# A Loss Reserving Model within the framework of Generalized Linear Models

Jun Zhou and José Garrido*
Concordia University, Montreal, Canada
E-mail: junzhou@mathstat.concordia.ca,
garrido@mathstat.concordia.ca

**Abstract**

Loss reserving is one of the most challenging tasks facing actuaries. Numerous approaches have been developed to give reasonable estimates. Generalized linear models (GLMs) are becoming quickly popular statistical analysis methods to estimate loss reserves. However, most of these models are aggregate reserving methods based on loss development triangles, without using information with regard to the actual claims processes. In this paper we establish a more sophisticated structural reserving method incorporating more detailed information, such as the premium exposure emergence pattern, the loss emergence pattern and the loss development pattern, within the framework of GLMs.

**Keywords:** GLMs, loss reserving, structural reserving method, loss function, loss development function

*This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant 36860–06.

# 1 Introduction

Loss reserving is one of the most challenging tasks facing actuaries. Numerous approaches have been developed to give reasonable estimates. Wiser et al. (2001) provides a detailed introduction of loss reserving. Schmidt (2006) gives a unifying survey of some of the most important methods and models of loss reserving which are based on loss development triangles. Haberman and Renshaw (1996) gives a comprehensive review of the application of generalized linear models (GLMs) to actuarial problems, including loss reserving. GLMs are becoming quickly popular statistical analysis methods to estimate loss reserves. Hoedemakers et al. (2005) constructs bounds for the discounted loss reserves within the framework of GLMs. Verrall (2004) uses a Bayesian parametric model within the framework of GLMs. Most of these models are aggregate reserving methods based on loss development triangles, without using information with regard to the actual claims processes. In this chapter we establish a more complex structural reserving method with more detailed information, such as the premium exposure emergence pattern, the loss emergence pattern and the loss development pattern, within the framework of generalized linear models (GLMs). This model has the following advantages:

- Theoretically it is more accurate than aggregate loss reserving methods based on loss development triangles, because more detailed information is used.
- It gives more flexibility in dealing with unusual or quickly changing situations, as variables are analyzed continuously rather than discretely.
- A discount factor can be added and adjusted to the model easily.
- It provides a mechanism to analyze the effects of each factor in loss reserving separately.
- The model connects the frequency and severity estimations, both in ratemaking and loss reserving, making the work of actuaries more consistent and explainable.

# 2 The Generalized Linear Models

This section provides a short summary of the main characteristics of generalized linear models (GLMs). McCullagh and Nelder (1989) provide a detailed introduction to GLMs. The books by Aitkin et al. (1989) and Dobson (1990) are also excellent references with many examples of applications of GLMs. Hardin and Hilbe (2007) provide a handbook for data analysis with GLMs and GLM extensions. Lee et al. (2007) is a comprehensive reference for GLMs with random effects. Anderson et al. (2005), Coutts (1984), Brockman and Wright (1992) are excellent references for the estimation of the pure risk premium by GLMs.

GLMs are a natural generalization of classical linear models that allow the mean of a population to depend on a linear predictor through a (possibly nonlinear) link function. This allows the response probability distribution to be any member of the exponential family (EF) of distributions.

A GLM consists of the following three components:

1. The response $Y$ has a distribution in the EF, with density function taking the form

$$f(y;\theta,\phi) = \exp\left\{ \int \frac{[y-\mu(\theta)]}{\phi V(\mu)}\, d\mu(\theta) + c(y,\phi) \right\}, \qquad (2.1)$$

   where $\theta$ is called the *natural* parameter, $\phi$ is a *dispersion* parameter, $\mu = \mu(\theta) = \mathbb{E}(Y)$ and $\mathbb{V}(Y) = \phi\, V(\mu)$, for a given variance function $V$ and known bivariate function $c$. The EF is very flexible and can model continuous, binary, or count data.

2. For a random sample $Y_1, \ldots, Y_n$, the linear component is defined as

$$\eta_i = \boldsymbol{X}_i'\boldsymbol{\beta}\,, \qquad i = 1, \ldots, n\,, \qquad (2.2)$$

for some vector of parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ and covariate $\boldsymbol{X}_i = (x_{i1}, \ldots, x_{ip})'$ associated to the observation $Y_i$.

3. A monotonic differentiable *link function* $g$ describes how the expected response $\mu_i = \mathbb{E}(Y_i)$ is related to the linear predictor $\eta_i$

$$g(\mu_i) = \eta_i\,, \qquad i = 1, \ldots, n\,. \tag{2.3}$$

**Example 2.1** *GLMs commonly used in insurance data*

Table 1 below gives the different model components of the GLMs most commonly used in insurance data for observed claim counts or claim severities.

| $Y \sim$ | Normal$(\mu, \sigma^2)$ | Gamma$(\alpha, \beta)$ | Poisson$(\lambda)$ | Bin.$(m, q)/m$ |
|---|---|---|---|---|
| $\mathbb{E}(Y) = \mu(\theta)$ | $\theta = \mu$ | $-\theta^{-1} = \frac{\alpha}{\beta}$ | $e^\theta = \lambda$ | $\frac{e^\theta}{1+e^\theta} = q$ |
| $\mathbb{V}(Y) = V(\mu)\phi$ | $\sigma^2$ | $\frac{1}{\theta^2\,\alpha} = \frac{\alpha}{\beta^2}$ | $e^\theta = \lambda$ | $\frac{q\,(1-q)}{m}$ |
| $V(\mu)$ | $1$ | $\theta^{-2}$ | $e^\theta = \lambda$ | $q(1-q)$ |
| $\phi$ | $\sigma^2$ | $\alpha^{-1}$ | $1$ | $1/m$ |
| $c(y, \phi)$ | $-\frac{1}{2}[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)]$ | $\alpha \ln \alpha y + \ln y - \ln \Gamma(\alpha)$ | $-\ln(y!)$ | $\ln \binom{m}{m\,y}$ |
| *Link g* | identity | reciprocal | log | logit |

Table 1: GLM Examples

Additional examples include inverse Gaussian and negative binomial observations, as well as multinomial proportions, etc. (for details see McCullagh and Nelder, 1989).

Maximum likelihood estimation (MLE) is used to estimate the parameter $\boldsymbol{\beta}$. For an observed independent random sample $y_1, \ldots, y_n$, consider the log–likelihood of $\boldsymbol{\beta}$:

$$l(\boldsymbol{\beta}) = \ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \Big\{ \int \frac{\big[y_i - \mu_i(\theta)\big]}{\phi V(\mu_i)} d\mu_i(\theta) + c(y_i, \phi) \Big\} \tag{2.4}$$

and its derivative:

$$\frac{dl(\boldsymbol{\beta})}{d\boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{dl(\boldsymbol{\beta})}{d\mu_i} \frac{d\mu_i}{d\boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{d\mu_i}{d\boldsymbol{X}_i'\boldsymbol{\beta}} \frac{d\boldsymbol{X}_i'\boldsymbol{\beta}}{d\boldsymbol{\beta}},$$

where

$$\frac{d\mu_i}{d\boldsymbol{X}_i'\boldsymbol{\beta}} = \frac{dg^{-1}(\boldsymbol{X}_i'\boldsymbol{\beta})}{d\boldsymbol{X}_i'\boldsymbol{\beta}} = \frac{1}{g'(\mu_i)}.$$

Hence

$$\frac{dl(\boldsymbol{\beta})}{d\boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{1}{g'(\mu_i)} \boldsymbol{X}_i'. \tag{2.5}$$

Note that if $Y_i$ has a normal distribution, then $g'(\mu_i) = 1$, and $V(\mu_i) = 1$ for all $i$. Setting $\frac{dl(\boldsymbol{\beta})}{d\boldsymbol{\beta}} = 0$ yields $\sum_{i=1}^{n} \boldsymbol{X}_i(y_i - \boldsymbol{X}_i'\boldsymbol{\beta}) = 0$. In other EF cases, no closed form solution is available to this system of $p$ equations. Instead, the maximum likelihood estimator (MLE) is obtained numerically, using iterative algorithms such as the Newton–Raphson or Fisher scoring methods (for details see McCullagh and Nelder, 1989).

# 3 A Loss Reserving Model within the Framework of GLMs

This section gives a detailed description of our loss reserving model within the framework of GLMs.

**Definition 3.1** *The loss function $l(t)$ is a stochastic process which represents the rate at which losses are occurring at time $t$.*

A loss function $l(t)$ tells us how losses occur, which is determined by in–force risk exposure and seasonality of the distribution of the risk exposures. A detailed reference of exposure bases is Bouska (1989). In practice we cannot observe $l(t)$ directly. However, we can approach the expected value of $l(t)$, by in–force exposure, as it directly depends on the premium emergence pattern.

**Definition 3.2** *An aggregate loss $L(t_1, t_2)$ occurred in the time period $(t_1, t_2)$ can be written as*

$$L(t_1, t_2) = \int_{t_1}^{t_2} l(t)dt. \tag{3.1}$$

**Definition 3.3** *The loss development function $D(t)$ is a stochastic process which represents the percentage of losses that are paid within t years after the loss occurrence.*

It is clear that $D(t) = 0$ for $t \leq 0$, and $\lim_{t\to\infty} D(t) = 1$ almost surely. For a given time $T > t$, then $l(t)D(T-t)$ represents the aggregate paid amount at $T$ for losses occurred at time $t$. Assuming that the process has continuous sample paths, by integrating these aggregate paid amounts over $(t_1, t_2)$, we have the following definition (see Figure 1 for a time–diagram reference).

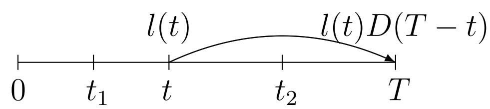

Figure 1: Time Scale

**Definition 3.4** *Given a loss function $l(t)$ and a stochastic loss development function $D(t)$, the aggregate paid losses from the losses incurred in period $(t_1, t_2)$, as developed to time $T \geq t_2 > t_1$, are defined as*

$$L(T, t_1, t_2) = \int_{t_1}^{t_2} l(t)\, D(T-t)\, dt\,. \tag{3.2}$$

The integrals in (3.1) and (3.2) give the ultimate losses and the paid losses incurred in period $(t_1, t_2)$. Their difference, $L(t_1, t_2) - L(T, t_1, t_2)$, is the unpaid losses, or also called loss reserves.

**Definition 3.5** *Given a loss function $l(t)$ and a loss development function $D(t)$, the loss reserves for claims incurred in period $(t_1, t_2)$, as developed to* *time* $T \geq t_2 > t_1$ *are defined as*

$$\begin{aligned} R(T) &= L(t_1, t_2) - L(T, t_1, t_2) \\ &= \int_{t_1}^{t_2} l(t)\,[1 - D(T-t)]\,dt\,. \end{aligned} \tag{3.3}$$

In the case of discounted reserves, we need to add a discount factor in the above analysis. Let $\delta(t)$ be the stochastic force of interest at time $t$. Again, assuming continuous sample paths, then $B(t) = \int_0^t \delta(s)\,ds$ defines the aggregate interest rate in the period of $(0, t)$, and more generally, $B(T+t) - B(t) = \int_T^{T+t} \delta(s)\,ds$ is the aggregate interest rate over $(T, T+t)$.

Here loss reserves are no longer obtained by difference. Instead, first consider a fixed time $t$, where $t_1 < t < t_2$, at which losses of $l(t)$ are incurred. Then $l(t)\,d(s-t)\,ds$ of these will develop at future instant $s > t$, where we assume that $d(t) = D'(t)$, almost surely. Hence, the discounted value at an evaluation date $T$ in $(t, s)$ (see Figure 2 for a time–diagram reference) is given by $e^{-[B(s)-B(T)]}\,l(t)\,d(s-t)\,ds$. Finally, integrating over all future development times $s \in (T, \infty)$ yields the definition below.

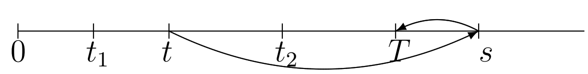


Figure 2: Evaluation Time $T$

**Definition 3.6** *Given a loss function $l(t)$, a stochastic loss development function $D(t)$, a stochastic aggregate interest rate $B(t)$, and any evaluation date $T$, the discounted value at time $T$ of the unpaid loss reserves from period $(t_1, t_2)$, as developed to $T \geq t_2 \geq t_1$, is given by*

$$Z(T) \quad = \quad \int_{t_1}^{t_2} l(t) \int_{T}^{\infty} e^{-[B(s)-B(T)]} \, \frac{\partial}{\partial t} D(s-t) \, ds \, dt \, , \tag{3.4}$$

*almost surely.*

Equations (3.3) and (3.4) give the formulas of the loss reserves and discounted loss reserves, respectively. In fact, note that when the aggregate interest rate $B(t) = 0$, i.e. without discounting, then (3.4) reduces to (3.3), since $\lim_{t\to\infty} D(t) = 1$ almost surely.

To conclude the definition of this loss reserving model, introduce the following assumption to calculate the expected value of the processes in (3.3) and (3.4).

(A1) The loss function $l(t)$, the loss development function $D(t)$ and the force of interest $\delta(t)$ are independent.

Assumption (A1) directly implies the following result.

**Theorem 3.1** *Given $\mathbb{E}[l(t)]$, $\mathbb{E}[D(t)]$ and $\mathbb{E}[B(s)]$ for fixed $t$ and $s$, then*

$$\mathbb{E}[R(T)] = \int_{t_1}^{t_2} \mathbb{E}[l(t)] \left\{1 - \mathbb{E}[D(T-t)]\right\} dt\,, \qquad T \geq t_2 > t_1\,, \quad (3.5)$$

$$\mathbb{E}[Z(T)] = \int_{T}^{\infty} e^{-\mathbb{E}[B(s)]} \int_{t_1}^{t_2} \mathbb{E}[l(t)]\, \mathbb{E}\big[\frac{\partial}{\partial s} D(s-t)\big]\, dt\, ds\,. \quad (3.6)$$

It is clear that the loss reserves only depend on the expected loss function and loss development function. We will use the GLMs to estimate these two functions.

In practice, loss and development functions can be very complex. Even when long historical data is available, the development process itself can change with time. The future force of interest is also unknown. In addition to the assumption (A1) above, the following additional assumptions are needed:

(A2) All policy periods are one year and the amount of exposure to risk of an insurance policy spreads uniformly over the policy period.

(A3) The expected value of the loss development function $D$ is of the form of $\mathbb{E}[D(t)] = 1 - a^{-t}$, where $a > 1$ is a constant.

(A4) The future force of interest is a known constant $\delta \geq 0$.

The average settlement time is a key parameter for loss development. Based on Assumption (A3), above, we can give estimate this parameter within the framework of GLMs.

**Lemma 3.1** *Given Assumption (A3), the expected average loss development time is* $\mathbb{E}(\tau) = \frac{1}{\ln a}$, *for* $a > 1$.

**Proof**: The expected average loss development time is given by:

$$\mathbb{E}(\tau) = \int_0^\infty [1 - D(t)]\, dt = \int_0^\infty \frac{dt}{a^t} = \frac{1}{\ln a}\,.$$

$\square$

The motivation for assumption (A1) is to estimate the expected values of the loss function, $\mathbb{E}[l(t)]$, of the loss development function, $\mathbb{E}[D(t)]$, and of the discount rate $\mathbb{E}[B(s)]$, separately. These can then be substituted into (3.5) and (3.6) to estimate the expected loss and discounted reserves. Assumption (A2) can be relaxed for seasonality or other distributional patterns. Assumption (A3) states that $\mathbb{E}[D(t)]$ takes the form of the cumulative distribution function (CDF) of an exponential distribution, which is appropriate for high–frequency/low–severity business lines, such as auto insurance. For heavy tail cases such as liability claims, the CDF of a Weibull or Pareto distribution are good candidates for the loss development function.

With all these assumptions, we could estimate the expected value of the loss function $l(t)$ and the loss development function $D(t)$ within the framework of GLMs. The key aspects are the modelling of the number of claims $n$ and the claim severity as independent responses of GLMs.

Consider a set of observed claims under some classification system. Let cell $i$ denote a generic class defined by this classification system. The GLMs for frequency and severity can be written as follows. Let

- $f_i$ be the frequency in cell $i$,
- $z_i$ be the claim severity in cell $i$,
- $\tau_i$ be the average settlement time in cell $i$,
- $w_i(t)$ be the number of exposures (policyholders) in cell $i$ at time $t$,
- $\eta_{fi}$, $\eta_{zi}$ and $\eta_{\tau i}$ be linear predictors of claim frequency, severity and average settlement time in cell $i$, respectively.
- $g_f$, $g_z$ and $g_\tau$ are the GLM link functions for the claim frequency, severity and average settlement time, respectively.

Then, for each cell $i$, the GLMs give the expected value of the claim frequency, severity and average settlement time as:

$$\mathbb{E}(f_i) = g_f^{-1}(\eta_{fi}), \tag{3.7}$$

$$\mathbb{E}(z_i) = g_z^{-1}(\eta_{zi}), \tag{3.8}$$

$$\mathbb{E}(\tau_i) = g_\tau^{-1}(\eta_{\tau i}). \tag{3.9}$$

Combining (3.9) with Assumption (A3) and Lemma 3.1 gives

$$g_\tau^{-1}(\eta_{\tau i}) = \frac{1}{\ln a_i} \qquad \Rightarrow \qquad a_i = \exp\left\{\frac{1}{g_\tau^{-1}(\eta_{\tau i})}\right\}.$$

Now with Assumption (A2), we get that the expected total loss rate $\mathbb{E}[l_i(t)]$ and loss development function $\mathbb{E}[D_i(t)]$ in cell $i$ at time $t$ are:

$$\mathbb{E}[l_i(t)] \;=\; w_i(t)\, g_f^{-1}(\eta_{fi})\, g_s^{-1}(\eta_{si})\,, \tag{3.10}$$

$$\mathbb{E}[D_i(t)] \;=\; \exp\left\{\frac{1}{g_\tau^{-1}(\eta_{\tau i})}\right\}. \tag{3.11}$$

Then (3.5) and (3.6) give the expected loss and discounted loss reserves in cell $i$:

$$\mathbb{E}[R_i(T)] \;=\; \int_{t_1}^{t_2} l_i(t)\left\{1 - \mathbb{E}[D_i(T-t)]\right\} dt\,. \tag{3.12}$$

$$\mathbb{E}[Z_i(T)] \;=\; \int_T^\infty e^\delta(s-t) \int_{t_1}^{t_2} \mathbb{E}[l_i(t)]\, \mathbb{E}[\frac{\partial}{\partial s} D(s-t)]\, dt\, ds\,. \tag{3.13}$$

Hence, summing over all cells in the portfolio we have the total loss and discounted loss reserves

$$\mathbb{E}[R(T)] \;=\; \sum_i \mathbb{E}[R_i(T)]\,, \tag{3.14}$$

$$\mathbb{E}[Z(T)] \;=\; \sum_i \mathbb{E}[Z_i(T)]\,. \tag{3.15}$$

# 4 Conclusion

This paper establishes a structural loss reserving model within the framework of GLMs. It enables the estimation of the loss reserves on an individual basis. Compared to traditional models based on loss development triangles, this GLM approach gives more detailed information, such as the premium exposure emergence pattern, the loss emergence pattern and the loss development pattern. The GLM model also gives a more detailed estimation of the loss reserves.

# References


[1] Aitkin, M., Anderson, D., Francis, B. and Hinde, J. (1989) *Statistical Modelling in GLIM*. Oxford Science Publications, Oxford.

[2] Dobson, A. (1990) *An Introduction to Generalized Linear Models*. Chapman and Hall, London.

[3] Haberman, S. and Renshaw, A.E. (1996) Generalized linear models and actuarial science. *The Statistician* **45**(4), 407–436.

[4] Hardin, J.W. and Hilbe, J.M. (2007) *Generalized Linear Models and Extensions*, 2nd Edition. Chapman and Hall/CRC, Boca Raton.

[5] Lee, Y., Nelder, J.A. and Pawitan, Y. (2007) *Generalized Linear Models With Random Effects*. Chapman and Hall/CRC, Boca Raton.

[6] McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*. Chapman and Hall, New York.

[7] Whitney, A.W. (1918) The theory of experience rating. *Proc. of the Casualty Actuarial Society* **4**, 274–292.

[8] Wolfinger, R. and O'Connell, M. (1993) Generalized linear mixed models: a pseudo–likelihood approach. *Journal of Statistical Computation and Simulation* **4**, 233–243.

[9] Wiser, R.F., J.E. Cockley and A. Gardner, 2001. Loss reserving. *Foundations of Casualty Actuarial Science*, 4th Edition. Casualty Actuarial Society, Virginia.

Corresponding author: Jose Garrido
Department of Mathematics and Statistics
Concordia University
1455 de Maisonneuve Blvd. West
Montreal, Qc
H3G 1M8, Canada