NUMBER OF ACCIDENTS OR NUMBER OF CLAIMS? AN APPROACH WITH ZERO-INFLATED POISSON MODELS FOR PANEL DATA

Jean-Philippe Boucher*,
Michel Denuit and Montserrat Guillén

*Département de mathématiques
Université du Québec à Montréal, Canada
http://www.math.uqam.ca/actuariat/boucher

39th ASTIN Colloquium, June 2, 2009

General Topics

- Count Data;
- Risk Classification;
- Panel Data (Longitudinal Data).

Insurance

- *A priori* Ratemaking;
- *A posteriori* Ratemaking;
- Hunger for Bonus: Number of Claims vs Number of Accidents.

Summary

Summary

1. Number of Claims versus Number of Accidents
   - Standard Approach
   - Lemaire's Model
2. Modeling Overview
   - Zero-Inflated Distribution
   - Cross-Section versus Panel Data
3. Panel Data Models
4. Potential Analysis based on the Zero-Inflated Models
   - Interpretation
5. Numerical Application
   - A priori Analysis
   - A posteriori Analysis
   - Distribution of the Number of Accidents
   - Models Comparison
6. Conclusion

Summary

1. Number of Claims versus Number of Accidents
   - Standard Approach
   - Lemaire's Model
2. Modeling Overview
   - Zero-Inflated Distribution
   - Cross-Section versus Panel Data
3. Panel Data Models
4. Potential Analysis based on the Zero-Inflated Models
   - Interpretation
5. Numerical Application
   - A priori Analysis
   - A posteriori Analysis
   - Distribution of the Number of Accidents
   - Models Comparison
6. Conclusion

Summary

Summary

Summary

Plan

1. Number of Claims versus Number of Accidents
   - Standard Approach
   - Lemaire's Model
2. Modeling Overview
   - Zero-Inflated Distribution
   - Cross-Section versus Panel Data
3. Panel Data Models
4. Potential Analysis based on the Zero-Inflated Models
   - Interpretation
5. Numerical Application
   - A priori Analysis
   - A posteriori Analysis
   - Distribution of the Number of Accidents
   - Models Comparison
6. Conclusion

Standard Approach

Overview

- A Poisson(λ) distribution models the number of accidents;
- A Bernoulli(p) variable models the probability p that the accidents will be filed.
- Formally, with M the number of accidents and N the number of claims:

$$N = \sum_{j=1}^{M} B_j \quad \Rightarrow N \sim Poisson(\lambda p)$$

where the B_j are i.i.d. Bernoulli(p) variables.

Lemaire's Model

Policyholders report their accidents only if they can obtain some benefit, i.e. if the claim cost exceeds the discount obtained for a no-claim record.

Lemaire's Model Assumptions

- Insureds are completely rational;
- Insureds know how a bonus-malus system works;
- Insureds can then calculate an optimal threshold from which it is profitable to claim.

Plan

Zero-Inflated Distribution

Overview

- A high number of zero-values is often observed in the fitting of count data;
- A finite mixture models of two distributions combining an indicator distribution for the zero case and a standard count distribution is appealing;
- For example, the probability function of the zero-inflated Poisson (ZIP) distribution is:

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

Cross-Section versus Panel Data

Data

- The data used for panel data analysis consist of N individual units, each having T observations;
- Cross-section data: each observation is considered to be mutually independent ($N \times T$ independent observations);
- Panel data: Each individual is assumed to be independent, but dependence between observations of the same individual is allowed.

Construction

- Missing of some important classification variables (swiftness of reflexes, aggressiveness behind the wheel, consumption of drugs, etc.) in the classification;
- Hidden features captured by an individual random heterogeneity term θ_i;
- Given 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 given by

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

- Models depend on the choices of the conditional distribution of the $N_{i,t}$ and the distribution of θ_i.

Plan

Multivariate Negative Binomial Distribution (MVNB)

When $N_{i,t}$ is conditionaly distributed as a Poisson distribution with random effects following a gamma distribution:

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

where $\lambda_i = exp(x_i'\beta)$.

Multivariate Zero-Inflated Models

- $N_{i,t}$ is conditionaly distributed as a ZI-Poisson distribution;
- Random effects on the individual term ϕ_i (beta(a,b) distributed);
- Random effects θ_i on the mean parameter of the Poisson distribution (gamma distributed);

$$\begin{aligned} Pr[N_{i,1},...,N_{i,T}] &= \int\int \prod_{t=1}^{T} \left(I_{(n_{i,t}=0)}\phi_i + (1-\phi_i)\frac{e^{-\lambda_{i,t}\theta_i}(\lambda_{i,t}\theta_i)^{n_{i,t}}}{n_{i,t}!} \right) g(\phi_i,\theta_i)d\phi_i d\theta_i \\ &= \sum_{j=0}^{T_0} \binom{T_0}{j} V_j^{NB}(n_{i,1},...,n_{i,T}) \frac{\beta(a+T_0-j, b+(T-T_0)+j)}{\beta(a,b)}, \end{aligned}$$

where T_0 is the number of insured periods without claims and:

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

Plan

Interpretation for Insurance Data

Hunger for Bonus

- The model linked to a reporting decision at the period level, and not at the accident level (as with Lemaire's model);
- Each year, a number of insureds will not claim at all, whatever the case;
- Why these insureds procure insurance ? Fear of insurance, minimal protection, mandatory insurance, etc.

Connections with Lemaire's model

- The first accident of each insured year indicates the way the insured will act for the rest of the year;
- Also assumes that the number of accidents is Poisson distributed;
- Model assumes that the insureds do not really know how a bonus-malus system works;
- Allows us to distinguish the underreporting from the driving behavior.

Predictive Distribution and Predictive Premiums

- Predictive distributions of panel data with random effects involve Bayesian analysis;
- At each insured period, the random effects can be updated for past claim experience, revealing some insured-specific informations.

$$\Pr[N_{i,T+1} = n_{i,T+1} | n_{i,1}, ..., n_{i,T}] = \int \int \Pr(n_{i,T+1} | \phi_i, \theta_i) g(\phi_i, \theta_i | n_{i,1}, ..., n_{i,T}) d\phi_i d\theta_i,$$

where $g(\phi_i, \theta_i | n_{i,1}, ..., n_{i,T})$ is the *a posteriori* distribution of the random effects ϕ_i, θ_i.

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

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

Plan

A priori Premiums

Models	Good Profile Mean	Good Profile Variance	Average Profile Mean	Average Profile Variance	Bad Profile Mean	Bad Profile Variance
MVNB	0.0549	0.0577	0.0648	0.0687	0.0899	0.0975
MZIP	0.0510	0.0577	0.0670	0.0729	0.0882	0.0965

Table: A priori Premiums

Predictive Premiums

Models	$T - T_0$	A priori	Sum of claims 0	1	2	3	4	10
MVNB	.	0.0648	0.0402	0.0781	0.1160	0.1538	0.1917	0.4188
MZIP	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

Table: Predictive Premiums

Number of Accidents

Kind of Insureds	Number of Accidents (Estimated) MZIP-Gamma	Number of Claims (Observed)
$v_2 = 0$	6,623	4,710
$v_2 = 1$	3,345	2,559
Total	9,968	7,269

Table: Number of Predicted Accidents and Number of Observed Claims

Models Comparison

- Specification tests;
- Information criteria (AIC, BIC);
- Vuong Test (adapted by Golden for panel data).

Plan

Conclusion

Summary of results

- Zero-inflated model for panel data;
- Number of claims versus number of accidents;
- Predictive analysis.

Future researches

- Hunger for Bonus;
- Time between claim.

Publications

J.-P. Boucher and M. Denuit and M. Guillén (2007).
Risk Classification for Claim Counts: Mixed Poisson, Zero-Inflated Mixed Poisson and Hurdle Models.
North American Actuarial Journal, 11-4, 110-131

J.-P. Boucher and M. Denuit and M. Guillén (2009).
Number of Accident or Number of Claims ? An Approach with Zero-Inflated Poisson Models for Panel Data.
to appear in Journal of Risk and Insurance.

J.-P. Boucher and M. Denuit (2008).
Credibility Premiums for the Zero-Inflated Poisson Model and New Hunger for Bonus Interpretation.
Insurance: Mathematics and Economics, 42, 727-735