# Prediction of Outstanding IBNR Liabilities Using Delay Probabilities

**Fatemeh Atatalab[1], Amir Teimour Payandeh Najafabadi[2]**

[1] Actuarial Science Department, Mathematics Faculty, Shahid Beheshti Uneversity, Tehran, Iran
f_atatalab@sbu.ac.ir

[2] Actuarial Science Department, Mathematics Faculty, Shahid Beheshti Uneversity, Tehran, Iran
amirtpayandeh@sbu.ac.ir

**Abstract:**
An important question in non life insurance research is the estimation of number of future payments and corresponding amount of them. A loss reserve is the money set aside by insurance companies to pay policyholders claims on their policies. The policyholder behavior for reporting claims after its occurrence have significant effect on the costs of the insurance company. This article considers the problem of predicting the amount and number of claims that have been incurred but not reported, say IBNR. Using the delay probabilities in monthly level, calculated by the Zero-Inflated Gamma Mixture distribution, it predicts IBNR's loss reserve. The model advantage in the IBNR reserve is insurers can predict the number of future claims for each future date. This enables them to change the claim reporting process. The practical applications of our findings are applied against a third party liability (TPL) insurance loss portfolio. Additional information about claim can be considered in the loss reserving model and making the prediction of amount more accurate.

*Keywords:* Insurance; Loss reserve; EM algorithm; Zero-Inflated Gamma Mixture distribution.
*JEL Classifications:* G22.

[2] Corresponding author
Received: 2021-05-02 Approved: 2021-05-30

# Introduction

A loss reserve is the money set aside by insurance companies to pay policyholders claims on their policies. Some of these claims may be settled long after the policy has expired. Insurance companies have to hold loss reserves for losses that are incurred but not yet paid. Therefore, we have two types of loss reserves, incurred but not reported (IBNR) and reported but not settled (RBNS).

In recent papers the double chain ladder approach is used to estimate the claim reserve through a micro-level approach of the claim development process, based on the number of reported claims and the amount of payments. For example, Verrall et al. (2010) used paid claims and also the numbers of reported claims to predict the IBNR and RBNS loss reserve separately. They built a model for aggregate paid claims from basic principles at the level of individual data. Martinez-Miranda et al. (2012) extended Verrall et al. (2010) model. They combined the upper triangles of count data with the paid data and introduced the Double Chain Ladder (DCL) model. Martinez-Miranda et al. (2013a) considered a Double chain ladder focusing on two specific types of prior knowledge: zero-claims for each underwriting year and relationship between the development of the claim and its mean severity.

Martinez-Miranda et al. (2013b) used a micro-level approach to predict the number of IBNR claims. Their continuous chain-ladder setting can be applied to data recorded in continuous time, although it is illustrated in the paper on data aggregated on a monthly level. Antonio and Plat (2014) proposed a micro-level model and used detailed information of the time of occurrence of the claim, the delay between occurrence and reporting to insurance company, the occurrences of payments and their sizes, and the final settlement. Verrall and Wüthrich (2016) constructed an inhomogeneous marked Poisson process with a monthly piecewise constant intensity and a weekday seasonal occurrence pattern.

Verbelen et al. (2019) presented a flexible regression framework to jointly estimate the occurrence and reporting of events from data at daily level. Badescu et al. (2019) proposed a marked Cox process and showed some desirable properties of Badescu et al. (2016)'s findings. They employed an EM algorithm and showed that the fitting algorithm can be implemented at a reasonable computational cost. Crevecoeur et al. (2018, 2019) considered IBNR claim reserve due to a delay and modelled the time between the occurrence and observation of the event. They proposed a granular model for the heterogeneity in the observation delay based on the occurrence day of the event and on calendar day effects in the observation process, such as weekday and holiday effects.

The main aim of this article is to obtain a computationally reasonable expression for predictors of both IBNR and RBNS loss reserve and their mean square errors of predictions based on a history up to today's time $t$. To solve the prediction problem, this article decomposes outstanding claims as IBNR and RBNS. Then, it considers a Zero-Inflated Gamma Mixture distribution for random reporting delay and a discrete random variable for the settlement delay. Using updated observation at time $t$ unknown parameters are estimated under the maximum likelihood approach, then both IBNR and RBNS outstanding claims are predicted.

The article is organized as follows: Section 2 presents theoretical foundation of the article. Section 3 shows how the previous section findings can be applied in practice. Suggestions and concluding remarks are given by Section 4.

# Theoretical foundation

It is common for claims to be reported to insurance company long after they were incurred. The reporting delay is an important driver in the risk management strategy of the insurer, whose core business is underwriting risks.

This article considers delay time on a monthly level, therefore, a considerable amount of reporting delay time will be zero, while some of them stand far from others. These two facts justified implementation of a zero-inflated and heavy distribution. For some practical reasons, we assume that the random reporting delay has been distributed according to the Zero-Inflated Gamma Mixture distribution, say ZIGM. We consider the ZIGM, as an appropriate distribution for the random reporting delay $U$ which is given by the following definition.

**Definition 0.1.** A random variable $U$ has the Zero-Inflated Gamma Mixture, say ZIGM, distribution if its density function is

$$g(u, \psi) = \pi I_{\{0\}}(u) + (1 - \pi) \sum_{h=1}^{k} w_h Gamma(\alpha_h, \theta_h) I_{(0,\infty)}(u), \qquad (1)$$

where $0 \leq \pi \leq 1$, $\sum_{h=1}^{k} w_h = 1$, $\psi = (\pi, w_1, \cdots, w_k, \alpha_1, \cdots, \alpha_k, \theta_1, \cdots, \theta_k)$ and $Gamma(\alpha_h, \theta_h)$ stands for the Gamma density function.$I_{\{A\}}$ denotes the indicator function of event A.

See Gharib (1995) for some properties of the Gamma Mixture distribution.

The following represents assumptions that we consider hereafter now. Assume that:

$A_1$) The total number of claims related to the accident time $i$, say $N_i$, follows from a non-homogeneous Poisson process with finite intensity $\lambda_i$;

$A_2$) Random reporting delay $U$ has been distributed according to the ZIGM distribution, given by Definition (0.1);

$A_3$) Discrete random settlement delay $D$ has probability mass function $q_l = P(D = l)$, for $l = 0, \cdots, d$;

$A_4$) The individual payments $Y^{(k)}_{i,j-l,l}$ are iid random variables with $E\left(Y^{(k)}_{i,j-l,l}\right) = \mu < \infty$ and $Var\left(Y^{(k)}_{i,j-l,l}\right) = \sigma^2 < \infty$;

$A_5$) $\mathcal{F}_t$ stands for the updated filtration based upon the past information at observation time $t$.

$A_6$) Claims are settled with a single payment.

As mentioned above, the outstanding claims represent claims which occurred at accident time $i$ and reported to the insurance company $j$ unit time later. But for some practical reasons, they paid (or settled) $l$ unit time after $j$. Let $N^{paid}_{i,j}$, for $i, j = 0, \cdots, I$, denotes total number of claims that occurred at accident time $i$ and fully paid before or at time $i + j$. $I$ denotes the last accident (or development) year. Under this setting

$$N^{paid}_{i,j} = \sum_{l=0}^{min(j,d)} N^{paid}_{i,j-l,l}$$

where $N^{paid}_{i,j-l,l}$, represents total number of claims that occurred at accident time $i$, reported at $i + j - l$ time and paid at $i + j$ time and $d$ is the maximum delay that insurance company settle a claim after reporting to it. Moreover suppose that, for $k = 1, \cdots, N^{paid}_{i,j-l,l}$, i.i.d random variable $Y^{(k)}_{i,j-l,l}$ stands for size of the $k^{th}$ individual payments that occurred at accident time $i$, reported at $i + j - l$ time and paid at $i + j$ time.

Therefore, the total payments at time $i + j$ is

$$X_{ij} = \sum_{l=0}^{\min\{j,d\}} \sum_{k=1}^{N^{paid}_{i,j-l,l}} Y^{(k)}_{i,j-l,l},$$

where the maximum payment delay $d$ can be chosen from evaluation process for each insurance companies.

For the $i+j \leq I$, the payments $X_{ij}$ are known. But for the $i+j > t \geq I$, the total payments $X_{ij}$ are unknown and one has to predict. For such situation, there is two types of unknown claims, one does not reported yet, say $X^{IBNR}_{ij}$, another one has been reported but not fully paid, say $X^{RBNS}_{ij}$. We focuse on IBNR part and we have

$$X^{IBNR}_{ij} = \sum_{l=0}^{i+j-I-1} \sum_{k=1}^{N^{paid}_{i,j-l,l}} Y^{(k)}_{i,j-l,l}, \quad \forall\ i + j \geq I.$$

*Remark* 0.2. It is worthwhile mentioning that in a situation that: one would like to make inference about number of outstanding claims rather than size of outstanding claims, he/she can consider $Y^{(k)}_{i,j-l,l} = 1$.

Under mentioned assumptions and using the Poisson process properties, one may conclude that: (**1**) the total number of claims that occurred at accident time $i$ and paid at time $i + j$, $N^{paid}_{i,j}$ follows from a nonhomogeneous Poisson process with intensity $\lambda_i p_j$; (**2**) the total number of payments at time $i + j$, related to accident time $i$, $N^{paid}_{i,j-l,l}$, follows from a nonhomogeneous Poisson process with intensity $\lambda_i p_{j-l} q_l$, where, for $l = 0, \cdots, d$, delay probability $p_j$ is

$$p_j = P(j \leq U \leq j + 1) = \int_j^{j+1} dG_{ZIGM}(u, \psi), \tag{2}$$

and $G_{ZIGM}(;)$ stands for the cumulative distribution function of the ZIGM distribution.

At observation time $t$, where $i+j > t \geq I$, conditional expectation of the total payments $X_{ij}$ given updated filtration $\mathcal{F}_t$ can be calculated as the following.

$$\begin{aligned} X_{ij}^{IBNR}(t) &= E(X_{ij}|\mathcal{F}_t) \\ &= E(\sum_{l=0}^{i+j-I-1} \sum_{k=1}^{N_{i,j-l,l}^{paid}} Y_{i,j-l,l}^{(k)}|\mathcal{F}_t) \\ &= \sum_{l=0}^{i+j-I-1} E(N_{i,j-l,l}^{paid}|\mathcal{F}_t)E(Y_{i,j-l,l}^{(k)}) \\ &= \sum_{l=0}^{i+j-I-1} \lambda_i^{(t)} p_{j-l}^{(t)} q_l^{(t)} \mu^{(t)}. \end{aligned}$$

To use the above finding, all model parameters must to be given, or has to be estimated based on available information in $\mathcal{F}_t$. The next section considers this issue.

## Parameter estimation

The log-likelihood function based upon the observed data up to observation date $t$ is

$$logL = \sum_i \sum_j (n_{i,j} \log(\lambda_i) + n_{i,j} \log(p_j) - \lambda_i p_j - log(n_{i,j}!)). \tag{3}$$

Therefore, the maximum likelihood estimator for $\lambda_i$ is

$$\hat{\lambda}_i = \sum_{j=i}^{t} n_{ij} / \sum_{j=i}^{t} p_j.$$

Substituting $\hat{\lambda}_i$ in the above log-likelihood function leads to

$$logL \propto \sum_{i=1}^{t} \sum_{j=i}^{t} [n_{ij}(log(\sum_{j=i}^{t} n_{ij}) - log(\sum_{j=i}^{t} p_j) - \sum_{j=i}^{t} n_{ij}] + \sum_{i=1}^{t} \sum_{j=i}^{t} n_{ij} log(p_j) \tag{4}$$

The above log-likelihood function can be understood as a log-likelihood function for truncated reporting delay random variable where truncation point $(t - i)$ is the maximal observed delay for a claim that incurred on accident time $i$.

The log-likelihood given by Equation (4) depends on the parameters of both the Poisson model for claim occurrence and the reporting delay distribution. It is not straightforward and reasonable to calculate the maximum likelihood estimation with respect to $p_j$. The complications of parameters estimation are simplified by applying the expectation-maximization, say EM, algorithm for delay probability.

The EM algorithm consists of two major steps:

**E-step)** Expectation step: Choose initial values for the set of parameters $\psi$ and compute the expected value of the $\mathbf{u}$ data. After the first iteration of the EM, the new estimators $\psi^1$ for $\psi$ are obtained. In this step, we estimate unobserved data.

**M-step)** Maximization step: Use the data arrived from the E-step, an updated maximum likelihood estimate of unknown parameters.

Please see da Silva and Yongacoglu (2015) or Moon (1996) for more details about the EM algorithm.

To estimate the $p_j$, at the first one has to estimate the unknown parameters $\psi$, given by Definition (0.1).

Consider the random variable $U$ has a mixture of k-Gamma distribution. Now, we introduce the EM algorithm in the context of Gamma mixture models. To find the maximum likelihood estimators with the EM algorithm, we can introduce a sample $\mathbf{v} = (v_1, \cdots, v_m)$ of the random variable $V$ which indicate which of the $k$ component densities was observed for each $m$; $v_m \in \{1, \cdots, k\}$. We call $\{U, V\}$ the complete data set, and we say $U$ is incomplete. Now suppose that $f_{U,V}$ stands for the joint density function of $U$ and $V$ have joint density. Therefore, the log-likelihood is given by

$$log[L(\psi|\mathbf{u}, \mathbf{v})] = log[\prod_{m=1}^{n} f_{U,V}(u_m, v_m; \psi)].$$

And for given $v_m$, we have

$$log[L(\psi|\mathbf{u}, \mathbf{v})] = \sum_{m=1}^{n} log[w_{v_m} f_{v_m}(v_m; \psi_{v_m})].$$

We guess the parameters $\psi_g = (w_1^g, \cdots, w_m^g, \alpha_1^g, \cdots, \alpha_m^g, \theta_1^g, \cdots, \theta_m^g)$ of the mixing density. Now, we use the EM algorithm to update the parameters at each step, i.e.

$$\begin{aligned} Q(\psi, \psi^g) &= E_V\{log[L(\psi|\mathbf{u}, \mathbf{v})]|\mathbf{u}, \psi^g\} \\ &= \sum_V \sum_{m=1}^{n} log[w_{v_m} f_{v_m}(u_m; \theta_{v_m})] \prod_{m=1}^{n} f_{V|U=u_m}(v_m; \psi^g) \end{aligned}$$

from a series of simplifying steps, the objective function is

$$\begin{aligned} Q(\psi, \psi^g) = &\sum_{h=1}^{k} \sum_{m=1}^{n} log(w_h) f_{V|U=u_m}(h, \psi^g) \\ &+ \sum_{h=1}^{k} \sum_{m=1}^{n} log[f_h(U_m; \theta_h)] f_{V|U=u_m}(h, \psi^g) \end{aligned}$$

where $f_{V|U=u_i}(h, \psi^g) = A_h / \sum_{j=1}^{k} A_j$ and $A_h = w_h \theta_{h(g)}^{\alpha_{h(g)}} u^{\alpha_{h(g)}-1} e^{-u\theta_{h(g)}} / \Gamma(\alpha_{h(g)})$.
Moreover, $log[f_h(u_m; \theta_h)] = \alpha_h log(\theta_h) + (\alpha_h - 1) log(u_m) - \theta_h u_m - log(\Gamma(\alpha_h))$.

This problem can be solved analytically. We differentiate $Q(h, \psi^g)$ with respect to each parameter, set the expressions equal to zero and solve for parameter. For each $w_h$ we have the restriction $\sum_{h=1}^{k} w_h = 1$. Thus, we employ a lagrangian method with lagrange multiplier parameter $\beta$ and obtain

$$\begin{aligned} \frac{\partial[Q(\psi, \psi^g) + \beta(\sum_{h=1}^{k} w_h - 1)]}{\partial w_h} = 0 \iff & \frac{1}{w_h} \sum_{m=1}^{n} f_{V|U=u_m}(h, \psi^g) + \beta = 0 \\ \iff & w_h = \frac{-\sum_{m=1}^{n} f_{V|U=u_m}(h, \psi^g)}{\beta}. \end{aligned}$$

Summing over $h$ leads to $\sum_{h=1}^{k} \sum_{m=1}^{n} f_{V|U=u_m}(h, \psi^g) = n$. Using the fact that $\sum_{h=1}^{k} w_h = 1$, we get $\beta = -n$, therefore, in each iteration g of the algorithm, for each $w_h$, the MLE is

$$\hat{w}_h = \frac{\sum_{m=1}^{n} f_{V|U=u_m}(h, \psi^g)}{n}$$

and

$$\hat{\theta}_h = \frac{\sum_{m=1}^{n} u_m f_{V|U=u_m}(h, \psi^g)}{\hat{\alpha}_h^{MLE} \sum_{m=1}^{n} u_m f_{V|U=u_m}(h, \psi^g)}.$$

The MLE for each $\alpha_h$ does not have an explicit solution, therefore it has to be found using numerically.

Now, we estimate other parameters, $q_l$ and $\phi$. Based on assumptions $A_1 - A_6$ , the mass function of $N_{ijl}^{paid}$ given $N_{ij}$ follows a multinomial distribution with probabilities $q_l$. The settlement delay probabilities, $q_l$, can be found through an MLE method. Given observed values, the log-likelihood function, denoted by $l$, is

$$l = \sum_{i=0}^{I}\sum_{j=0}^{I-1}\sum_{l=0}^{min\{j,I-1\}} log(n_{ij}!) - \sum_{i=0}^{I}\sum_{j=0}^{I-1}\sum_{l=0}^{min\{j,I-1\}} log(n_{i,j-l,l}!) + \sum_{i=0}^{I}\sum_{j=0}^{I-1}\sum_{l=0}^{min\{j,I-1\}} n_{i,j-l,l} log(q_l)$$

Using the fact that $\sum_{l=0}^{I-1} q_l = 1$, the above log-likelihood function can be restated in context of the lagrange method with lagrange multiplier parameter $\beta$ as the following,

$$l^* = \sum_{i=0}^{I}\sum_{j=0}^{I-1}\sum_{l=0}^{min\{j,I-1\}} log(n_{ij}!) - \sum_{i=0}^{I}\sum_{j=0}^{I-1}\sum_{l=0}^{min\{j,I-1\}} log(n_{i,j-l,l}!) + \sum_{i=0}^{I}\sum_{j=0}^{I-1}\sum_{l=0}^{min\{j,I-1\}} n_{i,j-l,l} log(q_l) - \beta(1 - \sum_{l=0}^{I-1} q_l).$$

Taking partial derivative with respect to $\beta$ and $q_l$, a straightforward calculation along the fact that $\sum_{l=0}^{I-1} q_l = 1$ lead to

$$\hat{\beta} = \sum_{i=1}^{I}\sum_{j=l}^{I-1}\sum_{l=0}^{min\{j,I-1\}} n_{i,j-l,l}$$

$$\hat{q}_l = \frac{\sum_{i=1}^{I}\sum_{j=l}^{I-1} n_{i,j-l,l}}{\sum_{i=1}^{I}\sum_{j=0}^{I-1}\sum_{l=0}^{min\{j,I-1\}} n_{i,j-l,l}}.$$

# A Practical Application

In this section we consider a material damage of motor third party liability insurance claim portfolio from a private insurance company of Iran. We

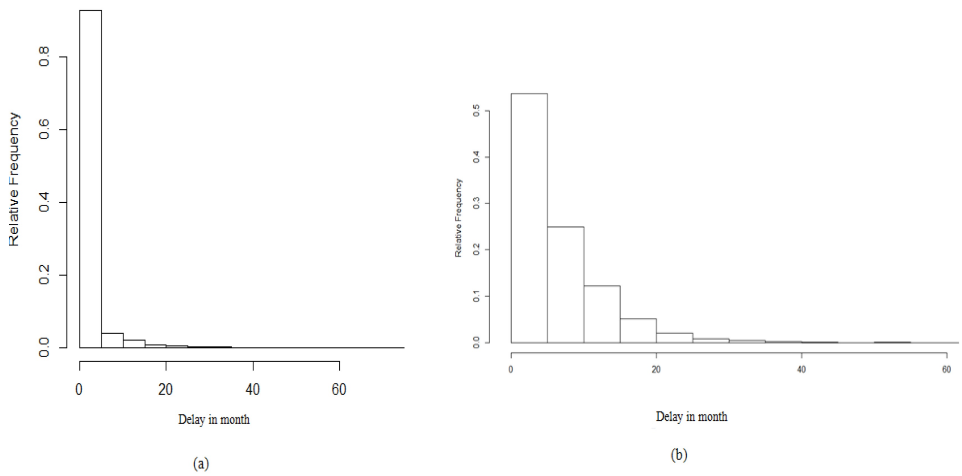


Figure 1: Reporting delay relative frequency histogram for material damage observed between from 2012-03-20 and 2019-03-19: Panel (a) represents all claims and Panel (b) represents claims that reported to the insurance company more than 30 days after occurrence time.

have 85649 claims in our data set. The observation period consists in calender years 20-March-2012 till 19-March-2019.

After a primary investigation and removing illogical events, such as transaction before occurrence, recovery of claims and allocated loss adjuster expenses, we just trusted information about 66346 claims.

As mentioned before, we calculate reporting (and settlement) delay in monthly scale. Figure (1.a and 1.b) illustrates such reporting delay in two cases, all claims and claims that reported more than 30 days. As Figure (1.a) illustrates, there is a considerable amount of zeros for claims which reported in the first month. This observation justifies using the zero-inflation distribution for random reporting delay.

Now, we employe two well known the Kolmogorove-Smirnov and the Cramér-von Mises test to make a decision about the following hypothesis test.

$$H_0 : \text{The random reporting delay has been distributed according to a ZIGM distribution}$$

The p-value of these two tests are 0.1695 and 0.1458, respectively, therefore the null hypothesis, $H_0$, has been accepted at confidence level 0.95.

Using *gammamixEM* function in *mixtools* package of the R software, releases that a zero-inflated two-Gamma mixture distribution is an appropriate distribution for the random reporting delay. More precisely, the density function for the random reporting delay $U$ is

$$0.56 * I_{\{0\}}(u) + 0.09 * Gamma(1.24, 0.38) I_{(0,\infty)}(u) \\ + 0.35 * Gamma(1.41, 5.53)) I_{(0,\infty)}(u).$$

Table 1 represents estimation for claim reporting delay probabilities $p_k$, non-homogeneous Poisson intensity $\lambda_k$ and settlement delay probabilities $q_{k-1}$.

The policyholder behavior for reporting claims after its occurrence have significant effect on the costs of the insurance company. As mentioned before, in our data set, the most of policyholders tend to report the claim to the insurance company less than 30 days. As a result, based on the above estimators the number of IBNR claims for next 12 months will be $\hat{N}^{IBNR} = \sum_{i,j} \hat{\lambda}_i \hat{p}_{ij} = 192$. We know the actual count for IBNR claims in calender year 20-March-2012 till 19-March-2020 is 248.

To predict the corresponding IBNR amount reserve we estimate the mean of an individual claims severity, $\hat{\mu} = 13$ million IRR. The third column in Table 1, we have the maximum likelihood estimator for the settlement delay probability, $q_l$ (the numbers round to four decimal). It shows that 67% of automobile material damage claims are settled in the one month after they reported to insurance company. These are the cheapest claims based on their average costs calculations.

Using these parameter estimates, the IBNR reserve for the next 12 months is 2270 million IRR and the actual amount is 6150 million IRR. This differences is because of the large economic inflation in claim amount severity.

# Conclusion and suggestion

An important question in non-life insurance research is the estimation of number of future payments and corresponding amount of them, i.e. loss reserves. In this article, we study IBNR claims using claims amounts and claim counts and propose a Zero-Inflated Gamma Mixture model for estimation of reporting delay probabilities in monthly level. Also, we consider settlement delay probabilities in our model. We belive that

Table 1: Estimation of claim reporting delay probabilities $p_k$, non-homogeneous Poisson intensity $\lambda_k$ and settlement delay probabilities $q_{k-1}$.

| $k$ | $p_k$ | $\lambda_k$ | $q_{k-1}$ | $k$ | $p_k$ | $\lambda_k$ | $q_{k-1}$ | $k$ | $p_k$ | $\lambda_k$ | $q_{k-1}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.5608 | 3.9437 | 0.6736 | 33 | 0 | 1704.662 | 0.0015 | 65 | 0 | 712.8507 | 0 |
| 2 | 0.0435 | 51.268 | 0.0602 | 34 | 0 | 1616.915 | 0.0014 | 66 | 0 | 1051.983 | 0 |
| 3 | 0.0287 | 105.4938 | 0.0222 | 35 | 0 | 1680.014 | 0.0014 | 67 | 0 | 1097.336 | 0 |
| 4 | 0.0212 | 222.8187 | 0.0142 | 36 | 0 | 1396.068 | 0.0009 | 68 | 0 | 936.6313 | 0 |
| 5 | 0.0153 | 194.227 | 0.0111 | 37 | 0 | 1353.673 | 0.0009 | 69 | 0 | 961.2813 | 0.0002 |
| 6 | 0.011 | 275.0727 | 0.0114 | 38 | 0 | 1218.602 | 0.0007 | 70 | 0 | 946.4949 | 0 |
| 7 | 0.0078 | 335.214 | 0.0116 | 39 | 0 | 1161.418 | 0.0007 | 71 | 0 | 918.8925 | 0 |
| 8 | 0.0055 | 416.0598 | 0.0115 | 40 | 0 | 962.2615 | 0.0005 | 72 | 0 | 967.209 | 0 |
| 9 | 0.0038 | 403.2428 | 0.0124 | 41 | 0 | 1147.615 | 0.0007 | 73 | 0 | 935.6667 | 0 |
| 10 | 0.0027 | 437.7501 | 0.0127 | 42 | 0 | 879.1205 | 0.0003 | 74 | 0 | 826.2512 | 0 |
| 11 | 0.0019 | 503.807 | 0.0124 | 43 | 0 | 1006.628 | 0.0003 | 75 | 0 | 738.5181 | 0 |
| 12 | 0.0013 | 627.0474 | 0.0117 | 44 | 0 | 905.0779 | 0.0004 | 76 | 0 | 842.0805 | 0 |
| 13 | 0.0009 | 509.7225 | 0.0116 | 45 | 0 | 914.9372 | 0.0003 | 77 | 0 | 840.1545 | 0 |
| 14 | 0.0006 | 601.4134 | 0.0108 | 46 | 0 | 887.3313 | 0.0002 | 78 | 0 | 869.8067 | 0 |
| 15 | 0.0004 | 589.5824 | 0.0101 | 47 | 0 | 909.0216 | 0.0002 | 79 | 0 | 766.3469 | 0 |
| 16 | 0.0003 | 681.2733 | 0.0106 | 48 | 0 | 926.7682 | 0.0003 | 80 | 0 | 781.2723 | 0 |
| 17 | 0.0002 | 656.6252 | 0.0087 | 49 | 0 | 799.5841 | 0.0004 | 81 | 0 | 815.0098 | 0 |
| 18 | 0.0001 | 766.0627 | 0.0082 | 50 | 0 | 738.4568 | 0.0001 | 82 | 0 | 700.7997 | 0 |
| 19 | 0.0001 | 770.9923 | 0.0075 | 51 | 0 | 866.6269 | 0.0001 | 83 | 0 | 656.7171 | 0 |
| 20 | 0.0001 | 855.7817 | 0.0068 | 52 | 0 | 820.2885 | 0.0002 | 84 | 0 | 636.4467 | 0 |
| 21 | 0 | 904.092 | 0.0064 | 53 | 0 | 871.5565 | 0.0001 | 85 | 0 | 620.3047 | 0 |
| 22 | 0 | 979.0222 | 0.0058 | 54 | 0 | 862.6832 | 0.0001 | 86 | 0 | 669.8068 | 0 |
| 23 | 0 | 1059.868 | 0.0051 | 55 | 0 | 915.9231 | 0 | 87 | 0 | 636.5304 | 0 |
| 24 | 0 | 1335.927 | 0.0052 | 56 | 0 | 845.9226 | 0.0001 | 88 | 0 | 716.256 | 0 |
| 25 | 0 | 1221.559 | 0.0047 | 57 | 0 | 926.7683 | 0.0001 | 89 | 0 | 736.5403 | 0 |
| 26 | 0 | 1369.448 | 0.0037 | 58 | 0 | 935.6417 | 0.0001 | 90 | 0 | 799.5841 | 0 |
| 27 | 0 | 1548.886 | 0.0034 | 59 | 0 | 929.7262 | 0 | 91 | 0 | 738.4568 | 0 |
| 28 | 0 | 1727.338 | 0.003 | 60 | 0 | 863.6694 | 0.0001 | 92 | 0 | 866.6269 | 0 |
| 29 | 0 | 1730.296 | 0.003 | 61 | 0 | 944.5153 | 0 | 93 | 0 | 820.2885 | 0 |
| 30 | 0 | 1879.171 | 0.0025 | 62 | 0 | 814.3734 | 0 | 94 | 0 | 871.5565 | 0 |
| 31 | 0 | 1948.185 | 0.0025 | 63 | 0 | 933.6705 | 0 | 95 | 0 | 862.6832 | 0 |
| 32 | 0 | 1708.606 | 0.0023 | 64 | 0 | 943.5302 | 0 | 96 | 0 | 915.9231 | 0 |

use reporting and settlement delay probabilities allows one to get exact prediction for future payments. Given the past observation, we study prediction of future payments (in number and amount) and their prediction errors and derived reasonable experssions for them. The model advantage in the IBNR reserve is insurers can predict the number of future claims for each future date. This enables them to change the claim reporting process.

The approach proposed in this paper can be improved with additional information about claim, such as accident year inflation rate, the seasons that claims occurs, the zone of accident, etc. These characteristics can be considered in the loss reserving model and making the prediction of amount more accurate.

## Bibliography


[1] K. ANTONIO, R. PLAT, *Micro-level stochastic loss reserving for general insurance*, Scandinavian Actuarial Journal, 7(2014), 649-669.

[2] A.L. BADESCU, X.S. LIN, D. TANG, *A marked Cox model for the number of IBNR claims: theory*, Insurance: Mathematics and Economics, 69(2016), 29-37.

[3] A.L. BADESCU, T. CHEN, X.S. LIN, AND D. TANG, *A marked COX model for the number of IBNR claims: estimation and application*, ASTIN Bulletin: The Journal of the IAA, 49(2019), 709-739.

[4] J. CREVECOEUR, K. ANTONIO, AND R. VERBELEN, *A Time Change Strategy to Model Reporting Delay Dynamics in Claims Reserving*, The 3rd Europen Actuarial Journal conference, (2018), KU Leuven, Belgium.

[5] J. CREVECOEUR, K. ANTONIO, AND R. VERBELEN, *Modeling the number of hidden events subject to observation delay*, European Journal of Operational Research, 277(2019), 930-944.

[6] V.R. DA SILVA, A. YONGACOGLU, *EM algorithm on the approximation of arbitrary pdfs by gaussian, gamma and lognormal mixture distributions*, The 7th IEEE Latin-American Conference on Communications, LATINCOM, IEEE, (2015).

[7] M. GHARIB, *Two characterisations of a gamma mixture distribution*, Bulletin of the Australian Mathematical Society, 52(1995), 353-358.

[8] M.D. MARTINEZ-MIRANDA, J.P. NIELSEN, AND R.J. VERRALL, *Double chain ladder*, ASTIN Bulletin: The Journal of the IAA, 42(2012), 59-76.

[9] M.D. MARTINEZ-MIRANDA, J.P. NIELSEN AND R.J. VERRALL, *Double chain ladder and Bornhuetter-Ferguson*, North American Actuarial Journal, 17(2013a), 101-113.

[10] M.D. MARTINEZ-MIRANDA, J.P. NIELSEN, S. SPERLICH, AND R.J. VERRALL, *Continuous Chain Ladder: Reformulating and generalizing a classical insurance problem*, Expert Systems with Applications, 40(2013b), 5588-5603.

[11] T.K. MOON, *The expectation-maximization algorithm*, IEEE Signal Processing Magazine, 13(1996), 48-60.

[12] R. VERBELEN, K. ANTONIO, G. CLAESKENS, AND J. CREVECOEUR, *Modeling the occurrence of events subject to a reporting delay via an EM algorithm*, arXiv:1909.08336, (2019).

[13] R.J. VERRALL, J.P. NIELSEN, AND A. H. JESSEN, *Prediction of RBNS and IBNR Claims using Claim Amounts and Claim Counts*, ASTIN Bulletin: The Journal of the IAA, 40(2010), 871-887.

[14] R.J. VERRALL, M.V. WÜTHRICH, *Understanding reporting delay in general insurance*, Risks, 4(2016),1-36.