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
Introduction

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

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

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

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
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
A Poisson($\lambda$) 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:

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

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

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
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:

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

\[= I_{(n=0)} \phi + (1 - \phi)\frac{e^{-\lambda}\lambda^n}{n!} \]
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 $\theta_i$;
- Given $\theta_i$, the annual claim numbers $N_{i,1}, N_{i,2}, \ldots, N_{i,T}$ are independent.
- The joint probability function of $N_{i,1}, \ldots, N_{i,T}$ is given by

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

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

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

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
Multivariate Negative Binomial Distribution (MVNB)

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

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

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

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

Boucher, Jean-Philippe (UQAM)  Accidents or Claims?  June 2, 2009  12 / 24
Multivariate Zero-Inflated Models

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

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

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

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

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

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

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

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

<table>
<thead>
<tr>
<th>Models</th>
<th>Good Profile</th>
<th>Average Profile</th>
<th>Bad Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
<td>Mean</td>
</tr>
<tr>
<td>MVNB</td>
<td>0.0549</td>
<td>0.0577</td>
<td>0.0648</td>
</tr>
<tr>
<td>MZIP</td>
<td>0.0510</td>
<td>0.0577</td>
<td>0.0670</td>
</tr>
</tbody>
</table>

Table: A priori Premiums
### Predictive Premiums

<table>
<thead>
<tr>
<th>Models</th>
<th>$T - T_0$</th>
<th>A priori</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVNB</td>
<td>.</td>
<td>0.0648</td>
<td>0.0402</td>
<td>0.0781</td>
<td>0.1160</td>
<td>0.1538</td>
<td>0.1917</td>
<td>0.4188</td>
</tr>
<tr>
<td>MZIP</td>
<td>0</td>
<td>0.0670</td>
<td>0.0434</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>0.0789</td>
<td>0.1151</td>
<td>0.1515</td>
<td>0.1882</td>
<td>0.4150</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>0.1138</td>
<td>0.1498</td>
<td>0.1860</td>
<td>0.4088</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>0.1482</td>
<td>0.1839</td>
<td>0.4029</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>0.1818</td>
<td>0.3972</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>0.3672</td>
</tr>
</tbody>
</table>

**Table:** Predictive Premiums
## Number of Accidents

<table>
<thead>
<tr>
<th>Kind of Insureds</th>
<th>Number of Accidents (Estimated) MZIP-Gamma</th>
<th>Number of Claims (Observed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_2 = 0$</td>
<td>6,623</td>
<td>4,710</td>
</tr>
<tr>
<td>$\nu_2 = 1$</td>
<td>3,345</td>
<td>2,559</td>
</tr>
<tr>
<td>Total</td>
<td>9,968</td>
<td>7,269</td>
</tr>
</tbody>
</table>

**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

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
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.
_North American Actuarial Journal, 11-4, 110-131_

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._

Credibility Premiums for the Zero-Inflated Poisson Model and New Hunger for Bonus Interpretation.
_Insurance: Mathematics and Economics, 42, 727-735_