Generalized Linear Models in Compound Risk Model with Dependent Structures

Jae Youn Ahn (with Woojoo Lee and Donghwan Lee)
Department of Statistics, Ewha Womans University, Korea

This presentation has been prepared for the Actuaries Institute 2015 ASTIN and AFIR/ERM Colloquium. The Institute Council wishes it to be understood that opinions put forward herein are not necessarily those of the Institute and the Council is not responsible for those opinions.
Table of Contents

- Introduction
- Compound GLM Model with Dependence
- Analysis of Statistical Estimation
- Conclusion
Compound Model

\[ S = \begin{cases} 
\sum_{i=1}^{N} Y_i, & N > 0 \\
0, & N = 0 
\end{cases} \]

with \( N, Y_1, Y_2, \cdots \) are independent and \( Y_i \) are identically distributed.

Likelihood function with observation \((n, \mathbf{y}) = (n, y_1, \cdots, y_n)\)

\[ f_{N, \mathbf{y}}(n, \mathbf{y}) = f_N(n)f_{\mathbf{y}}(y_1, \cdots, y_n) = f_N(n)f_{\mathbf{y}}(y_1) \cdots f_{\mathbf{y}}(y_n), \]

Individual data \( \mathbf{y} \) may not be available while summarized data

\[ s := \sum_{i=1}^{n} y_i \quad \text{or} \quad m := s/n \]

is available.
Gamma and Poisson Distribution

- Poisson($\theta$) Distribution
  - Density
  
  \[
  f_N(n) = \frac{\exp(-\theta) \theta^n}{n!}.
  \]

  - $\mathbb{E} [N] = \theta$
  - $\text{Var}(N) = \theta$

- Gamma($\nu, \xi$) Distribution
  - Density
  
  \[
  f_Y(y) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\xi}\right)^\nu y^{\nu-1} \exp\left(-\frac{y}{\xi/\nu}\right)
  \]

  - $\mathbb{E} [Y] = \xi$
  - $\text{Var}(Y) = \frac{\xi^2}{\nu}$
  - Dispersion Parameter: $\phi := 1/\nu$ such that

  \[
  \text{Var}(Y) = \phi \cdot (\mathbb{E} [Y])^2
  \]
Compound Model and Tweedie Model

- Aggregate observations: \((N, S)\)

\[(n_1, s_1), \cdots, (n_k, s_k)\]

or \((N, M)\) with average observation \(M := S/N:\)

\[(n_1, m_1), \cdots, (n_k, m_k)\]

- Tweedie Model, \(S\):
  - \(N \sim \text{Pois}(\theta)\)
  - \(Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \text{Gamma}(v, \xi)\)

Then, for given \(N > 0\),

- \(S \mid N \sim \text{Gamma}(Nv, N\xi)\)
- \(M \mid N \sim \text{Gamma}(Nv, \xi)\)

Likelihood function with an observation \((n_i, m_i)\)

\[
f_{N,M}(n_i, m_i; v, \xi, \theta) = \begin{cases} f_N(n_i; \theta)f_M(m_i \mid n_i; v, \xi), & n_i > 0, \\ f_N(n_i; \theta), & n_i = 0. \end{cases}
\]
Complete data:
- \((n_1, y_{1,1}, \cdots, y_{1,n_1})\) and explanatory variables \(X_{1}^{T}\)
- \((n_k, y_{k,1}, \cdots, y_{k,n_k})\) and explanatory variables \(X_{k}^{T}\)

Summarized data:
- \(n_1, \frac{\sum_{i=1}^{n_1} y_{1,i}}{n_1}\) and explanatory variables \(X_{1}^{T}\)
- \(n_k, \frac{\sum_{i=1}^{n_k} y_{k,i}}{n_k}\) and explanatory variables \(X_{k}^{T}\)
Generalized Linear Model: Data Structures

Complete data:

- \( (n_1, y_{1,1}, \ldots, y_{1,n_1}) \) and explanatory variables \( X_{1}^{T} \)
  
  \[ \vdots \]

- \( (n_k, y_{k,1}, \ldots, y_{k,n_k}) \) and explanatory variables \( X_{k}^{T} \)

Summarized data:

- \( (n_1, m_1) \) and explanatory variables \( X_{1}^{T} \)
  
  \[ \vdots \]

- \( (n_1, m_k) \) and explanatory variables \( X_{k}^{T} \)
Generalized Linear Model 0 (Independence)

**Model Assumption:**
- \( n_i \sim \text{Poisson}(\theta_i), \quad \log(\theta_i) = X_i^T \alpha \)
- \( y_{i,1}, \ldots, y_{i,n_i} | n_i \sim \text{i.i.d. Gamma}(v, \xi_i), \quad \log(\xi_i) = X_i^T \beta \)
- \( n_i \) and \( y_{i,1}, \ldots, y_{i,n_i} \) are independent.

**Under Model Assumption,**
- \( n_i \sim \text{Poisson}(\theta_i), \quad \log(\theta_i) = X_i^T \alpha \)
- \( m_i | n_i \sim \text{Gamma}(n_i v, \xi_i), \quad \log(\xi_i) = X_i^T \beta \)

Joint distribution function of \((n_i, m_i)\) can be written as

\[
f_{N,M}(n_i, m_i; \alpha, \beta) = \begin{cases} 
  f_N(n_i; \alpha) f_M(m_i | n_i; X_i, \beta), & n_i > 0. \\
  f_N(n_i; \alpha), & n_i = 0.
\end{cases}
\]

and

\[
\prod_{i=1}^{k} f_{N,M}(n_i, m_i; \alpha, \beta) = \prod_{i=1}^{k} f_N(n_i; \alpha) \prod_{n_i > 0} f_M(m_i | n_i; X_i, \beta)
\]
Independence assumption between severity and frequency may be unrealistic.

We want to model dependence between $N$ and $Y$.

Hernández-Bastida et al. (2009): Dependence between $n_i$ and $m_i$
- Dependence between parameters: $\mathbb{E}[N] = \theta$ and $\mathbb{E}[Y] = \xi$
- Gives dependence via Sarmanov-Lee family distributions
- Estimation procedure is not provided.

Czado et al. (2012): Dependence between $n_i$ and $m_i$
- Gives dependence via Gaussian copula

$$F_{N,M}(n, m) = C(F_N(n), F_M(m))$$

Complicated likelihood function:

$$f_{N,M}(n, m) = \frac{\partial}{\partial m} F_{N,M}(n, m)f_m(m) - \frac{\partial}{\partial m} F_{N,M}(n-, m)f_m(m)$$

Dependence between $n_i$ and $m_i$ is somehow artificial.
Modelling Dependence between $N$ and $M$?

Example

- **Frequency:**
  
  $n = \begin{cases} 
  1, & \text{with probability } 1/2; \\
  2, & \text{with probability } 1/2; 
  \end{cases}$

- **Severity:**
  
  $y_i \sim_{\text{i.i.d.}} N(0, 1^2)$

We think severity and frequency are independent: $n$ and $y_1, y_2, \cdots$, are independent

Question: Are $n$ and $m$ independent?

$n, m$ are **not independent**:

- $\mathbb{P}\left( \frac{s_i}{n_i} \leq 1.645 \mid n_i = 1 \right) = 0.95$
- $\mathbb{P}\left( \frac{s_i}{n_i} \leq 1.645 \mid n_i = 2 \right) = 0.99$
Table of Contents

- Introduction
- Compound GLM Model with Dependence
- Analysis of Statistical Estimation
- Conclusion
Dependence between Frequency \((N)\) and Average Severity \(M\): not natural

Dependence between Frequency \((N)\) and Individual Severity \((Y_1, \cdots, Y_N)\).

Data: \((N_1, M_1), \cdots, (N_k, M_k)\)
Mean modelling with dependence (Lee and Ahn, 2015):  
- **Model Assumption:**
  - \( n_i \sim \text{Poisson}(\theta_i) \), \( \log(\theta_i) = X_i^T \alpha \)
  - \( y_{i,j} \mid n_i \sim \text{Gamma}(\nu, \xi_i) \), \( \log(\xi_i) = X_i^T \beta + n_i \beta^* \). Then
- **Under Model Assumption**
  - \( n_i \sim \text{Poisson}(\theta_i) \), \( \log(\theta_i) = X_i^T \alpha \)
  - \( m_i \mid n_i \sim \text{Gamma}(n_i \nu, \xi_i) \), \( \log(\xi_i) = X_i^T \beta + n_i \beta^* \)
- Define \( \beta^+ := (\beta, \beta^*) \)

Joint distribution function of \((n_i, m_i)\) can be written as

\[
f_{N,M}(n_i, m_i; \alpha, \beta^+) = \begin{cases} 
  f_N(n_i; \alpha)f_M(m_i \mid n_i; \beta^+), & n_i > 0. \\
  f_N(n_i; \alpha), & n_i = 0. 
\end{cases}
\]

and

\[
k \prod_{i=1}^{k} f_{N,M}(n_i, m_i; \alpha, \beta^+) = \prod_{i=1}^{k} f_N(n_i; \alpha) \prod_{n_i > 0} f_M(m_i \mid n_i; \beta^+)
\]

- See also Shi et al. (2015).
• Is dispersion parameter \((1/v)\) constant?
• Evidence from literature: Dispersion is not a constant (GORDON, 2002).
• Under the dependence assumption, which model do you prefer?
  • \(\log(1/v_i) = X_i^T \beta\) or
  • \(\log(1/v_i) = X_i^T \beta + n_i \beta^*\)
Modelling Dispersion Parameter: Shared Random Effect Example

- Frequency (assume $\alpha = 0$):
  \[ n_i \mid X_i^T, R_i \sim \text{Poisson}(\theta_i), \quad \theta_i = X_i^T \alpha + R_i \alpha^*. \]

- Severity:
  \[ y_{i,1}, \cdots, y_{i,n_i} \mid X_i^T, R_i \sim \text{Gamma}(\nu, \xi_i), \quad \log(\xi_i) = X_i^T \beta + R_i \beta^*. \]

  \[ \Rightarrow \text{Conditional mean} \]
  \[ \mathbb{E} [y_{i,j} \mid X_i, n_i] = \exp (X_i^T \beta' + \psi_N(n_i) \tau) \]

  \[ \Rightarrow \text{Conditional variance} \]
  \[ \text{Var} (y_{i,j} \mid X_i, n_i) = (\mathbb{E} [y_{i,j} \mid X_i, n_i])^2 \frac{1}{\nu_i} \]

  where $\log(1/\nu_i) := \left(\frac{1}{\nu_i} + 1\right) \left(\frac{(b(\alpha^* - \beta^*) + 1)^2}{(\alpha^* b + 1)(b(\alpha^* - 2 \beta^*) + 1)}\right)^{n_i + a} - 1$
Modelling Dispersion Parameter: Shared Random Effect Example

- Frequency (assume $\alpha = 0$):
  
  $$n_i \mid X^T_i, R_i \sim \text{Poisson} (\theta_i), \quad \theta_i = X^T_i \alpha + R_i \alpha^*.$$ 

- Severity:

  $$y_{i,1}, \cdots, y_{i,n_i} \mid X^T_i, R_i \sim \text{Gamma} (v, \xi_i), \quad \log(\xi_i) = X^T_i \beta + R_i \beta^*.$$ 

  $\Rightarrow$ Conditional mean

  $$\mathbb{E} [y_{i,j} \mid X_i, n_i] = \exp (X^T_i \beta' + \psi_N(n_i) \tau)$$ 

  $\Rightarrow$ Conditional variance

  $$\text{Var} (y_{i,j} \mid X_i, n_i) = \left( \mathbb{E} [y_{i,j} \mid X_i, n_i] \right)^2 \frac{1}{v_i}$$ 

  where $\log(1/v_i) := \psi_R(n_i) \beta^*$.
Model Assumption:
- \( n_i \sim \text{Poisson}(\theta_i) \), \( \log(\theta_i) = \mathbf{X}_i^T \alpha \)
- \( y_{i,j} \mid n_i \overset{i.i.d.}{\sim} \Gamma(n_i \nu_i, \xi_i) \), \( \log(\xi_i) = \mathbf{X}_i^T \beta + n_i \beta^* \).
- \( \log(1/\nu_i) = \mathbf{X}_i^T \gamma + n_i \gamma^* \)

Under Model Assumption:
- \( n_i \sim \text{Poisson}(\theta_i) \), \( \log(\theta_i) = \mathbf{X}_i^T \alpha \)
- \( m_i \mid n_i \sim \Gamma(n_i \nu_i, \xi_i) \), \( \log(\xi_i) = \mathbf{X}_i^T \beta + n_i \beta^* \)
- \( \log(1/\nu_i) = \mathbf{X}_i^T \gamma + n_i \gamma^* \)

Define \( \beta^+ := (\beta, \beta^*) \)

Joint distribution function of \((n_i, m_i)\) can be written as
\[
\prod_{i=1}^{k} f_{N,M}(n_i, m_i \mid \mathbf{X}_i; \alpha, \beta^+) = \begin{cases} 
  f_N(n_i \mid \mathbf{X}_i; \alpha) f_M(m_i \mid n_i, \mathbf{X}_i; \beta^+, \gamma), & n_i > 0. \\
  f_N(n_i \mid \mathbf{X}_i; \alpha), & n_i = 0.
\end{cases}
\]

Lee, W., Lee, D. and Ahn, J.

Generalized Linear Models in Compound Risk Model with Dependent Structures
Effect of Dispersion Parameter: Example

- **Frequency**
  \[ n_i \sim \text{Poisson}(\theta), \quad \theta = 10, \]

- **Severity**
  \[ y_{i,1}, \ldots, y_{i,n_i} \mid n_i \overset{\text{i.i.d.}}{\sim} \Gamma(v_i, \xi_i) \]

which is equivalent with

\[ m_i \mid n_i \sim \Gamma(n_i v_i, \xi_i) \]

- **Mean**
  \[ \log(\xi_i) = \beta_0 + \beta_1 n_i \]

- **Dispersion**
  \[ \log \left( \frac{1}{\nu_i} \right) = \alpha_1 + \alpha_1 n_i \]
Effect of Dispersion Parameter: Example

(a) Without Dispersion Modelling (Model I)

(b) With Dispersion Modelling (Model II)
Table of Contents

- Introduction
- Compound GLM Model with Dependence
- Analysis of Statistical Estimation
- Conclusion
Data Reduction and Sufficiency Issues

- **Model Assumption:**
  - \( n_i \sim \text{Poisson}(\theta_i) \), \( \log(\theta_i) = X_i^T \alpha \)
  - \( y_{i,j} \mid n_i \text{ i.i.d. } \sim \text{Gamma}(\nu_i, \xi_i) \), \( \log(\xi_i) = X_i^T \beta + n_i \beta^* \).

\[
\log(1/\nu_i) = X_i^T \gamma + n_i \gamma^*
\]

- **Two different set of data**
  - Complete data
    \[ \{(n_1, y_1), \cdots, (n_k, y_k)\} \]
  - Summarized data
    \[ \{(n_1, m_1), \cdots, (n_k, m_k)\} \]

- Question: What is efficiency difference between using Complete data and Summarized data?
Data Reduction and Sufficiency Issues

- Equivalent(?) Model:
  - $n_i \sim \text{Poisson}(\theta_i)$, \quad $\log(\theta_i) = X_i^T \alpha$
  - $m_i | n_i \sim \text{Gamma}(n_i v_i, \xi_i)$, \quad $\log(\xi_i) = X_i^T \beta + n_i \beta^*$

\[
\log(1/v_i) = X_i^T \gamma + n_i \gamma^*
\]

- Two different set of data
  - Complete data
    \[
    \{(n_1, y_1), \ldots, (n_k, y_k)\}
    \]
  - Summarized data
    \[
    \{(n_1, m_1), \ldots, (n_k, m_k)\}
    \]

- Question: What is efficiency difference between using Complete data and Summarized data
Information Loss

- **On the mean parameter \( \xi_i \) (or \( \beta^+ \))**
  - No information loss:

\[
\{(n_1, m_1), \ldots, (n_k, m_k)\}
\]

is a sufficient statistic for mean parameter \( \xi_i \)

- **On the dispersion parameter \( 1/\nu_i \) (or \( \gamma^+ \))**
  - Information loss:

\[
\{(n_1, m_1), \ldots, (n_k, m_k)\}
\]

is a not sufficient statistic for dispersion parameter \( 1/\nu_i \)

- Minimal sufficient statistic for dispersion parameter \( 1/\nu_i \) (or \( \gamma^+ \)).

\[
\{(n_1, y_1), \ldots, (n_k, y_k)\}
\]

- Assumption: \( \{(n_1, m_1), \ldots, (n_k, m_k)\} \) is available data
Simulation Study: Comparison of Efficiency

- $n_i \sim \text{Pois}(\theta_i)$ with $\log(\theta_i) = 0.5 + 1 \cdot Z_1$.
- True Parameter
  
  $$(\beta_0, \beta_1, \beta_2) = (0.2, 0.1, 0.1) \quad \text{and} \quad (\gamma_0, \gamma_1) = (-0.6, -0.1)$$

- $k = 500$ samples
- 500 repetitions

\[
\sqrt{E[N]} = 1.683142
\]

<table>
<thead>
<tr>
<th></th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean in Model 1</td>
<td>0.198</td>
<td>0.101</td>
<td>0.099</td>
<td>-0.583</td>
<td>-0.107</td>
</tr>
<tr>
<td>Mean in Model 2</td>
<td>0.197</td>
<td>0.101</td>
<td>0.099</td>
<td>-0.593</td>
<td>-0.102</td>
</tr>
<tr>
<td>sd1, sd in Model 1</td>
<td>0.056</td>
<td>0.062</td>
<td>0.009</td>
<td>0.132</td>
<td>0.038</td>
</tr>
<tr>
<td>sd2, sd in Model 2</td>
<td>0.056</td>
<td>0.062</td>
<td>0.009</td>
<td>0.082</td>
<td>0.019</td>
</tr>
<tr>
<td>sd1/sd2</td>
<td>1.002</td>
<td>1.004</td>
<td>1.003</td>
<td>1.605</td>
<td>2.022</td>
</tr>
</tbody>
</table>

Table: Note that $\sqrt{E[N]} = \sqrt{e^{0.5}(e-1)} = 1.683142$
Simulation Study: Comparison of Efficiency

Figure: Comparison of Efficiency in Two Models

Lee, W., Lee, D. and Ahn, J.
Generalized Linear Models in Compound Risk Model with Dependent Structures
Model:
- \( n_i \sim \text{Poisson}(\theta_i) \), \( \log(\theta_i) = X_i^T \alpha \)
- \( m_i | n_i \sim \text{Gamma}(n_i v_i, \xi_i) \), \( \log(\xi_i) = X_i^T \beta + n_i \beta^* \)
  \[ \log(1/v_i) = X_i^T \gamma + n_i \gamma^* \]

Estimation:
- Find \( \alpha, \beta^+, \gamma^+ \) which maximize
  \[ l(\alpha, \beta^+, \gamma^+) = \prod_{i=1}^{k} f_N(n_i | X_i; \alpha) \prod_{n_i > 0} f_M(m_i; X_i, \beta^+, \gamma^+) \]
  \[ = l_1(\alpha) l_2(\beta^+, \gamma^+) \]

Separation of optimization problem:
- Find \( \alpha \) in
  \[ n_i \sim \text{Poisson}(\theta_i), \quad \log(\theta_i) = X_i^T \alpha \]
- Find \( \beta^+, \gamma^+ \) in
  \[ m_i | n_i \sim \text{Gamma}(n_i v_i, \xi_i), \quad \log(\xi_i) = X_i^T \beta + n_i \beta^* \]
  \[ \log(1/v_i) = X_i^T \gamma + n_i \gamma^* \]
Find $\alpha$ in

$$n_i \sim Poisson(\theta_i), \quad \log(\theta_i) = X_i^T \alpha$$

> glm1<-glm(M~ X, family=poisson)

Find $\beta^+$, $\gamma^+$ in

$$m_i | n_i \sim Gamma(n_i v, \xi_i), \quad \log(\xi_i) = X_i^T \beta + n_i \beta^*$$

$$\log(1/v_i) = X_i^T \gamma + n_i \gamma^*$$

> glm2<-dglm(M~ X+N, ~X+N, weights=N,
  family=Gamma(link="log"), dlink="log")
Actual Data Analysis

A motor insurance data from Zhang (2013).
- \( k = 10 \), 296 records, 29 variables
- Data: \((n_i, m_i, X_i)\)

Several (selected) covariates as follows:
- \( n_i, m_i \): the number of claims and average claim amount in the past 5 years.
- travel time \((X1_i)\), car value \((X2_i, \text{ in million})\), income \((X7_i, \text{ in million})\), education \((X8_i)\),
  \(1(\text{low education}) \sim 5(\text{High Education})\), mvr pts \((X9_i)\),
  \(0 \sim 13\)
- \( n_i \sim Poisson(\theta_i) \) where
  \[
  \log(\theta_i) = -0.66 - 3.55X2 - 85.3X7 + 0.03X8 + 0.2X9
  \]
- \( m_i | n_i \sim Gamma(n_i \nu, \xi_i) \) with
  \[
  \log(\xi_i) = 11.5 - 0.002X1 - 4.3N \\
  \log(1/\nu_i) = -0.53 + 0.49N
  \]
Table of Contents

- Introduction
- Compound GLM Model with Dependence
- Analysis of Statistical Estimation
- Conclusion
Conclusion

- Proposed compound model is simple and intuitive.
- Proposed compound model is flexible as it can be divided into severity ($S/N$) part and frequency ($N$) part.
- In the recent literatures: Simple Shared Random Effect (Baumgartner et al., 2015).
- Flexible version of Shared Random Effect.


