Credibility in Loss Reserving

Peng Shi
School of Business
University of Wisconsin - Madison
Email: pshi@bus.wisc.edu

Brian M. Hartman
Department of Mathematics
University of Connecticut
Email: brian.hartman@uconn.edu

Abstract

This article proposes using credibility theory in the context of stochastic claims reserving. We consider the situation where an insurer has access to the claims experience of its peer competitors and has the potential to improve prediction of outstanding liabilities by incorporating information from other insurers. Based on the framework of Bayesian linear models, we show that the development factor in the classical chain-ladder setting has a credibility expression, i.e. a weighted average of the prior mean and the best estimate from the data. In the empirical analysis, we examine loss triangles for the line of commercial auto insurance from a portfolio of insurers in the US. We employ hierarchical model for the specification of prior and show that prediction could be improved through borrowing strength among insurers based on a hold-out sample validation.

Keywords: Bayesian Modeling, Chain-ladder method, Hierarchical model

1 Introduction

General insurance (also known as property-casualty insurance in the U.S. and non-life insurance in other countries) protects a person or business against the losses to its physical property or legal liability through injury property damage. General insurance is a stable cornerstone of and makes significant contributions to any developed economy. In 2011, as the largest insurance market in the world the U.S. underwrote over $0.66 trillion U.S. dollars of premium in property and casualty insurance, which account for about 4.45% of the nation's GDP (*International Insurance Fact Book 2013*). Because of its critical role in the economy, the general insurance industry is usually highly regulated to monitor and ensure its financial health. For example, each insurer is required to provide sufficient technical provisions, also known as loss reserves, to support its potential outstanding liabilities.

Loss reserves represent the best estimate of an insurer's outstanding loss payments. In general insurance potential reporting lags, the settlement process, and potentially reopened claims can all lengthen the time to close a claim. For the purposes of valuation and financial reporting, the insurer predicts the ultimate payment amount for all the claims arising from past exposures. This includes estimates of both incurred but not reported and reported but not settled claims. The loss reserve is then built up based on the best estimate and updated at each valuation.

There is an extensive literature on the prediction of outstanding losses and quantification of associated predictive uncertainty. See, for example, Taylor (2000) and Wüthrich and Merz (2008) for comprehensive reviews. One approach worth mentioning is the chain-ladder method which is the current industry benchmark and is also the building block of the hierarchical model employed in this study. Think of a run-off triangle of cumulative payments, where aggregated paid losses are arranged in a triangular fashion to reflect the occurrence and development over years. The chain-ladder algorithm uses year-to-year development factors to project cumulative payments for each accident year. This simple algorithm is further justified by a variety of statistical models which also provide the foundation to quantify reserving variability. Several commonly used variations include the Mack chain ladder (Mack (1993, 1999)), the Munich chain ladder (Quarg and Mack (2008)), and bootstrap chain ladder (England and Verrall (2002)). Additionally, the chain-ladder model can be easily implemented in the statistical package R (see Sturtz et al. (2005)).

Incorporating the experience of loss payment from peers could add value to the prediction of an insurer's own liabilities. First, an insurer's own claim experience might not be reliable, especially for small insurers. In this case, the insurer might want to give less credibility to its own experience but more to the industry-level information. Second, an insurer could borrow strength in the prediction of its own outstanding claims by combining experiences from other companies that share similar claim payment patterns, often in the same line of business. Third, the claim experience of all insurers are influenced by certain common factors whether macroeconomic or due to a change in regulation, pooling experience from multiple insurers can better capture and measure such factors.

In this work, we develop a formal structure to incorporate claim information from peer insurers with an insurer's own information for reserving purposes. We focus on the chain-ladder approach and using the theory of Bayesian linear models, we show that the development factor in the claim-ladder method has a credibility expression, i.e. a weighted average of prior knowledge and an estimate from data. Furthermore, through hierarchical models we explore the impact of prior specification. The Bayesian approach is a natural choice to blend collateral information with an insurer's own claim experience. Additionally, Bayesian models naturally incorporate parameter uncertainty in the prediction. Bayesian methods have a long history in the loss reserving literature, with the earliest efforts traced back to 1990s (see, for example, Jewell (1989, 1990) and Verrall (1990)). Partly because of the development of the Markov chain Monte Carlo (MCMC) techniques, the loss reserving literature has observed an increasing number of applications from the Bayesian perspective. Some recent examples include Antonio and Beirlant (2008), de Alba and Nieto-Barajas (2008), Peters et al. (2009), Meyers (2009), Merz and Wüthrich (2010), Shi et al. (2012), and Zhang and Dukic (2012) among others.

Apart from the above literature, two recent studies incorporate information from multiple insurers for reserving. Zhang et al. (2012) employed a hierarchical growth curve to predict insurers' outstanding liabilities for a single business line. Extending this idea, Shi (2013) proposed a Bayesian copula regression model for determining reserves for dependent lines of business. Different from these studies, we focus on the classical chain-ladder model and derive a credibility estimate. Note that although credibility is widely used in ratemaking, to the best of our knowledge, it has not been studied in reserving. Furthermore, both Zhang et al. (2012) and Shi (2013) focused on prediction for the portfolio of insurers. In contrast, we emphasize the value of external information for individual insurers.

The rest of the article is structured as follows: Section 2 formulates the Bayesian linear model and presents the credibility results in reserving prediction. Section 3 describes the loss triangle data. Section 4 introduces the hierarchical model and proposes alternative choices for the prior specification. Model inferences are discussed as well. Section 5 demonstrates the prediction using the Bayesian model and compares model performance using out-of-sample validation. Section 6 concludes the paper.

2 Model

Credibility is a technique for incorporating relevant outside data and is widely used in ratemaking. Studies on credibility begin with Mowbray (1914) and Whitney (1918). The theoretical foundation for credibility ratemaking is due to Bühlmann (1967) where traditional credibility formulas are derived in a distribution-free setup using a least-squares criterion. The approach was subsequently extended and popularized by a series of studies (see Bühlmann and Gisler (2005) for a comprehensive review). Despite of its long history in ratemaking, credibility is rarely used in reserving even though the goal is prediction as well.

We investigate credibility in loss reserving based on the framework of Bayesian linear models and show the credibility results for the chain-ladder method. Bayesian credibility was introduced by Bailey (1950) and further extended by Mayerson (1964), Miller and Hickman (Miller and Hickman), and Luo et al. (2004) among others. Our study is unique because instead of focusing on a single insurer we show the credibility results for a group of insurers. We argue that an insurer could borrow predictive strength from the claims experience of peer insurers.

Consider N run-off triangles, each from an individual insurer. Assume all triangles are of the same dimension with I accident years and $J(=I)$ development years. Let $C_{i,j}^{n}$ denote the cumulative paid loss in the ith ($i=1,\cdots,I$) accident year and the jth ($j=0,\cdots,I-1$) development lag of the nth ($n=1,\cdots,N$) insurer. Define $\mathbf{C}_j^{(n)}=\left(C_{1,j}^{(n)},\cdots,C_{I,j}^{(n)}\right)'$ for $j=0,\cdots,I-1$ and $n=1,\cdots,N$. Denote $\mathbf{C}_{U,j}^{(n)}=\left(C_{1,j}^{(n)},\cdots,C_{I-j,j}^{(n)}\right)'$ and $\mathbf{C}_{L,j}^{(n)}=\left(C_{I-j+1,j}^{(n)},\cdots,C_{I,j}^{(n)}\right)'$ as the vector of cumulative payment in the upper triangle (realized loss) and lower triangle (outstanding payment), respectively.

For the purposes of brief presentation, we further define

$$\mathbf{Y}_j = \begin{pmatrix} \mathbf{C}_{U,j}^{(1)} \\ \vdots \\ \mathbf{C}_{U,j}^{(N)} \end{pmatrix}, \ \mathbf{X}_{j-1} = \begin{pmatrix} \mathbf{C}_{U,j-1}^{(1)} & & \\ & \ddots & \\ & & \mathbf{C}_{U,j-1}^{(N)} \end{pmatrix} \tag{1}$$

for $j = 1, \cdots, I-1$. To determine reserve, we follow the spirit of classic chain-ladder method and focus on the year-to-year development factors in the triangle. Specifically we examine the following linear model:

$$\mathrm{E}(\mathbf{Y}_j|\boldsymbol{\beta}_j) = \mathbf{X}_{j-1}\boldsymbol{\beta}_j \tag{2}$$

$$\mathrm{Var}(\mathbf{Y}_j|\boldsymbol{\beta}_j) = \mathbf{R}_j \tag{3}$$

where $\boldsymbol{\beta}_j = (\boldsymbol{\beta}_j^{(1)}, \cdots, \boldsymbol{\beta}_j^{(N)})'$ represents the vector of development factors from lag $j-1$ to j, and $\mathbf{R}_j$ denote the (conditional) covariance matrix for the jth development year.

We adopt a Bayesian approach for predicting outstanding payments and quantifying reserve variability. Using a conjugate multivariate normal prior $\boldsymbol{\beta}_j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Omega}_j)$, we have

$$\begin{pmatrix} \boldsymbol{\beta}_j \\ \mathbf{Y}_j \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_j \\ \mathbf{X}_{j-1}\boldsymbol{\mu}_j \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Omega}_j & \boldsymbol{\Omega}_j\mathbf{X}_{j-1}' \\ \mathbf{X}_{j-1}\boldsymbol{\Omega}_j & \mathbf{R}_j + \mathbf{X}_{j-1}\boldsymbol{\Omega}_j\mathbf{X}_{j-1}' \end{pmatrix}\right) \tag{4}$$

It is straight forward to derive the posterior distribution of $\boldsymbol{\beta}_j$ with

$$\mathrm{E}(\boldsymbol{\beta}_j|\mathbf{Y}_j) = \boldsymbol{\mu}_j + \boldsymbol{\Omega}_j\mathbf{X}_{j-1}'(\mathbf{R}_j + \mathbf{X}_{j-1}\boldsymbol{\Omega}_j\mathbf{X}_{j-1}')^{-1}(\mathbf{Y}_j - \mathbf{X}_{j-1}\boldsymbol{\mu}_j) \tag{5}$$

$$\mathrm{Var}(\boldsymbol{\beta}_j|\mathbf{Y}_j) = \boldsymbol{\Omega}_j - \boldsymbol{\Omega}_j\mathbf{X}_{j-1}'(\mathbf{R}_j + \mathbf{X}_{j-1}\boldsymbol{\Omega}_j\mathbf{X}_{j-1}')^{-1}\mathbf{X}_{j-1}\boldsymbol{\Omega}_j \tag{6}$$

Credibility Result 1: The posterior mean of development factor is a matrix-weighted average of the prior mean and the generalized least squares estimator, i.e. $\mathrm{E}(\boldsymbol{\beta}_j|\mathbf{Y}_j) = (\mathbf{I} - \boldsymbol{\zeta}_\beta)\boldsymbol{\mu}_j + \boldsymbol{\zeta}_\beta\boldsymbol{\beta}_j^{GLS}$, where $\boldsymbol{\zeta}_\beta = (\boldsymbol{\Omega}_j^{-1} + \mathbf{X}_{j-1}'\mathbf{R}_j^{-1}\mathbf{X}_{j-1})^{-1}\mathbf{X}_{j-1}'\mathbf{R}_j^{-1}\mathbf{X}_{j-1}$ and $\boldsymbol{\beta}_j^{GLS} = (\mathbf{X}_{j-1}'\mathbf{R}_j^{-1}\mathbf{X}_{j-1})^{-1}\mathbf{X}_{j-1}'\mathbf{R}_j^{-1}\mathbf{Y}_j$.

Proof.

$$
\begin{aligned}
&\boldsymbol{\Omega}_j\mathbf{X}_{j-1}^{'}(\mathbf{R}_j+\mathbf{X}_{j-1}\boldsymbol{\Omega}_j\mathbf{X}_{j-1}^{'})^{-1}\mathbf{Y}_j\\
=&\{\boldsymbol{\Omega}_j\mathbf{X}_{j-1}^{'}\mathbf{R}_j^{-1}-\boldsymbol{\Omega}_j\mathbf{X}_{j-1}^{'}\mathbf{R}_j^{-1}\mathbf{X}_{j-1}(\boldsymbol{\Omega}_j^{-1}+\mathbf{X}_{j-1}^{'}\mathbf{R}_j^{-1}\mathbf{X}_{j-1})^{-1}\mathbf{X}_{j-1}^{'}\mathbf{R}_j^{-1}\}\mathbf{Y}_j\\
=&\boldsymbol{\Omega}_j\{\mathbf{I}-\mathbf{X}_{j-1}^{'}\mathbf{R}_j^{-1}\mathbf{X}_{j-1}(\boldsymbol{\Omega}_j^{-1}+\mathbf{X}_{j-1}^{'}\mathbf{R}_j^{-1}\mathbf{X}_{j-1})^{-1}\}\mathbf{X}_{j-1}^{'}\mathbf{R}_j^{-1}\mathbf{Y}_j\\
=&\boldsymbol{\Omega}_j\{\boldsymbol{\Omega}_j^{-1}\}(\boldsymbol{\Omega}_j^{-1}+\mathbf{X}_{j-1}^{'}\mathbf{R}_j^{-1}\mathbf{X}_{j-1})^{-1}\mathbf{X}_{j-1}^{'}\mathbf{R}_j^{-1}\mathbf{Y}_j\\
=&\boldsymbol{\zeta}_\beta\boldsymbol{\beta}_j^{GLS}\\
&\boldsymbol{\mu}_j-\boldsymbol{\Omega}_j\mathbf{X}_{j-1}^{'}(\mathbf{R}_j+\mathbf{X}_{j-1}\boldsymbol{\Omega}_j\mathbf{X}_{j-1}^{'})^{-1}\mathbf{X}_{j-1}\boldsymbol{\mu}_j\\
=&\boldsymbol{\mu}_j-(\boldsymbol{\Omega}_j^{-1}+\mathbf{X}_{j-1}^{'}\mathbf{R}_j^{-1}\mathbf{X}_{j-1})^{-1}\mathbf{X}_{j-1}^{'}\mathbf{R}_j^{-1}\mathbf{X}_{j-1}\boldsymbol{\mu}_j\\
=&(\mathbf{I}-\boldsymbol{\zeta}_\beta)\boldsymbol{\mu}_j
\end{aligned}
$$

It is straightforward to see that when $\boldsymbol{\Omega}_j \to \infty$ or $\mathbf{0}$, $\boldsymbol{\zeta}_\beta \to \mathbf{I}$ or $\mathbf{0}$, respectively. That is, if one knows the true value of development factors, then zero credibility is given to the data. Otherwise, if one has no prior information on the development factors, full credibility is given to the data.

Credibility Result 2: The predictive mean of cumulative payment in each lower triangle is a weighted average of the prior mean and the best prediction, i.e. $\mathrm{E}(\mathbf{C}_{L,j-1}^{(n)}\boldsymbol{\beta}_j|\mathbf{Y}_j) = (1-\zeta_\beta^{(n)})\mathbf{C}_{L,j-1}^{(n)}\mu_j^{(n)} + \zeta_\beta^{(n)}\mathbf{C}_{L,j-1}^{(n)}\beta_j^{(n)GLS}$, iff

$$
\mathbf{R}_j = \begin{pmatrix} \mathbf{R}_j^{(1)} & & \\ & \ddots & \\ & & \mathbf{R}_j^{(N)} \end{pmatrix} \text{ and } \boldsymbol{\Omega}_j = \begin{pmatrix} \left(\omega_j^{(1)}\right)^2 & & \\ & \ddots & \\ & & \left(\omega_j^{(N)}\right)^2 \end{pmatrix} \tag{7}
$$

where $\zeta_\beta^{(n)} = \frac{\mathbf{C}_{U,j-1}^{(n)'}\left(\mathbf{R}_j^{(n)}\right)^{-1}\mathbf{C}_{U,j-1}^{(n)}}{\left(\omega_j^{(n)}\right)^{-2}+\mathbf{C}_{U,j-1}^{(n)'}\left(\mathbf{R}_j^{(n)}\right)^{-1}\mathbf{C}_{U,j-1}^{(n)}}$, $\beta_j^{(n)GLS} = \frac{\mathbf{C}_{U,j-1}^{(n)'}\left(\mathbf{R}_j^{(n)}\right)^{-1}\mathbf{C}_{U,j}^{(n)}}{\mathbf{C}_{U,j-1}^{(n)'}\left(\mathbf{R}_j^{(n)}\right)^{-1}\mathbf{C}_{U,j-1}^{(n)}}$, and $\mu_j^{(n)}$ is the nth element in $\boldsymbol{\mu}_j$. Furthermore, if $\mathbf{R}_j^{(n)} = diag\left(\left(\sigma_j^{(n)}C_{1,j-1}^{(n)}\right)^2, \cdots, \left(\sigma_j^{(n)}C_{I-j,j-1}^{(n)}\right)^2\right)$, the predictive mean of the outstanding payment is a weighted average of the prior mean and the chain-ladder prediction.

Proof. The first part of the result follows from conditional assumption among triangles. The second part of the result is due to $\beta_j^{(n)GLS} = \sum_{i=1}^{I-j} C_{i,j}^{(n)} / \sum_{i=1}^{I-j} C_{i,j-1}^{(n)}$ $(n = 1, \cdots, N)$, which is the chain-ladder development factor.

Note that the above results could be derived for each individual triangle. We emphasize that by pooling triangles from multiple insurers we allow an insurer to blend its own claim experience with its peers. The information sharing could be achieved by allowing triangles from different insurers to be correlated with each other. Explicitly, that correlation could be introduced through the sampling distribution. In this study, we focus on an implicit strategy, a hierarchical prior specification (see Section 4). The hierarchical model is more natural and intuitive for this application, allowing an insurer to adjust its priors based on the information borrowed from other insurers.

3 Data

In the empirical analysis, we consider run-off triangles of commercial automobile insurance from a group of property-casualty insurers in the US. The data are from Schedule P of the National Association of Insurance Commissioners (NAIC) database. The triangles are available in terms of both incurred and paid losses. Our analysis uses the 1997 paid losses. Each triangle contains payments for the claims in ten accident years from 1988 to 1997, and for each accident year up to ten development lags. Table 1 illustrates organization of the data. For example, the first row contains payments for claims which occurred in 1988. Because of the reporting and settlement lags, we observe payments from 1988 through the valuation year, 1997. In contrast, for accident year 1997, we only have one year of payments by the valuation year.

Table 1: Run-off triangle from Schedule P of NAIC

Accident Year	0	1	2	3	4	5	6	7	8	9		
1988	×	×	×	×	×	×	×	×	×	×		
1989	×	×	×	×	×	×	×	×	×		←	1998
1990	×	×	×	×	×	×	×	×			←	1999
1991	×	×	×	×	×	×	×				←	2000
1992	×	×	×	×	×	×					←	2001
1993	×	×	×	×	×						←	2002
1994	×	×	×	×							←	2003
1995	×	×	×								←	2004
1996	×	×									←	2005
1997	×										←	2006

The goal of reserving practice is to identify the payment pattern based on realized paid losses and to predict outstanding future payments. Using the example in Table 1, and assuming that all claims will be settled in ten years, we predict the unpaid losses represented by the cells in the highlighted lower triangle. To validate the model, we use a hold-out sample to evaluate the prediction. In our analysis, we will use the data from 1997 to develop the model and use realizations of future payments in lower triangles to examine the predictive performance of alternative models. The validation data are extracted from the Schedule P in the NAIC database of subsequent years 1998-2006. Specifically, the paid losses of accident year 1989 are from the Schedule P of year 1998, the paid losses of accident year 1990 are from the Schedule P of year 1999, and so on. This process is also demonstrated in Table 1 where the last column indicates the year from which the future payments in lower triangles are gathered.

Schedule P contains firm level run-off triangles of aggregated claims for major business lines of U.S. property-casualty insurers. Examples include personal auto liability, commercial auto liability, worker's compensation, general liability, and medical malpractice. The settlement periods for liability insurance could be lengthy due to late reporting, protracted negotiations, or judicial proceedings. However, the triangle data of Schedule P only contains payments for the most recent ten years. Because of this drawback, we focus on commercial auto liability where, compared with other casualty lines, the loss payments have relatively shorter tails and take fewer years to close.

In our analysis, we examine fifteen insurers with large commercial auto liability books. We expect that insurers could borrow more from peers of similar size. In selecting the group of insurers, we also make sure that there is no major merger and acquisition in this particular line of business over the study period. Specifically, the Schedule P of years 1998-2006 contains paid losses in the upper triangles that are already extracted from the Schedule P of year 1997 as well. We use observations in overlapping years to cross-validate the data quality of the selected insurers. To visualize the data, Figure 1 displays the development of cumulative payments for each insurer by accident year. Each curve connects the paid losses over time corresponding to a single accident year. As anticipated, the curve flattens in later development years. In particular, there is no substantial increase in the payment from the eighth to the ninth development lag for accident year 1988, which supports our assumption that it takes about ten years to close all the claims. Notice that the volume of business written varies over years and there is substantive heterogeneity across insurers.

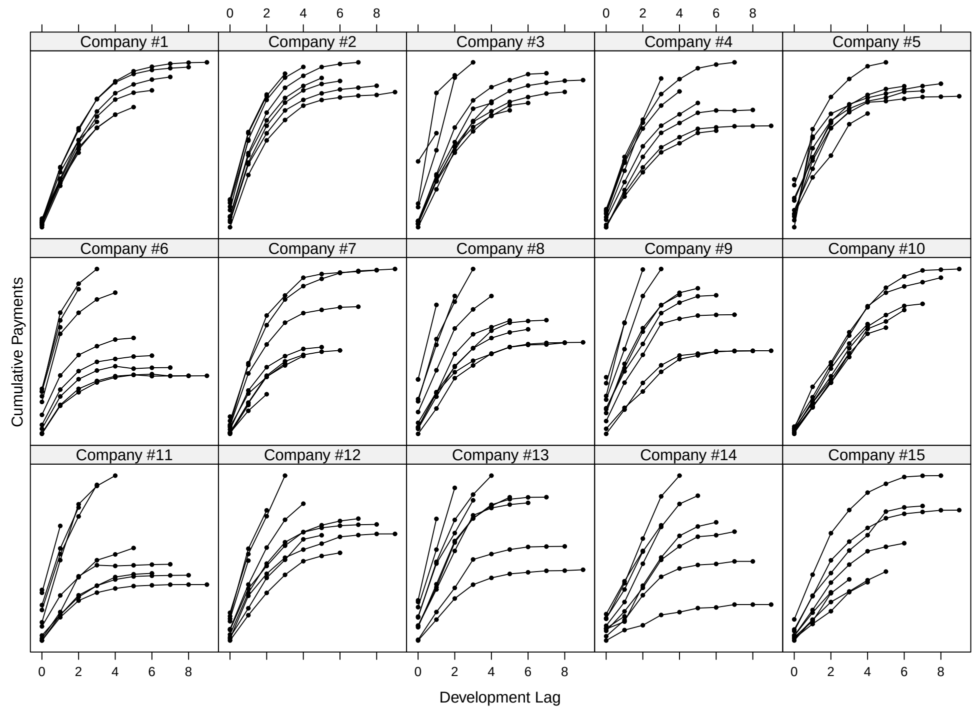

Figure 1: Multiple time series plot of cumulative paid loss

4 Bayesian Inference

In the empirical analysis, we introduce correlation between insurers and thus enable information borrowing simply from the hierarchical specification. Specifically, we start with the model

$$C_{ij}^{(n)} \sim N\left(C_{ij-1}^{(n)}\beta_j^{(n)}, \left(\sigma_j^{(n)}C_{ij-1}^{(n)}\right)^2\right)$$
$$\beta_j^{(n)} \sim N(\mu_j, \theta^2)$$

Here, we assume that the development factors in the jth year, $\beta_j^{(n)}$, have the same prior distribution with mean μ_j. an insurer is expected to incorporate experience of payment development from other insurers into its own experience. The parameter θ^2 is fixed and known. It determines the degrees of shrinkage among multiple insurers in that smaller values will increase the shrinkage and larger values will weaken it. We employ an empirical Bayes estimates for $\sigma_j^{(n)}$ from the classical chain-ladder model. This allows for fair comparison to the chain-ladder prediction and demonstrates the value added by credibility.

There are different ways to specify the prior distribution for hyperparameter μ_j. We discuss two alternatives that are particularly useful in reserving applications. The first and a natural choice is a conjugate prior. We use

$$\mu_j \sim N(a, b^2)$$

where b^2 controls the precision of prior knowledge that one has on μ_j and also create shrinkage in the development factors over years. We use $a = 1$ and impose a diffuse prior $b = +\infty$ assuming that an insurer has no prior knowledge on the development factor and the only way to gather information is to learn through its peers. The diffuse prior also guarantees the heterogeneity in development factors over time, which is desirable because we do not expect shrinkage over time though we anticipate shrinkage across insurers.

Alternatively, we know that as payments develop over time the development factors will tend to one. We can think of it as a change point where at some development time k, the claims are settled and all later factors are one. Specifically, the model is written as follows:

$$\mu_j \sim \begin{cases} N(a, b^2) & \text{if } j < k \\ N(1, 0.0001^2) & \text{if } j \geq k \end{cases}$$
$$k \sim DU(1, 10)$$

Here we assume that there are two states for hyperparameter μ_j. The posterior of parameter k determines the time period that it takes to close all claims such that the development factor is essentially one. Assuming no prior knowledge, we use a discrete uniform prior. Note that it is possible that it takes longer to close all claims than the window period of the triangle. In our application, claims might continue to develop after ten years. In this case, k will be 10. In practice, the domain knowledge of the reserving actuaries will determine the priors. Another choice that serves a similar purpose is to think of the prior of μ_j as a mixture of a normal distribution and 1, then the value of the weight for the normal distribution is the posterior probability that the development factor is significantly different from 1 (π_j):

$$\mu_j \sim \begin{cases} N(a, b^2) & \text{with probability } \pi_j \\ N(1, 0.0001^2) & \text{with probability } 1 - \pi_j \end{cases}$$

$$\pi_j \sim Unif(0, 1)$$

We estimate the hierarchical model using 50,000 MCMC iterations with the first 40,000 iterations discarded as a burn-in sample. Though not reported here, we generate multiple chains from different initial values, and the convergence for each parameter is confirmed with the Gelman-Rubin statistic. The posterior of k appears to be 8, indicating that the hyperparameter μ_j will transit to the absorbing state (=1) in the eighth development year.

To compare between the conjugate prior and the change point prior, Figure 2 presents the posterior distribution of μ_j. The left and right panel represents the posterior when using the conjugate and change point prior, respectively. Each box-plot corresponds to the prior mean of the development factor in each accident year. As anticipated, we observe relative larger development factors in the early stage and the the rate at which claims develop decreases over time. The two panels display similar patterns in the development factors. The subtle difference is that the development factors in the last two development years are equal to one under the change point process, however, they follow normal distribution under the conjugate prior. It is not surprising to see the little difference because when essentially all the claims are closed in the last two development years as suggested by the change point process, the normal distribution could not pick up much variability in the data.

The Bayesian linear model in Section 2 is based on the normality assumption. We employ residual analysis to validate this assumption. Note that residuals are not well defined in a Bayesian context. We follow the classic definition and calculate residuals as $e_{ij}^{(n)} = \left(C_{ij}^{(n)} - \hat{\beta}_j^{(n)} C_{ij-1}^{(n)} \right) / \left(\hat{\sigma}_j^{(n)} C_{ij-1}^{(n)} \right)$, where $\hat{\beta}_{ij}^{(n)}$ is the the posterior mode and $\hat{\sigma}_j^{(n)}$ is the empirical Bayes estimates. We present the normal qq plot in Figure 3. The agreement with the 45 degree line is consistent with the normality assumption. Also reported in Figure 3 is the plot of residual versus fitted value, where no particular pattern is detected. Note that because there is little difference between the conjugate hyperprior and the change point hyperprior, we only report the results from one of the two models.

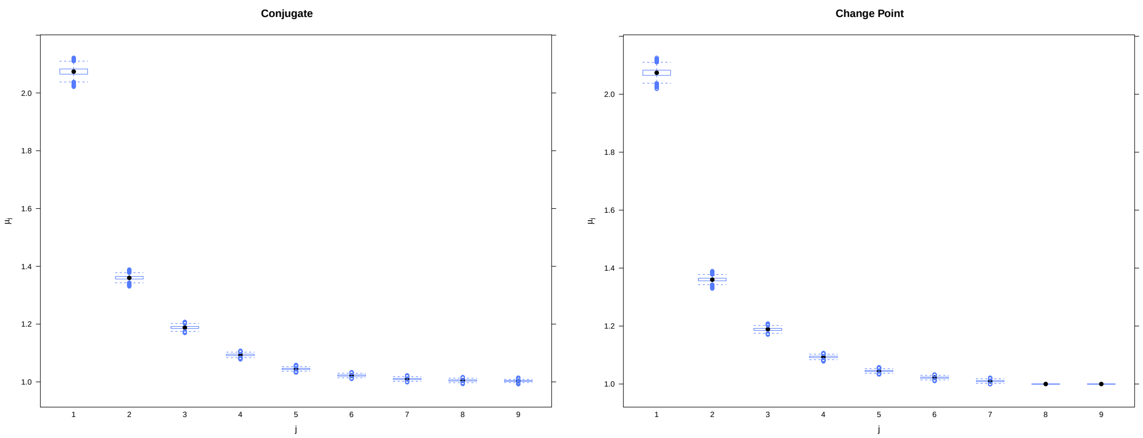

Figure 2: Posterior distribution of μ_j under conjugate and change point priors

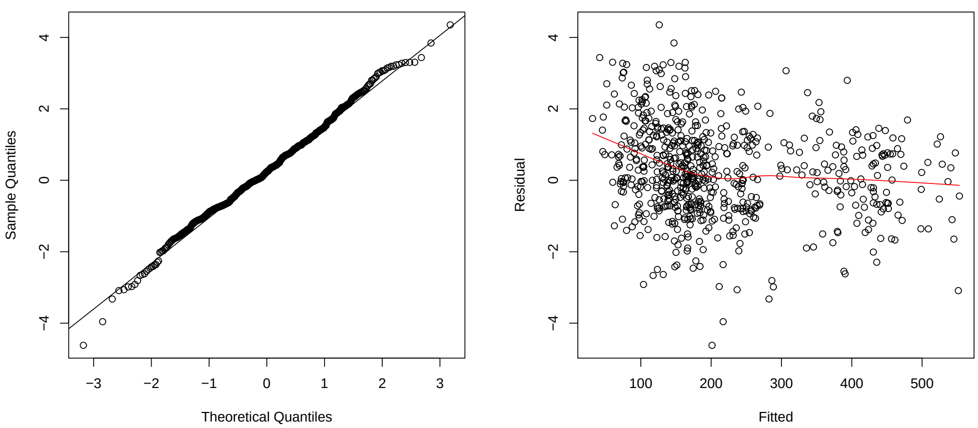

Figure 3: Normal qq plot and residual plot

By pooling multiple triangles, an insurer could gain predictive power by borrowing strength from other insurers. This is reflected by the shrinkage effect on the development factors, which is illustrated in Figure 4. Each panel reports the development factors in a particular year. Recall that parameter θ^2 controls the degrees of shrinkage. We estimate the model at $\theta = 1,\ 0.1,$ and 0.01. Within a panel, each curve connects the development factors estimated at various shrinkage for a single insurer. For comparison, we also report the development factor in the chain ladder model. As anticipated we see that a smaller θ shrinks the development factors of all insurers toward the group average. We also observe a larger shrinkage effect on the development factors in early years but smaller effect for later years. This is explained by the weak heterogeneity across insurers in later development years and the small variability in their posterior mean as shown in Figure 2. In the extreme case, the change point process even suggests that the expected development factors in the most recent two years are equal to one. There is no shrinkage effect in the chain ladder approach. As indicated in Section 2, the Bayesian linear model will reproduce the chain-ladder prediction when diffuse prior is used for inference. Finally, we stress that the degrees of shrinkage, i.e. whether to rely on an insurer's own claim experience or to adjust the prediction toward the industry average, requires the expert knowledge of reserving actuaries.

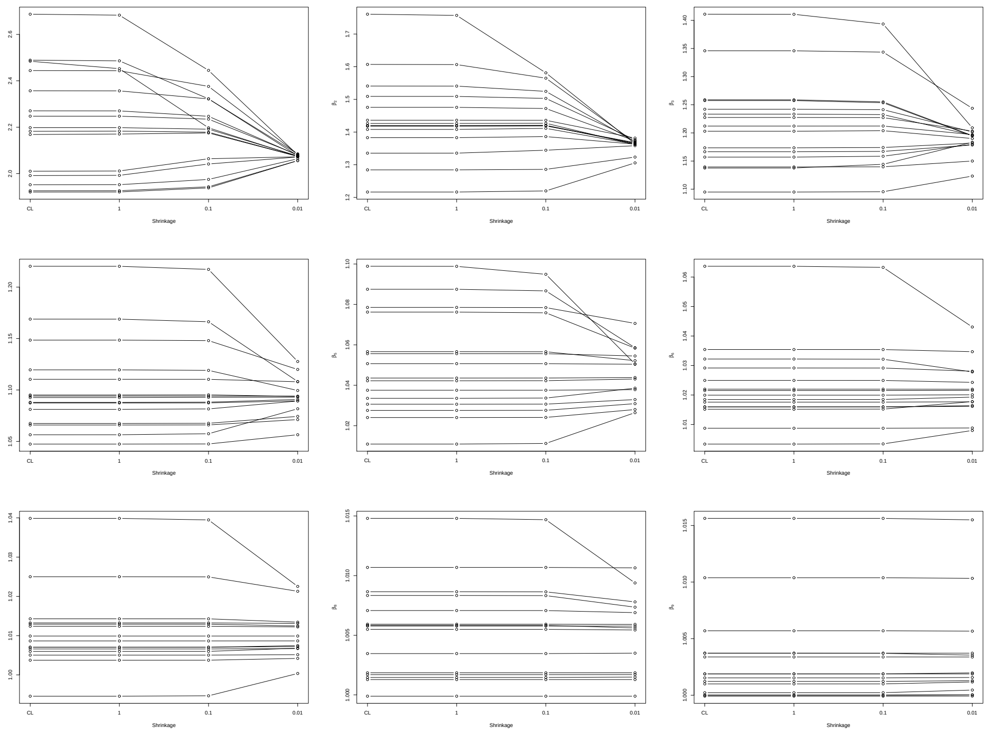

Figure 4: Shrinkage effects on development factors

5 Prediction

In this study we use chain-ladder method as the benchmark from which we employ a hierarchical model to introduce credibility. The inference has focused on the development factor in the chain-ladder framework. Therefore, the prediction follows in a straightforward way. The reserve (outstanding payments) for insurer n, $R^{(n)}$, is estimated by

$$\widehat{R}^{(n)} = \sum_{i=1}^{I} C_{iI-i}^{(n)} \left(\hat{\beta}_{I-i}^{(n)} \cdots \hat{\beta}_{I-1}^{(n)} - 1 \right)$$

where $\hat{\beta}_j^{(n)}$ $(j = 0, \cdots, I-1)$ are the best estimates i.e. the posterior mode from the hierarchical model. Wisely, reserving actuaries are more interested in a credible predictive range than a single point prediction. A commonly used measure of reserve variability is the mean squared prediction error that combines both uncertainty in the stochastic model and the unknown parameters.

Straightforward calculation shows similar decoupling results in a Bayesian context:

$$\text{MSEP}_{R^{(n)}} = \text{E}\left[(R^{(n)} - \widehat{R}^{(n)})^2\right] = \text{E}\left[\text{Var}(R^{(n)}|\boldsymbol{\Theta})\right] + \text{Var}\left[\text{E}(R^{(n)}|\boldsymbol{\Theta})\right]$$

The total variance (TV) is decomposed into average process variance (PV) and estimation error (EE) as in the classical analysis.

Tables 2 and 3 summarize the prediction results using the conjugate hyperprior and the change point hyperprior respectively. We report in each table the best estimate of firm-level reserves and the associated variability with the decoupling components under different degrees of shrinkage. The panel with $\theta = +\infty$ is equivalent to the chain ladder prediction. As θ decreases, the shrinkage effect strengthens. The amount of shrinkage has a pretty significant impact on the reserve, especially when $\theta = 0.01$. Note that the hierarchical specification drives the development factor not necessarily the reserve toward the group average, because the development factor and reserve could be negatively correlated. For example, a small firm might have a larger development factor, thus the shrinkage prediction would lower the reserve prediction. In addition, we also observe the effect of shrinkage on the reserving variability, especially for the estimation uncertainty. This is expected because the uncertainty in hyperpriors will be added to the parameter estimates. The average process variance is small because the conditional process variance is calculated following the chain-ladder approach and the estimation uncertainty is subdued by the averaging process. Consistent with results in Section 4, the predictions under the conjugate hyperprior and the change point hyperprior are quite similar.

In the above analysis, we have used a diffuse prior ($a = 1,\ b = +\infty$) for the hyperparameter μ_j, assuming that no prior knowledge is available at the point of valuation. The Bayesian approach allows expert opinions into the inference process. This could also be viewed as a downside because management could manipulate loss reserves through prior beliefs to manage earnings or hide solvency issues, though this is somewhat true under standard models depending on how the development factors are chosen or which method is used. We perform a prior sensitivity analysis of the reserve predictions to determine the extent of that control. Specifically, we consider the six combinations of $a = 0.5,\ 1,\ 2$ and $b = 0.1,\ 1$. The reserve estimates, total variance, process variance and estimation error are calculated under each specification. Along with the base case,

Table 2: Reserve prediction using conjugate prior

Company	Reserve	$\sqrt{TV}$	$\sqrt{PV}$	$\sqrt{EE}$	Reserve	$\sqrt{TV}$	$\sqrt{PV}$	$\sqrt{EE}$
	$\theta = +\infty$				**$\theta = 1$**			
1	498,645	28,222	24,953	13,184	498,808	28,244	24,956	13,226
2	410,216	18,174	14,975	10,298	410,150	18,273	14,975	10,472
3	490,000	90,432	80,935	40,341	490,151	90,694	80,958	40,880
4	463,987	30,850	25,902	16,758	463,906	31,072	25,901	17,164
5	157,824	46,527	41,740	20,555	156,413	45,899	41,743	19,085
6	67,497	6,874	4,633	5,078	67,543	6,808	4,634	4,988
7	93,136	10,601	9,447	4,810	93,046	10,538	9,446	4,672
8	145,421	11,175	8,435	7,331	145,446	11,216	8,436	7,391
9	99,618	9,445	7,376	5,899	99,765	9,465	7,380	5,927
10	83,508	7,952	5,818	5,420	83,536	7,979	5,819	5,459
11	84,934	9,971	7,916	6,062	85,222	9,957	7,924	6,029
12	88,281	7,525	5,807	4,785	88,347	7,534	5,809	4,798
13	239,553	21,880	17,005	13,768	239,992	21,913	17,016	13,807
14	82,357	12,395	10,691	6,271	82,179	12,364	10,681	6,228
15	42,301	6,207	5,248	3,316	42,212	6,208	5,243	3,323
	$\theta = 0.1$				**$\theta = 0.01$**			
1	497,245	28,031	24,938	12,800	450,286	25,672	23,996	9,125
2	413,177	18,227	15,001	10,354	455,377	18,261	15,500	9,656
3	491,968	88,191	81,058	34,747	444,800	79,260	77,503	16,594
4	458,874	30,734	25,808	16,690	379,806	26,294	24,102	10,509
5	146,799	43,454	42,125	10,665	153,487	44,558	44,190	5,717
6	68,806	6,796	4,648	4,958	95,992	6,648	5,024	4,353
7	88,571	10,233	9,386	4,075	80,285	9,472	9,173	2,362
8	146,181	11,180	8,449	7,322	145,011	9,639	8,501	4,544
9	101,446	9,428	7,405	5,835	95,709	7,968	7,224	3,362
10	82,514	7,897	5,789	5,372	52,722	5,954	4,923	3,348
11	87,309	9,791	7,972	5,684	92,342	8,722	8,185	3,013
12	87,964	7,435	5,803	4,648	76,680	6,294	5,548	2,973
13	232,388	21,211	16,845	12,889	179,319	16,735	15,582	6,103
14	75,427	11,404	10,161	5,177	45,127	8,357	7,977	2,491
15	39,045	5,946	5,103	3,051	26,129	4,646	4,474	1,251

Table 3: Reserve prediction using change point prior

Company	Reserve	$\sqrt{TV}$	$\sqrt{PV}$	$\sqrt{EE}$	Reserve	$\sqrt{TV}$	$\sqrt{PV}$	$\sqrt{EE}$
	$\boldsymbol{\theta = +\infty}$				$\boldsymbol{\theta = 1}$			
1	498,870	28,317	24,955	13,382	498,910	28,393	24,959	13,535
2	410,435	18,217	14,978	10,369	410,483	18,189	14,979	10,319
3	490,445	91,108	80,969	41,768	489,781	90,693	80,918	40,959
4	464,048	31,019	25,903	17,065	463,516	31,043	25,890	17,127
5	157,670	46,508	41,728	20,537	156,102	46,060	41,725	19,507
6	67,681	6,863	4,635	5,061	67,663	6,818	4,635	5,000
7	93,040	10,576	9,441	4,766	92,919	10,526	9,437	4,662
8	145,140	11,165	8,430	7,320	145,104	11,212	8,430	7,393
9	99,813	9,437	7,381	5,880	99,940	9,427	7,384	5,859
10	83,543	8,011	5,820	5,506	83,692	8,021	5,823	5,517
11	84,922	9,924	7,916	5,986	84,863	9,882	7,914	5,918
12	88,409	7,511	5,810	4,760	88,209	7,551	5,806	4,827
13	239,901	21,956	17,012	13,880	240,141	22,015	17,019	13,964
14	82,378	12,396	10,696	6,266	82,236	12,350	10,685	6,193
15	42,236	6,216	5,245	3,336	42,196	6,235	5,242	3,376
	$\boldsymbol{\theta = 0.1}$				$\boldsymbol{\theta = 0.01}$			
1	497,308	28,205	24,936	13,180	450,274	25,654	23,996	9,074
2	413,434	18,341	15,003	10,549	454,381	18,193	15,491	9,540
3	491,147	88,324	80,998	35,220	445,029	79,329	77,528	16,808
4	458,812	30,773	25,806	16,763	379,380	26,257	24,095	10,433
5	147,018	43,513	42,132	10,873	152,649	44,506	44,139	5,700
6	68,799	6,870	4,648	5,059	96,024	6,605	5,024	4,288
7	88,497	10,203	9,385	4,003	80,238	9,468	9,171	2,353
8	146,199	11,119	8,449	7,228	144,272	9,592	8,488	4,468
9	101,371	9,434	7,403	5,848	95,695	7,968	7,224	3,362
10	82,402	7,942	5,787	5,440	51,792	5,837	4,906	3,162
11	87,259	9,824	7,972	5,741	92,359	8,718	8,186	2,998
12	87,938	7,459	5,803	4,687	76,402	6,285	5,543	2,963
13	232,652	21,360	16,853	13,124	179,148	16,699	15,578	6,016
14	75,096	11,400	10,143	5,204	45,057	8,355	7,975	2,491
15	39,236	5,951	5,113	3,046	25,911	4,638	4,467	1,248

we present the reserve estimates and the predictive uncertainty in Figure 5. Each line in the figure represents an individual insurer. The predictions and associated variability are relatively robust to the prior specification, suggesting that data are informative enough for model inference.

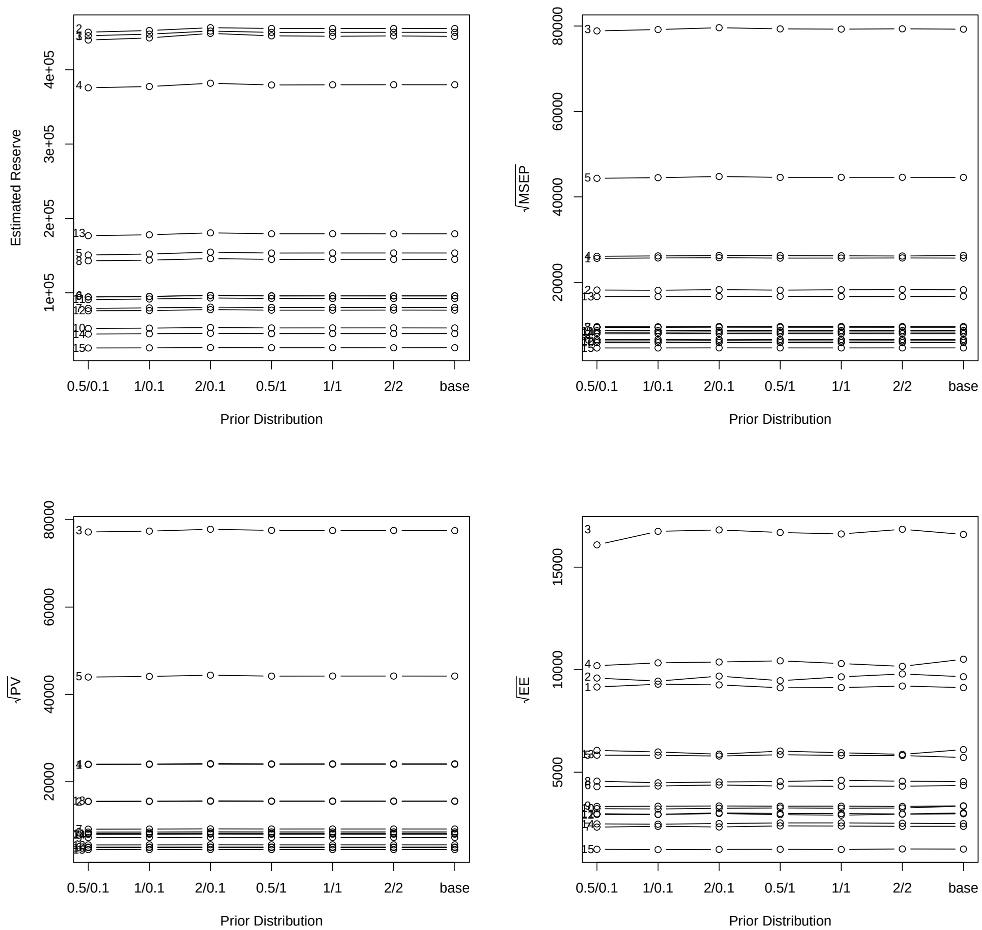

Figure 5: Robust analysis of reserve prediction

We employ out-of-sample validation to examine the value added by the information pooling. Recall that we have access to the Schedule P in years 1998-2006, and thus can calculate the actual amount of future payments (paid losses in the lower triangle) for each insurer, denoted by $Q^{(n)}$. Based on the predictive distribution of reserve $R^{(n)}$, we compute the two-sided p-value $\min\left\{\Pr(R^{(n)} < Q^{(n)}),\ \Pr(R^{(n)} > Q^{(n)})\right\}$. A smaller p-value indicates a more extremal outcome, i.e. the realized outcome is further away from the center of prediction. Because both under and over reserving could be detrimental to the insurer, a small p-value implies poor predictive performance.

The p-value is calculated using both conjugate prior and change point prior and using different degrees of shrinkage for each prior. Results are presented in Table 4 with the largest p-value highlighted. Recall that the diffuse prior $\theta = +\infty$ reproduce the chain ladder predictions. The small p-values for this scenario suggest that using some degrees of shrinkage to borrow information from peer insurers, an insurer is as least as good as the chain ladder method and as the amount of shrinkage increases, the model improves.

Table 4: p-values from out-of-sample validation

	Conjugate Prior				Change Point Prior			
Company	$\theta = +\infty$	$\theta = 1$	$\theta = 0.1$	$\theta = 0.01$	$\theta = +\infty$	$\theta = 1$	$\theta = 0.1$	$\theta = 0.01$
1	0.093	0.093	0.100	**0.333**	0.093	0.093	0.101	**0.332**
2	0.001	**0.001**	0.001	0.000	0.001	**0.001**	0.001	0.000
3	0.025	0.025	0.021	**0.048**	0.025	0.025	0.022	**0.048**
4	0.045	0.046	0.062	**0.113**	0.046	0.048	0.063	**0.109**
5	**0.250**	0.237	0.164	0.211	**0.248**	0.236	0.166	0.205
6	0.015	0.015	**0.023**	0.020	0.016	0.015	**0.024**	0.020
7	**0.000**	0.000	0.000	0.000	**0.000**	0.000	0.000	0.000
8	0.094	**0.094**	0.083	0.069	0.098	**0.099**	0.081	0.078
9	0.151	0.148	0.109	**0.231**	0.146	0.142	0.111	**0.232**
10	0.155	0.155	**0.185**	0.000	0.156	0.152	**0.190**	0.000
11	0.234	0.242	0.310	**0.492**	0.232	0.230	0.308	**0.492**
12	0.003	0.003	0.004	**0.082**	0.003	0.004	0.004	**0.088**
13	0.007	0.006	0.013	**0.362**	0.006	0.006	0.013	**0.358**
14	0.044	0.045	**0.106**	0.027	0.044	0.044	**0.111**	0.027
15	0.291	0.296	**0.489**	0.003	0.295	0.297	**0.476**	0.003
Selected	2	2	4	7	2	2	4	7

6 Conclusion

In this paper, we investigated credibility in reserving. We started with the classical chain-ladder method and, based on Bayesian linear models, we showed credibility results for both development factors and reserve estimates, i.e. a weighted average of prior knowledge and best estimates from the data. Further, we employed a hierarchical model for the prior specification such that an insurer could blend its own experience with claim experience from peer insurers. The hierarchical specification also leads to a shrinkage effect on the information across insurers. We emphasized that the degree of shrinkage used in the prediction is a judgement call of the reserving actuaries, allowing for more flexibility in the model.

In the empirical analysis, we examined a portfolio of fifteen large US property-casualty insurers' commercial auto insurance lines. We explored alternative approaches for prior specification, including conjugate and change point priors. The former is a natural choice for hierarchical model, and the latter is particularly useful if one is more interested in the payment pattern in the tails. We illustrated the advantage of the Bayesian approach to quantify reserve variability. Without loss of interpretability, the total variance can still be decomposed into the process variance and estimation error. Through out-of-sample validation, we showed that prediction for individual insurers can be improved by borrowing strength from peer insurers.

References

Antonio, K. and J. Beirlant (2008). Issues in claims reserving and credibility: a semiparametric approach with mixed models. Journal of Risk and Insurance 75(3), 643–676.

Bailey, A. L. (1950). Credibility procedures: Laplace's generalization of bayes' rule and the combination of collateral knowledge with observed data. Proceedings of the Casualty Actuarial Society 37(67), 7–23.

Bühlmann, H. (1967). Experience rating and credibility. ASTIN Bulletin 4(3), 199–207.

Bühlmann, H. and A. Gisler (2005). A Course in Credibility Theory and Its Applications. Springer.

de Alba, E. and L. Nieto-Barajas (2008). Claims reserving: a correlated Bayesian model. Insurance: Mathematics and Economics 43(3), 368–376.

England, P. and R. Verrall (2002). Stochastic claims reserving in general insurance. British Actuarial Journal 8(3), 443–518.

Jewell, W. (1989). Predicting IBNyR events and delays I. Continuous time. ASTIN Bulletin 19(1), 25–55.

Jewell, W. (1990). Predicting IBNyR events and delays II. Discrete time. ASTIN Bulletin 20(1), 93–111.

Luo, Y., V. R. Young, and E. W. Frees (2004). Credibility ratemaking using collateral information. Scandinavian Actuarial Journal 2004(6), 448–461.

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. ASTIN Bulletin 23(2), 213–225.

Mack, T. (1999). The standard error of chain-ladder reserve estimates, recursive calculation and inclusion of a tail factor. ASTIN Bulletin 29(2), 361–366.

Mayerson, A. L. (1964). A bayesian view of credibility. Proceedings of the Casualty Actuarial Society 51(95), 7–23.

Merz, M. and M. Wüthrich (2010). Paid-incurred chain claims reserving method. Insurance: Mathematics and Economics 46(3), 568–579.

Meyers, G. (2009). Stochastic loss reserving with the collective risk model. Variance 3(2), 239–269.

Miller, R. B. and J. C. Hickman. Insurance credibility theory and bayesian estimation. In P. M. Kahn (Ed.), Credibility-Theory and Applications. Academic Press.

Mowbray, A. H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium? Proceedings of the Casualty Actuarial Society 1(1), 24–30.

Peters, G., P. Shevchenko, and M. Wüthrich (2009). Model uncertainty in claims reserving within Tweedie's compound Poisson models. ASTIN Bulletin 39(1), 1–33.

Quarg, G. and T. Mack (2008). Munich chain ladder: a reserving method that reduces the gap between ibnr projections based on paid losses and ibnr projections based on incurred losses. Variance 2(2), 266–299.

Shi, P. (2013). A multivariate analysis of intercompany loss triangles. Working Paper.

Shi, P., S. Basu, and G. Meyers (2012). A bayesian log-normal model for multivariate loss reserving. North American Actuarial Journal 16(1), 29–51.

Sturtz, S., U. Ligges, and A. Gelman (2005). R2winbugs: a package for running winbugs from r. Journal of Statistical Software 12(3), 1–3.

Taylor, G. (2000). Loss Reserving: An Actuarial Perspective. Kluwer Academic Publishers.

Verrall, R. (1990). Bayes and empirical Bayes estimation for the chain ladder model. ASTIN Bulletin 20(2), 217–243.

Whitney, A. (1918). Theory of experience rating. Proceedings of the Casualty Actuarial Society 4(10), 274–292.

Wüthrich, M. and M. Merz (2008). Stochastic Claims Reserving Methods in Insurance. John Wiley & Sons.

Zhang, Y. and V. Dukic (2012). Predicting multivariate insurance loss payments under the bayesian copula framework. Journal of Risk and Insurance. Forthcoming.

Zhang, Y., V. Dukic, and J. Guszcza (2012). A bayesian nonlinear model for forecasting insurance loss payments. Journal of the Royal Statistical Society: Series A 175(1), 1–20.