Jinadasa Gamage, Professor of Mathematics, Illinois State University, Normal, IL, e-mail: jina@ilstu.edu

Jed L. Linfield, FSA, MAAA, Health Actuary, Kaiser Permanente, Rockville, MD, e-mail: Jed.L.Linfield@kp.org

Krzysztof Ostaszewski, FSA, CFA, MAAA, Actuarial Program Director and Professor of Mathematics, Illinois State University, Normal, IL, e-mail: krzysio@ilstu.edu

An Alternative Approach to Calculation of IBNR Reserve in Health Insurance

Recent decades have brought about an increased use of statistical methodologies in property/casualty reserving, as an alternative to the deterministic approach of the completion factor method. This has also resulted in increased interest in statistical methodologies for the purpose of health insurance reserving. For the calculation of the incurred but not reported (IBNR) reserves in health insurance, practicing health actuaries often rely on a two-step process: first, they calculate completion factors or some other deterministic measure of IBNR, and then they complete the IBNR reserve calculation with a regression analysis or actuarial judgment on the most recent experience. In this work, we investigate the use of a statistical approach to the calculation of IBNR reserve directly, and in an integrated fashion, without an intermediate step of deriving the completion factors. Our approach allows for an integrated confidence interval estimate for IBNR, as well. We illustrate the methodology with a use of a realistic, but hypothetical data set, and derive confidence bounds for the IBNR estimate (with the use of a Taylor series approximation).

Introduction

Practicing health actuaries often rely on a traditional completion factor method for the calculation of the incurred but not reported (IBNR) reserves in health insurance. This traditional methodology, deterministic in nature, uses estimates of completion factors derived from historical data, to calculate IBNR reserve as the sum of all ultimate amounts to be paid for yet uncompleted claims, minus amounts already paid on those claims. Claims are generally grouped by months, and estimates are made on per-month-per-member (PMPM) basis. This methodology is presented and explained, for example, by Bornhuetter and Ferguson (1972), Fearrington and Lynch (2004), Lloyd (2005), and Sutton and Sorbo (1987) (this paper provides an excellent review of the most commonly used practical approach). We should note that this methodology is also used in property/casualty reserving, commonly referred to as chain-ladder method, with completion factors replaced by development factors.

The deterministic completion factor method has come under some criticism, for example by Barnett and Zehnwirth (2000), but especially Lynch (2004). Such criticism points out that the approach of using some form of weighted-average of historical completion factors as the current completion factor is not grounded in any statistical or scientific methodology, but rather in convenient approximation. Any problems with the approximation are often believed to be alleviated by the necessary process of adjustment over time, as claims are paid and completed. The passage of the Sarbanes-Oxley Act of 2002 made this issue even more significant, as the objections to the traditional approach often originate in observed volatility of estimates of completion factors, and such volatility may lead to wide swings in reported reserves, resulting in an appearance of non-compliance with the provisions of the Sarbanes-Oxley Act, which require certification of the financial statement by a firm's management.

Over the recent decades, there have been significant inroads made into the possible applications of statistical methodologies in IBNR reserving. They have been, generally, limited to property/casualty insurance. Zehnwirth (1994, 1996) and Mosley (2004) are notable entries in the body of actuarial literature on this subject. In this work, we investigate a possible new approach to statistical estimation of IBNR, and its practical implications.

IBNR reserves are the largest item in the liability portion of a health insurance company financial statement, and their estimation is a key task of the valuation actuary. IBNR reserve equals the difference between the projected total amounts of claims to be paid (often referred to as incurred claims), for claims not yet paid in full, and the total of all payments already made on those claims. In the completion factor method, an actuary estimates the percentage of the projected total amount to be paid that the amount already paid represents, and uses that estimate to derive IBNR. The estimate is done for each claim month, in aggregate, on a PMPM basis. In this work, we propose instead to estimate the entire amount yet unpaid (i.e., the entire IBNR reserve) in one combined process. Let us present the underlying statistical methodology proposed.

1. Statistical approach to IBNR

One of the drawbacks of the traditional chain ladder method of calculating IBNR is that it is not possible to invoke statistical arguments directly to estimate the errors in the calculated IBNR values. We consider a method of estimating IBNR using a parametric statistical model building approach. We use monthly claim PMPM values, and not the cumulative paid values. Let $Y(i,t)$ be the PMPM amount for the claim incurred in the month i and reported in the month $t,$ with $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, n.$ If all the data values are available then this produces a square array of data. But in actuarial practice, only the upper triangle of the data is available to the actuary, and the IBNR estimate is the sum of the values in the yet unknown triangle. The available set of data, also forming a triangle, will be called the *claim triangle.* The task of the actuary is to estimate the missing values in the square array using the values in the claim triangle, and report the cumulative IBNR.

The method we propose uses the historical data in the near history as well as the claim triangle for the present year to estimate the IBNR values. We illustrate our methodology with a specific data set created for this research, but based on a realistic data. Table 1 presents the data. The rows in Table 1 show the process of *completion* (also called *development* in property-casualty reserving) of the claims originating in the month, for which the row is created, while the columns represent subsequent months of completion. The process of completion is done in the first two years (2001 and 2002), and the data is shaped into a triangle for the last year (2003). The amounts given are actual amounts paid in a given month of completion for a specified claim month, not the cumulative amounts. In particular, there are some months where the amounts paid are negative – these represent adjustments, or third party recoveries for claims where previous payments have turned out to be excessive.

The methodology we propose is general in nature, and applicable even in the case when no completed claims are available and one has to work only with an incomplete data triangle.

Table 1

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	180	436082	933353	116978	42681	41458.5	5088	22566	4750.5	3280.5	-187.5	1464	1696.5
1	5161.68	940722	561967	21694.4	171659	11007.8	19088	5212.94	4337.08	7843.52	2972.79	4061.21	10235.9
2	42263.5	844293	720302	94633.8	182077	32215.8	12936.9	22815	1754.41	4695.1	1325.66	757.52	2177.11
3	20780.6	762302	394625	78043	157950	46173.2	126254	4838.86	336.5	1572.86	9572.67	1947.42	5936.67
4	20345.9	772404	392330	315888	39196.7	21360.4	8720.76	5452.19	16627.4	2117.83	4119.37	5666.43	-1977.1
5	20490.7	831793	738087	65525.9	27767.9	12184.6	1493.28	11265	1805.47	29278.1	13019.7	2966.56	-83.05
6	37954.5	1126675	360514	89316.7	40126	16576.2	16701.4	2443.54	8265.69	11310.5	8006.03	1403.37	3123.59
7	138558	806362	589304	273117	36912.5	16831.5	19941.2	13310.2	8619.23	4678.52	3094.16	4608.62	236.1
8	28331.7	954543	246571	205528	60060.2	15198.5	42208	17568.2	1685.95	9897.12	3367.21	2062.16	421.49
9	104160	704796	565939	323789	45307.3	32517.9	26226.7	7976.11	3363.65	991.52	33962.8	2199.55	1292.75
10	40746.9	927158	425794	146145	66663.3	31214.1	12807.6	15858.6	373.68	3079.32	411.52	936.57	1874.71
11	10860.8	847338	272165	134798	71804.3	27800.1	17917	3929.79	2793.64	846.17	1961.73	1879.33	16059.9

12	77938.2	896195	544372	173606	41595.4	4209.16	16473.3	756000	-65.62	-1880.5	-4053.9	84232.7	4920.96
13	38041.4	1035439	438153	115587	12488.7	22260.1	13203.1	6394.65	2056.12	-3323.4	33397.4	3478.9	-1624.7
14	39409.5	1022024	255002	169881	35230.1	40307.5	21067.2	5377.81	5508.42	17606.1	-24320	1298.16	1362.02
15	68252.7	1414379	317110	91880.3	53969.6	10887.9	3170.9	11659.6	20860.7	1033.06	-21670	2633.57	148.7
16	124824	1053972	516876	145954	25171.2	12608.9	7704.25	29632.9	4555.18	6203.3	3872.11	1115.66	665.74
17	49725.3	1119099	533444	80181.9	32202.9	23204.7	18806.7	7944.07	4151.88	-910.3	3663.86	608.33	528
18	44317.5	1297335	385789	141155	150726	35075.2	16176.1	8070.24	67.03	14216.6	2325.84	7090.56	687.22
19	134152	1111151	493175	101439	46656.6	22824.1	12818.3	3780.94	1265.42	2466.91	-62165	246.82	-8689.3
20	29968.2	1382043	178587	71030.3	25708.1	15068.3	3145.03	-4058.2	-1919.6	4984.06	-1523	-3538.9	-477.91
21	210377	999963	528880	201410	58003	26174.1	-9371.4	2016.79	9794.64	6688.15	-40.13	453.36	-73.44
22	56654.4	1206370	376504	56321.8	19590.5	12054.8	21076.9	11572.8	4038.96	821.64	6611.88	-9677.9	714.96
23	89180.6	1240938	279553	57163.5	75343.9	12665	71741.4	9048.55	1297.94	12164.4	19615.6	-4603.9	-3183.9
24	131568	1301927	716180	150253	110031	78147.5	4609.85	19855	18447.7	14432.4	118.97	2747.78	
25	76262	1130312	692736	174283	38890.6	41810.9	8834.16	18122.5	4268.11	-290.76	2119.3		
26	159575	1313809	704116	68411.5	30184.6	64401.6	19228.7	-3020.5	3220.03	1994.04			
27	76312.7	1505842	437084	50872.4	116723	18159.6	10974.7	12663.7	8804.83				
28	104028	1667823	360676	153274	37529.2	34840	17479.1	9373.82					
29	79687.9	1235573	776240	65302.8	18722.7	10778.6	10614.6						
30	76394.6	1689354	442965	234171	36806.5	22351.3							
31	110460	1492980	589184	93366.4	180095								
32	196687	2011979	313416	166839									
33	268365	1027925	897097										
34	58510.1	1225307											
35	96378.3												

We also present a scatterplot of the data in Figure 1.

Figure 1

3D Scatterplot of Original Y vs t vs i
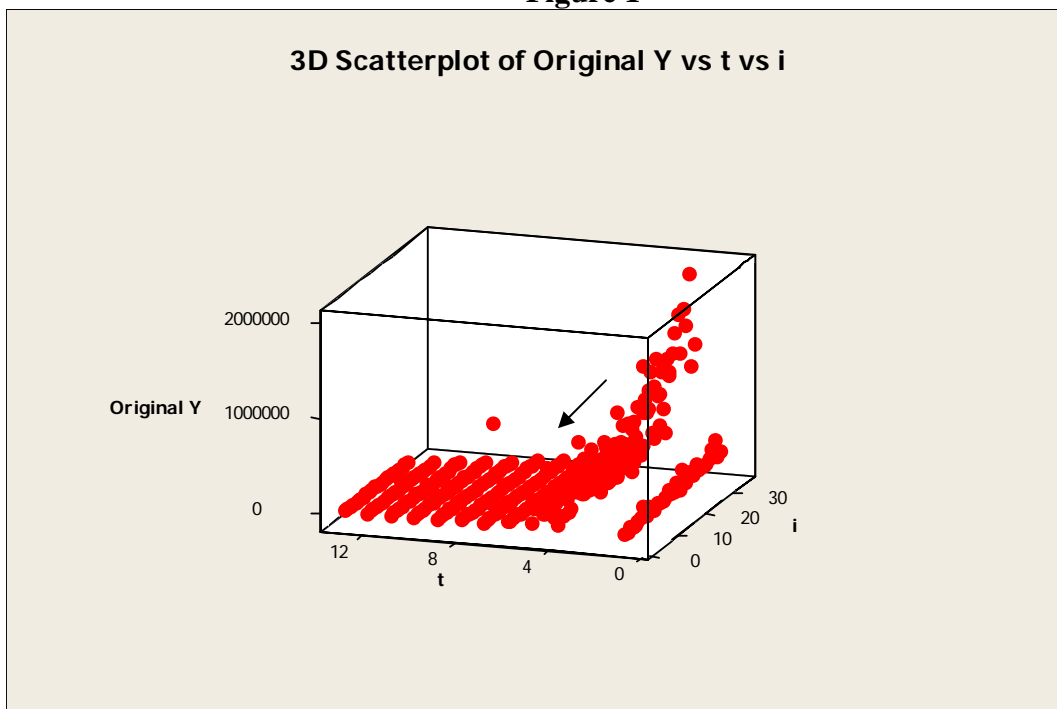
This plot indicates an outlier (marked with an arrow) in the data set. We have identified this to the observation corresponding to the claim month $i = 13$ and the month of development $t = 8$. The plot also indicates that the variable i has almost no effect on the response (dependent) variable Y, and the relationship between Y and t suggests an exponential function. However, fitting an exponential model poses a difficulty because some of the observations on the response variable are negative, and that is an impossibility if Y is a result of applying an exponential function to a real number. In order to make an exponential model feasible, we have decided to add an appropriately large positive number to all observations. At the end we will subtract this value form the estimates so that the estimates are not inflated. The largest negative value of Y is –62,165, and we have added 62,170 to all the observations. After this transformation, we conclude that the lowest observation of the response variable is a clear outlier in the data. In the Figure 2 below, we indicate this outlier with an arrow, although the small value of the response variable, i.e., 5, makes the point somewhat difficult to notice.
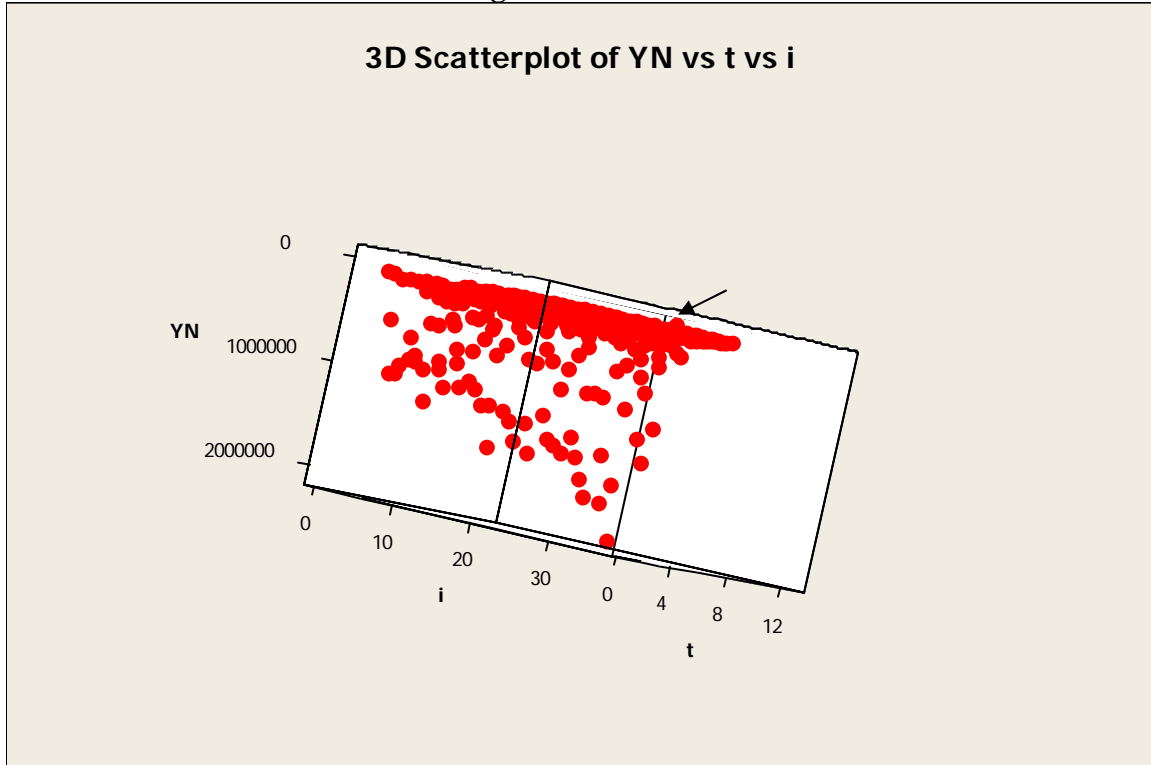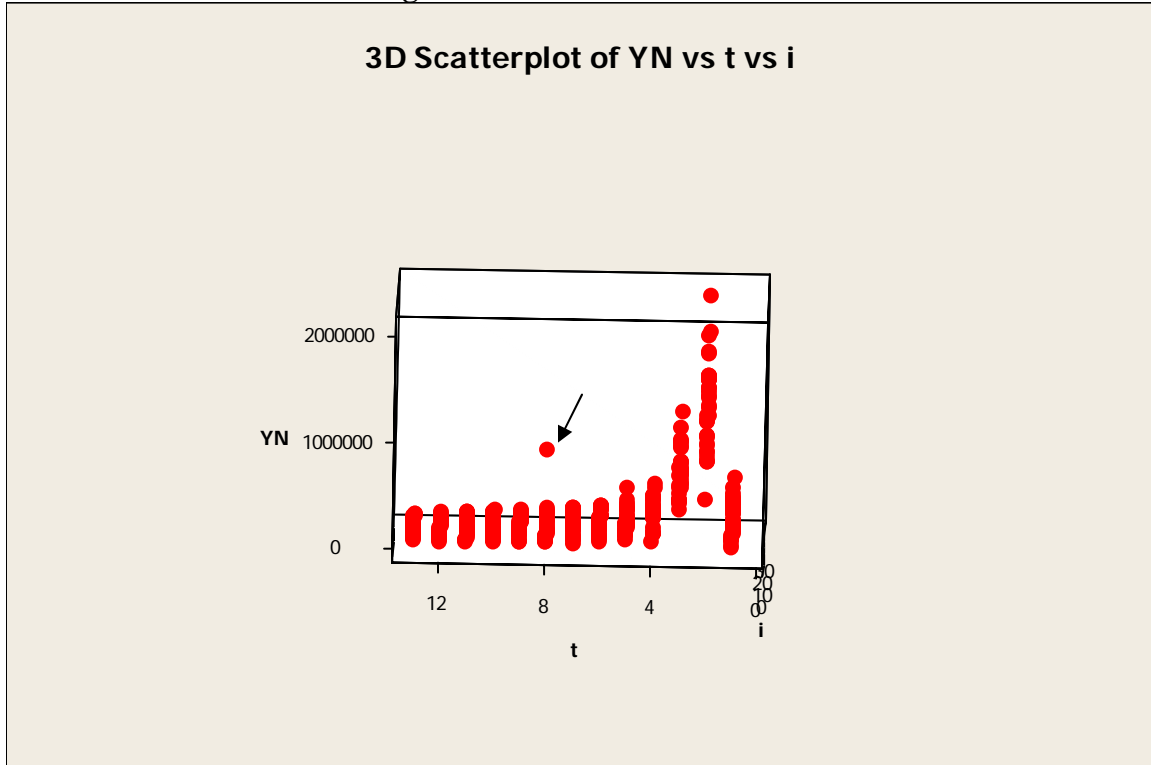
Figure 2

3D Scatterplot of YN vs t vs i

Figure 3 represents another view of the data, to indicate the second outlier, which we had noted earlier.

Figure 3

3D Scatterplot of YN vs t vs i

We have decided to remove these two clearly visible outliers from further analysis. The initial model we tested was:

$$\ln Y = \beta_0 + \beta_1 i + \beta_2 t + \varepsilon,$$

i.e., a basic exponential model. But we also considered several refinements of this model, and after analyzing their fit, and the residuals and finally we found that the following model is satisfactory in describing the data:

$$\ln Y = \beta_0 + \beta_1 i + \beta_2 t + \beta_3 t^2 + \beta_4 \ln t + \varepsilon.$$

We have obtained the following specific results with this model. The regression equation is

$$\ln Y = 14.2 + 0.00431 i + 2.33 t + 0.0918 t^2 + 4.19 \ln t. \tag{1}$$

We have used 388 data points, with two outliers removed.

	Coefficient	Standard Error	t-value	p-value
Constant	14.2018	0.1076	132.05	0.000
i	0.004310	0.002728	1.58	0.115
t	−2.2330	0.1062	−21.02	0.000
t^2	0.091842	0.004712	19.49	0.000
$\ln t$	4.1909	0.2299	18.23	0.000

For the whole model, we get the following values:

S = 0.476721
R-Square = 74.70%
Adjusted R-Square = 74.50%

Analysis of Variance

Source	DF	SS	MS	F-value	p-value
Regression	4	257.235	64.309	282.97	0.000
Residual Error	383	87.042	0.227		
Total	387	344.277			

As we have guessed by observing the 3-D scatter-plot, the independent variable i is not significant and all the variables t, t^2, and $\ln t$ are statistically significant. Given this situation, even though the variable i is not statistically significant, for practical reasons we decided to keep this variable in the model.

Figure 4 shows the normal probability plot of residuals along with a p-value, for testing for normality of the residuals using the Shapero-Wilk test.

Figure 4
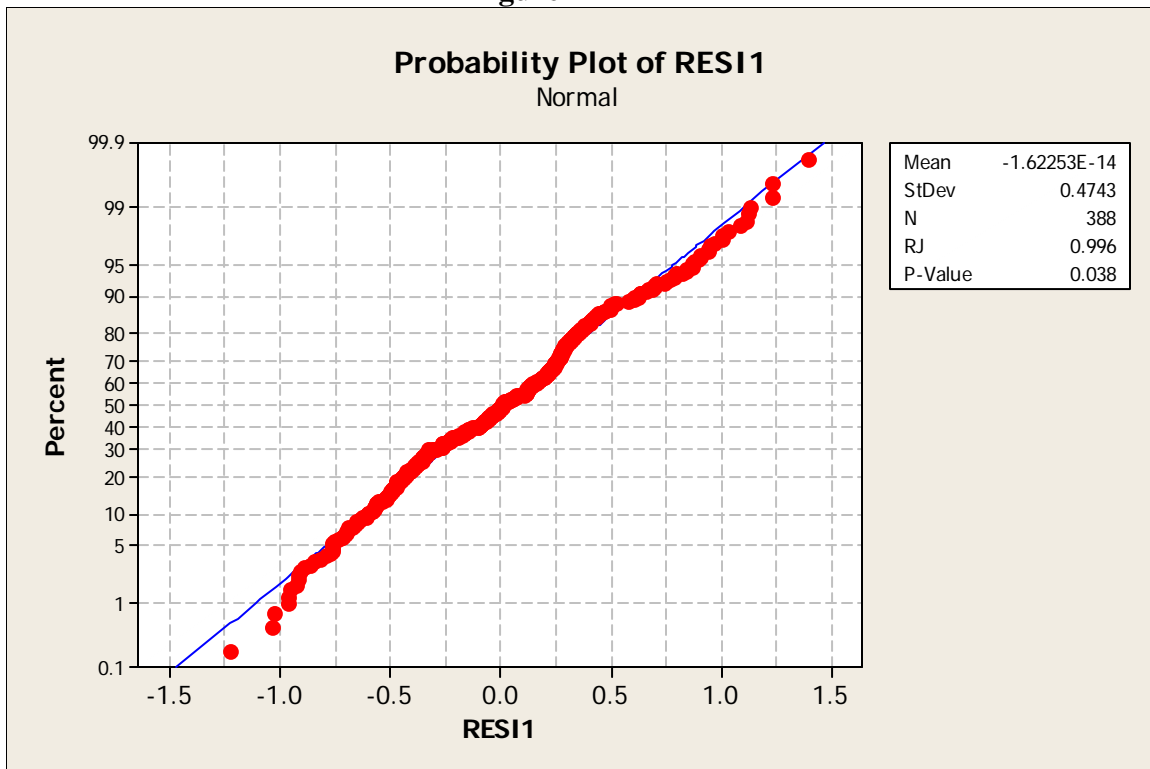
The p-value is 0.038 is not ideal, but we consider it acceptable. It is possible to achieve higher a p-value for testing for normality by removing other observations. We have decided that this level is satisfactory since the graph shows a very good linear pattern.

Estimated total IBNR values for each yet incomplete month, using our model after subtracting 62,170, are as follows:

Table 2

IBNR	= sum of
41692	1-estimate
49998	2-estimates
43686	3-estimates
32832	4-estimates
24823	5-estimates
27641	6-estimates
52977	7-estimates
120543	8-estimates
263819	9-estimates
531469	10-estimates
955589	11-estimates
1421315	12-estimates

Note that the last value is a sum of 12 estimates and the first value is just one observation. Note also that these estimates are maximum likelihood estimates and the underlying distribution is a log-normal distribution. More precisely

$$\ln \hat{\theta} \sim N\left(\vec{l}_i^{\,T} \cdot \vec{\beta}, \sigma^2\right),$$

where $\hat{\theta} = e^{\vec{l}_i^{\,T} \cdot \vec{\hat{\beta}}}$, and $\vec{\hat{\beta}}$ is the vector of regression parameter estimates from our model
(1), while $\vec{l}_i^{\,T}$ is a vector of appropriate constants that gives particular estimates. The
vector $\vec{l}_i^{\,T}$ is calculated as follows: each of the IBNR estimates in Table-2 is a sum of two or more individual predicted values except the first, and any of the estimated predicted values from a multiple linear regression model can be written as $\vec{l}_i^{\,T} \vec{\hat{\beta}}$ for some known vector $\vec{l}_i^{\,T}$.

In order to obtain an estimate of the entire IBNR, we suppose that the corresponding vectors are $\vec{l}_1, \vec{l}_2 \cdots \vec{l}_k$ and the sum of the estimates is

$$\vec{l}_1^{\,T} \cdot \vec{\hat{\beta}} + \vec{l}_2^{\,T} \cdot \vec{\hat{\beta}} + \cdots + \vec{l}_k^{\,T} \cdot \vec{\hat{\beta}} = \vec{u}^{\,T} \cdot \vec{\hat{\beta}},$$

where

$$\vec{u} = \vec{l}_1 + \vec{l}_2 + \cdots + \vec{l}_k.$$

The variance of this sum is

$$\mathrm{Var}\left(\vec{u}^{\,T} \vec{\hat{\beta}}\right) = \vec{u}^{\,T} \left(X^T X\right)^{-1} \vec{u} \sigma^2,$$

and any standard statistical computer package will produce the standard error of this estimate. But in our situation we have to estimate the variance of a quantity like

$$W = e^{A_1} + e^{A_2} + \cdots + e^{A_k},$$

where each A_i is a predicted value from the regression model and is of the form $\vec{l}_i^T \cdot \hat{\vec{\beta}}$. To develop a variance estimate, we use the Taylor series expansion of W about the vector of the means of the random variables $A_1, A_2, \ldots, A_k$ to get the following:

$$W \approx \left(e^{\mu_1} + e^{\mu_2} + \cdots + e^{\mu_k} \right) + \left(A_1 - \mu_1 \right) e^{\mu_1} + \left(A_2 - \mu_2 \right) e^{\mu_2} + \cdots + \left(A_k - \mu_k \right) e^{\mu_k},$$

where

$$\mu_i = E\left(A_i \right) = \vec{l}_i^T \cdot \hat{\vec{\beta}}$$

for $i = 1, 2, \ldots, k$. Thus

$$\mathrm{Var}\left(W \right) \approx \mathrm{Var}\left(\sum_{i=1}^{k} e^{\mu_i} A_i \right)$$

and if μ_i's are known then

$$\sum_{i=1}^{k} e^{\mu_i} A_i = \sum_{i=1}^{k} \vec{m}_i^T \cdot \hat{\vec{\beta}} = \vec{m}^T \cdot \hat{\vec{\beta}}$$

where $\vec{m}_i = e^{\mu_i} \vec{l}_i$ and $\vec{m} = \sum_{i=1}^{k} \vec{m}_i$. We proceed to estimate the variance of W as follows. We begin by getting the estimates of $\hat{\mu}_i = \vec{l}_i^T \cdot \hat{\vec{\beta}}$ for $i = 1, 2, \ldots, k$. Then we calculate the vector $\vec{m} = \sum_{i=1}^{k} e^{\hat{\mu}_i} \vec{l}_i$. Finally, an estimate of $\mathrm{Var}\left(W \right)$ is

$$\widehat{\mathrm{Var}\left(W \right)} = \vec{m}^T \left(X^T X \right)^{-1} \vec{m} \cdot MSE,$$

where X is the design matrix and MSE is the mean square error of the regression model.

In order to get an approximate confidence interval for IBNR, we use the critical values from the standard normal distribution. Since our estimates are maximum-likelihood estimates, and are asymptotically normal, the error in using standard normal critical values is minimal. Thus for each IBNR estimate W, an approximate $100(1 - \alpha)\%$ confidence interval for IBNR is

$$\widehat{W} \pm z_{\alpha/2} \sqrt{\widehat{\mathrm{Var}\left(W \right)}}.$$

Table 3 below gives the estimates and approximate 95% confidence intervals for the total IBNR:

Table 3

IBNR	Margin of Error	Lower Bound	Upper Bound
41692	6602.131	35089.87	48294.13
49998	7656.956	42341.04	57654.96
43686	7352.862	36333.14	51038.86
32832	7108.382	25723.62	39940.38
24823	6911.282	17911.72	31734.28
27641	7406.455	20234.54	35047.46
52977	8020.362	44956.64	60997.36
120543	13117.34	107425.7	133660.3
263819	28647.66	235171.3	292466.7
531469	61178.87	470290.1	592647.9
955589	116865	838724	1072454
1421315	178945.6	1242369	1600261
3566384	417530.3	3148854	3983914

The first column is the total IBNR estimate, the second one is an amount added or subtracted to reach the bounds of the confidence interval, and the lower and upper bounds are given in the last two columns.
Note that the last row gives an estimate of the total IBNR and a confidence interval.

2. PMPM calculation

We can use an analogous statistical approach for the paid PMPM data to calculate approximate 95% confidence interval for the total IBNR. Table 4 presents the paid PMPM data.

Table 4

-	1	2	3	4	5	6	7	8	9	10	11	12
0.02	39.10	83.68	10.49	3.83	3.72	0.46	2.02	0.43	0.29	-0.02	0.13	0.15
0.46	84.61	50.55	1.95	15.44	0.99	1.72	0.47	0.39	0.71	0.27	0.37	0.92
3.82	76.27	65.07	8.55	16.45	2.91	1.17	2.06	0.16	0.42	0.12	0.07	0.20
1.88	68.87	35.65	7.05	14.27	4.17	11.41	0.44	0.03	0.14	0.86	0.18	0.54
1.83	69.40	35.25	28.38	3.52	1.92	0.78	0.49	1.49	0.19	0.37	0.51	-0.18
1.83	74.44	66.05	5.86	2.49	1.09	0.13	1.01	0.16	2.62	1.17	0.27	-0.01
3.39	100.78	32.25	7.99	3.59	1.48	1.49	0.22	0.74	1.01	0.72	0.13	0.28
12.13	70.61	51.60	23.92	3.23	1.47	1.75	1.17	0.75	0.41	0.27	0.40	0.02
2.49	83.73	21.63	18.03	5.27	1.33	3.70	1.54	0.15	0.87	0.30	0.18	0.04
9.09	61.52	49.40	28.26	3.95	2.84	2.29	0.70	0.29	0.09	2.96	0.19	0.11
3.56	81.02	37.21	12.77	5.83	2.73	1.12	1.39	0.03	0.27	0.04	0.08	0.16
0.94	73.33	23.55	11.67	6.21	2.41	1.55	0.34	0.24	0.07	0.17	0.16	1.39
6.66	76.57	46.51	14.83	3.55	0.36	1.41	0.51	-0.01	-0.16	-0.35	7.20	0.42
3.22	87.58	37.06	9.78	1.06	1.88	1.12	0.54	0.17	-0.28	2.82	0.29	-0.14
3.35	86.96	21.70	14.45	3.00	3.43	1.79	0.46	0.47	1.50	-2.07	0.11	0.12
5.86	121.36	27.21	7.88	4.63	0.93	0.27	1.00	1.79	0.09	-1.86	0.23	0.01
10.67	90.06	44.17	12.47	2.15	1.08	0.66	2.53	0.39	0.53	0.33	0.10	0.06
4.29	96.64	46.07	6.92	2.78	2.00	1.62	0.69	0.36	-0.08	0.32	0.05	0.05

3.83	112.06	33.32	12.19	13.02	3.03	1.40	0.70	0.01	1.23	0.20	0.61	0.06
11.51	95.34	42.31	8.70	4.00	1.96	1.10	0.32	0.11	0.21	-5.33	0.02	-0.75
2.55	117.77	15.22	6.05	2.19	1.28	0.27	-0.35	-0.16	0.42	-0.13	-0.30	-0.04
17.70	84.11	44.48	16.94	4.88	2.20	-0.79	0.17	0.82	0.56	0.00	0.04	-0.01
4.74	100.94	31.50	4.71	1.64	1.01	1.76	0.97	0.34	0.07	0.55	-0.81	0.06
7.35	102.29	23.04	4.71	6.21	1.04	5.91	0.75	0.11	1.00	1.62	-0.38	-0.26
10.76	106.48	58.57	12.29	9.00	6.39	0.38	1.62	1.51	1.18	0.01	0.22	0.0891
6.25	92.64	56.78	14.28	3.19	3.43	0.72	1.49	0.35	-0.02	0.17	0.1486	0.0895
13.16	108.31	58.05	5.64	2.49	5.31	1.59	-0.25	0.27	0.16	0.2462	0.1492	0.0898
6.37	125.63	36.47	4.24	9.74	1.52	0.92	1.06	0.73	0.4045	0.2472	0.1498	0.0902
8.72	139.84	30.24	12.85	3.15	2.92	1.47	0.79	0.6582	0.4061	0.2482	0.1504	0.0906
6.75	104.59	65.71	5.53	1.58	0.91	0.90	1.0584	0.6609	0.4078	0.2492	0.151	0.0909
6.48	143.32	37.58	19.87	3.12	1.90	1.6763	1.0627	0.6636	0.4094	0.2502	0.1516	0.0913
9.45	127.73	50.41	7.99	15.41	2.6025	1.6831	1.067	0.6663	0.4111	0.2512	0.1523	0.0917
16.77	171.51	26.72	14.22	3.9322	2.6131	1.6899	1.0714	0.669	0.4128	0.2522	0.1529	0.092
22.66	86.80	75.75	5.7121	3.9482	2.6236	1.6967	1.0757	0.6717	0.4144	0.2532	0.1535	0.0924
4.92	102.95	77.1406	5.7352	3.9642	2.6343	1.7036	1.0801	0.6744	0.4161	0.2543	0.1541	0.0928
8.14	40.3386	77.4531	5.7584	3.9802	2.6449	1.7105	1.0844	0.6771	0.4178	0.2553	0.1547	0.0932

There are again negative values in the PMPM data. We present the data in a three-dimensional scatterplot below.
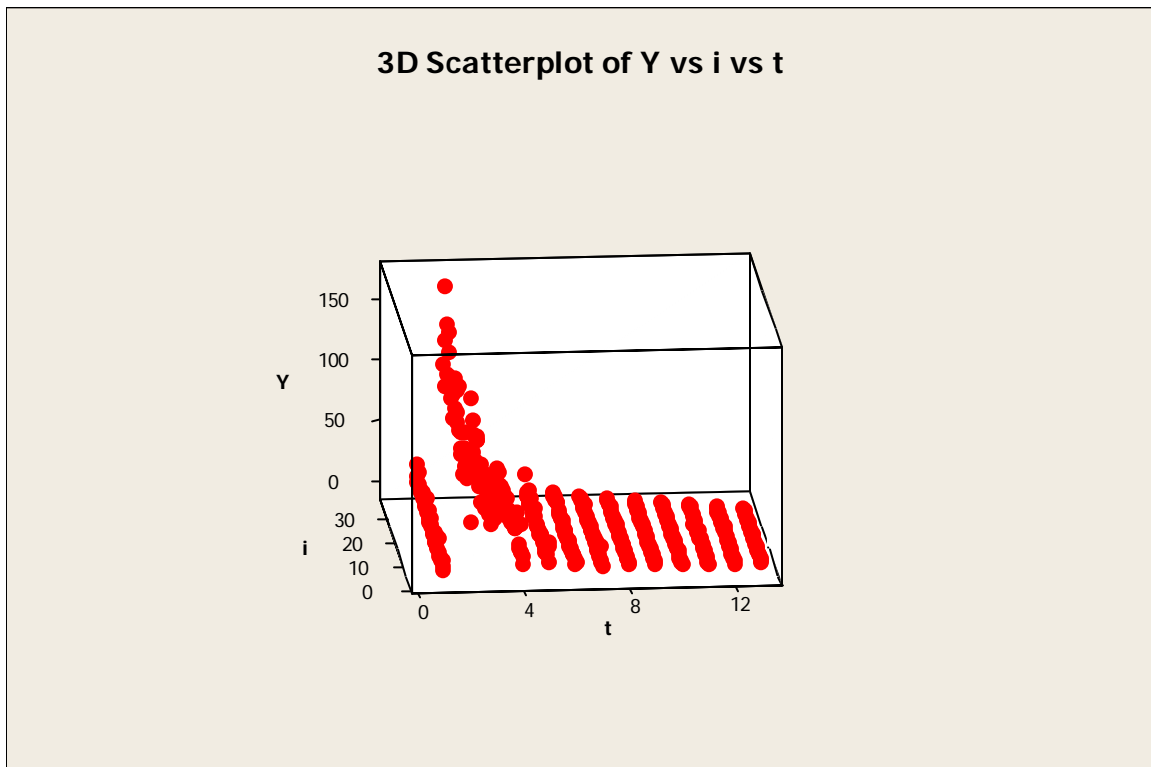
This graph shows a gamma function pattern. But there are negative values in the data. The distribution of these values cannot be explained by a pattern of a gamma function. We have decided to analyze the positive values only and fit a gamma function to the data.

Below we show a three-dimensional scatter plot of the natural logarithm of only positive values of PMPM values against i and t.

3D Scatterplot of LGYPOS vs i vs t
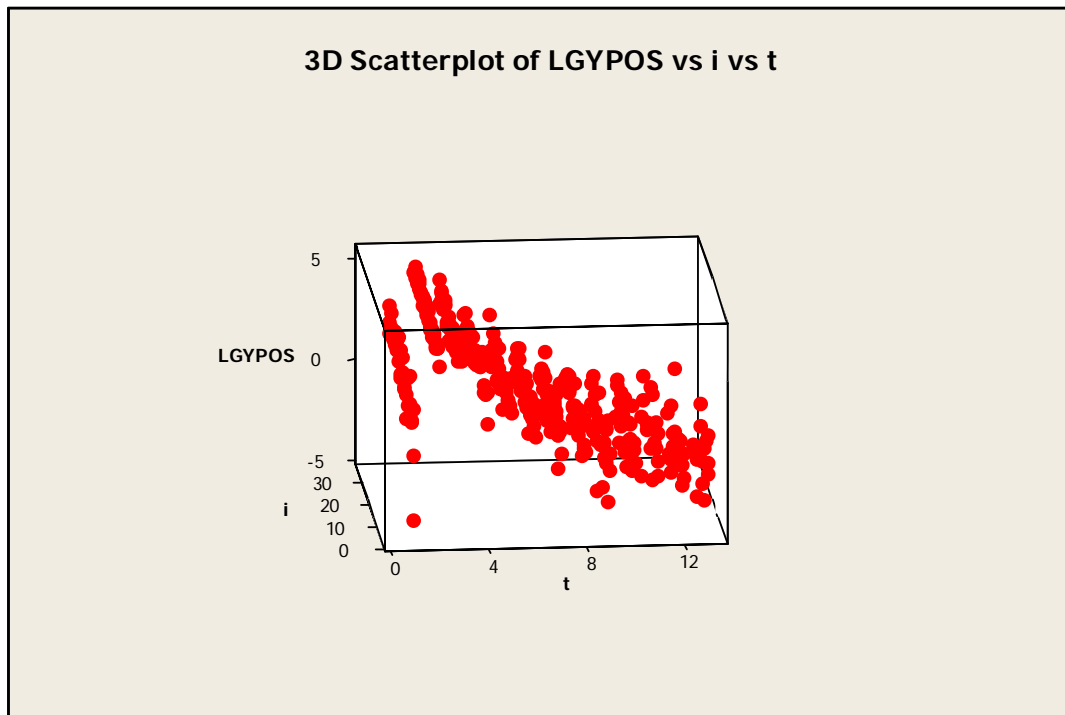
Regression analysis

We use the following model, with the dependent variable LGYPOS defined as the natural logarithm of those values of PMPM that are positive (and only those historical values for which PMPM is positive are used to develop the model)

$$LGYPOS = 2.77 + 0.00375i - 0.719t + 1.48\ln t + 3.03INDLGT.$$

The variable IND3LGT is defined as the natural logarithm of if t is less than or equal to 3, and 0 otherwise. We used 360 cases of data, with 30 cases involving missing values. Results are summarized below:

Predictor	Coef	SE of Coef	t-statistic	p-value
Constant	2.7725	0.1846	15.02	0.000
i	0.003752	0.005861	0.64	0.523
t	−0.71939	0.04305	−16.71	0.000
$\ln t$	1.4830	0.2062	7.19	0.000
INDLGT	3.0311	0.2764	10.97	0.000

S = 0.989331
R-Sq = 78.9%
R-Sq(adj) = 78.7%

Analysis of Variance

Source	Degrees of Freedom	SS	MS	F-statistic	p-value
Regression	4	1303.06	325.76	332.83	0.000
Residual Error	355	347.47	0.98		
Total	359	1650.52			

Below we present the residual plot corresponding to the above regression calculations.
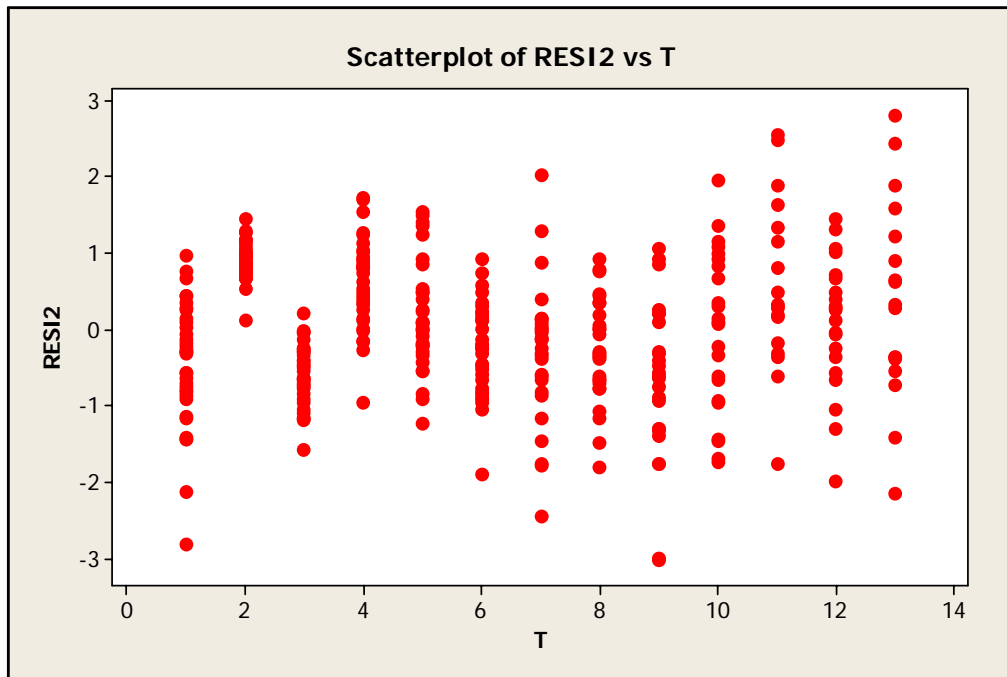
Normal probability plot of the residuals is given below.
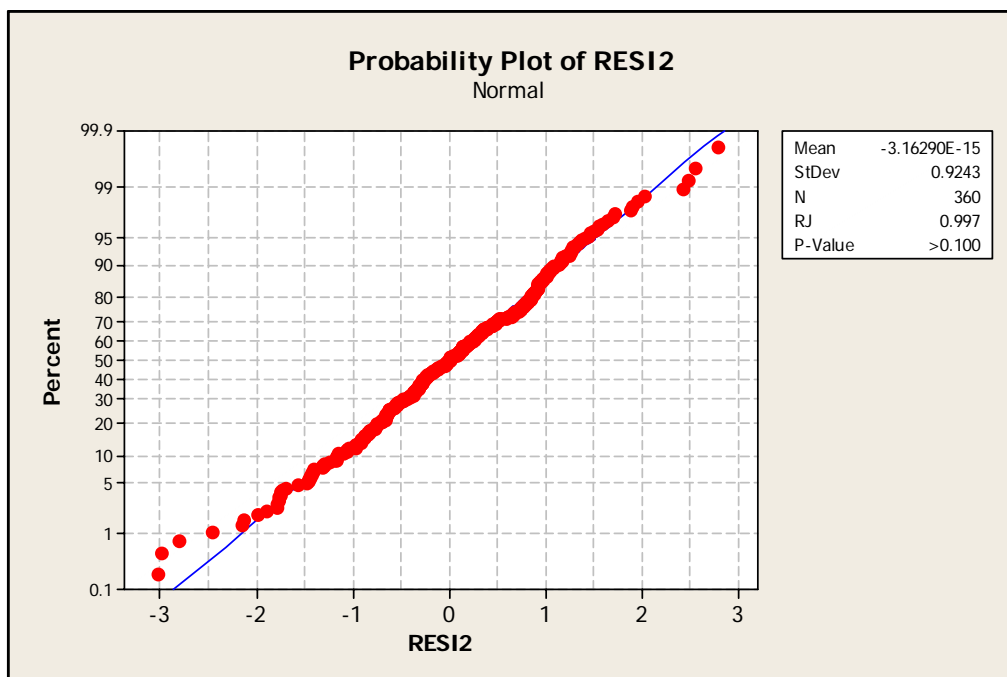
We present the PMPM table with projected values filled in below

0	1	2	3	4	5	6	7	8	9	10	11	12
0.02	39.10	83.68	10.49	3.83	3.72	0.46	2.02	0.43	0.29	-0.02	0.13	0.15
0.46	84.61	50.55	1.95	15.44	0.99	1.72	0.47	0.39	0.71	0.27	0.37	0.92
3.82	76.27	65.07	8.55	16.45	2.91	1.17	2.06	0.16	0.42	0.12	0.07	0.20
1.88	68.87	35.65	7.05	14.27	4.17	11.41	0.44	0.03	0.14	0.86	0.18	0.54
1.83	69.40	35.25	28.38	3.52	1.92	0.78	0.49	1.49	0.19	0.37	0.51	-0.18
1.83	74.44	66.05	5.86	2.49	1.09	0.13	1.01	0.16	2.62	1.17	0.27	-0.01
3.39	100.78	32.25	7.99	3.59	1.48	1.49	0.22	0.74	1.01	0.72	0.13	0.28
12.13	70.61	51.60	23.92	3.23	1.47	1.75	1.17	0.75	0.41	0.27	0.40	0.02
2.49	83.73	21.63	18.03	5.27	1.33	3.70	1.54	0.15	0.87	0.30	0.18	0.04
9.09	61.52	49.40	28.26	3.95	2.84	2.29	0.70	0.29	0.09	2.96	0.19	0.11
3.56	81.02	37.21	12.77	5.83	2.73	1.12	1.39	0.03	0.27	0.04	0.08	0.16
0.94	73.33	23.55	11.67	6.21	2.41	1.55	0.34	0.24	0.07	0.17	0.16	1.39
6.66	76.57	46.51	14.83	3.55	0.36	1.41	0.51	-0.01	-0.16	-0.35	7.20	0.42
3.22	87.58	37.06	9.78	1.06	1.88	1.12	0.54	0.17	-0.28	2.82	0.29	-0.14
3.35	86.96	21.70	14.45	3.00	3.43	1.79	0.46	0.47	1.50	-2.07	0.11	0.12
5.86	121.36	27.21	7.88	4.63	0.93	0.27	1.00	1.79	0.09	-1.86	0.23	0.01
10.67	90.06	44.17	12.47	2.15	1.08	0.66	2.53	0.39	0.53	0.33	0.10	0.06
4.29	96.64	46.07	6.92	2.78	2.00	1.62	0.69	0.36	-0.08	0.32	0.05	0.05
3.83	112.06	33.32	12.19	13.02	3.03	1.40	0.70	0.01	1.23	0.20	0.61	0.06
11.51	95.34	42.31	8.70	4.00	1.96	1.10	0.32	0.11	0.21	-5.33	0.02	-0.75
2.55	117.77	15.22	6.05	2.19	1.28	0.27	-0.35	-0.16	0.42	-0.13	-0.30	-0.04
17.70	84.11	44.48	16.94	4.88	2.20	-0.79	0.17	0.82	0.56	0.00	0.04	-0.01
4.74	100.94	31.50	4.71	1.64	1.01	1.76	0.97	0.34	0.07	0.55	-0.81	0.06
7.35	102.29	23.04	4.71	6.21	1.04	5.91	0.75	0.11	1.00	1.62	-0.38	-0.26
10.76	106.48	58.57	12.29	9.00	6.39	0.38	1.62	1.51	1.18	0.01	0.22	0.0891
6.25	92.64	56.78	14.28	3.19	3.43	0.72	1.49	0.35	-0.02	0.17	0.1486	0.0895
13.16	108.31	58.05	5.64	2.49	5.31	1.59	-0.25	0.27	0.16	0.2462	0.1492	0.0898
6.37	125.63	36.47	4.24	9.74	1.52	0.92	1.06	0.73	0.4045	0.2472	0.1498	0.0902
8.72	139.84	30.24	12.85	3.15	2.92	1.47	0.79	0.6582	0.4061	0.2482	0.1504	0.0906
6.75	104.59	65.71	5.53	1.58	0.91	0.90	1.0584	0.6609	0.4078	0.2492	0.151	0.0909
6.48	143.32	37.58	19.87	3.12	1.90	1.6763	1.0627	0.6636	0.4094	0.2502	0.1516	0.0913
9.45	127.73	50.41	7.99	15.41	2.6025	1.6831	1.067	0.6663	0.4111	0.2512	0.1523	0.0917
16.77	171.51	26.72	14.22	3.9322	2.6131	1.6899	1.0714	0.669	0.4128	0.2522	0.1529	0.092
22.66	86.80	75.75	5.7121	3.9482	2.6236	1.6967	1.0757	0.6717	0.4144	0.2532	0.1535	0.0924
4.92	102.95	77.1406	5.7352	3.9642	2.6343	1.7036	1.0801	0.6744	0.4161	0.2543	0.1541	0.0928
8.14	40.3386	77.4531	5.7584	3.9802	2.6449	1.7105	1.0844	0.6771	0.4178	0.2553	0.1547	0.0932

The following table shows the estimated yet unpaid PMPM values from the model, number of members and the estimated IBNR values.

Est. PMPM	Members	IBNR
$0.09	12,227	1089.426
$0.24	12,201	2905.058
$0.49	12,130	5885.476
$0.89	11,986	10687.92
$1.55	11,927	18528.59
$2.62	11,814	30931.41
$4.31	11,787	50744.21
$6.93	11,689	80948.66
$10.89	11,731	127697.8
$16.64	11,843	197085.3
$93.85	11,902	1116999
$134.57	11,844	1593826

The estimated total IBNR is $3237329.00 and the standard error of the estimate is 853191.7 and an approximate 95% confidence interval for the total IBNR is [$2384137.00, $4090520.00].

Conclusions
The approach we propose in this work uses a multivariate regression model to calculate IBNR for a set of health insurance data. The approach proposed by us produces an integrated IBNR estimate, as well as a unified confidence interval for the entire IBNR. The approach makes it also possible to produce confidence intervals for each incomplete month separately. The statistical methodology proposed is a standard statistical procedure, but our approach in using it for an IBNR estimate is innovative. Let us also note that the model can be generalized to include additional factors, such as the number of days or working days in a month (by dividing the data by the number of effective work days in a month).

References

Barnett Glen, and Ben Zehnwirth, "Best Estimates for Reserves," *Proceedings of the Casualty Actuarial Society 87,* Part 2, No. 167, November 12-15, 2000, pp. 245-321. Available online at: http://www.casact.org/pubs/proceed/proceed00/00245.pdf

Belsey, David A., Kuh Edwin, and Welsch, *Regression Diagnostics; Identifying Influential Data and Sources of Collinearity*, John Wiley, New York, 1980

Bornhuetter, and Ferguson "The Actuary and IBNR" *Proceedings of the Casualty Actuarial Society,* vol. 59, No. 112, 1972.

Fearrington, Doug and Robert Lynch, "Approaches to Determining Unpaid Claim Liabilities: Old and New," *The Records of the Society of Actuaries,* Valuation Actuary Symposium, Boston, Massachusetts, September 20-21, 2004, Session 39TS.

Little, R. J. A. and Rubin, D. B., *Statistical Analysis with Missing Data,* John Wiley, New York, 1987.

Lloyd, John C., "Health Reserves", *Society of Actuaries Study Note 8GM-305-00,* Society of Actuaries, Schaumburg, Illinois, 2005.

Lynch, Robert G., "Gerbils on Espresso: A Better Way to Calculate IBNR Reserves with Low Variance," *Contingencies,* January-February 2004, pp. 28-38.

Mosley, Roosevelt C., "Estimating Claim Settlement Values Using GLM," *Casualty Actuarial Society, Discussion Paper Program - Applying and Evaluating Generalized Linear Models Including Research Papers on the Valuation of P&C Insurance Companies,* 2004.
Available online at: http://www.casact.org/pubs/dpp/dpp04/

Pindyck, R. S. and Rubinfeld, D. L., *Econometric Models and Economic Forecasts, Fourth Edition,* Irwin McGraw-Hill, New York, 1998.

Rubin, D. B., *Multiple Imputations for Non-Response in Surveys,* John Wiley, New York, 1986

Sen Ashis and Srivastava Muni, *Regression Analysis: Theory, Methods, and Applications,* Springer-Verlag, New York, 1990.

Sutton, Harry L. and Sorbo, Allen J., *Actuarial Issues in the Fee-For-Service/Prepaid Medical Group*, Center for Research in Ambulatory Health Care Administration, Denver, Colorado, 1987, pages 62-67.

Zehnwirth, Ben, "Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals, and Risk Based Capital," *Casualty Actuarial Society Forum* 1994, pp. 510-605.
Available online at: http://www.casact.org/pubs/forum/94spforum/94spf447.pdf

Zehnwirth, Ben, "Three Powerful Diagnostic Models for Loss Reserving," *University of Melbourne, Centre for Actuarial Studies, Australia, Working Paper Series,* 1996.
Available online at: http://www.economics.unimelb.edu.au/actwww/No34.pdf