Bayesian Modelling for the Prediction of Outstanding Automobile Claim Amounts

Ioannis Ntzoufras* and Petros Dellaportas**

*Department of Statistics and Actuarial Science,

School of Sciences, University of the Aegean,

e-mail: ntzoufras@aegean.gr

**Department of Statistics

Athens University of Economics and Business,

e-mail:petros@aueb.gr

2001, April, Skiathos

For postcript file: http://stat-athens.aueb.gr/~jbn/ntzoufras.html

Contents/Plan of the Presentation

1 The Problem

- Insurance companies delay payments of automobile claims.
- Outstanding claims are being paid gradually.
- An insurance company has outstanding claims from years in the past.
- Problem of Interest: Estimate claim amounts to be paid within next years.

In this work we...

- Use a Bayesian approach
- Investigate various models
- Use MCMC to generate from each posterior distribution of interest.
- Incorporate information about both claim amounts and counts.

2 The Data

	B				
	Y_{11}	Y_{12}	$\cdots$	$\cdots$	Y_{1r}
	Y_{21}	Y_{22}	$\cdots$	$Y_{2,r-1}$	
A	$\vdots$	$\vdots$	$\vdots$		
	$\vdots$	$\vdots$			
	Y_{r1}				

Table 1:

Structure of Outstanding Claim Amounts Data

	B					
	n_{11}	n_{12}	$\cdots$	$\cdots$	n_{1r}	T_1
	n_{21}	n_{22}	$\cdots$	$n_{2,r-1}$		T_2
A	$\vdots$	$\vdots$	$\vdots$			T_i
	$\vdots$	$\vdots$				$\vdots$
	n_{r1}					T_r

Table 2:

Structure of Outstanding Claim Counts Data

For $i, j = 1, \ldots, r$

- A_i: Accident year
- B_j: Delay payment of a claim.
- Y_{ij}: Claim Amount paid with a delay of $j-1$ years for accidents happened in the year i.
- n_{ij}: Number of accidents paid with a delay of $j-1$ years for accidents happened in the year i.
- T_i: Total number of claims for each accident year
- inf_{ij}: Inflation factor for cell ij.
- We use deinflated Claim Amounts to make valid comparisons through different years.

3 Data from a major Greek insurance company

		B						
	Year	1	2	3	4	5	6	7
	1989	527003	220645	130250	84085	72182	21656	49868
	1990	715247	341364	166001	99845	108648	91958	
	1991	1166119	428365	166410	155376	191644		
A	**1992**	1686294	647331	335290	427069			
	1993	2780948	961010	444610				
	1994	3619446	1328151					
	1995	4002087						

Table 3: Table of Outstanding Claim Amounts

		B								Inflation
	Year	1	2	3	4	5	6	7	Total	(%)
	1989	6622	1943	489	138	61	223	66	9542	100.0
	1990	6943	2133	632	154	162	390		10496	120.4
	1991	8610	2216	736	651	256			12601	143.9
A	**1992**	9791	3167	1570	624				15565	166.6
	1993	11722	3192	1773					17735	190.6
	1994	13684	3664						19746	214.2
	1995	13068							18600	235.6

Table 4: Table of Outstanding Claim Counts

4 Related Work and Selected Publications

- Poisson Log-Linear and/or Log-Normal model:
  - Renshaw (1989, JIA),
  - Verrall (1991-JIA, 1993-JIA) and
  - Renshaw and Verrall (1998, BAJ).
- Bayesian and Empirical Bayes estimates of Log-Normal model: Verrall (1990,ASTIN)
- Generalized additive models:
  - Verrall (1996, IME)
- Log-Normal state space model:
  - De Jong (1983, JIA) and
  - Verrall (1989-JIA, 1993-JIA, 1994-ASTIN)
- Bootstrap Estimation:
  - England and Verrall (1999, IME).

5 Models Implemented Using MCMC

1. Log-Normal model for claim amounts.
2. Log-Normal for claim amounts & Poisson model for counts.
3. State space modelling of claim amounts .
4. State space modelling of average claim amount per accident (using claim amounts and counts).

5.1 Model 1: Log-Normal model for claim amounts

- Simple two-way Anova type model with log-Normal Error.

Formulation of the model

$$Z_{ij} = log\frac{Y_{ij}}{inf_{ij}},$$

$$Z_{ij} \sim N(\mu_{ij}, \sigma^2), \quad \mu_{ij} = \mu + \alpha_i + \beta_j$$

Sum to zero constraints :

$$\Sigma\alpha_i = \Sigma\beta_j = 0$$

Prior distributions used in example:

$$\mu \sim N(0, 1000), \quad \sigma^2 \sim IG(0.001, 0.001)$$

$$\alpha_i \sim N(0, 100),\ j = 2, ..., r,$$

$$\beta_j \sim N(0, 100),\ j = 2, ..., r$$

5.2 Model 2: Log-Normal & Poisson model

- Incorporate available information from claim counts (accidents).
- Model average claim amount per accident instead of total claim amount.
  - Simple Log-Normal model for claim amounts data,
  - Multinomial logit with appropriate restrictions for claim counts.
- Adjust predicted claim amounts according to counts.
- Predict amounts when we have large number of accidents (e.g. due to a catastrophe such as earthquake or flood).

Formulation of the model

$$Z_{ij} \sim N(\mu_{ij}, \sigma^2), \quad \mu_{ij} = \mu + \alpha_i + \beta_j + log(n_{ij})$$

$$\mathbf{n}_i \sim Multinomial(\boldsymbol{\pi}; T_i),\ log(\pi_j/\pi_1) = \beta_j^*$$

$$\mathbf{n}_i = (n_{i1}, n_{i2}, \ldots, n_{ir})^T$$

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_r)^T$$

Constraints :

$$\sum_{i=1}^{r} n_{ij} = n_{i.} = T_i$$

$$\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{r} \beta_j = \beta_1^* = 0$$

Prior distributions used in example:

$$\mu \sim N(0, 1000), \quad \sigma^2 \sim IG(0.001, 0.001)$$

$$\alpha_i \sim N(0, 100),\ j = 2, ..., r,$$

$$\beta_j \sim N(0, 100),\ j = 2, ..., r$$

$$\beta_j^* \sim N(0, 100),\ j = 2, ..., r$$

5.3 Model 3: State space modelling of claim amounts.

- Log-Normal Anova type model.
- Recursive equations on α_i, β_{ij}.

<u>Formulation of the model</u>

$$Z_{ij} \sim N(\mu_{ij}, \sigma^2), \qquad \mu_{ij} = \mu + \alpha_i + \beta_{ij}$$

<u>Recursive associations:</u>

$$\beta_{ij} \sim N(\beta_{i-1,j}, sw), \quad i = 2, ..., r$$

$$\alpha_i \sim N(\alpha_{i-1}, sw_a), \quad i = 3, ..., r$$

<u>Constraints:</u>

$$\alpha_1 = \beta_{i1} = 0, \ \forall \ i = 1, 2, ..., r$$

<u>Prior distributions used in example:</u>

$$\mu \sim N(0, 1000), \ \sigma^2 \sim IG(0.001, 0.001)$$

$$\alpha_i \sim N(0, 100), \ \beta_{ij} \sim N(0, 100), \ i, j = 2, ..., r$$

$$sw \sim IG(10^{-10}, 10^{-10}), \ sw_a \sim IG(10^{-10}, 10^{-10}).$$

5.4 Model 4: State space modelling of average claim amount per accident.

- Log-Normal Anova type model using also the claim counts table.
- Recursive equations on α_i, β_{ij}.

<u>Formulation of the model</u>

$$Z_{ij} \sim N(\mu_{ij}, \sigma^2), \quad \mu_{i,j} = \mu + \alpha_i + \beta_{ij} + log(n_{ij})$$

$$\mathbf{n}_i \sim Multinomial(\boldsymbol{\pi}; T_i),\ log(\pi_j/\pi_1) = \beta^*_j$$

$$\mathbf{n}_i = (n_{i1}, \ldots, n_{ir})^T, \boldsymbol{\pi} = (\pi_1, \ldots, \pi_r)^T$$

<u>Recursive associations:</u>

$$\beta_{ij} \sim N(\beta_{i-1,j}, sw), i = 2, ..., r$$

$$\alpha_i \sim N(\alpha_{i-1}, sw_a), i = 2, ..., r$$

<u>Constraints:</u>

$$\alpha_1 = \beta_{i1} = \beta^*_1 = 0,\ \forall\ i = 1, 2, ..., r$$

Prior distributions as in Models 2 and 3

6 Results of illustrated example

- Recursive Associations improve the fit of Model 2 but not the fit of Model 1.
- Standard errors of predictions in Model 4 are much lower.
- Model Selection in this Problem is not straightforward
  1. The information used in Model 2&4 is larger and of different quality
  2. The dimension of likelihood in Models 2&4 is different than in Models 1&3.
  3. We are interested in the predictive performance and not in fitness of the data.

We Measure Predictive Performance Using

$$\hat{L}_m^* = \sqrt{\frac{1}{|\mathcal{L}|}\hat{L}_m^2}$$

$$\hat{L}_m^2 = \frac{1}{K-b} \sum_{k=b+1}^{K} \sum_{(i,j)\in\mathcal{L}} \left([y_{ij}^{(m)}]_k - y_{ij}\right)^2$$

where

- $[y_{ij}^{(m)}]_k$: MCMC generated missing claim amounts of cell ij in iteration k resulted from model m,
- K: Total number of iterations,
- b: Number of iterations discarded as burn-in and
- $|\mathcal{L}| = r(r-1)/2 - 2 = 19$.

This measure is based on the L-criterion of Laud and Ibrahim (1995, JRSSB).

Model	Total	$\hat{L}_m^*$
1	2911(665)	114,197
3	2317(730)	126,130
2	2943(905)	121,266
4	2040(229)	74,119

Table 5: L-Criterion.

	Year					
Model	1990	1991	1992	1993	1994	1995
1	34(17)	66(24)	213(64)	404(117)	764(233)	1426(562)
2	31(21)	13(7)	96(41)	302(117)	638(249)	1235(632)
3	30(14)	71(27)	210(71)	443(171)	774(293)	1428(518)
4	28(3)	22(2)	99(9)	335(32)	495(59)	1062(169)

Table 6: Posterior Mean and Standard Deviation for Total Outstanding Claim Amounts of Each Accident Year (million drachmas; adjusted for inflation).

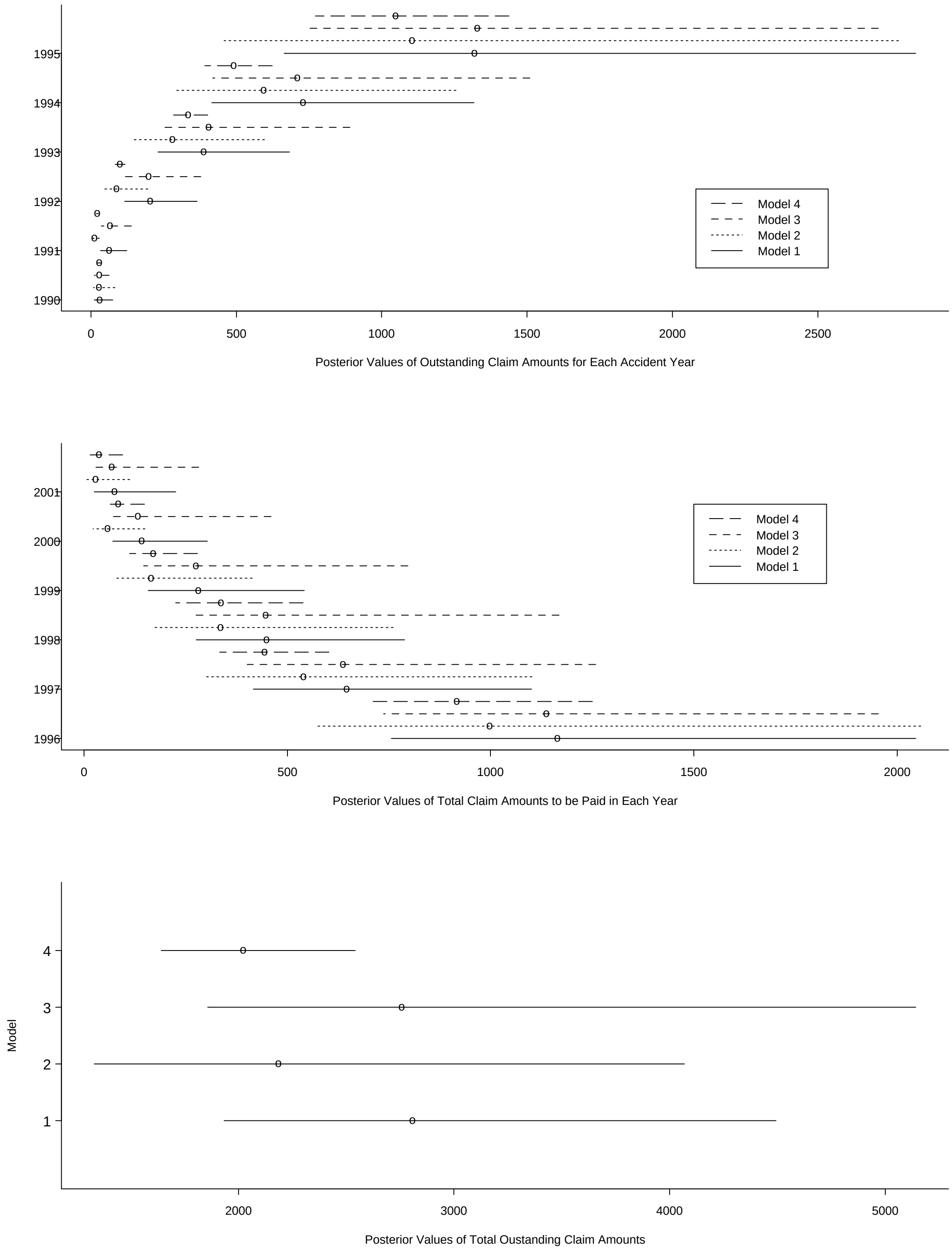


Figure 1: 95% Credible Intervals for Total Outstanding Deinflated Claim Amounts.

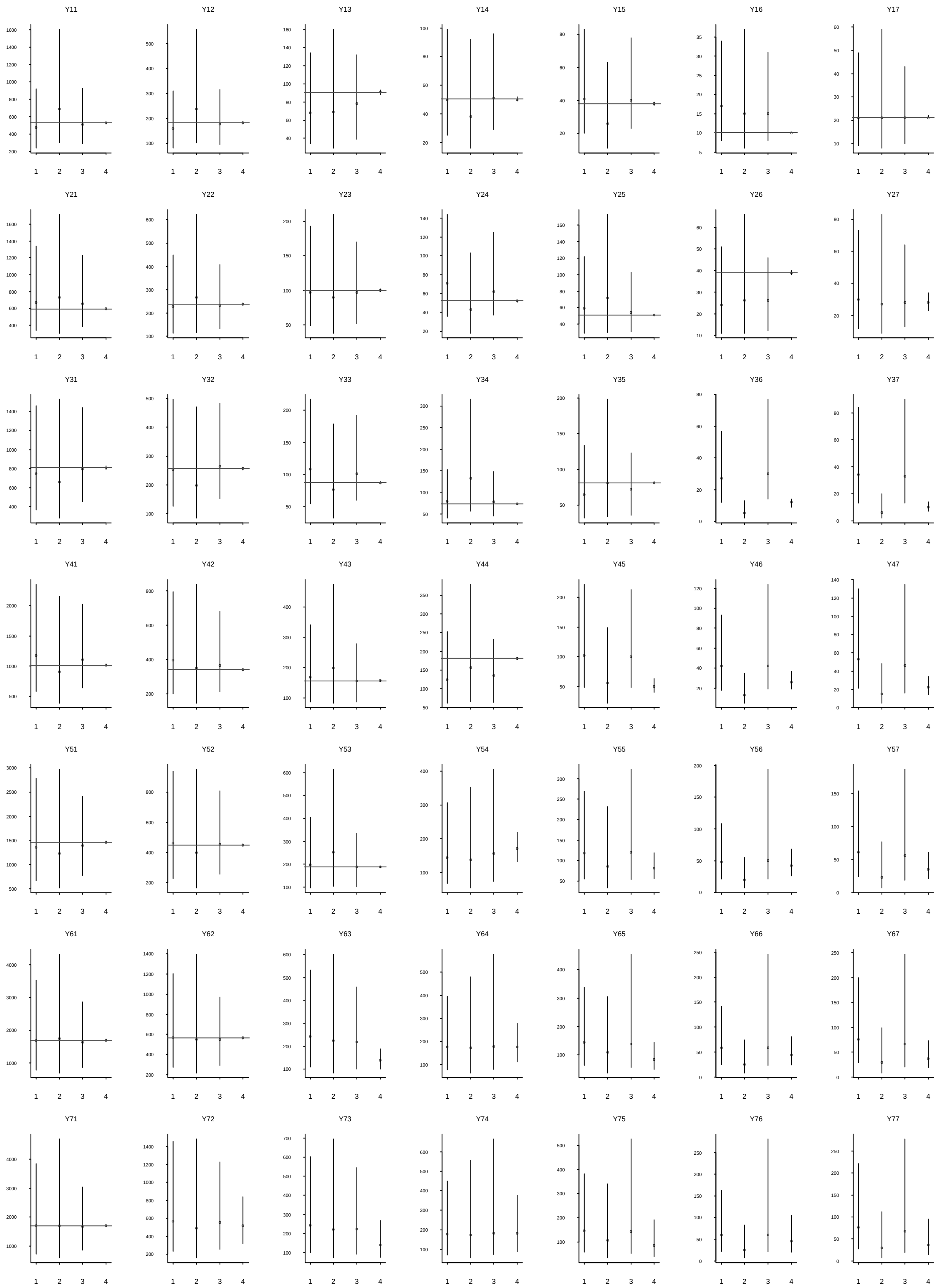


Figure 2: 95% Credible Intervals for Outstanding Deinflated Claim Amounts for Each Cell of Table 3 (Dots Indicate Posterior Medians; X-axis: Model; Y-axis: Claim Amounts in Million Drachmas; The Upper Triangle was Produced Using Predictive Posterior Distribution $f(Y_{ij}^{pred}|\boldsymbol{Y})$; Horizontal Line in Upper Triangle Indicate Actual Data Claim Amount; The Lower (Missing) Triangle was Produced Using Posterior Distribution of Missing Parameters).

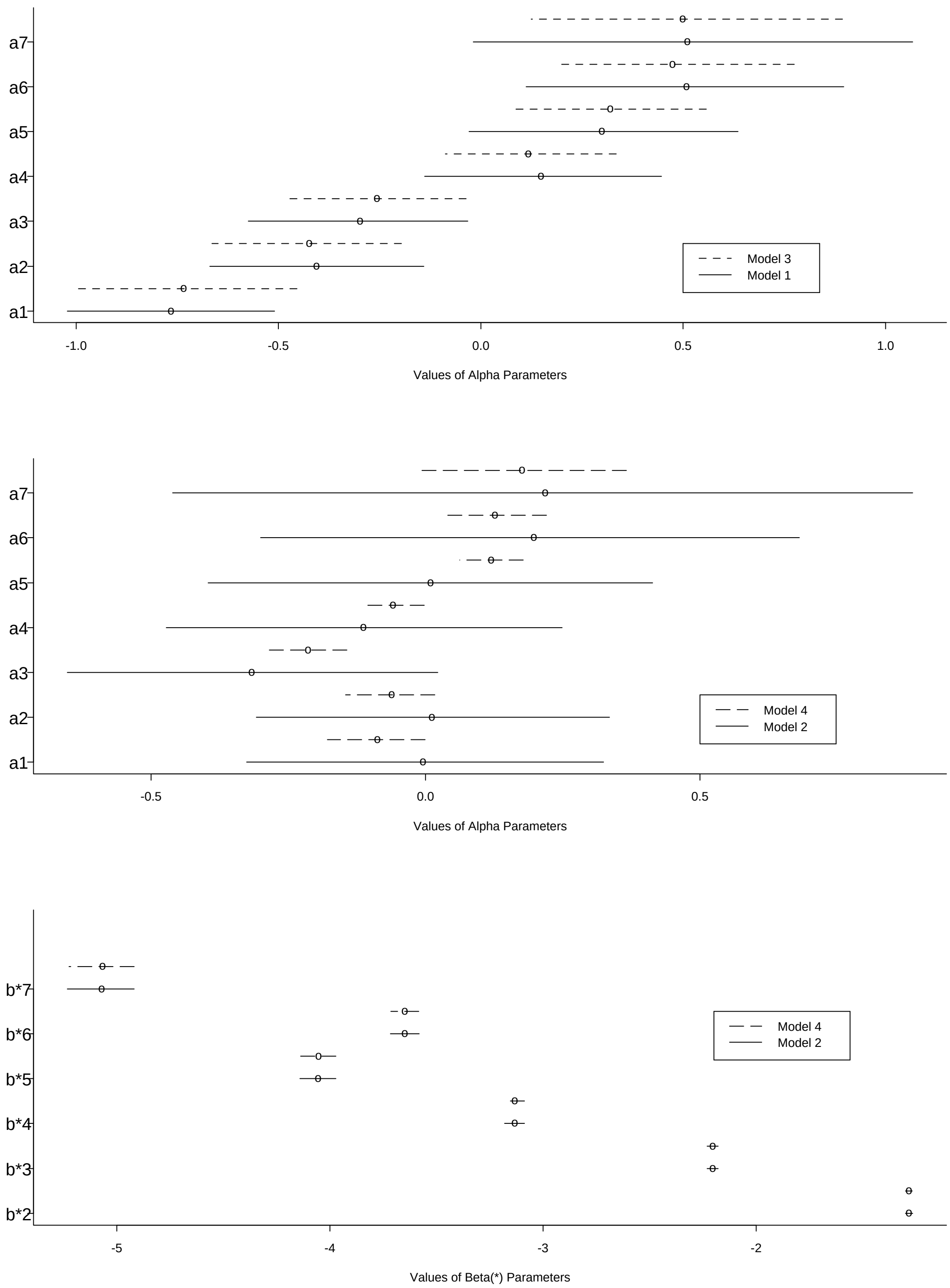
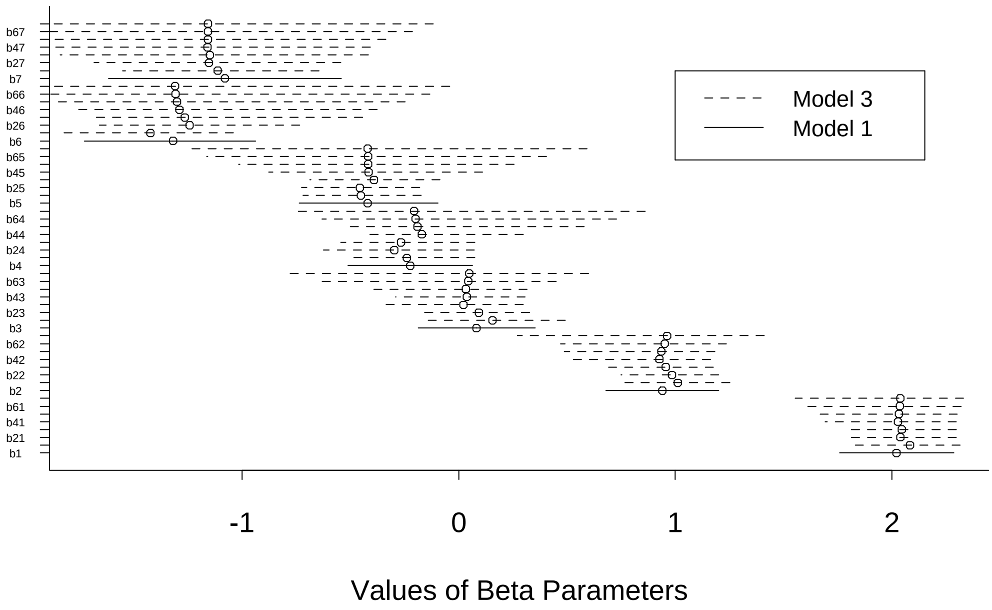


Figure 3: 95% Credible Intervals for Parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}^*$ (Dots Indicate the Posterior Median).



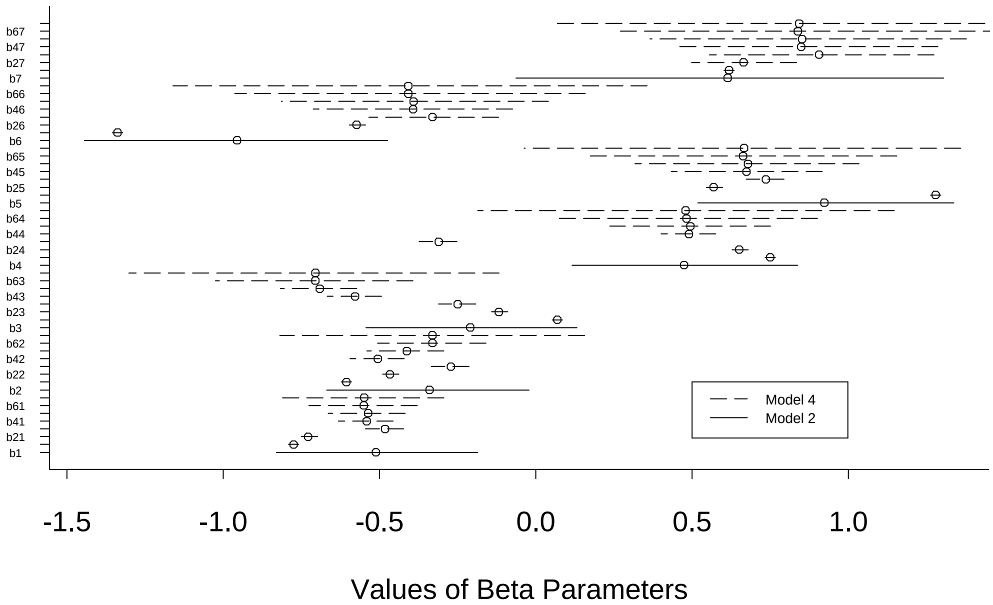


Figure 4: 95% Credible Intervals for Parameter Vector $\boldsymbol{\beta}$.

7 Conclusions

- Log-Normal model seems to give good results but does not adjust for claim counts.
- Modelling average claim amount per accidents demands more complicated models.
- State space model for average claim amount seems better than simple models used in the past.
- Bayesian statistics and MCMC can help researchers in Actuarial science to fit more complicated models and predict more accurately outstanding claims.