N_{i,t},T_0=T] = \lambda_i \frac{(1-\phi_i)\left(\frac{1/\alpha}{T\lambda_i+1/\alpha}\right)^{1+1/\alpha}}{\phi_i+(1-\phi_i)\left(\frac{1/\alpha}{T\lambda_i+1/\alpha}\right)^{1/\alpha}}, \quad (4.24)$$

from which we can see that the premium of insured i goes to zero, which is his average number of reported claims, if the number of insured periods increases significantly.

If we wish to obtain the predictive and the posterior distributions of the model with correlated random effects, no closed-form equations can be obtained. In this situation, Markov chain Monte Carlo simulations can be used. The idea of MCMC simulations is to simulate realizations from a Markov chain that converges to the joint distribution of the random effects. The resulting random draws are no longer independent, but under mild regularity conditions (as described in the Appendix of Smith and Roberts (1993), for example), the value of the draw tends in distribution to that of a random draw from the joint distribution as the number of draws becomes moderately large. See Boucher et al. (2006a) for an example of MCMC simulations with similar panel count data to that used with multivariate hurdle distribution.

5. Numerical Application

We applied the presented models to the number of reported claim in third-liability for a sample of portfolio from a major insurance company operating in Spain. It is well known that the number of reported claims in insurance exhibits a high number of zeros. In most of the bonus-malus scheme throughout the world (see Lemaire Lemaire (1995) for an overview), a reported claim causes an increase in the premium for the following years. These types of system induce a "hunger for bonus" (Lemaire Lemaire (1977)), where there is an incentive to not report all incurred claims since the resulting increase of subsequent premiums can be greater than the claim indemnity. This fear of the bonus-malus system or the presence of a deductible are possible explanations for the presence of high zero count (Boucher et al. (2006c)). The reasons behind the good fit of the zero-inflated models for cross-section data can also be used to justify the random effects introduce into the zero-inflated term and the count distribution. Indeed, this hunger of bonus can be seen as individually specific, while the absence of some important classification variables (swiftness of reflexes, aggressiveness

behind the wheel, consumption of drugs, etc.) can be used to justify the random effects on the count distribution.

5.1 Assumptions

As seen in the presentation of the zero-inflated models for panel data, covariates are time independent, that is to say that they do not change over all the observations of an individual unit. This assumption was also made in by Gourieroux (1999) for insurance data. Even if this assumption causes additional heterogeneity in the data, and therefore a certain amount of anti-selection, it is not as restricting as it might seem. Indeed, the majority of the time dependent covariates that are used in insurance involves the age of the insured or the length of stay in the company. These variables do not evolve randomly in time since the change can already be known in advance. Consequently, the estimated coefficient of a parameter can still be interpretable since the evolution of the parameter is known. Other covariates can also change in time, such as the city or the type of vehicle belonging to the driver but this kind of major change often involves the creation of an other policy.

Another required assumption was linked to the transformation of the cross section data model into a panel data model. No correlation between the regressors and the random effect term was allowed. In linear regression, this correlation leads to inconsistency of the estimated parameters (Mundlak (1978) or Hsiao (2003) for a general review). The same problem exists for the count data regression when $E[\theta|x_i] \neq E[\theta]$ (Mullahy (1997)) and it leads to biased parameter estimates. In insurance, a correlation between regressors and error term is often present (Boucher and Denuit (2006)) and may be caused by omitted variables that are correlated with those included in the model.

As noted in Winkelmann (2003a), consistent estimates may be found if corrections are made to the standard estimation procedures. However, as shown by Boucher and Denuit (2006), for insurance applications, standard estimation methods can still be used. Indeed, the resulting parameter estimates, although being biased, represent the apparent effect on the frequency of claim, which is the main point interest when the correlated omitted variables cannot be used in classification.

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. 5.1 – Exogenous variables

5.2 Data used

In this paper, we worked with a sample of the automobile portfolio of a major company operating in Spain. Only cars for private use were considered in this sample. The panel data contains information for the period from 1991 to 1998. Our sample contains 15,179 policyholders who remained with in the company for seven complete periods. We used six complete years to estimate the parameters and reserved the last year to compare the predictions of the models with the observed results. Five exogeneous variables (see table 5.1) were kept in the panel plus the annual number of accidents. For every insured we used the initial information from his first policy of period. The total number of claims at fault that took place within each annual period was also used. The average claim frequency of the portfolio is 6.73% for 91,074 contracts.

In this paper, the exogenous variables, shown in table 5.1, were used to model the distribution parameters. The characteristics of the insureds were expressed through some functions $h(\beta_0 + x_i'\beta)$, where β_0 is the intercept and $\beta' = (\beta_1, ..., \beta_p)$ is a vector of regression parameters for explanatory variables $x_i = (x_{i,1}, ..., x_{i,p})$.

5.3 A Priori Analysis

5.3.1 Estimated Parameters

In our numerical example, the MVNB distribution, uses gamma random effects of mean 1 and variance α (as well as the other models with count distribution random effects). The estimated parameters of this distribution are shown in table 5.2, while the parameters of the zero-inflated models for panel data are shown in table 5.3. A model assuming a correlation between the random effects was also estimated. However, as expected, the results were of no use here as the estimation of the model did not converge for a value of ρ different from zero.

Parameter	MVNB	
β_0	-2.6811	(0.0378)
β_1	0.1145	(0.0438)
β_2	-0.1703	(0.0351)
β_3	-0.2210	(0.0398)
β_4	0.0599	(0.0370)
β_5	0.0980	(0.0339)
α	0.9407	(0.0508)
Loglike.	-22,619.58	

TAB. 5.2 – Estimated Parameters for Poisson-Gamma model

The analysis of the results is particularly interesting. We can see that the some covariates are included in the zero-inflated process while others are only used in the count distribution. Since this is a major variable in actuarial classification, we included the sex of the driver in the estimated parameters even if it was not statistically significant for almost all models. The values of all γ parameters are shown to be very different, but remember that they do not represent the same thing since it can represent either ϕ_i or the variable a_i used with the beta distribution for random effects. Nevertheless, in the cases where a comparison can be made (MZIP-Beta vs MZIP-BetaGamma, MZIP-Gamma vs ZI-MVNB), the values are shown to be very different. This is caused by the presence (or absence) of the random effects of the count distribution that directly affect all the probabilities (as opposed to the hurdle model for panel data where a random effect of a given process does not affect the other process (Boucher et al. (2006a))). Yet even if we see that the presence of random effects has an impact on all covariates, it can also be observed that the sign of all estimated parameters remains the same, meaning that the impact of the covariates is fairly similar depending on the model.

5.3.2 **Premiums**

Difference between models can be analyzed through the mean and the variance of selected insured profiles. The mean can be seen as the frequency component of the insurance premium, while the other part involves an analysis of the amount of claims reported. Several profiles were selected and are described in Table 5.4. The first profile is classified as a good driver, while the

Parameter	MZIP-Beta		MZIP-Gamma		MZIP-BetaGamma		ZI-MVNB	
γ_0	0.0114	(0.0642)	−0.7299	(0.1755)	2.1993	(2.0980)	−1.9749	(0.4240)
γ_1	−0.2223	(0.0550)	−0.6391	(0.2193)	−0.6092	(0.2163)	−1.6402	(0.9953)
γ_3	0.3124	(0.0937)	0.6827	(0.1328)	0.6585	(0.1358)	1.2859	(0.3949)
γ_4	−0.1120	(0.0472)	−0.2499	(0.1527)	−0.2430	(0.1470)	−0.8115	(0.6014)
γ_5	−0.1541	(0.0579)	−0.4193	(0.1183)	−0.4013	(0.1197)	−0.9291	(0.2919)
β_0	−1.6799	(0.0318)	−2.3062	(0.0522)	−2.2905	(0.0627)	−2.5514	(0.0344)
β_2	−0.1563	(0.0370)	−0.1698	(0.0345)	−0.1696	(0.0345)	−0.1419	(0.0337)
α	.	.	0.8304	(0.0526)	0.8039	(0.0759)	0.7678	(0.0693)
b	0.5766	0.0169	.	.	17.8185	(38.9070)	.	.
Loglike.	- 22,624.30		-22,589.59		-22,589.55		-22,622.11	

Tab. 5.3 – Estimated Parameters

Profile Number	Kind of Profile	v1	v2	v3	v4	v5
1	Good	0	0	1	0	0
2	Average	1	1	0	0	0
4	Bad	1	0	0	1	1

Tab. 5.4 – Profiles analyzed

last one usually exhibits bad loss experience. The other profile is medium risk. The results are given in Table 5.5. This table shows that the expected values exhibit some small differences for the five models studied. For example, we can see that the ZI-MVNB model offers a smaller insurance premium for the higher risk profile. The same trend can be seen in the analysis of the variance estimates.

5.4 **Predictive Distributions**

Besides the fitting of past observations, the comparison of the predictive distributions can also be interesting. Table 5.6 shows the predictive premiums for a medium risk-profile. The values depend on the total reported claims and on the number of insured periods with at least one reported claim $(T - T_0)$. To illustrate this, we selected a loss experience of 10 years, although other situations can easily be illustrated since closed-form formulas have been found to computed the predictive

Models	1^{st} Profile		2^{nd} Profile		3^{rd} Profile	
	Mean	Variance	Mean	Variance	Mean	Variance
MVNB	0.0549	0.0577	0.0648	0.0687	0.0899	0.0975
MZIP-Beta	0.0549	0.0621	0.0663	0.0725	0.0898	0.0984
MZIP-Gamma	0.0510	0.0577	0.0670	0.0729	0.0882	0.0965
MZIP-BetaGamma	0.0512	0.0579	0.0670	0.0728	0.0884	0.0968
ZI-MVNB	0.0519	0.0540	0.0659	0.0692	0.0776	0.0822

Tab. 5.5 – A priori Premiums

premiums of each model.

One of the first observations concerning the MZIP-Beta is, as expected, that it does not depend on the number of reported claims but only on the number of insured periods during which a claim is reported. When compared with the premiums expected for the MVNB model, we can see that this model is very likely to underestimate the premium of a large number of insureds.

Even if the random effects were only introduce into the count distribution, the results of the MZIP-Gamma distribution indicate that T_0 has also an impact on the future premiums. Depending on the pattern of claim reporting, the future premiums show significant differences. The MZIP-BetaGamma models assuming random effects on the number of reported claims as well as on the zero-inflated component gave results that are fairly close to the premiums obtained with the MZIP-Gamma model.

The ZI-MVNB model could be an interesting alternative to the MVNB model since it lessens the impact of the random effects to some extent. Indeed, since another part of the model adjusts the zeros, it reduces the variance of the random effects distribution, leading to predictive premiums that will be lower than those obtained with the MVNB model. This conclusion can be seen in the numerical results since the values of the experimental model are less than those of the MVNB model in both tails of the distribution.

A major difference that can be seen between the models lies in the variance values of the predictive distribution. These results are shown in the next table :

Models	$T - T_0$	A priori	Sum of claims 0	1	2	3	4	10
MVNB	1	0.0648	0.0402	0.0781	0.1160	0.1538	0.1917	0.4188
MZIP-Beta	0	0.0663	0.0438	.	.	.	.	.
	1	.	.	0.0888	0.0888	0.0888	0.0888	0.0888
	2	.	.	.	0.1116	0.1116	0.1116	0.1116
	3	.	.	.	.	0.1242	0.1242	0.1242
	4	.	.	.	.	.	0.1320	0.1320
	10	.	.	.	.	.	.	0.1481
MZIP-Gamma	0	0.0670	0.0434	.	.	.	.	.
	1	.	.	0.0789	0.1151	0.1515	0.1882	0.4150
	2	.	.	.	0.1138	0.1498	0.1860	0.4088
	3	.	.	.	.	0.1482	0.1839	0.4029
	4	.	.	.	.	.	0.1818	0.3972
	10	.	.	.	.	.	.	0.3672
MZIP-BetaGamma	0	0.0670	0.0436	.	.	.	.	.
	1	.	.	0.0786	0.1136	0.1488	0.1839	0.3963
	2	.	.	.	0.1134	0.1484	0.1835	0.3950
	3	.	.	.	.	0.1482	0.1831	0.3938
	4	.	.	.	.	.	0.1828	0.3928
	10	.	.	.	.	.	.	0.3892
ZI-MVNB	0	0.0659	0.0426	0.0787	0.1129	0.1471	0.1813	0.3864

Tab. 5.6 – Predictive Premiums

Models	$T - T_0$	A priori	Sum of claims 0	1	2	3	4	10
MVNB	.	0.0687	0.0418	0.0811	0.1204	0.1596	0.1989	0.4347
MZIP-Beta	0	0.0725	0.0489	.	.	.	.	.
	1	.	.	0.0951	0.0951	0.0951	0.0951	0.0951
	2	.	.	.	0.1169	0.1169	0.1169	0.1169
	3	.	.	.	.	0.1286	0.1286	0.1286
	4	.	.	.	.	.	0.1356	0.1356
	10	.	.	.	.	.	.	0.1497
MZIP-Gamma	0	0.0729	0.0458	.	.	.	.	.
	1	.	.	0.0840	0.1237	0.1643	0.2059	0.4792
	2	.	.	.	0.1223	0.1623	0.2033	0.4709
	3	.	.	.	.	0.1604	0.2008	0.4631
	4	.	.	.	.	.	0.1983	0.4556
	10	.	.	.	.	.	.	0.4166
MZIP-BetaGamma	0	0.0728	0.0461	.	.	.	.	.
	1	.	.	0.0838	0.1224	0.1618	0.2022	0.4681
	2	.	.	.	0.1218	0.1609	0.2009	0.4610
	3	.	.	.	.	0.1602	0.1998	0.4551
	4	.	.	.	.	.	0.1988	0.4502
	10	.	.	.	.	.	.	0.4317
ZI-MVNB	.	0.0692	0.0453	0.0799	0.1147	0.1494	0.1841	0.3924

Tab. 5.7 – Predictive Variance

The same conclusions that can be made about the variance are essentially the same as those made about the predictive premiums. The MZIP-Beta seems to significantly underestimate the variance, while the values of the variance obtained with the other models are similar to those obtained with the MVNB model. With regard to the expected predictive value, the observation should be noted that the number of insured periods with a claim has a greater impact on subsequent premiums than the total number of reported claims.

5.4.1 Prediction versus Real

The last insured period is reserved for comparison betwee, the predictive values obtained with our model and the real values of the same insureds. These results can be seen in the table below, with the first line showing a comparison between the total number of reported claims expected for each model and the real number of claims observed. In order to compare the number of claims by classes, the remainder of the table was adjusted to obtain a total number of claims equal to the observed number. Note that the profiles have been ranked from the low risk to high risk, based on the expected value from the MVNB model.

The comparison of the total number of reported claims is not itself particularly useful since the future insurance experience depends on too many factors that were not considered here (inflation, unemployment, weather, etc.). What is important to note is that each model predict approximately the same total number of claims. When the analysis is based on the comparison by classes, we can see that the differences between each model are marginal, and even the MZIP-Beta, which has been shown to be incorrectly constructed since it only depends on T_0, gives results that are very close to the ones obtained using the other models.

5.5 Models Comparison

As seen in the previous sections, the differences between models produce, for example, different *a priori* and predictive premiums. To distinguish between these models, this section will present intuitive comparisons as well as more formal testing procedures which determine the model that is thought to be the best fit for our data.

5.5.1 Specification Tests

Of the models considered previously, certain are equivalent for some simple parameters restrictions. This can be tested using classical hypothesis with two standard tests : the log-likelihood ratio and the Wald tests. These tests are asymptotically equivalent. In some cases, the parameter restric-

v1	v2	v3	v4	v5	Freq.	# sin	MVNB	MZIP-Beta	MZIP-Gamma	MZIP-BetaGamma	ZI-MVNB
Total					15179	1143	1020.99	1021.00	1020.94	1020.88	1030.65
0	0	1	0	0	1020	64	60.84	61.31	57.58	57.74	62.49
0	1	0	0	0	1002	57	63.82	64.02	63.02	63.03	65.22
0	0	1	1	0	113	11	6.93	7.11	6.90	6.90	7.44
0	0	1	0	1	2204	165	151.62	152.19	152.65	152.55	161.50
0	1	0	1	0	225	15	15.84	15.68	15.77	15.76	16.16
1	0	1	0	0	240	18	16.22	17.12	17.21	17.17	17.99
0	1	0	0	1	3094	220	221.03	220.40	221.37	221.28	220.46
0	0	1	1	1	178	13	12.83	13.16	13.22	13.22	13.82
1	1	0	0	0	216	11	16.11	16.51	16.49	16.48	16.14
1	0	1	1	0	40	4	2.99	3.13	3.19	3.19	3.27
0	1	0	1	1	645	45	49.21	49.17	49.07	49.10	47.86
1	0	1	0	1	191	17	15.10	15.37	16.17	16.13	16.07
0	0	0	0	0	825	70	63.36	62.71	62.48	62.45	63.46
1	1	0	1	0	65	6	5.48	5.56	5.52	5.52	5.26
1	1	0	0	1	322	20	25.59	25.52	25.75	25.79	24.15
1	0	1	1	1	40	1	3.27	3.30	3.49	3.49	3.38
0	0	0	1	0	376	31	32.72	32.14	32.40	32.40	32.64
0	0	0	0	1	2386	209	201.41	199.26	201.71	201.60	196.84
1	1	0	1	1	114	10	8.74	9.22	8.70	8.74	8.03
1	0	0	0	0	169	12	15.44	15.01	15.72	15.71	15.08
0	0	0	1	1	1116	84	98.45	98.16	98.41	98.46	94.10
1	0	0	1	0	136	12	12.38	12.63	12.56	12.57	11.78
1	0	0	0	1	262	31	24.17	24.34	24.37	24.40	22.45
1	0	0	1	1	200	17	19.45	19.98	19.26	19.30	17.41

Tab. 5.8 – Predicted Values by Profile

tion concerns to the boundary of the parameter space. One problem with the Wald or log-likelihood ratio tests arises 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 performed. Results form Chernoff (1954) for the log-likelihood ratio statistic and Moran (1971) for the Wald Test 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, a one-sided test must be used. Analogous results 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, the usual one-sided test critical value of $z_{1-\partial}$ is still used.

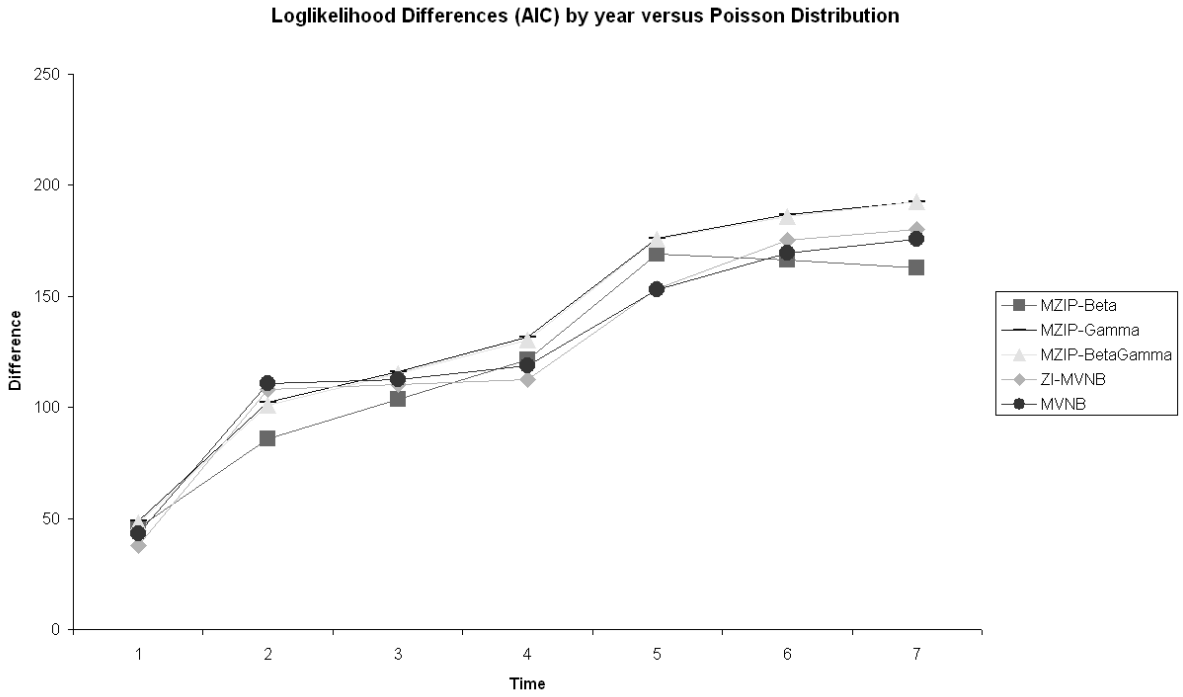
Obviously, the MZIP-Beta and the MZIP-Gamma model are nested with the MZIP-BetaGamma model. The gamma random effects can be tested with the log-likelihood ratio and the Wald test, while beta random effects are better tested with the log-likelihood ratio test since no additional parameter added with the random effects can be tested directly. Using the precautions noted above, the data indicate that the MZIP-Beta is rejected against the MZIP-BetaGamma model (*p-value* ≤ 0.001), while no statistical difference is shown between the MZIP-Beta and the MZIP-BetaGamma (*p-value* of 0.4988).

The ZI-MVNB is nested to the MVNB model for $\phi \to 0$. No statistical difference is shown between the ZI-MVNB and the MVNB model (*p-value* greater than 0.95 !). This is not altogheter surprising given that the same conclusion was obtained in Boucher et al. (2006c), where the authors found that the ZI-Poisson model cannot be fitted with any heterogeneity distributions. Neverthless, we think that the ZI-MVNB can be interesting for small panel data sample. Indeed, since the zero-inflated term needs to be larger for small T.

5.5.2 AIC and Intuitive Comparison

The remaining models could not be compared directly since they are non-nested models. A standard method of comparing non-nested models (and also nested models) is to use the information criteria, such as the Akaike Information Criteria $(AIC) = -2\log(L) + 2k$ or the Bayesian Information Criteria $(BIC) = -2\log(L) + 2\log(n)k$, where k represents the number of parameters of the model and n the total number of observations.

Before performing this AIC/BIC comparison between models, we can do an intuitive analysis of the log-likelihood by models. The analysis presented here can be seen as a generalization of the comparison of the predictive values given in the last section. Using the estimated parameters for the first six years, we computed the log-likelihood for each model and for all seven years. We also estimated a standard Poisson distribution with the first six years of data. The Poisson distribution does not imply dependence between contracts of the same insured but the distribution can be useful in obtaining comparable results between models since it will smooth the random variations between years by comparing only the difference of fit. The graph below shows the values of a pseudo-AIC by year (versus the Poisson distribution) for each of the models presented in this paper :



As the number of past experiences grows, we can see that the fits of our four models are clearly better than the Poisson distribution. Moreover, we can see that the MZIP-Gamma and the MZIP-BetaGamma models show approximately the same results (as expected from the results of a previous specification test), while the MZIP-Beta model does not seem to provide any advantages in comparison with the MVNB model. Indeed, as the past experience increases, this model shows its limited predictive capacity that is based only on the number of time periods with at least one claim. For the last year that was not used for the parameters estimation, we can see that the MZIP-Gamma and the MZIP-BetaGamma distributions show the better fit, while the other models do

not show significant differences from the standard MVNB model.

Using only the remaining models (those not rejected by the specification tests), we can compare the MVNB and the MZIP-Gamma models with the information criteria (using the first six years) :

Model	Log-likelihood	# Param.	AIC	BIC
MVNB	-22,619.58	7	45,253.16	45,308.59
MZIP-Gamma	-22,589.59	8	45,195.19	45,258.54

In both situations, the MZIP-Gamma shows better information criteria than the MVNB model.

5.5.3 Vuong Test

To see if the differences in the log-likelihood and the information criteria between the models are statistically significant, a test based on the difference in the log-likelihood can be performed. For independent observations, a log-likelihood ratio test for non-nested models, developed by Vuong (1989) and generalized by Rivers and Vuong (2002) can be used to see whether the MZIP-Gamma model is statistically better than the MVNB model. This test is based on the likelihood ratio approach of Cox (1961), with a correction applied to the standard deviation of this difference. This test cannot be applied directly to our models since some observations - all contracts of the same insured - are not independent. However, as proposed by Golden (2003) and applied for instance in Boucher et al. (2006b), this non-nested models test can be generalized fairly directly to correlated observations. It can be expressed as :

$$T_{LR,NN} = \frac{\left(\sum_i \sum_t^T \ell(f_{i,t}, \hat{\beta}_1) - \ell(g_{i,t}, \hat{\beta}_2)\right)}{\sqrt{nT}\ \sigma_{T_{LR,NN}}}, \tag{5.1}$$

where the log-likelihood function for the distribution f is defined as :

$$\ell(f_{i,t}, \hat{\beta}_1) = -\log\left(Pr_f[N_{i,t} = n_{i,t} | n_{i,t-1}, ..., n_{i,1}]\right). \tag{5.2}$$

Then, for an observation at time t, the conditional log-likelihood must be expressed based on experience $1, 2, ..., t-1$, using predictive distributions. The parameter $\sigma_{T_{LR,NN}}$ is the square root of the *discrepancy variance* and is expressed as :

$$\sigma^2_{T_{LR,NN}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{T}\left(\ell(f_{i,t}, \hat{\beta}_2) - \ell(g_{i,t}, \hat{\beta}_2)\right)\left(\ell(f_{i,j}, \hat{\beta}_2) - \ell(f_{i,j}, \hat{\beta}_2)\right). \tag{5.3}$$

The variance expression is closely related to the expression obtained using the standard Vuong test, except that the covariances between correlated observations are considered. Since the generalization of this test for correlated observations is very similar to the Vuong test, French and Jones (2004) applied it directly. However, Golden (2003) showed that intermediary tests must be performed to ensure that the test is valid. First, a necessary but not sufficient condition for establishing the asymptotic properties of this test involves the calculation of the following matrix :

$$B = \begin{bmatrix} B_{f,f} & B_{f,g} \\ B_{g,f} & B_{g,g} \end{bmatrix}, \tag{5.4}$$

where

$$B_{f,g} = \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{j=1}^{T} \nabla \ell(f_{i,t}, \hat{\beta}_1) \nabla \ell(g_{i,j}, \hat{\beta}_2). \tag{5.5}$$

The matrix B must be positive definite or, at least, must converge to it (see Golden (2003) for details). Since the variance of the difference in log-likelihood includes covariance terms, the second intermediary step is performed to avoid a situation in which $\sigma_{T_{LR,NN}}$ could be equal to zero. This involves the calculation of the *estimated discrepancy autocorrelation coefficient* and is defined as :

$$\hat{r} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{j=1, j \neq t}^{T} \Big(\ell(f_{i,t}, \hat{\beta}_1) - \ell(g_{i,t}, \hat{\beta}_2) \Big) \Big(\ell(f_{i,j}, \hat{\beta}_1) - \ell(g_{i,j}, \hat{\beta}_2) \Big)}{2T \sum_{i=1}^{n} \sum_{t=1}^{T} (\ell(f_{i,t}, \hat{\beta}_1) - \ell(g_{i,j}, \hat{\beta}_2))^2} \tag{5.6}$$

The adapted test of Golden (2003) assumes that $\hat{r} \neq -1/(2T)$, so this assumption must be verified to ensure that the test is correctly specified. The last intermediary test is to verify that the two models cannot be equal. Since we work with non-nested models and we already ensure that models cannot be equal by introducing simultaneous parameter restrictions, this step is not needed in our case.

To apply the adapted Vuong test for correlated observation, neither of the two models has to be true. The null hypothesis of the test is that the two model are equivalent, expressed as $H_0 : E[\ell(f, \hat{\beta}_1) - \ell(g, \hat{\beta}_2)] = 0$. Under the null hypothesis, the test converge to a standard normal distribution. Rejection of the test in favour of the distribution f occurs when $T_{LR,NN} > c$, or in favour of g if $T_{LR,NN} < c$, where c represents the critical value for some significance level.

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

$$\hat{C}(\theta) = C(\theta) + K(f, g),$$

where $K(f, g)$ is a correction factor, such as those used in the information criteria test seen in the previous section.

For the MZIP-Gamma against the MVNB models, the Vuong test, adapted for correlated observations, shows that the MZIP-Gamma model is statistically different from the MVNB model. Indeed, the resulting test involve *p-values* of less than 0.01% for the adapted tests (AIC and BIC). Intermediary tests show that the matrix B is positive definite, while the value of the *estimated discrepancy autocorrelation coefficients* is far from $-1/12$, being approximately -0.003.

6. Conclusion

We presented new models for panel count data that can be used to model the data exhibiting a large number of zero. The joint distribution and their predictive distributions can be expressed analytically, which can be a significant advantage. The models offer an interesting alternative to the standard multivariate negative binomial.

Comparison between models shows that our insurance data were better fitted by the multivariate zero-inflated Poisson gamma distribution. Unlike the other new distributions seen in this paper, the MZIP-Gamma can be generalized to model time dependent covariates, which can be an advantage in the analysis of the number of reported claims. Future analysis of these distributions based on generalizations of negative binomial distribution could be interesting.