

Handling Sparse Data for Reserving Using Bayesian MCMC

Mike Larsen

Abstract

A sparse data set may consist of either a limited number of exposures (a small number of policyholders), a limited number of accident years or a combination of both. This paper will demonstrate, through a series of reserve modeling examples that will focus on displaying the results using graphs, how the Bayesian MCMC (Markov Chain Monte Carlo) modeling environment can assist an actuary in providing plausible reserve estimates to a client, even with sparse data.

Actuaries have dealt with the issue of limited data in practice by applying credibility weighting. Bayesian MCMC offers the means to incorporate a form of credibility weighting in a regression type model and this paper will provide examples that show how that combination can be used to model reserve estimates given sparse data.

The theory behind Bayesian MCMC, as well as background on how STAN (the software used in the paper to run Bayesian MCMC models) operates, can be found in textbooks and articles cited in the bibliography. Explaining the theory underlying Bayesian MCMC is outside the scope of this paper.

The code used to create the reserve modeling examples in the paper is available via a Rmarkdown file. Instructions on installing the software required to run the examples are included as an attachment to this paper.

I. Introduction

The benefits of using Bayesian MCMC for a reserve analysis are described in the monograph “Stochastic Loss Reserving Using Bayesian MCMC 2nd Edition” by Meyers. In that monograph, he demonstrated that Bayesian MCMC provides a useful platform for doing a reserve analysis, since those estimates tend to be more accurate than those estimated via link ratio based methods. As Meyers noted in his monograph, with the availability of the STAN software and related tools, it has become more practical to use Bayesian MCMC.

The benefits of using a smooth curve, developed via some form of regression, to model reserve estimates are described in the monograph by Taylor and McGuire, “Stochastic Loss Reserving Using Generalized Linear Models.” At the end of their monograph, they noted that one could fit a smooth curve to describe a development pattern rather than

model development one year at a time, which reduces the risk of overfitting. The approach Taylor and McGuire suggested moves the modeling problem from fitting a series of categorical variables to fitting a set of smooth curves (one may need a small number of categorical variables early in the development year pattern). Using a GLM though, requires enough data to train the model parameters which may not be possible with a sparse data set.

The prior distribution parameter feature of Bayesian MCMC enables an actuary to fit a smooth curve to a development pattern when working with a sparse data set by incorporating accumulated knowledge to supplement the available data. By utilizing this feature, an actuary can assign a distribution with a specified mean and standard deviation to each of the coefficients in a regression model.

The Bayesian MCMC routine will weight the prior distribution values for the regression parameters with results from fitting the loss triangle to create a posterior distribution for each coefficient. In credibility weighting terms, the prior distributions serve as a compliment of credibility. The reserve modeling examples given in this paper will illustrate how setting the prior distribution parameters can produce plausible ultimate loss reserve distribution estimates when combined with the available data.

A practical way to view the prior distribution feature of Bayesian MCMC is that it allows the actuary to set guard rails for the MCMC machinery in its search for the optimal set of parameter estimates, eliminating the need to search the entire real number line. By narrowing the range of plausible results for the coefficient estimates, the actuary makes the MCMC task more manageable and significantly reduces the likelihood of the process producing nonsensical results for the forecast loss reserve distribution.

If the data set contains sufficient information, the results derived from the data, under normal circumstances, will override the prior distributions. This is similar to our credibility weighting approach, where for data sets with ample information, the indications from the data set receive nearly 100% credibility, while the compliment of credibility receives 0% or a small percentage. However, it is possible to configure the prior distribution to prevent the MCMC routine from exploring certain regions of potential answers, which is why the caveat “under normal circumstances” is necessary.

This paper is centered around a series of reserve modeling examples that were created with the goal of illustrating the concepts described above. Before we can get to those examples, some introductory comments are necessary.

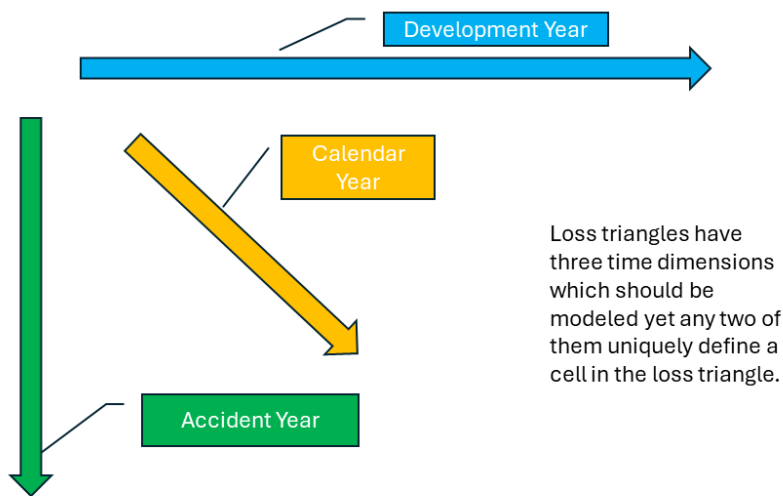
I.1 Linking Credibility Weighting to Prior Distributions

Most actuaries have some exposure to regression modeling, and many CAS members may recall studying Bayesian conjugate prior credibility weighting and least squares credibility weighting concepts for their exams. However, the idea of developing a set of prior distributions to describe the likely range of values for the coefficients in a regression

equation, as a means of incorporating a form of credibility weighting into a model, may be unfamiliar to a significant portion of the actuarial community.

Loss triangles have three time dimensions: accident year, calendar year and development year. In the reserve models that follow, the type of credibility weighting will vary by time dimension.

Loss Triangle Time Dimension

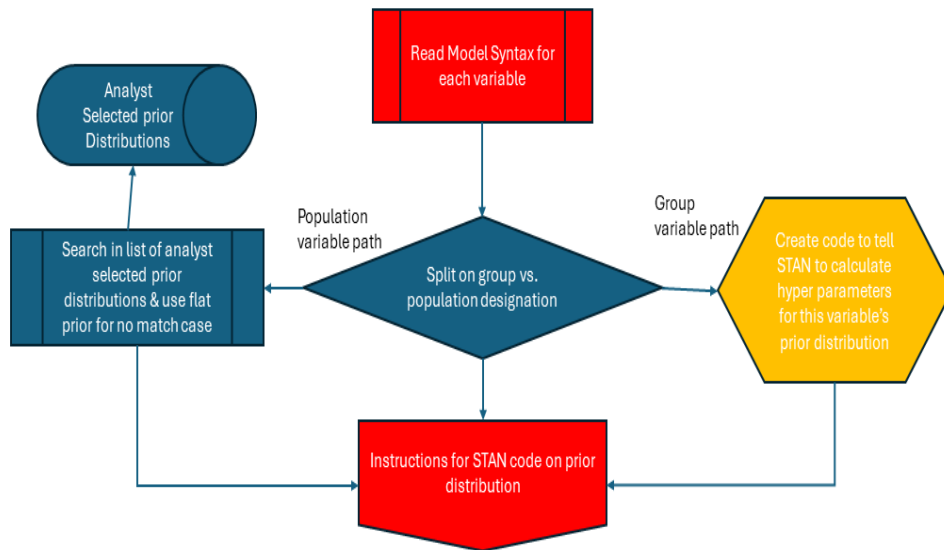


The regression coefficients for development year and calendar year will have prior distributions with selected parameters that are adjusted by the data to form a posterior distribution for each coefficient. These regression coefficients are for variables that fall under the population variable category.

The regression coefficients for accident year will be classified as group variables and will be modified using a least squares-type credibility weighting. This means that the prior distribution parameters, which control the weighting between a group's estimate and the overall estimate, will be calculated based on the data set.

Each variable in a Bayesian MCMC model will have a prior distribution and the instructions to set up the prior distribution will be contained in the code used to call STAN.

Selecting Prior Distribution by Variable



Please note that the prior distributions used in a Bayesian MCMC model differ from those in the credibility weighting concepts we read about on the exams for Bayesian conjugate prior combinations, as they are not limited to the distributions that form conjugate priors. The development of efficient algorithms in the MCMC environment has freed us from this constraint by eliminating the need to calculate the weighted average denominator required to move from the prior to the posterior distribution.

The way Bayesian MCMC handles group variables in terms of credibility weighting shares some similarities with least squares credibility weighting, but the terminology differs. Least squares credibility weighting examines how the data behaves in terms of variance of the group means relative to process variance. This means the credibility weighting between the group estimate and the overall estimate is driven by the data's behavior rather than a selected prior distribution.

In the Bayesian MCMC environment, achieving a similar effect involves using a model structure where the prior distribution for the group variable depends on another layer of priors (hyperparameters) forming a hierarchical model. The hyperparameters are estimated from the data at hand. In the Bayesian MCMC literature, this approach is described as partial pooling, which is a compromise between pooling (assuming no difference in by group behavior) and modeling each group separately (no pooling). Partial

pooling is useful when information about one group's behavior can improve the estimates for the other groups' behavior, even though they are not identical. We will revisit this topic when we set up the first reserve model.

I.2 Reserve Analysis Example Overview

The reserve analysis examples will be set up to forecast incremental loss payments. The reserve estimate will be the sum of the forecasted future incremental loss payments.

The incremental payments will be normalized prior to modeling to make it easier to focus on common development patterns across the accident years. The normalization process will be reversed when the modeling process goes from estimating the modeling parameters to forecasting the future incremental payments.

The normalization process will have two parts:

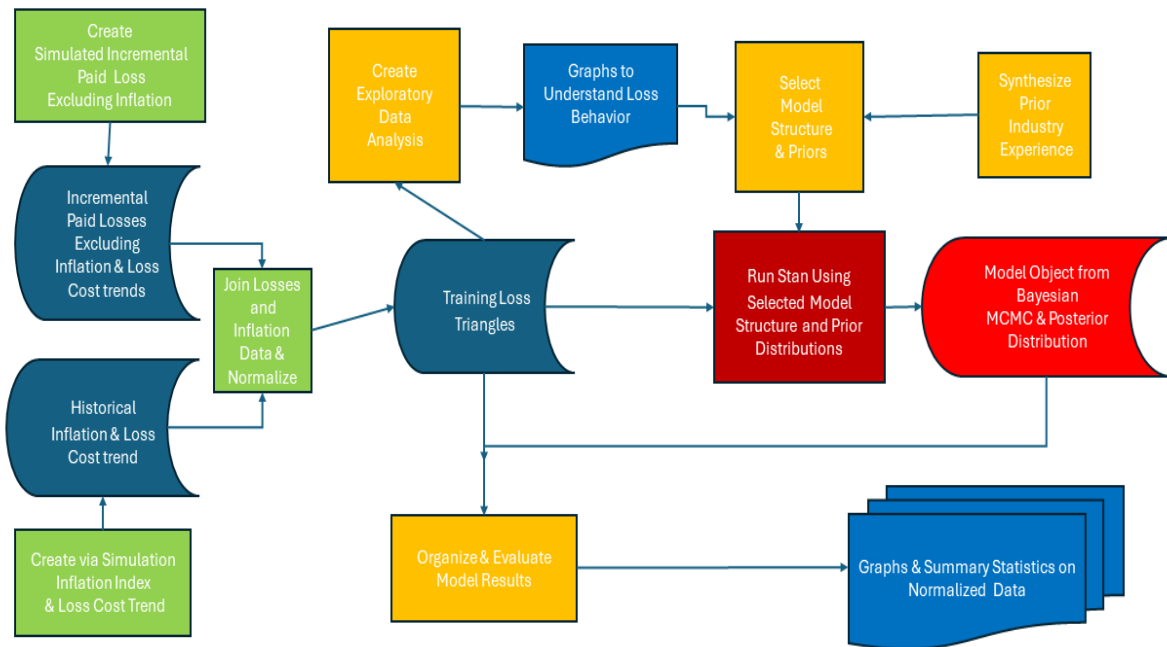
- Dividing the incremental payments by an exposure base attached to the accident year. The exposure base in this case will be reported claim counts at 12 months of development.
- Dividing the incremental payments by an inflation index. We will accomplish this by the use of a simulated inflation index.

Normalizing the incremental payments will make it easier to view the underlying development pattern across accident years by removing the effect of varying business volumes and putting the dollar values of payments on a common basis.

The normalized incremental payments will be modeled using the lognormal distribution and will be referred to as deflated pure premium values in the sections that talk about the reserve modeling results. The mu and sigma values for the lognormal distribution will be modeled as a function of accident year, development year and calendar year. The formulas used to fit mu and sigma will be described once we move into the reserve model examples.

An outline of the sequence of steps used to build the reserve model examples in this paper is given below:

Model Building Sequence



The sequence is similar to the usual sequence of steps in building a GLM model:

- Build the data sets.
- Do some exploratory analysis to understand the data.
- Set up the model.
- Run the model.
- Evaluate the model results.

There are some differences:

- The loss triangle data sets are built via simulation.
- There is a simulation routine to generate inflation and loss cost trend.
- Inclusion of prior distributions.
- Posterior distribution of results rather than point estimates

The data sets for the incremental loss triangles were built via simulation for the purpose of illustrating when selecting the prior distribution parameters using insight developed from on the job reserving experience can be useful. The loss triangles were created so as to mimic a generic pattern for liability lines where incremental payments start slowly, peak after a few development years and then taper to close to zero dollars per development period as the accident year ages with the development pattern becoming more unstable at later development periods. For this paper, simulating the data sets had the advantages of

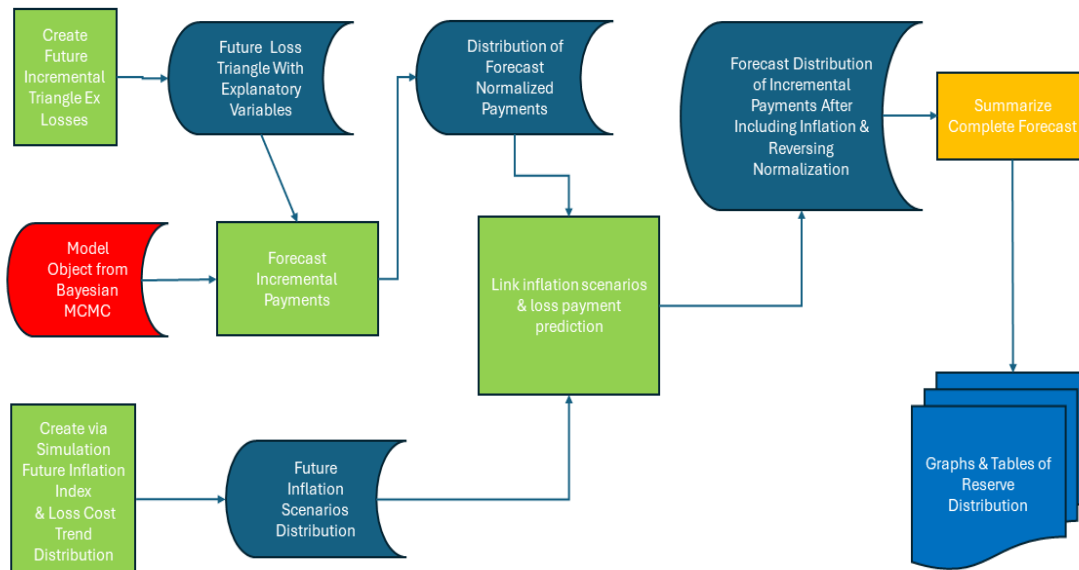
eliminating the need to obtain data from a company and allowing anyone who opens up the code to see an example of how one can use simulation to build data sets and experiment with building a Bayesian MCMC model. The data sets were simulated by inverting the Poisson distribution to generate incremental paid counts and inverting the lognormal distribution for the severity amounts with the parameters for these distributions set in the code. More details on the data set construction can be found in Appendix A.

The simulation routine to generate inflation and loss cost trend sets up the ability to attach a distribution of inflation forecasts to the distribution of incremental loss cost estimates to give a realistic picture of the variability of the reserve estimates.

The Bayesian MCMC routines generates a random sample of the parameters for a posterior distribution rather than point estimates. Point estimates of the coefficient values can be obtained by summarizing the information in the posterior distribution data set. This approach contrasts with the GLM process, where one would need to bootstrap the model results to obtain a distribution.

The process used to generate the reserve model forecast is summarized below:

Forecasting Sequence



One can apply the model results to future explanatory variables (future development years for the set of accident years) to obtain a forecast which is similar to a GLM modeling routine, but the forecast will be in the form of a distribution of future results.

More detail on the steps behind creating the forecast distribution can be found in Appendix B.

Information on the software selected to create the examples can be found in Appendix C.

I.3 Reserve Analysis Example Scope

This paper focuses on how Bayesian MCMC can assist actuaries in dealing with sparse data in reserving through the prior distribution feature of Bayesian MCMC. In a typical reserve analysis, one would investigate multiple models with different explanatory variables to evaluate their forecasting reliability. However, this paper will not delve into the analysis required to compare different potential model forms (different explanatory variables) using metrics such as a form of the AIC statistic or cross-validation adapted for Bayesian MCMC models. Additionally, while the Bayesian MCMC modeling process usually involves examining and commenting on diagnostics to ensure that the MCMC machinery is functioning correctly, these comments will be omitted from the discussion in this paper.

There are five data sets used in the reserve model examples, each with a decreasing amount of information in the loss triangles to illustrate how the benefit of selecting a prior distribution set of parameters increases as information decreases. The same set of frequency and severity distribution parameters were used to simulate the incremental losses for all data sets. Each data set will be modeled with and without selected prior distributions parameters, using the same set of explanatory variables for each case. The data sets will be labeled by case number when reviewing the reserve modeling results:

- Case 1 has 22 accident years (2000 through 2021) with a uniform reported claim count of 1000 reported claims at 12 months for each accident year and a maximum of 22 years of development time.
- Case 2 has 22 accident years (2000 through 2021) with a uniform reported claim count of 500 reported claims at 12 months for each accident year and a maximum of 22 years of development time.
- Case 3 has 22 accident years (2000 through 2021) with a uniform reported claim count of 100 reported claims at 12 months for each accident year and a maximum of 22 years of development time.
- Case 4 has 22 accident years (2000 through 2021) with a uniform reported claim count of 50 reported claims at 12 months for each accident year and a maximum of 22 years of development time.
- Case 5 has 11 accident years (2011 through 2021) with a uniform reported claim count of 100 reported claims at 12 months for each accident year and a maximum of 22 years of development time.

Comparing modeling results for different data sets with and without selected prior distributions will illustrate how the value of using selected prior distributions varies based on the information in the loss triangle data set. While there will be a limited review of the model's reasonableness in fitting the data, the examples will primarily focus on how using selected prior distributions can reduce forecast uncertainty. Please note that if the analyst does not select a prior distribution for a given parameter of a population variable, the program will insert flat priors by default which implies the entire real number line provides a set of equally likely results for that parameter.

II. Case 1 Modeling Results

Case 1 has 22 accident years (2000 through 2021) with a uniform reported claim count of 1000 reported claims at 12 months for each accident year and a maximum of 22 years of development time.

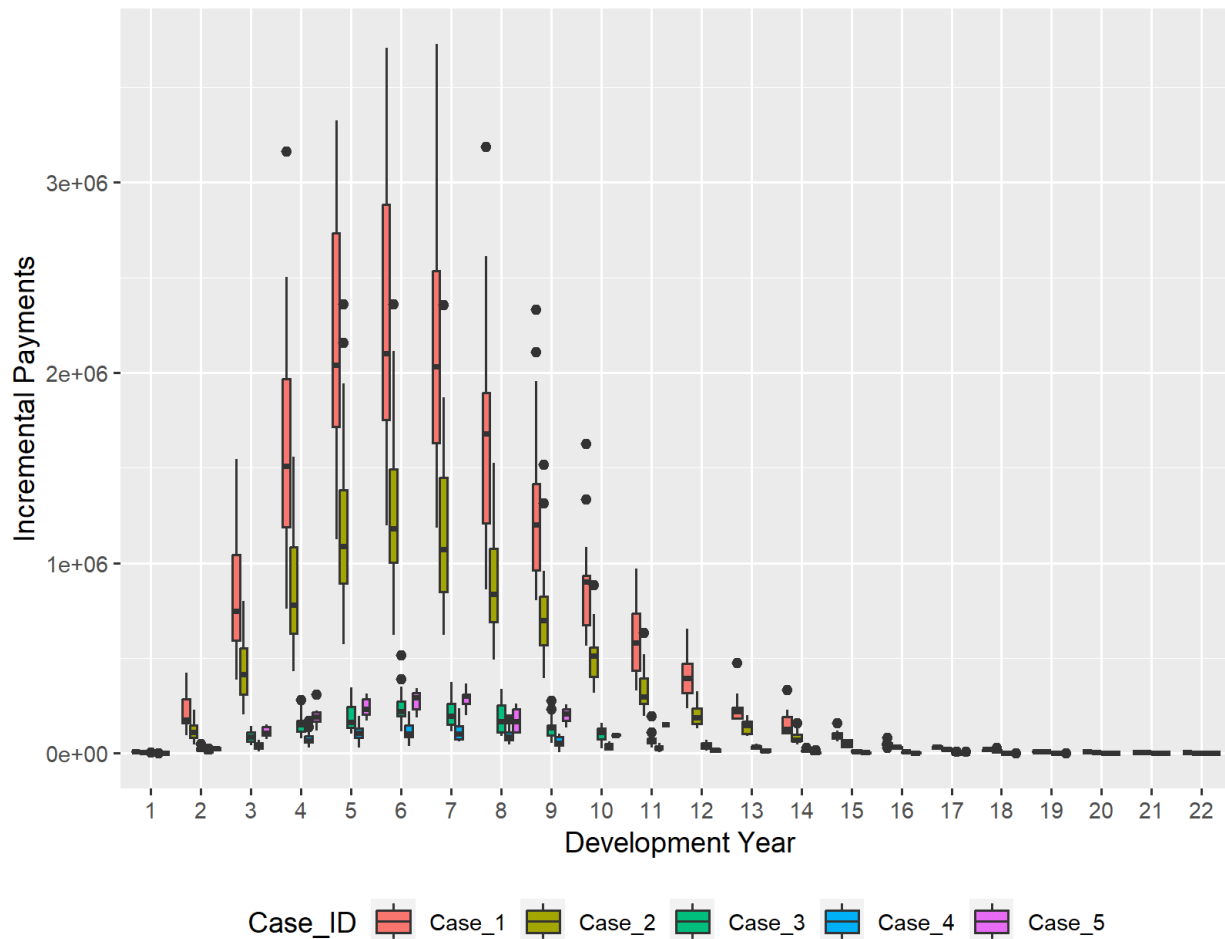
II.1 Exploratory Data Analysis

EDA exhibits are included to illustrate the usual and customary process of building a Bayesian MCMC model by starting with EDA exhibits. The graphs are a starting point for formulating potential regression equations for the distribution parameters and checking for outliers.

The Exploratory Data Analysis (EDA) exhibits for the different cases are combined in most of the graphs below since the underlying frequency and severity parameters are the same for each case. The different data sets are denoted by color in the graphs that combine different data sets with the legend at the bottom of the graph giving the color code that identifies the Case number.

The first graph shows the incremental payments prior to normalization. Subsequent graphs illustrate that as adjustments are made to normalize the values and then transform them to the log scale, the inherent pattern of how incremental payments behave as an accident year ages becomes clearer.

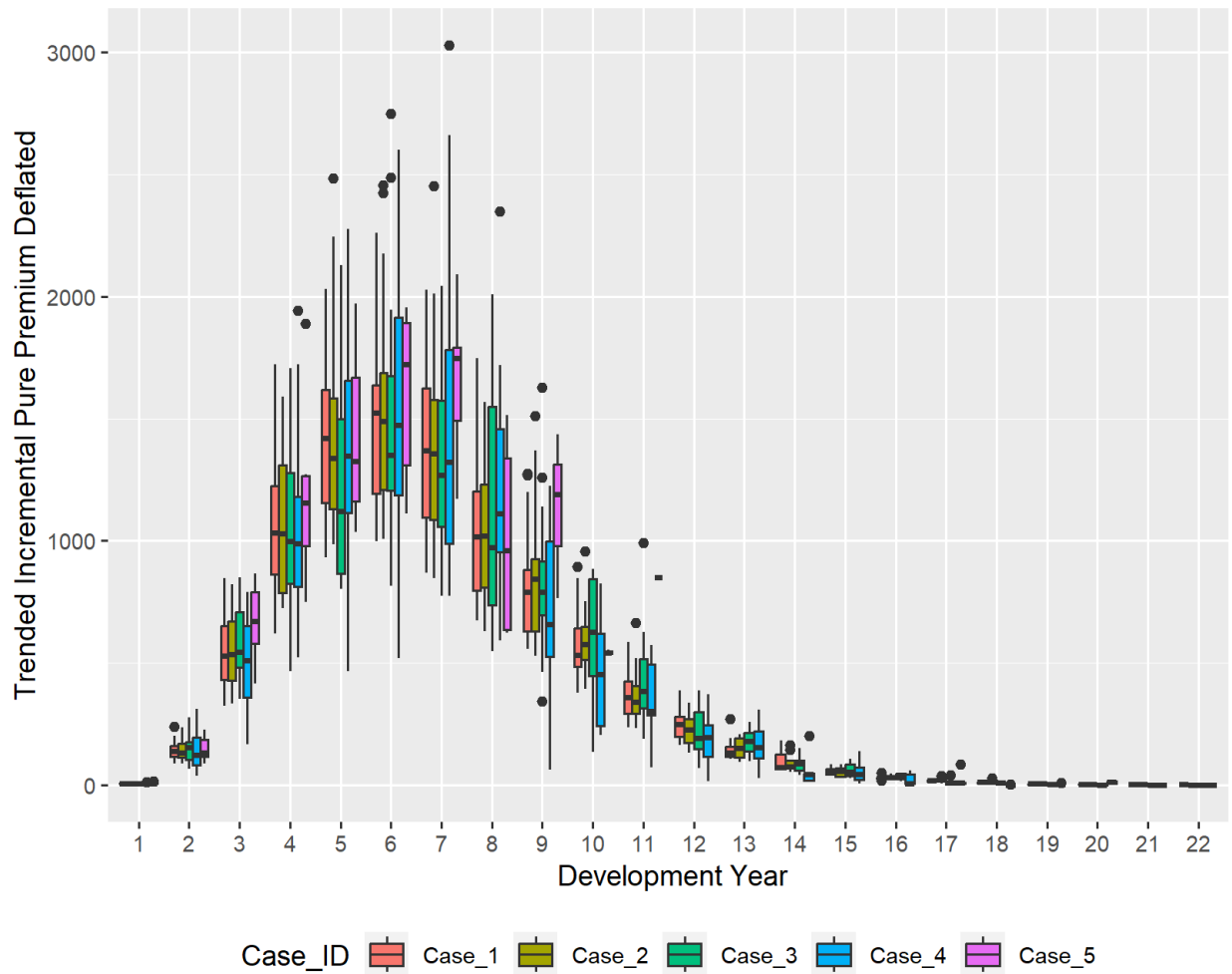
Box Plot of Incremental Payments Across Accident Year by Development Year by Case Number



By normalizing using the reported claim count to account for the volume of business and adjusting for inflation to put the dollar values on the same scale, the similarity in the pattern across the cases becomes clearer. Please note that the additional loss cost trend beyond general inflation is still reflected in the normalized pure premium values below.

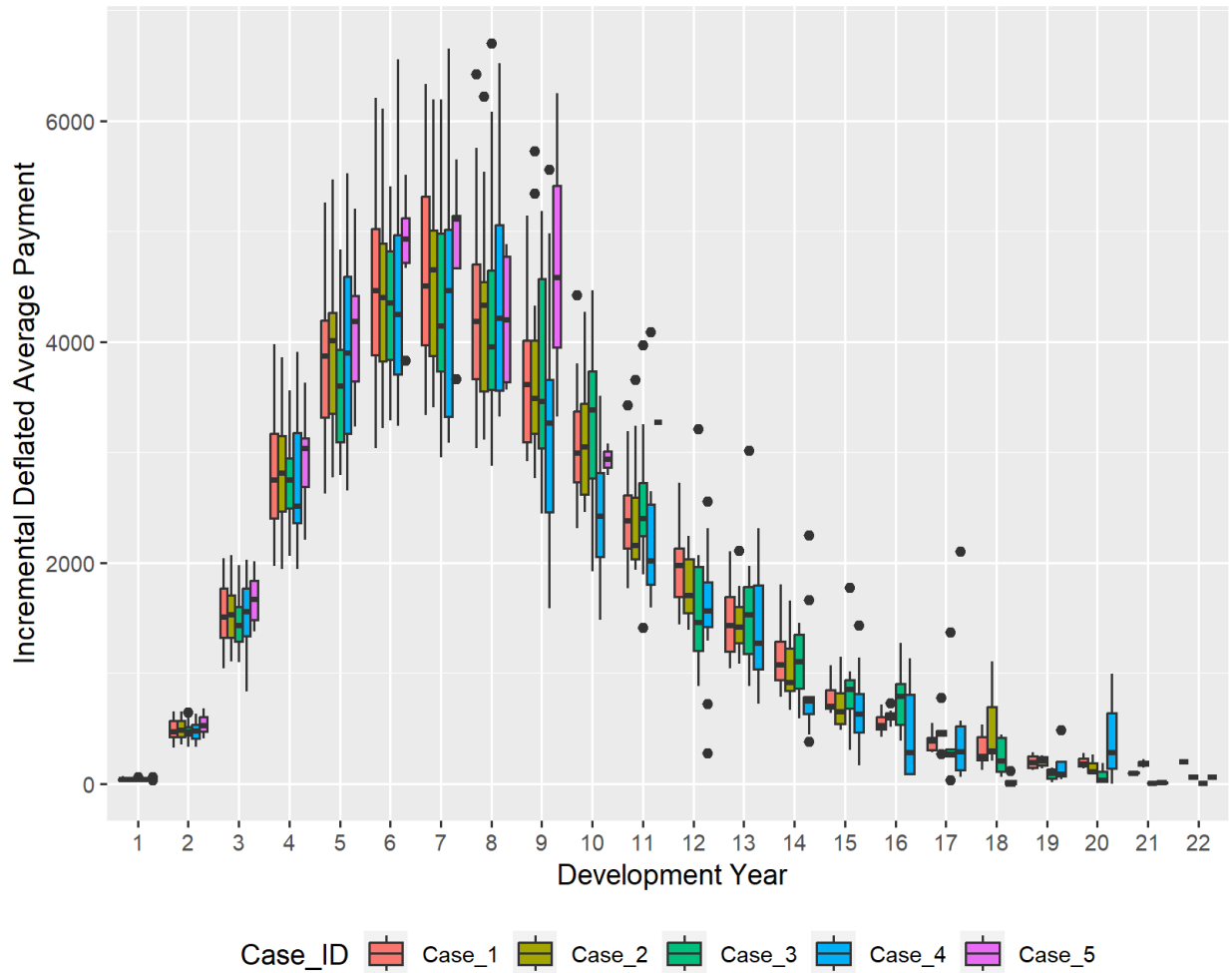
Since Case 5 does not include any development years beyond eleven years, you will not see any data for Case 5 beyond eleven years in the EDA exhibits.

Box Plot of Deflated Incremental Pure premium
Across Accident Year by Development Year



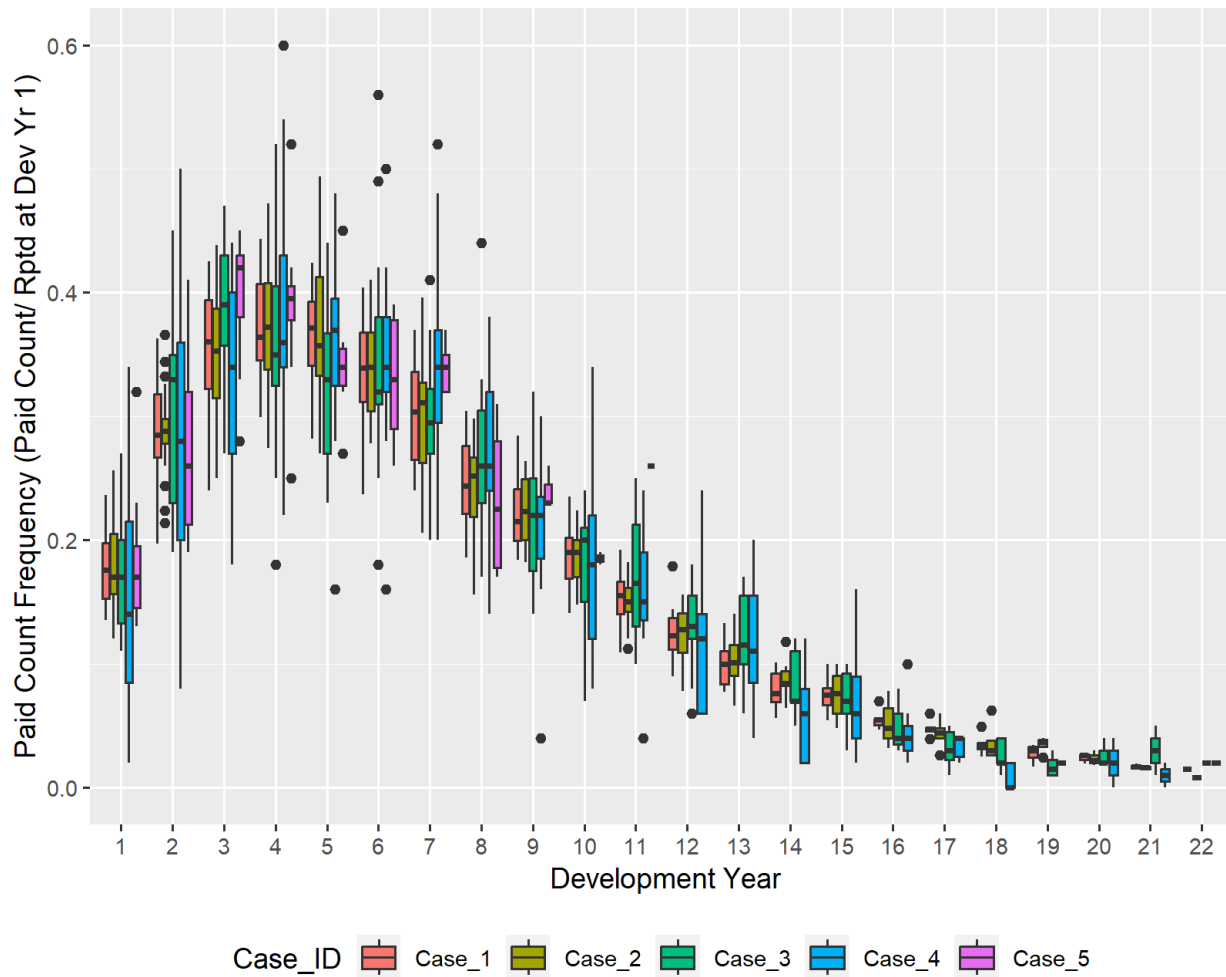
The next set of graphs examine the frequency and severity components of the pure premium graph above. For the incremental average severity, it is important to note that Case #5 includes a higher proportion of observations from more recent calendar years compared to the other cases. This is because it only encompasses the most recent ten accident years, which tends to skew the incremental average severity higher than the other cases.

Box Plot of Incremental Deflated Average Payment Across Accident Year by Development Year



The incremental frequency plot is given below. Incremental frequency tends to peak at either development year three, four or five while incremental average severity tends to peak at age six or seven.

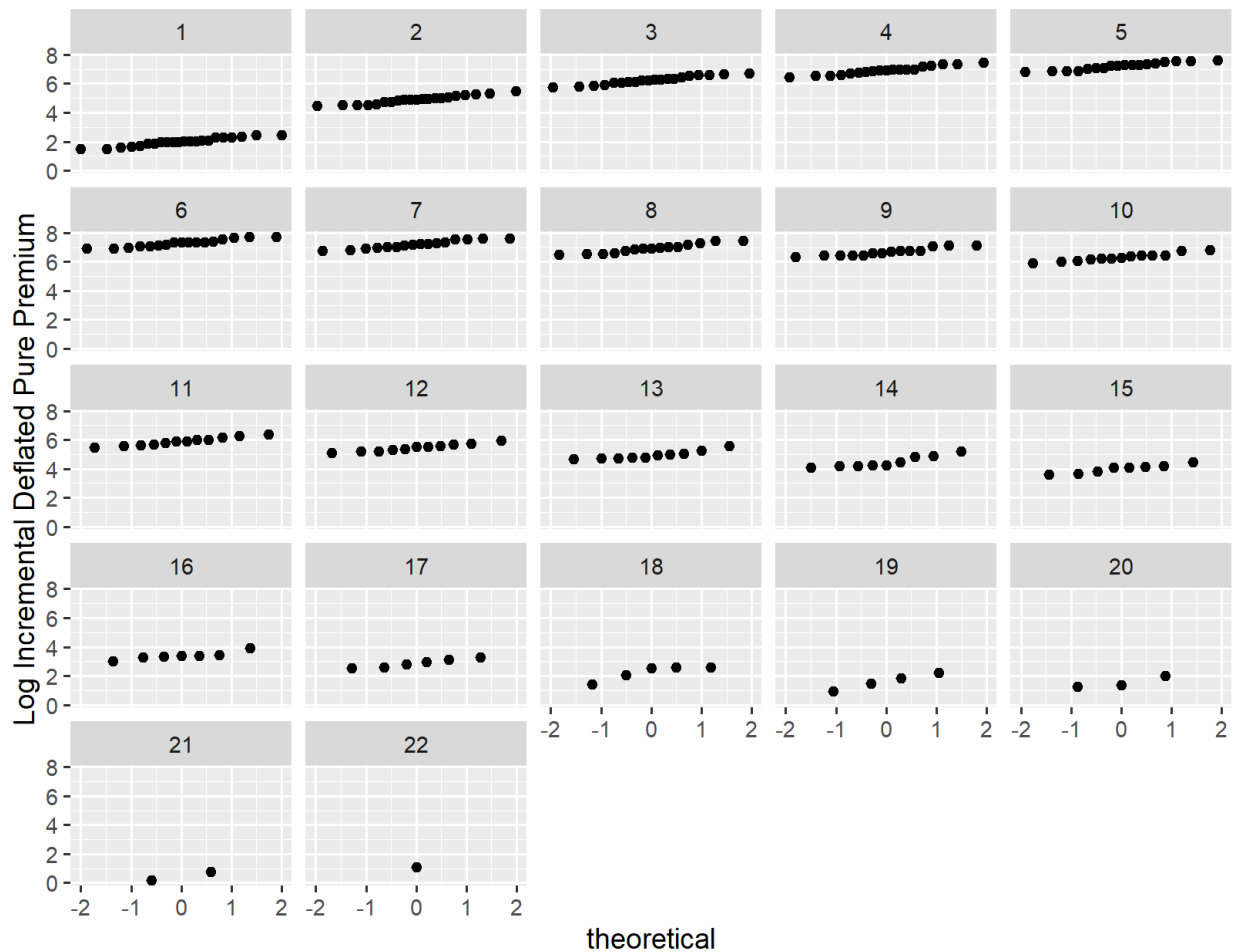
Box Plot Incremental Paid Frequency Across Accident Year by Development Year



If we look at the incremental normalized pure premiums on the natural log scale using a QQ plot for Case #1 as a representative sample, the graph below indicates that using a lognormal distribution for modeling the incremental normalized pure premiums is a reasonable choice.

QQ Plot Log Incremental Deflated Pure Premium Across Accident Year by Development Year

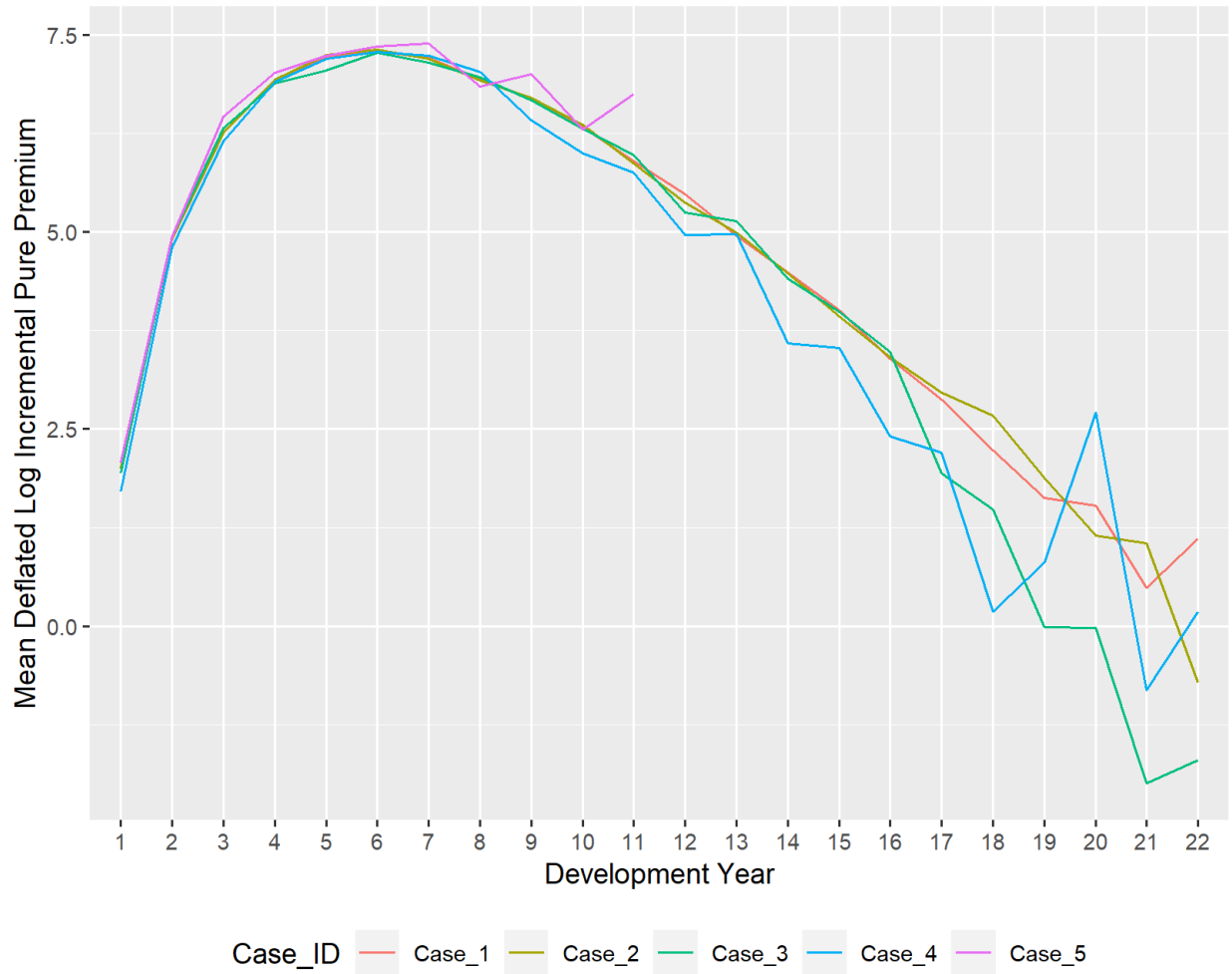
Case 1



The line graph of the average natural log of the normalized pure premium values highlights a few key points:

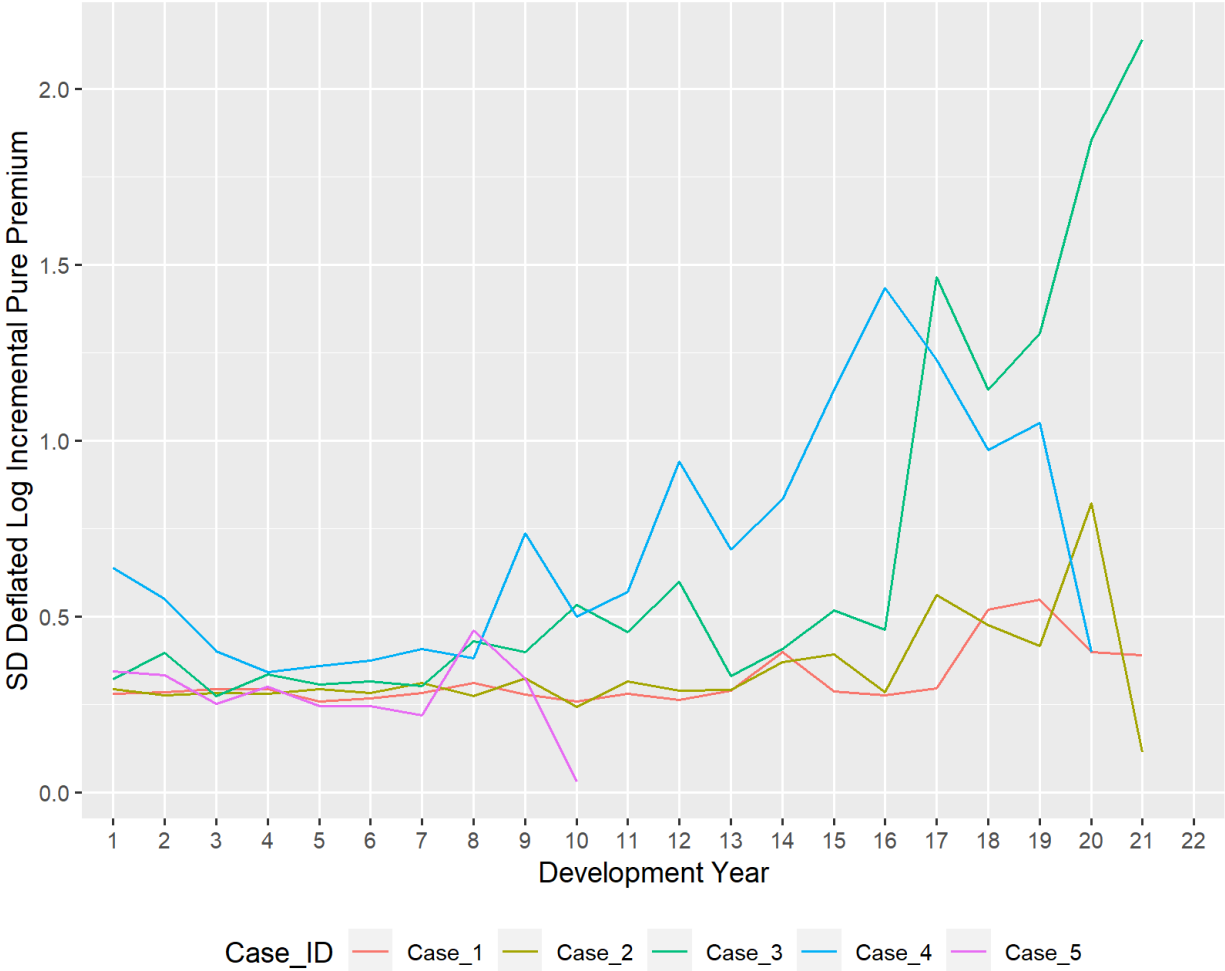
- The results for Case 1 and Case 2 are similar across time and have smooth pattern.
- Case 3 and Case 4 have a similar pattern to the two cases with higher reported claim counts, but the pattern is not as smooth.

Mean Natural Log of Incremental PP Deflated
Across Accident Year by Development Year



The sigma value for the lognormal distribution varies across development years. Unlike the mu parameter, the indications for sigma are less consistent across cases. Cases with fewer reported claims exhibit greater instability as the accident year ages.

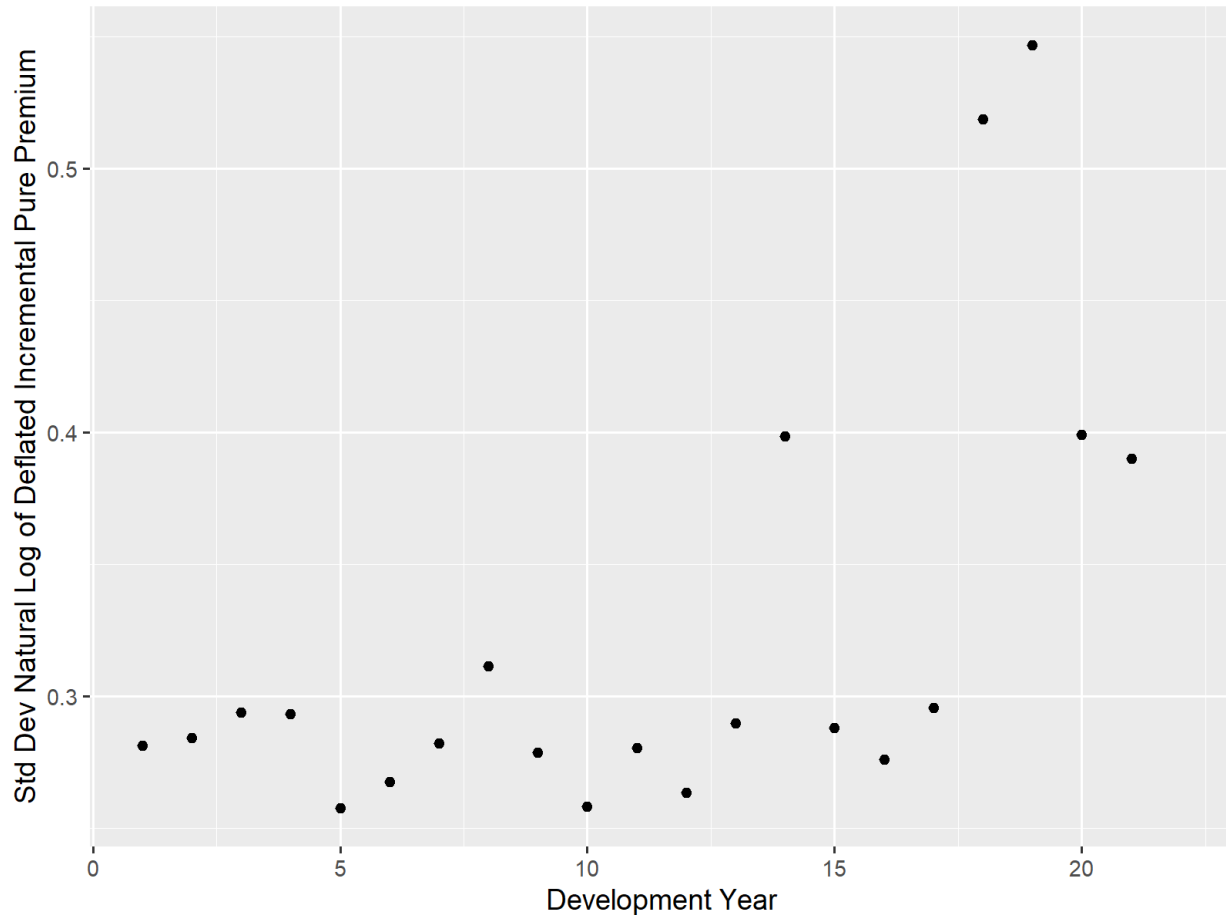
Standard Deviation Natural Log of Incremental PP Deflated Across Accident Year by Development Year



The difference in scale between the cases for sigma makes it necessary to look at Case 1 by itself to identify a likely starting point for a pattern to specify in the Bayesian MCMC models. It looks like sigma is roughly constant for the first ten years then starts to gradually increase.

Standard Deviation
Natural Log of Incremental PP Deflated
Across Accident Year by Development Year

Case 1



II.2 Parameter Estimates Based Solely on Prior Distribution Selection

This section has two goals:

- Illustrate what it means to select prior distributions for the coefficients in the Bayesian MCMC model equations.
- Provide an example of how to review the results of selecting a set of prior distributions for a development curve formula and parameters beyond the observed development period in a new line of business before running the Bayesian MCMC model.

Before addressing those two goals, we need to introduce some of the terminology that will be used in the reserve model results throughout the rest of this paper.

The same explanatory variables will be used throughout this paper. The variables were selected to follow the patterns illustrated in the EDA exhibits for mu and sigma on the lognormal scale. A description of the variables used follows:

- Dev_Yr_2_Factor: A categorical variable for the first three development years
- Dev_Yr_6_Cap: A continuous variable that is capped at 6 for all years after development year 6.
- Dev_Yr_6_Cap_Sqrd: $\text{Dev_Yr_6_Cap} * \text{Dev_Yr_6_Cap}$
- Dev_Yr_6_Spline: A continuous variable equal to 0 if development year <7 and development year - 6 otherwise
- Dev_Yr_6_Spline_Sqrd: $\text{Dev_Yr_6_Spline} * \text{Dev_Yr_6_Spline}$
- Cal_Yr_Time: A continuous variable equal to Accident Year + Development Year – 1 - 2000
- Dev_Yr_10_Spline: A continuous variable equal to 0 if development year <11 and development year -10 otherwise
- Dev_Yr_10_Spline_Ln: A continuous variable equal to 0 if development year <11 and $\log(\text{development year} - 10)$ otherwise
- (1||Acc_Yr): A group variable (the prior variables are population variables) that adjusts the intercept by accident year after a form of credibility weighting to reflect a given group's effect on the Intercept. Please note that the "||" within the parentheses is included to tell the program to assume that the accident years are not correlated. This set of examples assumes all observations are uncorrelated, but one may need to account for correlation in modeling work on the job.

Each example will include an extract of the summary of the modeling posterior distribution results for the population variables (those not receiving the least squares type credibility weighting). A brief description of the terms used in these summary exhibits is provided below:

- Coefficient: The name of the explanatory variable being modeled
- Estimate: mean of posterior distribution for that parameter

- Std_Err: standard deviation of the parameter estimate for the posterior distribution
- LB_CI_95: lower bound of the 95% confidence interval for the parameter estimate from the posterior distribution
- UB_CI_95: upper bound of the 95% confidence interval for the parameter estimate from the posterior distribution

The results for each model will also display the prior distribution definitions used, along with instructions on how to build the model to generate the Bayesian MCMC results. A summary of selected columns from the prior definitions table includes:

- Prior_Distribution: prior distribution for the coefficient and selected parameters for that distribution.
- Variable_Class: type of coefficient either a **b** for a beta in a regression equation or **sd** for standard deviation (used for group variable).
- Coefficient: name of explanatory variable attached to the coefficient
- Group_Indicator: identifies a group variable (blank for population variables)
- Distribution_Parameter: identifies distribution parameter variables besides the mean (in this case sigma) with a blank indicating that coefficient is part of the regression equation for the mean (mu in this case). This particular model assumes that sigma should be modeled separately and simultaneously with mu for the lognormal distribution. There are other distributional models one could use to match the behavior of a different data set.
- Source: either “user” (user selects prior distribution) or “default” (flat prior)

If you choose to run the code in the attached files, please note that I renamed the labels from the summary exhibit and the prior distribution exhibits produced by running the models in Rstudio. This was done to make it easier to follow the discussion.

The results presented in this section were generated by running the model without using the loss triangle data set. This feature allows for the evaluation and documentation of the initial assumptions for the reserve analysis.

As an example of reviewing starting assumptions, the results below show the implied development curves for mu and sigma based solely on the prior distribution selections used in modeling Case 1 and 2. Please note that the parameter results shown under the “Estimate” column are just the selected prior distribution figures along with their standard errors, shown under the “Std_Err” column, in this particular run. This will not generally be

true in general. The prior distribution selections for Case 3, 4 and 5 will be described when those model results are presented.

For this example, the prior distribution for the Dev_Yr_6_Cap coefficient is set to a Normal distribution with mean equal to 1.5 and a standard deviation of 0.2. These figures are reflected in the results summary for this example, with a slight variation due to simulation noise. By examining the figures for other explanatory variable coefficients, one can observe the same pattern, illustrating how to review the implications of the selected prior distributions. Normally, the posterior distributions of the parameters will differ from the prior distributions.

There are separate regression equations for mu and sigma. In the table below, the coefficient labels for sigma start with “sigma_” followed by the name of the explanatory variable appended. The coefficient labels for mu are simply the names of the explanatory variables included in the regression equation for mu.

In practice, prior parameter estimates do not have to be calibrated to match the data to provide useful guard rails for the convergence routines in STAN, rather they just need to be plausible. In this case, plausible would mean you would feel comfortable presenting the results to another actuary experienced in reserving.

No Data Case Population Parameters

Coefficients	Estimate	Std_Err	LB_CI_95	UB_CI_95
Intercept	0.002	0.245	-0.484	0.483
Dev_Yr_2_Factor2	1.001	0.25	0.523	1.491
Dev_Yr_2_Factor3	2.003	0.248	1.523	2.486
Dev_Yr_6_Cap	1.501	0.204	1.091	1.899
Dev_Yr_6_Cap_Sqrd	-0.1	0.1	-0.297	0.098
Dev_Yr_6_Spline	-0.1	0.1	-0.296	0.098
Dev_Yr_6_Spline_Sqrd	-0.05	0.1	-0.247	0.148
Cal_Yr_Time	0.02	0.01	0	0.04
Dev_Yr_10_Spline	-0.05	0.101	-0.25	0.149
sigma_Intercept	-1.498	0.255	-2.003	-0.996
sigma_Dev_Yr_10_Spline_Ln	0.5	0.024	0.457	0.542

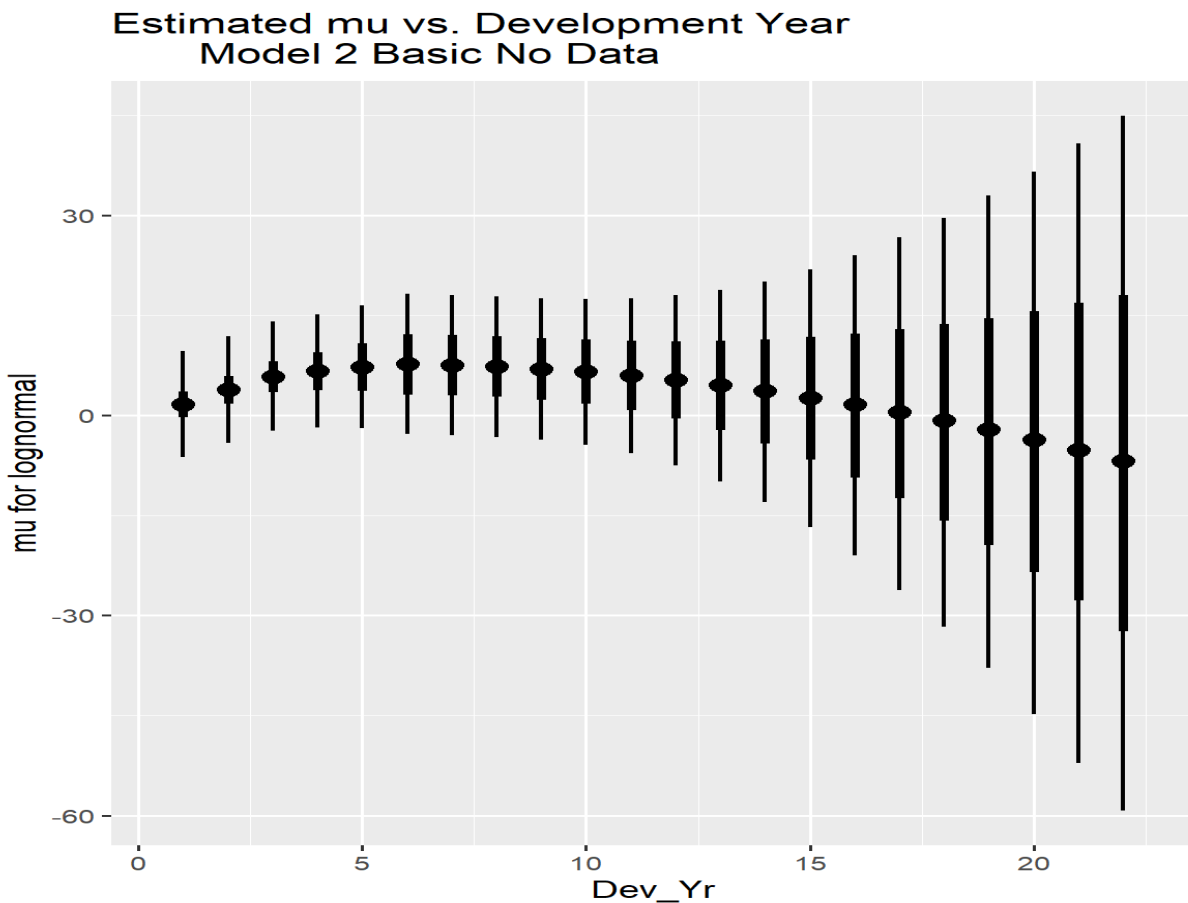
Prior_Distribution	Variable_Class	Coefficient	Group_Indicator	Distribution_Parameter	Source
	b				default
normal(0.02, 0.01)	b	Cal_Yr_Time			user
normal(-0.05, 0.1)	b	Dev_Yr_10_Spline			user
normal(1, 0.25)	b	Dev_Yr_2_Factor2			user
normal(2, 0.25)	b	Dev_Yr_2_Factor3			user
normal(1.5, 0.2)	b	Dev_Yr_6_Cap			user
normal(-0.1, 0.1)	b	Dev_Yr_6_Cap_Sqrd			user
normal(-0.1, 0.1)	b	Dev_Yr_6_Spline			user
normal(-0.05, 0.1)	b	Dev_Yr_6_Spline_Sqrd			user
normal(0, 0.25)	b	Intercept			user
	b			sigma	default
student_t(2, 0.5, 0.01)	b	Dev_Yr_10_Spline_Ln		sigma	user
normal(-1.5, 0.25)	b	Intercept		sigma	user
student_t(3, 0, 2.5)	sd				default
	sd		Acc_Yr		default
	sd	Intercept	Acc_Yr		default

There are a variety of approaches to setting up the prior distributions for a reserve model. One could run a GLM model focusing on the development pattern for a loss triangle for a similar line of business or, if sufficient data is available, use the loss triangle to be modeled. Sometimes, experimenting with different equations and parameters in EXCEL can provide a logical starting point for setting the mean values in the prior distributions, with the standard deviation set by informed judgement on the degree to which the actuary believes the compliment of credibility should influence the posterior distribution coefficient estimates.

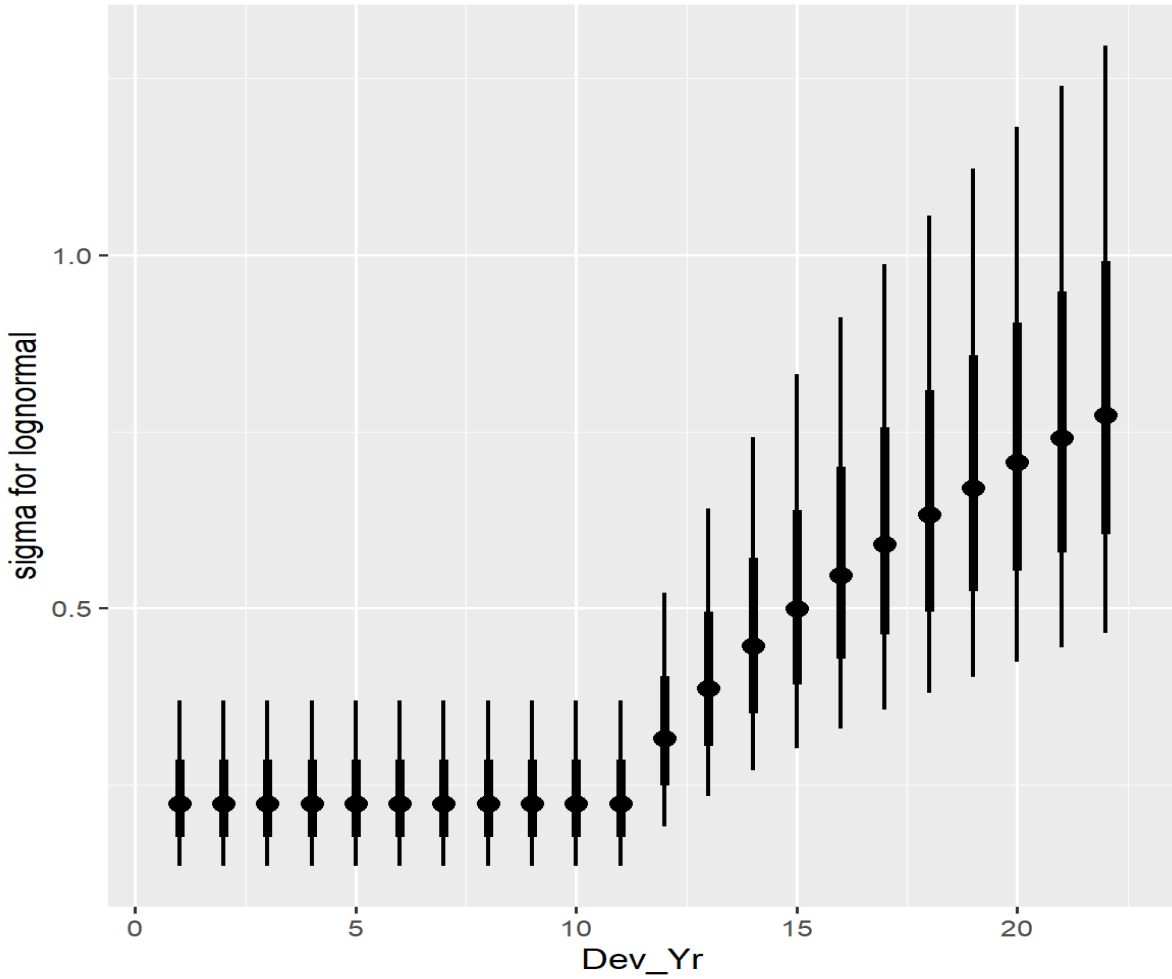
Only the graphs for mu and sigma modeled for the lognormal distribution and the accident year intercepts are shown below. Normally, there would be some comparison of modeled to observed results but given there was no data used in running this model such a comparison is not possible. The graph for mu demonstrates that the pattern implied by the prior distributions follows a roughly parabolic path. In contrast, the graph for sigma

indicates it remains relatively constant for the first ten years and then develops upward with some variation around these average patterns. The accident year graph shows that the intercepts are all centered around the same intercept value, zero, which is consistent with the instructions to STAN to gather information to set the relative weight for the group-level variable effect from the data, despite no data being provided.

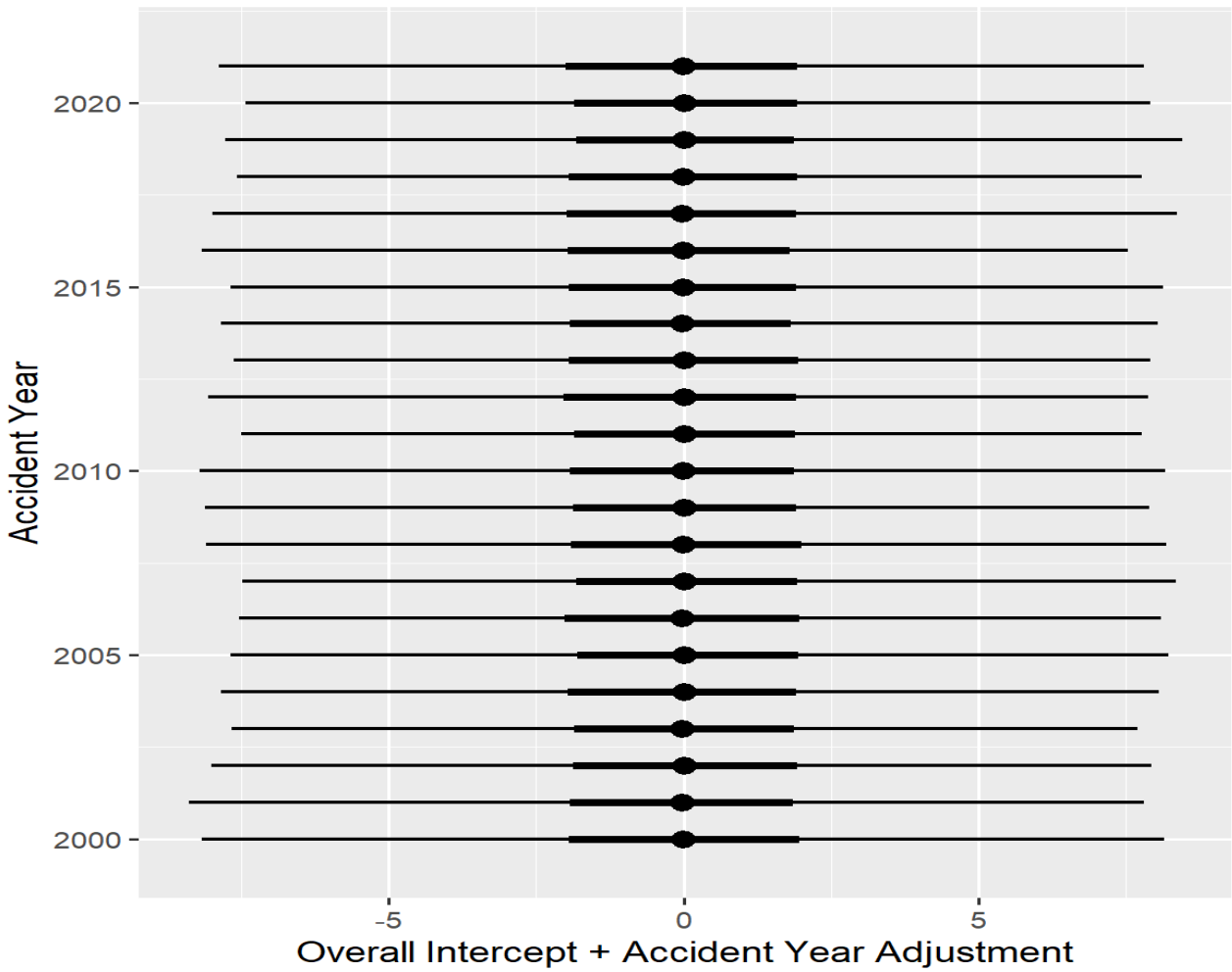
All of the graphs for the reserve modeling results that are not box plots will show the median result as a dark dot in the center of two sets of lines. The broader inner line represents the 66% confidence interval, and the narrower outer line represents the 95% confidence interval, as drawn from the information in the posterior distribution data set created by the model.



Estimated sigma vs. Development Year Model 2 Basic No Data



Median fitted Adjusted Intercept by Accident Year
 Deflated Incremental Pure Premium Model
 No Data With Selected Prior Distributions



II.3 Case 1 Results with Prior Distribution Selected

There are a few additional preliminary items to address before commenting on the Bayesian MCMC modeling results.

The models used in this paper make the following assumptions:

- A smooth curve describes how the normalized incremental payments (deflated pure premium) develop as an accident year ages, and this pattern is consistent across accident years.
- Beyond the overall inflation effect on loss costs, there is a loss cost trend with an average annual compound effect across all calendar years.

- The insured population’s makeup is similar across accident years, but not necessarily identical and that effect can be measured via an intercept value that varies across accident years. The least squares type approach to credibility weighting (partial pooling) is an appropriate way to reflect the idea that the population as grouped by accident year is similar but not the same across accident years.

Each Case presented will follow the same format:

- Summary exhibits on the fit of the model using selected prior distributions.
- Summary exhibits on the fit of the model using flat priors.
- A comparison of the lognormal parameters μ and σ on the training data set for the models with and without selected prior distribution parameters.
- A comparison of the forecast lognormal parameters μ and σ for the future development years within an accident year for the models with and without selected prior distribution parameters.
- A comparison of forecast reserve estimates between the model that used selected prior distributions and the one that did not.

The summary exhibits on the fit of the model will be limited in scope relative to what one would typically review when fitting a model. The intent is to provide sufficient information for the reader to evaluate if the model provides a reasonable fit to the observations before moving on to the primary focus of the paper: comparing the forecast variability with and without prior distributions as the volume of information in the loss triangles changes.

The summary for Case 1 shows that all population variables likely have a non-zero effect, as their 95% confidence intervals do not include zero. The graph for observed vs. modeled indicate that the model provides a reasonable fit to the data since the modeled results tend to follow the observed data.

Regarding the prior distribution versus the posterior distribution, we can look at the results for Dev_Yr_6_Cap as an example. The prior distribution specified that the coefficient for this variable follows a Normal Distribution with a mean of 1.5 and a standard deviation of 0.2. The summarized posterior distribution shows a mean of 1.657 and a standard deviation of 0.069. The data set indications shifted the prior distribution mean, and the standard deviation around that parameter estimate became smaller which is a typical result.

In the “Prior_Distribution” column below, you will see that the Normal distribution was selected in all cases but one for the population variables. Sometimes, using the Student distribution can be an effective approach to preventing outliers from weighing too heavily on the fit for a coefficient.

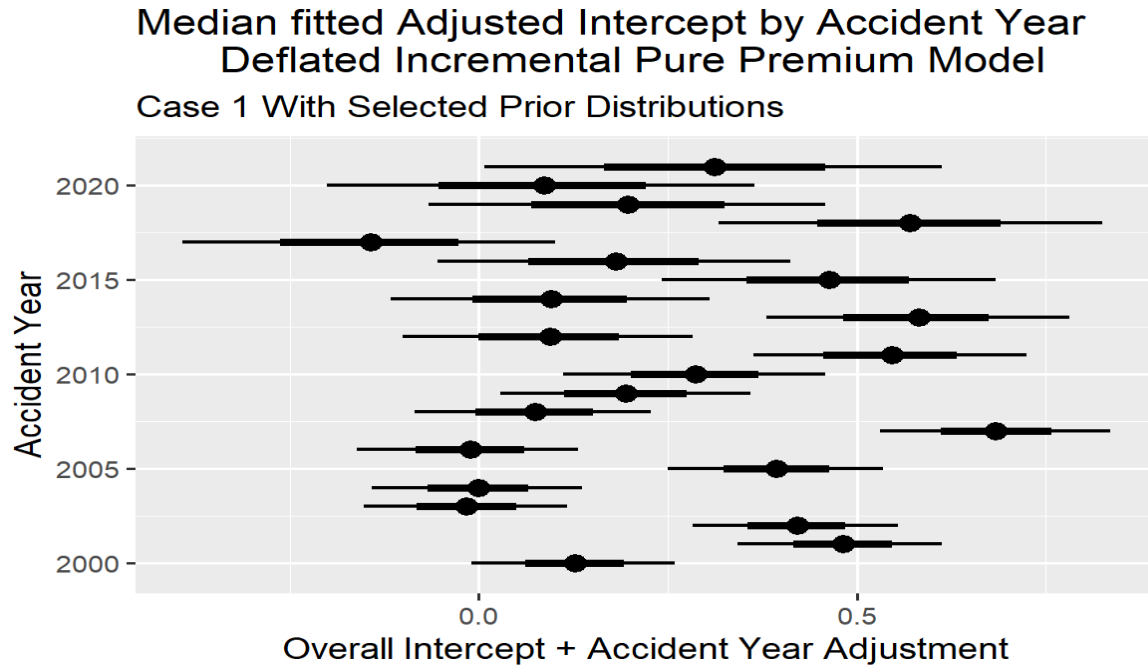
Case 1 With Selected Prior Distribution

Coefficients	Estimate	Std_Err	LB_CI_95	UB_CI_95
Intercept	0.244	0.1	0.045	0.438
Dev_Yr_2_Factor2	1.676	0.053	1.571	1.778
Dev_Yr_2_Factor3	2.075	0.089	1.898	2.248
Dev_Yr_6_Cap	1.657	0.069	1.523	1.795
Dev_Yr_6_Cap_Sqrd	-0.145	0.008	-0.16	-0.131
Dev_Yr_6_Spline	-0.217	0.011	-0.238	-0.196
Dev_Yr_6_Spline_Sqrd	-0.009	0.002	-0.014	-0.005
Cal_Yr_Time	0.023	0.006	0.01	0.035
Dev_Yr_10_Spline	-0.147	0.027	-0.198	-0.092
sigma_Intercept	-2.468	0.062	-2.597	-2.354
sigma_Dev_Yr_10_Spline_Ln	0.566	0.072	0.489	0.739

Prior_Distribution	Variable_Class	Coefficient	Group_Indicator	Distribution_Parameter	Source
	b				default
normal(0.02, 0.01)	b	Cal_Yr_Time			user
normal(-0.05, 0.1)	b	Dev_Yr_10_Spline			user
normal(1, 0.25)	b	Dev_Yr_2_Factor2			user
normal(2, 0.25)	b	Dev_Yr_2_Factor3			user
normal(1.5, 0.2)	b	Dev_Yr_6_Cap			user
normal(-0.1, 0.1)	b	Dev_Yr_6_Cap_Sqrd			user
normal(-0.1, 0.1)	b	Dev_Yr_6_Spline			user
normal(-0.05, 0.1)	b	Dev_Yr_6_Spline_Sqrd			user
normal(0, 0.25)	b	Intercept			user
	b			sigma	default
student_t(2, 0.5, 0.01)	b	Dev_Yr_10_Spline_Ln		sigma	user
normal(-1.5, 0.25)	b	Intercept		sigma	user
student_t(3, 0, 2.5)	sd				default
	sd		Acc_Yr		default
	sd	Intercept	Acc_Yr		default

The graph below for the Adjusted Intercept shows the predicted intercept adjusted by accident year with the adjustment calculated using partial pooling. The mean estimate is 0.244 for the Intercept and one can see there is some variation in the Intercept value across the accident years. The least squares approach allows one to recognize that the intercepts are a little different by accident year without running into the multi-collinearity problem created by trying to model in three dimensions (accident year, development year

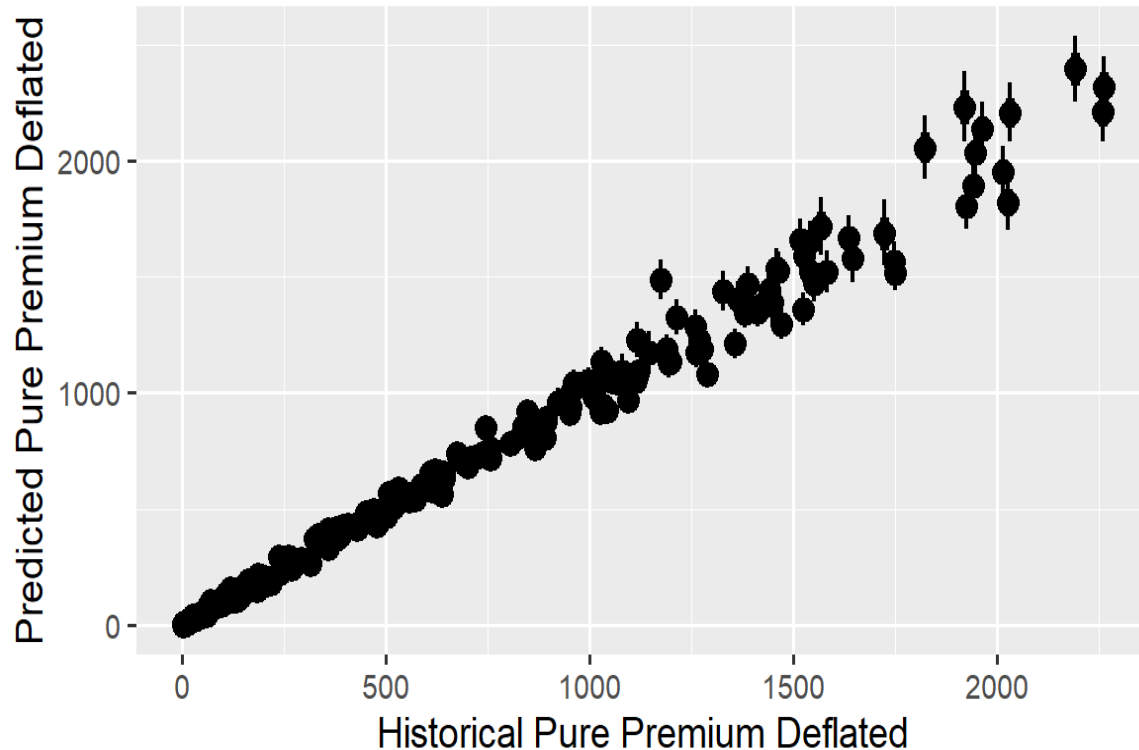
and calendar year) on a two-dimensional data set that one can experience when modeling in a GLM environment.



The graph below indicates that the modeled results track reasonably well with the observed deflated pure premium.

Historical Pure Premium Deflated vs. Distribution of Predicted

Case 1 With Selected Prior Distributions



II.4 Case 1 With Flat Prior Distribution

The results below were created using the same model as above for Case 1 data but with the priors set to the default value of a flat prior for the real number line. Electing to use the default flat prior means that the prior distribution column will show blanks (without a prior distribution selection) for the rows for the regression equation coefficients in the exhibits showing the prior distributions. In a sense, one cannot avoid selecting a prior distribution for each coefficient but selecting a flat prior means electing to use a prior distribution without specifying the specific distribution and the parameters for that distribution.

The intent is to allow the reader to see that with a large data set (one with lots of information) the Bayesian MCMC routine still produces sensible results without the benefit of selecting a plausible set of prior distribution parameters. The parameter estimates shown in the summary exhibit differ from those from the model, which included selected prior distributions. Although the standard deviation around these estimates can be slightly larger, the residual patterns still indicate the model produces usable results. Additionally, the 95% confidence intervals suggest that the coefficients are reliably non-zero.

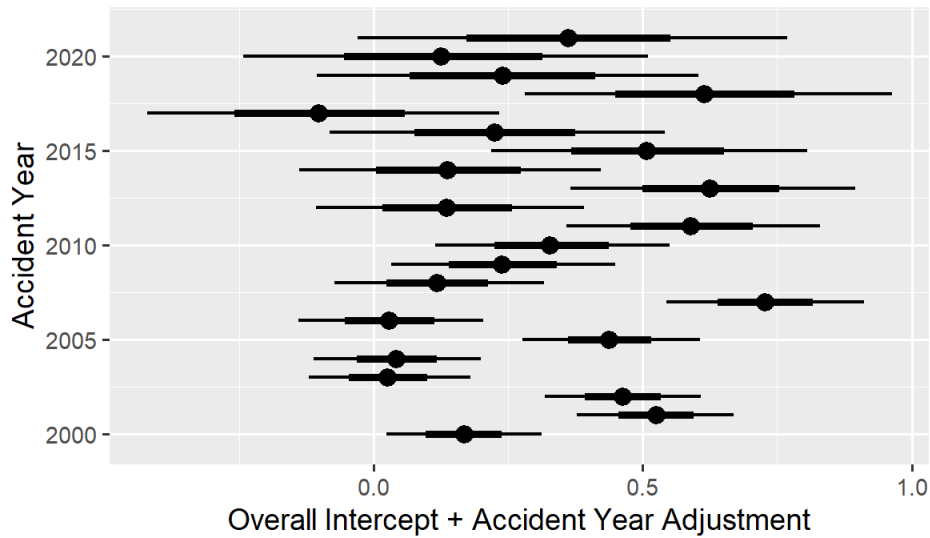
Case 1 Without Selected Prior Distribution

Coefficients	Estimate	Std_Err	LB_CI_95	UB_CI_95
Intercept	0.299	0.125	0.062	0.547
Dev_Yr_2_Factor2	1.718	0.056	1.608	1.83
Dev_Yr_2_Factor3	2.144	0.099	1.952	2.341
Dev_Yr_6_Cap	1.606	0.077	1.454	1.758
Dev_Yr_6_Cap_Sqrd	-0.14	0.008	-0.156	-0.123
Dev_Yr_6_Spline	-0.212	0.013	-0.238	-0.186
Dev_Yr_6_Spline_Sqrd	-0.011	0.002	-0.016	-0.006
Cal_Yr_Time	0.023	0.009	0.004	0.04
Dev_Yr_10_Spline	-0.132	0.029	-0.186	-0.076
sigma_Intercept	-2.591	0.056	-2.697	-2.479
sigma_Dev_Yr_10_Spline_Ln	0.73	0.067	0.604	0.867

Prior_Distribution	Variable_Class	Coefficient	Group_Indicator	Distribution_Parameter	Source
	b				default
	b	Cal_Yr_Time			default
	b	Dev_Yr_10_Spline			default
	b	Dev_Yr_2_Factor2			default
	b	Dev_Yr_2_Factor3			default
	b	Dev_Yr_6_Cap			default
	b	Dev_Yr_6_Cap_Sqrd			default
	b	Dev_Yr_6_Spline			default
	b	Dev_Yr_6_Spline_Sqrd			default
	b	Intercept			default
	b			sigma	default
	b	Dev_Yr_10_Spline_Ln		sigma	default
	b	Intercept		sigma	default
student_t(3, 0, 2.5)	sd				default
	sd		Acc_Yr		default
	sd	Intercept	Acc_Yr		default

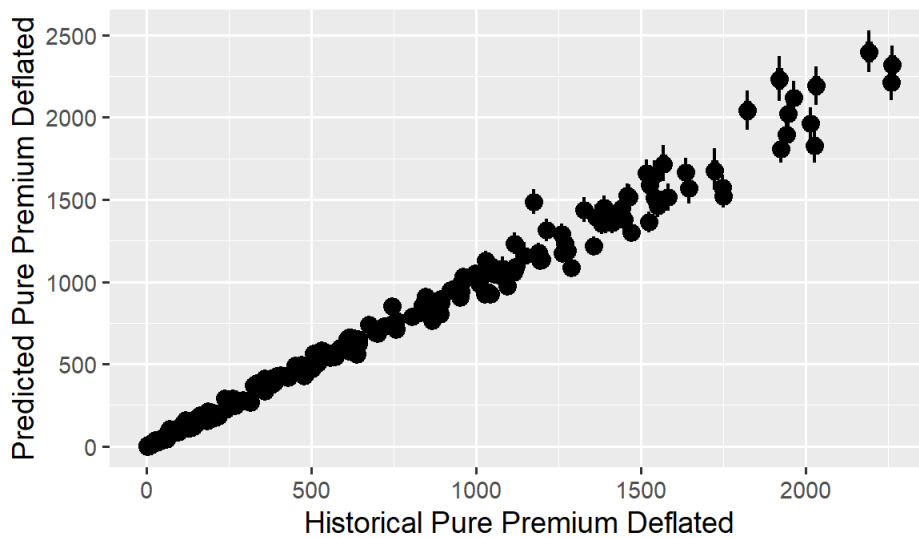
The graph below shows the effect of using partial pooling on the intercept by accident year. This illustrates that one can use partial pooling (least squares weighting) in a Bayesian MCMC model even if one does not elect to select the prior distribution parameters for the population variables. Please note that the more recent accident years tend to have larger confidence intervals below, since they have fewer development year observation which translates to more uncertainty in the adjustment forecast.

Median fitted Adjusted Intercept by Accident Year
 Deflated Incremental Pure Premium Model
 Case 1 Without Selected Prior Distributions



The graph below demonstrates that, even without selecting prior distribution parameters for the population variables, the triangle data set contains sufficient information to fit the observations reasonably well given the volume of information in Case 1.

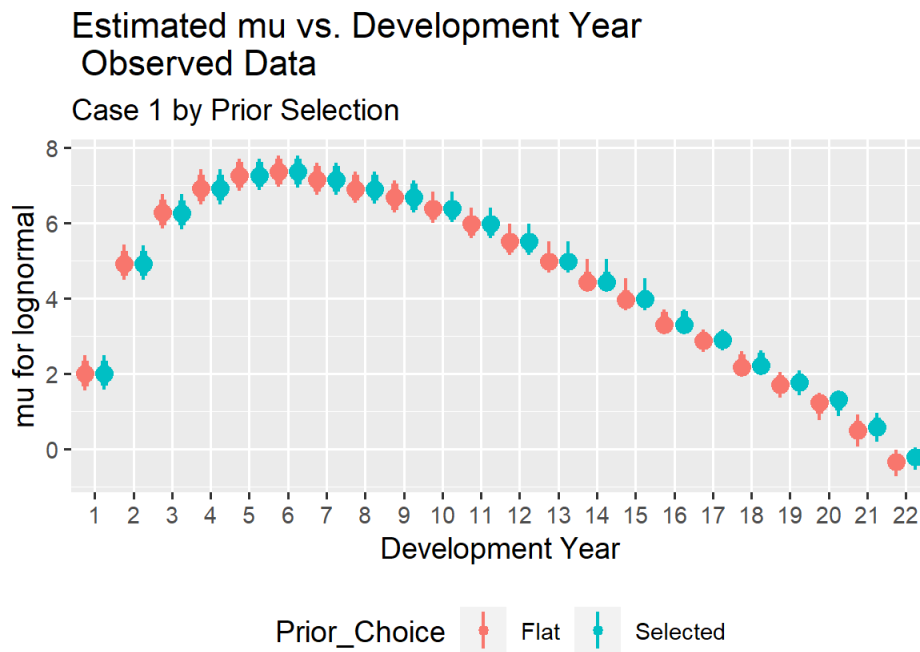
Historical Pure Premium Deflated
 vs. Distribution of Predicted
 Case 1 Without Selected Prior Distributions



II.5 Case I Compare Reserve Distribution Estimate With and Without Selected Priors

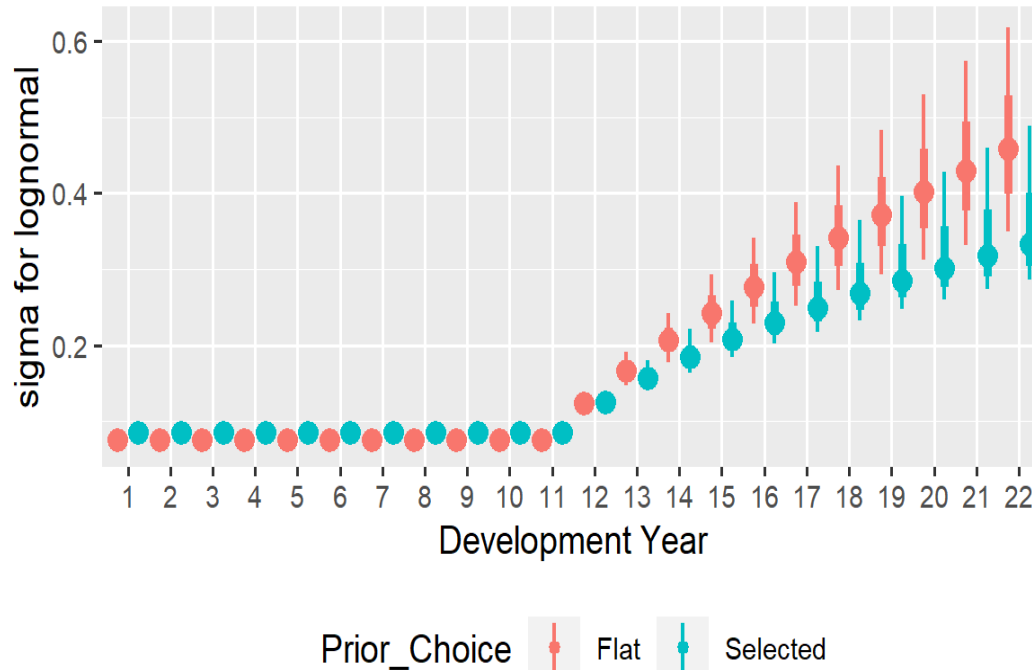
The end result for the total indicated reserve distribution indicates that for Case 1 there is no appreciable difference in the estimated reserve distribution between the model with selected prior distributions and the model that took the default (flat) prior distributions.

It appears that with 1,000 reported claims, the loss triangle contains enough information to outweigh the influence of the prior distribution parameters (close to 100 % credibility).



Estimated sigma vs. Development Year Observed Data

Case 1 by Prior Selection

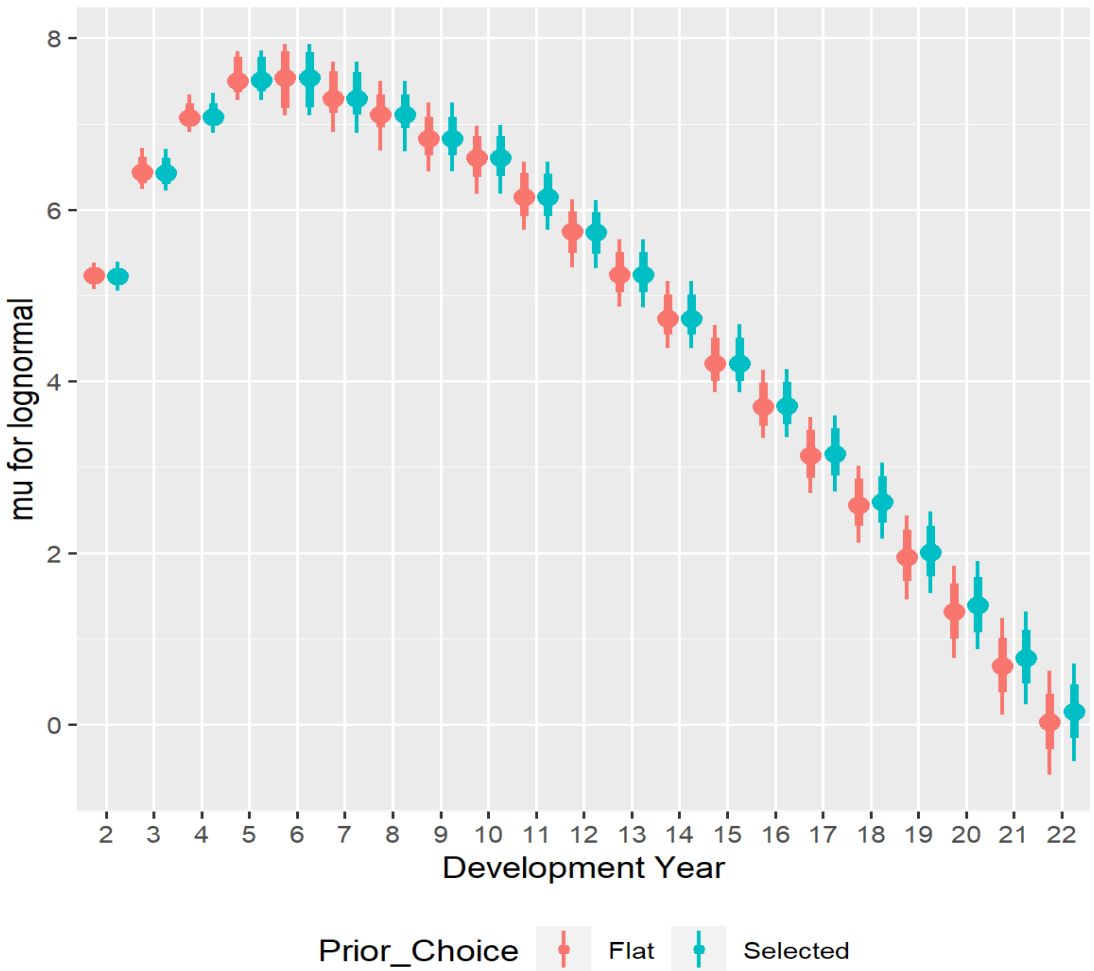


Given that there is less information for older development ages in the observed loss triangle, there is some difference in the estimates for sigma in the tail.

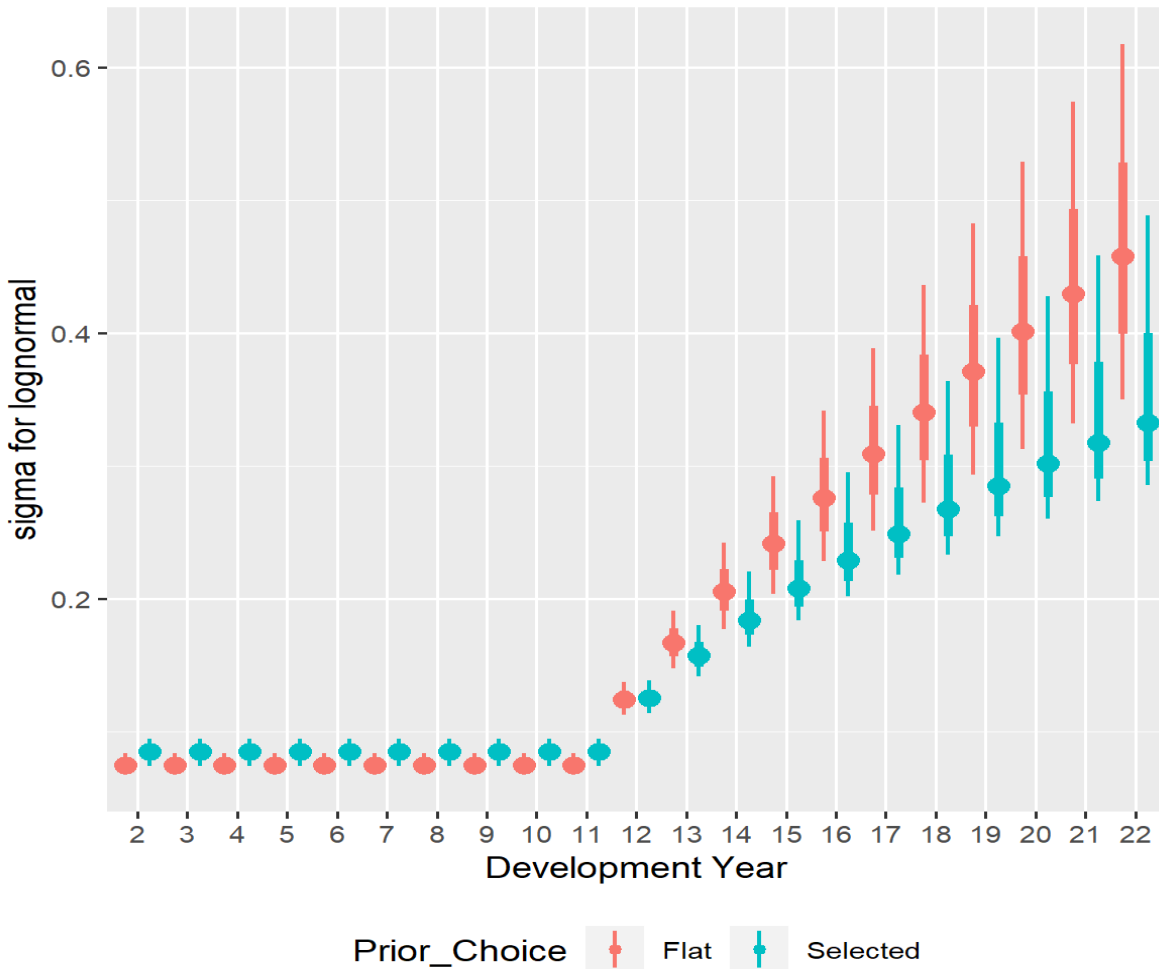
We can also examine the mu and sigma values for the future time periods in the forecast. It is important to note that when the posterior distribution for the forecast loss payments is created, the routines will invert the lognormal distribution using those parameters rather than simply calculating the expected values, as is done for the time periods with observations. In other words, the loss reserve forecasts (the sum of the future incremental loss payments) will account for both process variance and parameter uncertainty.

If you look at the two graphs below, you can see that the development year starts at two rather than one because all of the future development years for the accident years in the forecast are greater than one.

Forecast mu vs. Development Year Future Accident & Development Years Case 1 by Prior Selection



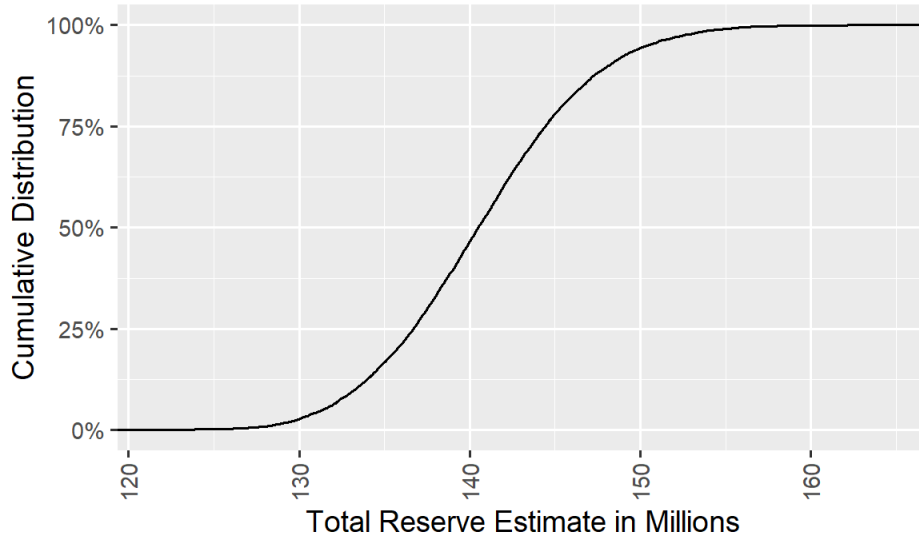
Forecast sigma vs. Development Year Future Accident & Development Years Case 1 by Prior Selection



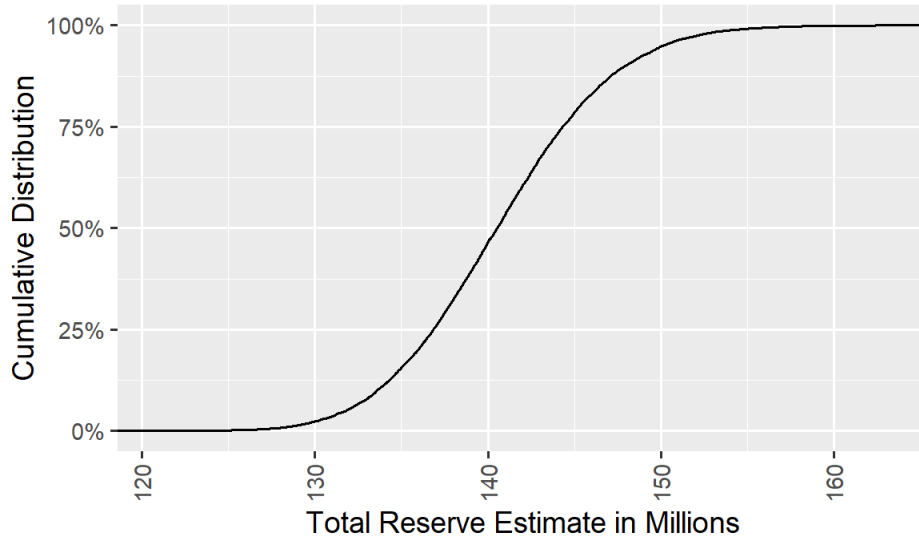
Although the parameters differ slightly with and without selected priors, the total reserve distributions with and without selected prior distribution parameters are close. The estimated reserve distributions are shown both separately and overlaid for comparison, as they diverge noticeably in later cases.

The accident year comparison of forecast loss reserves for Case 1 shows that the reserve forecast distributions are virtually identical with and without selecting the prior parameters. However, this will change as the information in the loss triangle decreases.

Total Reserve Estimate
Case 1 With Selected Prior Distribution
Simulated Future Inflation

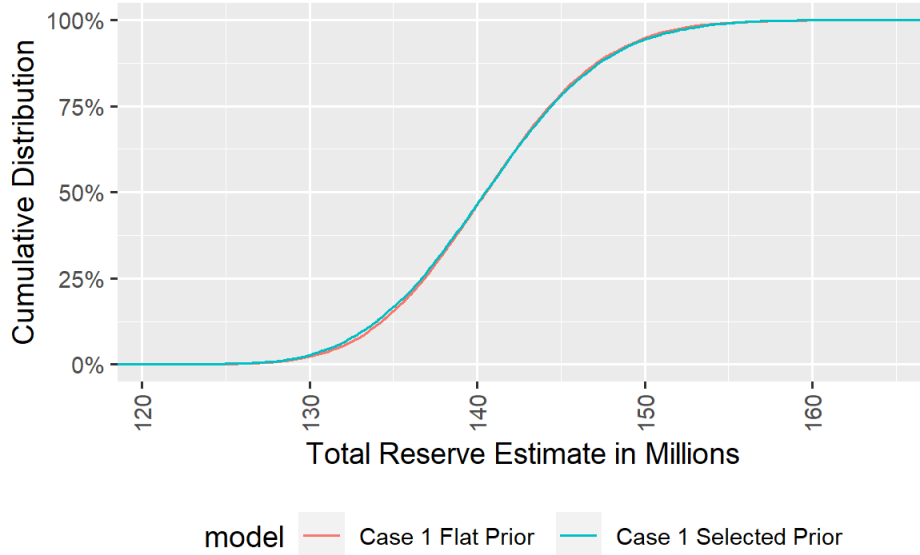


Total Reserve Estimate
Case 1 No Selected Prior Distribution
Simulated Future Inflation



The total reserve forecasts with and without selected prior distribution parameters are shown separately, even though the overlaid forecasts appear virtually the same. This approach maintains a consistent presentation since at some point, the forecasts without selected prior distribution parameters will exhibit such high tail values that overlaying the two forecast loss distributions would not be useful.

Total Reserve Estimate Case 1 Models With and Without Selected Prior



Accident Year Ultimate Reserve Estimates Case 1 by Prior Selection



III. Case 2 Modeling Results

Case 2 has 22 accident years (2000 through 2021) with a uniform reported claim count of 500 reported claims at 12 months for each accident year and a maximum of 22 years of development time.

The reason for presenting modeling results for Case 2 is to demonstrate how systematically reducing the information in a loss triangle affects the forecast loss reserve distribution.

III.1 Case 2 Results With Prior Distribution Selected

As the exposure count dropped from 1,000 to 500 reported claims per accident year, the amount of information in the triangle decreased, leading to an increase in the standard error around the parameter estimates. This occurred even though the prior distributions remained the same as for those used in Case 1. The mean parameter estimates also changed slightly, but the 95% confidence intervals still indicate that the effects are non-zero.

Case 2 With Selected Prior Distribution

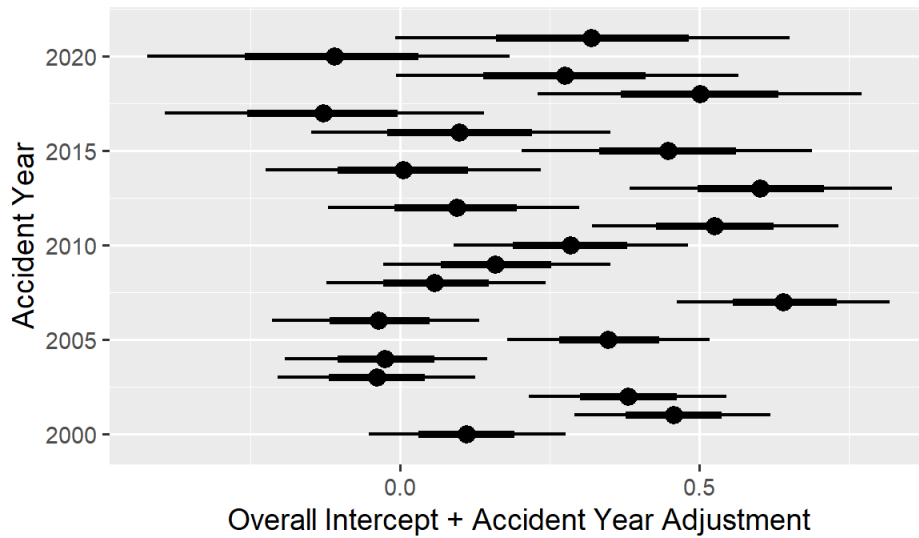
Coefficients	Estimate	Std_Err	LB_CI_95	UB_CI_95
Intercept	0.213	0.107	0.007	0.419
Dev_Yr_2_Factor2	1.659	0.063	1.538	1.784
Dev_Yr_2_Factor3	2.022	0.106	1.818	2.232
Dev_Yr_6_Cap	1.7	0.082	1.538	1.861
Dev_Yr_6_Cap_Sqrd	-0.149	0.009	-0.167	-0.132
Dev_Yr_6_Spline	-0.226	0.012	-0.251	-0.202
Dev_Yr_6_Spline_Sqrd	-0.008	0.002	-0.012	-0.003
Cal_Yr_Time	0.021	0.007	0.008	0.034
Dev_Yr_10_Spline	-0.155	0.032	-0.218	-0.094
sigma_Intercept	-2.192	0.052	-2.299	-2.09
sigma_Dev_Yr_10_Spline_Ln	0.525	0.041	0.484	0.648

Prior_Distribution	Variable_Class	Coefficient	Group_Indicator	Distribution_Parameter	Source
	b				default
normal(0.02, 0.01)	b	Cal_Yr_Time			user
normal(-0.05, 0.1)	b	Dev_Yr_10_Spline			user
normal(1, 0.25)	b	Dev_Yr_2_Factor2			user
normal(2, 0.25)	b	Dev_Yr_2_Factor3			user
normal(1.5, 0.2)	b	Dev_Yr_6_Cap			user
normal(-0.1, 0.1)	b	Dev_Yr_6_Cap_Sqrd			user
normal(-0.1, 0.1)	b	Dev_Yr_6_Spline			user
normal(-0.05, 0.1)	b	Dev_Yr_6_Spline_Sqrd			user
normal(0, 0.25)	b	Intercept			user
	b			sigma	default
student_t(2, 0.5, 0.01)	b	Dev_Yr_10_Spline_Ln		sigma	user
normal(-1.5, 0.25)	b	Intercept		sigma	user
student_t(3, 0, 2.5)	sd				default
	sd		Acc_Yr		default
	sd	Intercept	Acc_Yr		default

The effect of partial pooling by accident year for the intercept is shown on the graph below.

Median fitted Adjusted Intercept by Accident Year Deflated Incremental Pure Premium Model

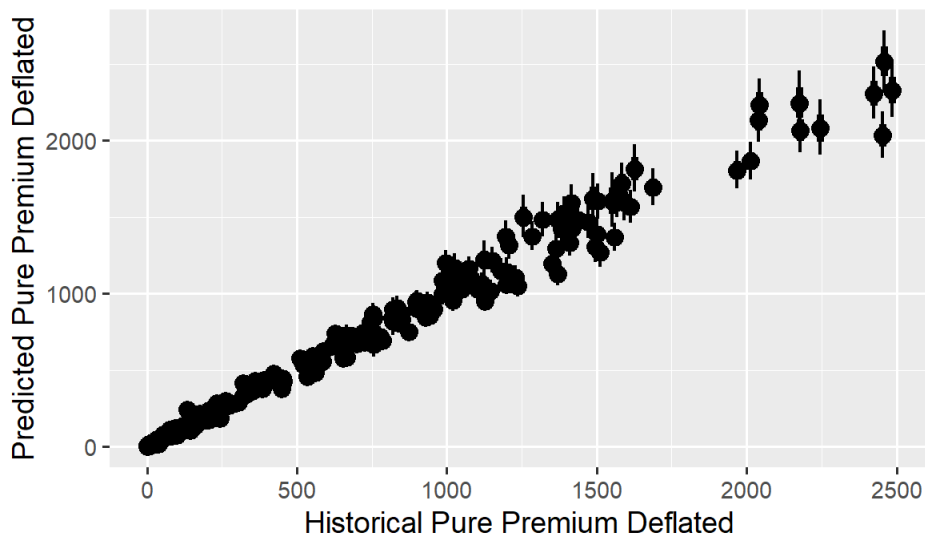
Case 2 With Selected Prior Distributions



The graphs indicate that the model provides a usable fit to the loss triangles. The predicted vs. observed pure premium graph shows that the model results track reasonably well with the observed values.

Historical Pure Premium Deflated vs. Distribution of Predicted

Case 2 With Selected Prior Distributions



III.2 Case 2 With Flat Prior Distribution

The results are similar to those observed in Case 1. Although the standard errors around the parameters are larger when using the flat priors (prior distributions without selected distribution parameters), the model still produces usable results. The summary of results indicate that the parameters can be safely assumed to be non-zero since the 95% confidence interval does not include zero. Additionally, the graphs indicate the model provides a reliable fit to the observed loss triangle.

Case 2 Without Selected Prior Distribution

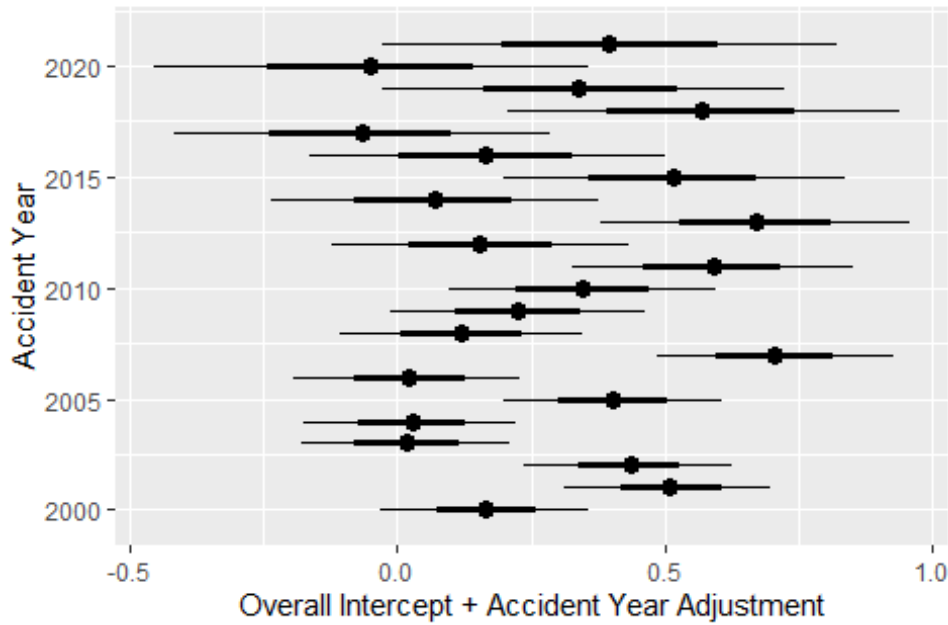
Coefficients	Estimate	Std_Err	LB_CI_95	UB_CI_95
Intercept	0.286	0.141	0	0.554
Dev_Yr_2_Factor2	1.713	0.075	1.563	1.859
Dev_Yr_2_Factor3	2.108	0.133	1.844	2.363
Dev_Yr_6_Cap	1.637	0.103	1.441	1.843
Dev_Yr_6_Cap_Sqrd	-0.143	0.011	-0.165	-0.121
Dev_Yr_6_Spline	-0.228	0.015	-0.257	-0.199
Dev_Yr_6_Spline_Sqrd	-0.007	0.003	-0.012	-0.001
Cal_Yr_Time	0.02	0.009	0.002	0.038
Dev_Yr_10_Spline	-0.171	0.036	-0.24	-0.101
sigma_Intercept	-2.295	0.056	-2.403	-2.183
sigma_Dev_Yr_10_Spline_Ln	0.677	0.064	0.557	0.81

Prior_Distribution	Variable_Class	Coefficient	Group_Indicator	Distribution_Parameter	Source
	b				default
	b	Cal_Yr_Time			default
	b	Dev_Yr_10_Spline			default
	b	Dev_Yr_2_Factor2			default
	b	Dev_Yr_2_Factor3			default
	b	Dev_Yr_6_Cap			default
	b	Dev_Yr_6_Cap_Sqrd			default
	b	Dev_Yr_6_Spline			default
	b	Dev_Yr_6_Spline_Sqrd			default
	b	Intercept			default
	b			sigma	default
	b	Dev_Yr_10_Spline_Ln		sigma	default
	b	Intercept		sigma	default
student_t(3, 0, 2.5)	sd				default
	sd		Acc_Yr		default
	sd	Intercept	Acc_Yr		default

The graph below shows the effect of partial pooling on the intercept for the equation for the lognormal parameter mu.

Median fitted Adjusted Intercept by Accident Year Deflated Incremental Pure Premium Model

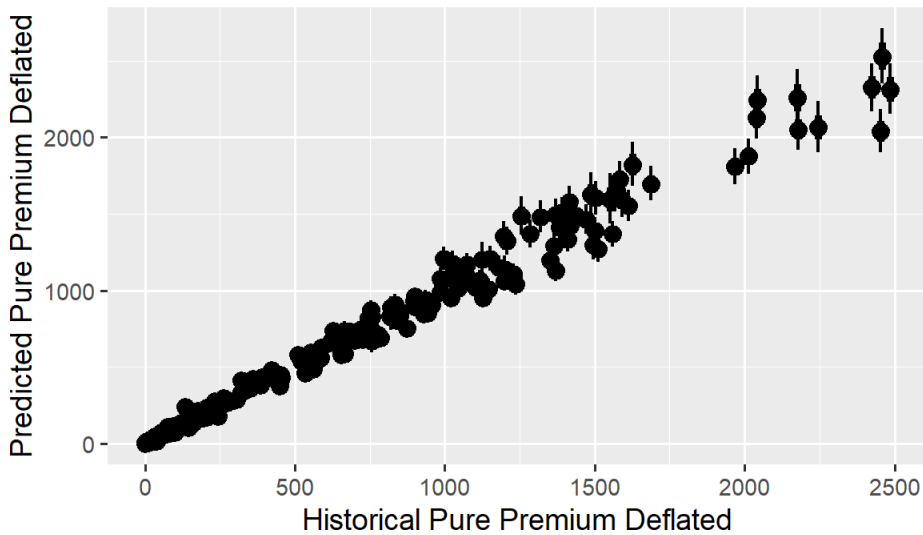
Case 2 Without Selected Prior Distributions



The graph showing the observed deflated pure premium compared to the posterior distribution estimates is shown below and indicates that using a flat prior still produces reasonable results.

Historical Pure Premium Deflated vs. Distribution of Predicted

Case 2 Without Selected Prior Distributions



III.3 Case 2 Compare Reserve Distribution Estimate With and Without Selected Priors

The results below were generated using the same process for Case 2 as for Case 1. The results indicate that reducing the number of exposures from 1,000 to 500 did not significantly impact the comparison between using a selected a set of prior distribution parameters and using a flat set of prior distributions.

There is still sufficient information in the loss triangles used to fit the model effectively, outweighing the influence of the selected prior distributions. While using a chosen set of prior distribution parameters produced different posterior estimates for the coefficients, the net difference is negligible. However, there is a slight difference in the total reserve forecast when overlaying the curve from the model with selected prior distribution parameters on the curve from the model with flat priors.

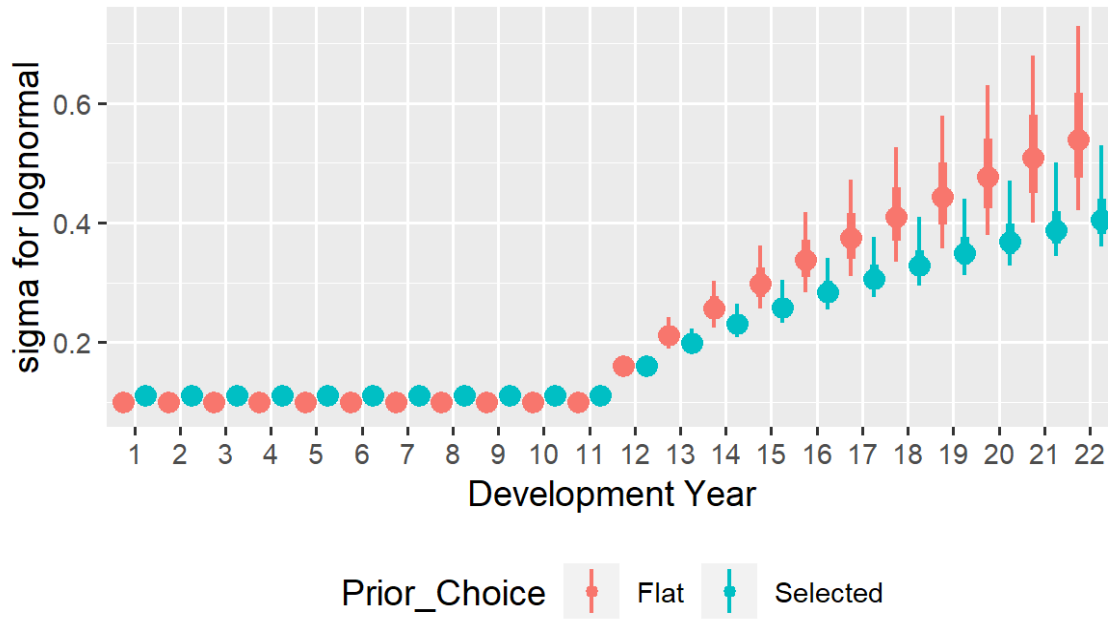
It is important to note that this result is specific to this set of examples. In practice, you would need to experiment with a data set tailored to match observations within your company to evaluate the effect of reducing the exposure count.



Estimated sigma vs. Development Year

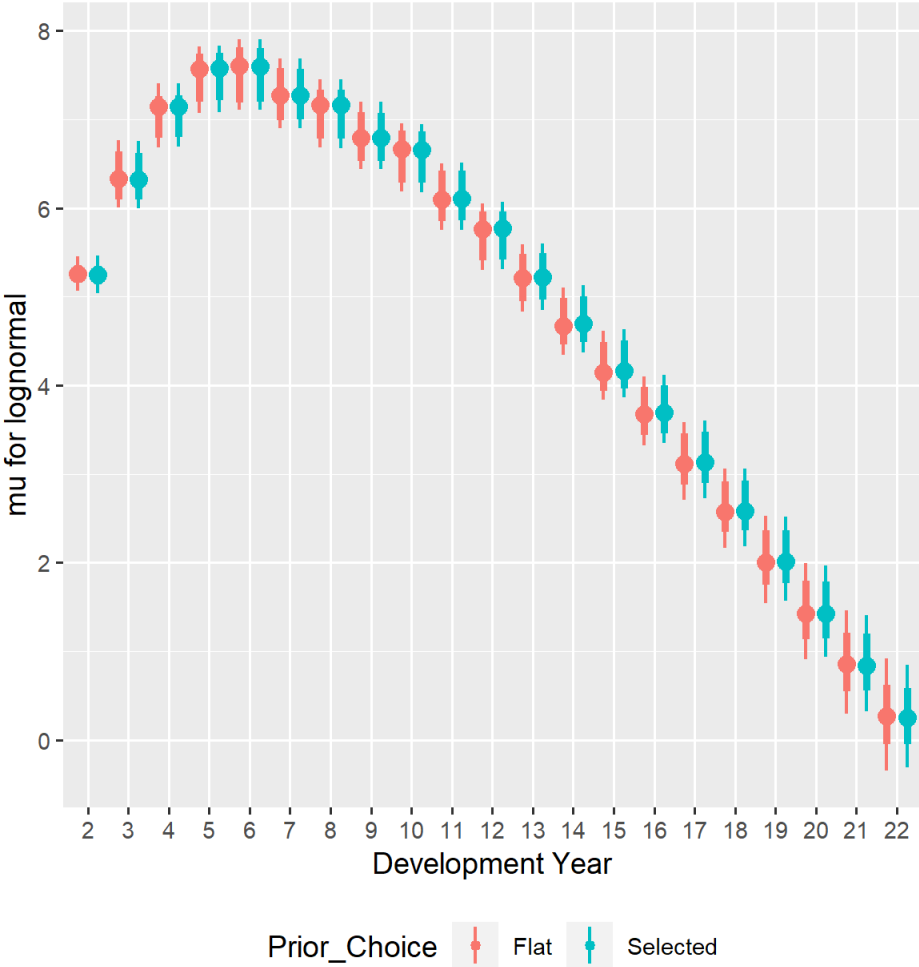
Observed Data

Case 2 by Prior Selection

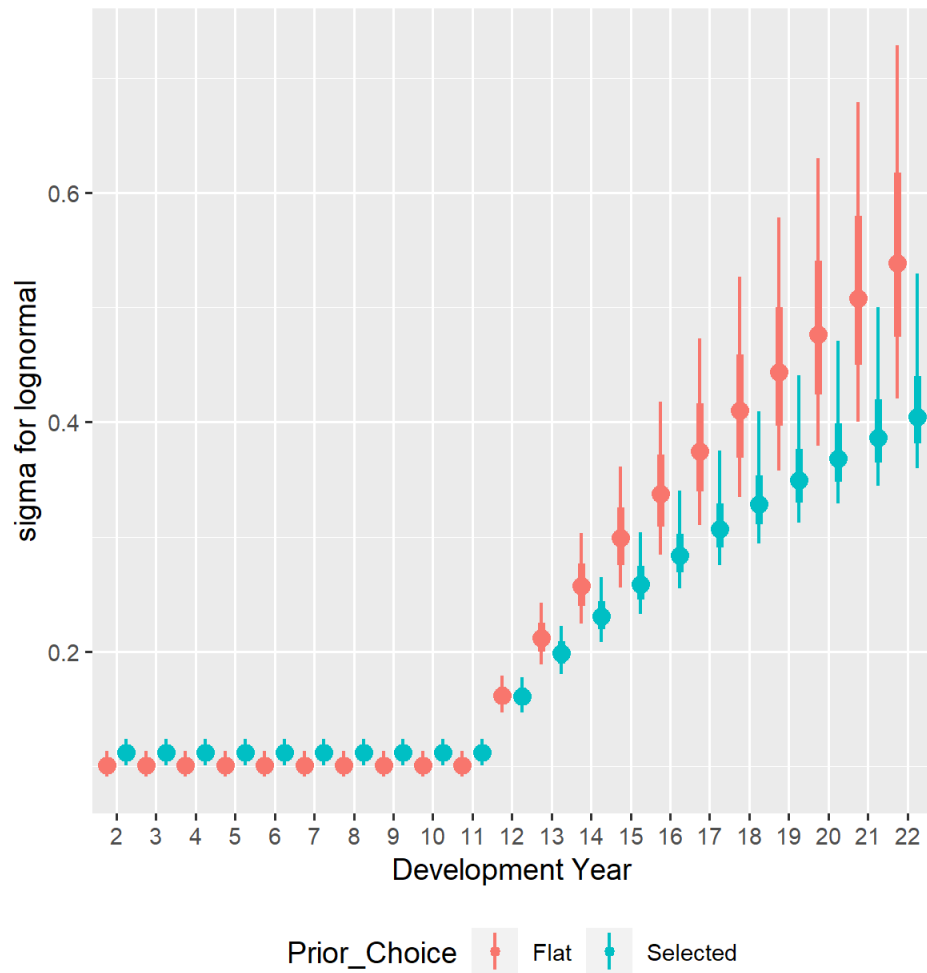


Forecast mu vs. Development Year Future Accident & Development Years

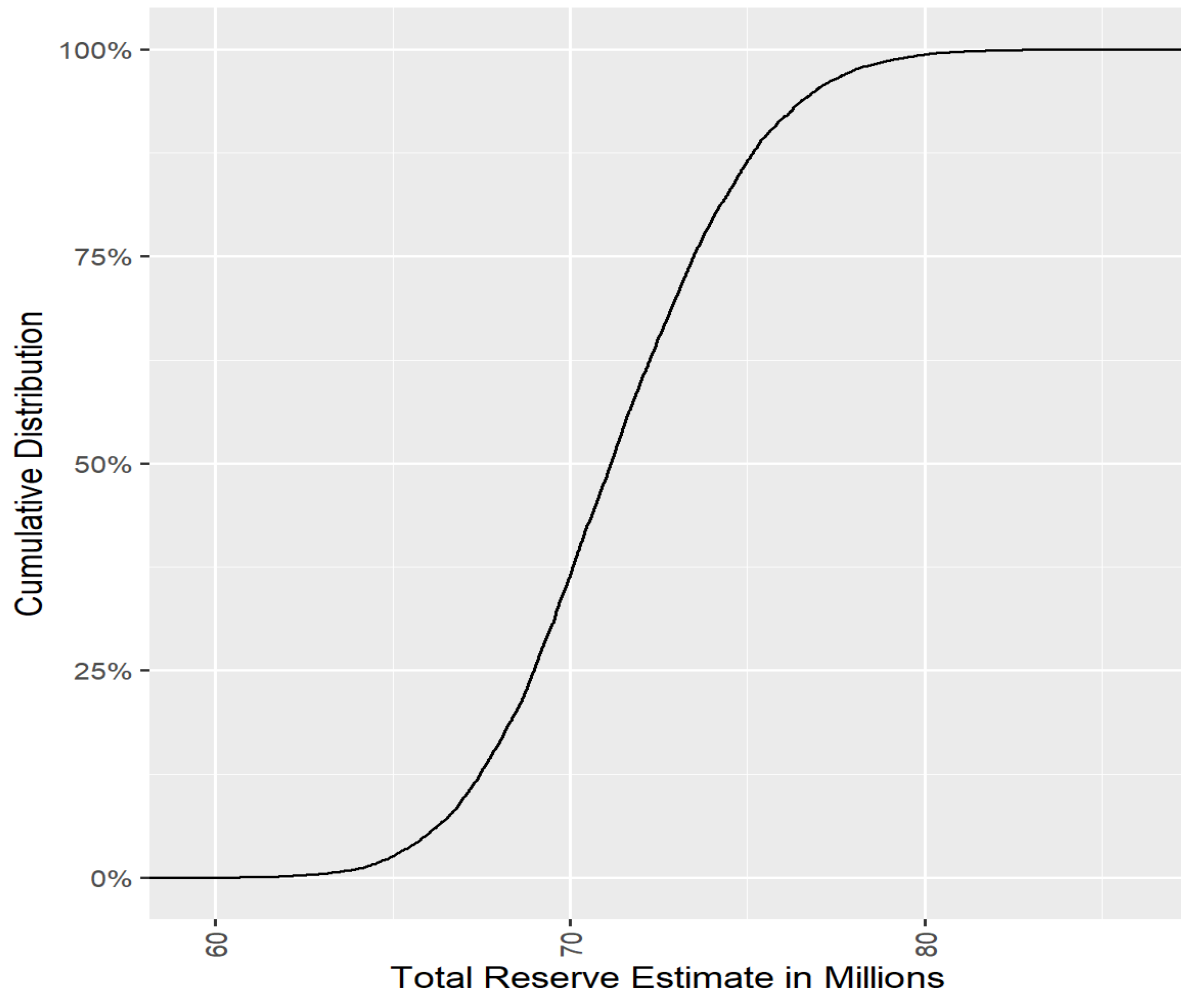
Case 2 by Prior Selection



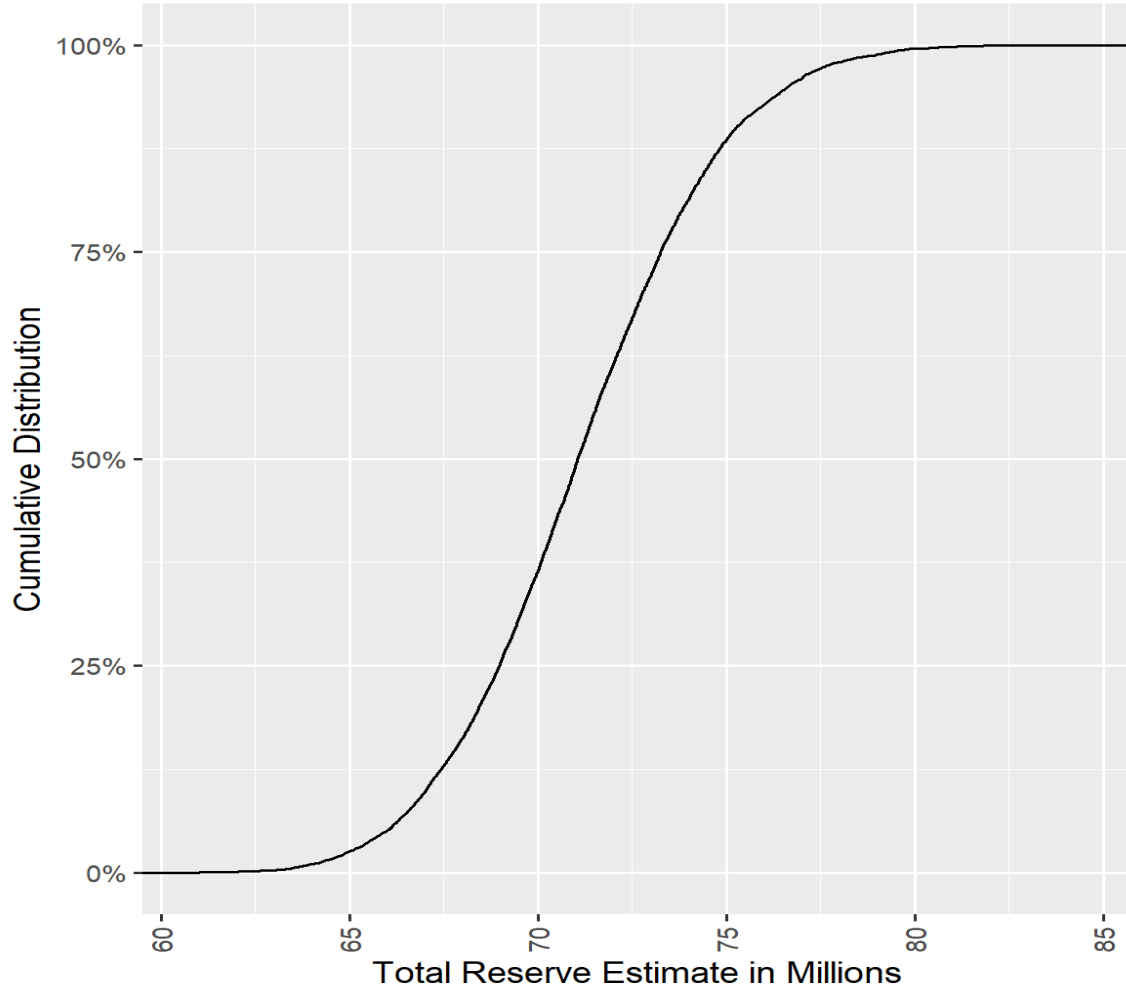
Forecast sigma vs. Development Year Future Accident & Development Years Case 2 by Prior Selection



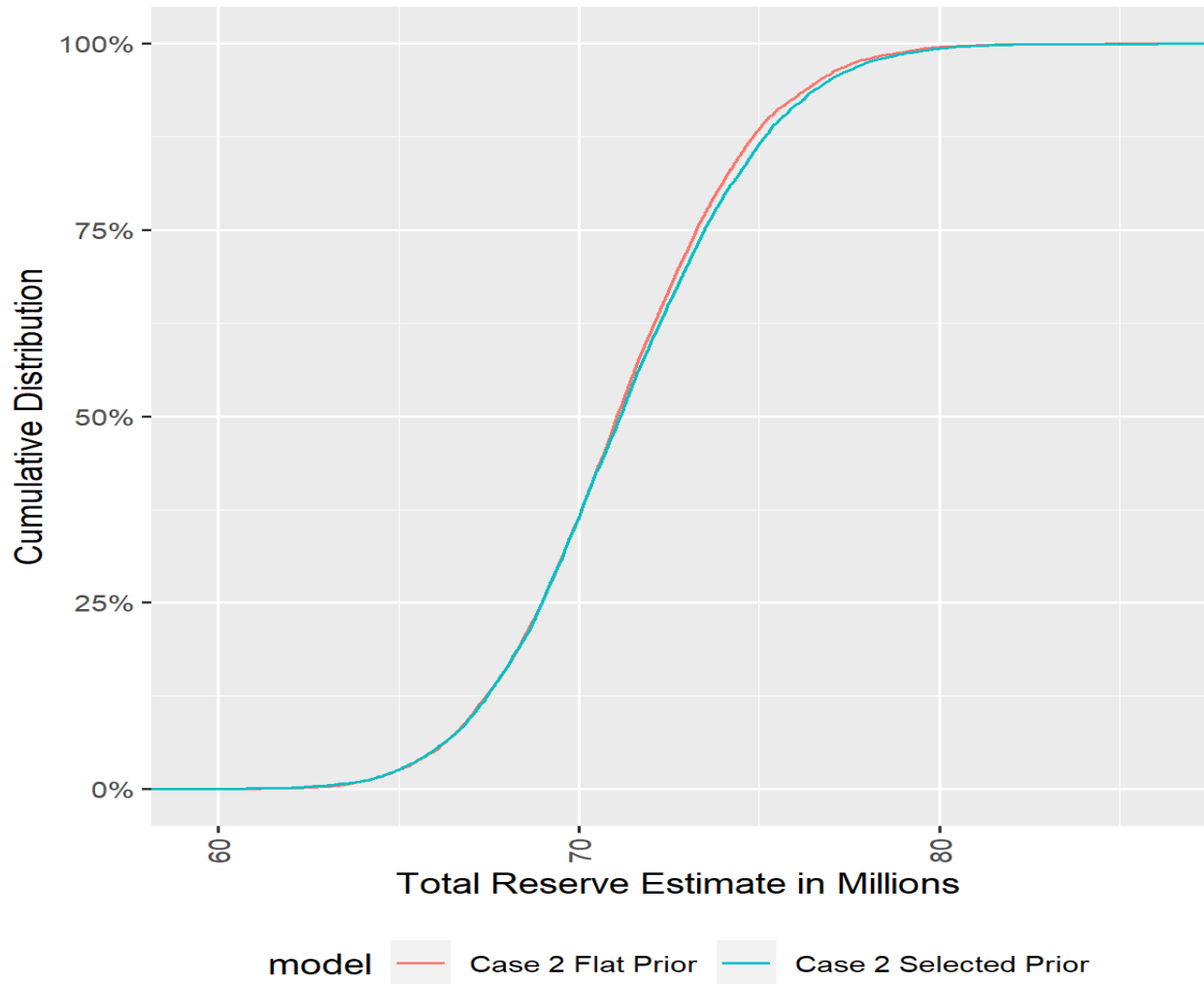
Total Reserve Estimate
Case 2 With Selected Prior Distribution
Simulated Future Inflation



Total Reserve Estimate
Case 2 No Selected Prior Distribution
Simulated Future Inflation

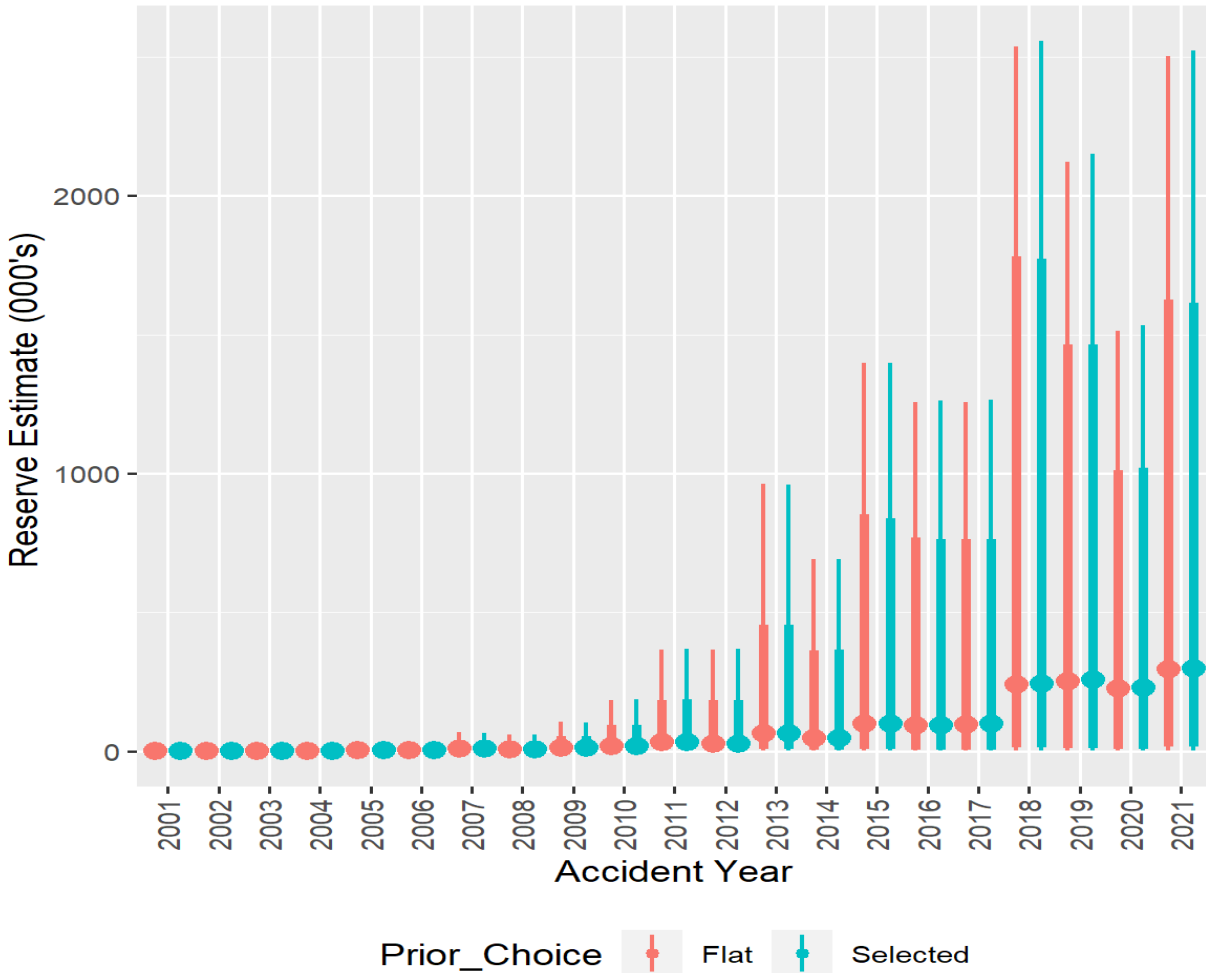


Total Reserve Estimate Case 2 Models With and Without Selected Prior



Accident Year Ultimate Reserve Estimates

Case 2 by Prior Selection



IV. Case 3 Modeling Results

Case 3 has 22 accident years (2000 through 2021) with a uniform reported claim count of 100 reported claims at 12 months for each accident year and a maximum of 22 years of development time.

The purpose of presenting the modeling results for Case 3 is to further illustrate how reducing the amount of information in a loss triangle impacts the effectiveness of using a prior distribution in a Bayesian MCMC model.

IV.1 Case 3 Results with Prior Distribution Selected

The results for Case 3 will be produced in the same manner as for the first two cases. A similar progression is observed when reducing the exposure base from 500 reported claims to 100, as seen when reducing from 1,000 to 500 reported claims. The standard error around the estimated parameters increased due to the reduced information in the loss triangle, but the actual-to-predicted graph indicates that the model results are still usable.

When moving from 500 to 100 claims, the coefficients for development years beyond six years became unstable. That led to altering the prior distributions for the coefficients used to predict development beyond six years to put in tighter guard rails around the estimation process.

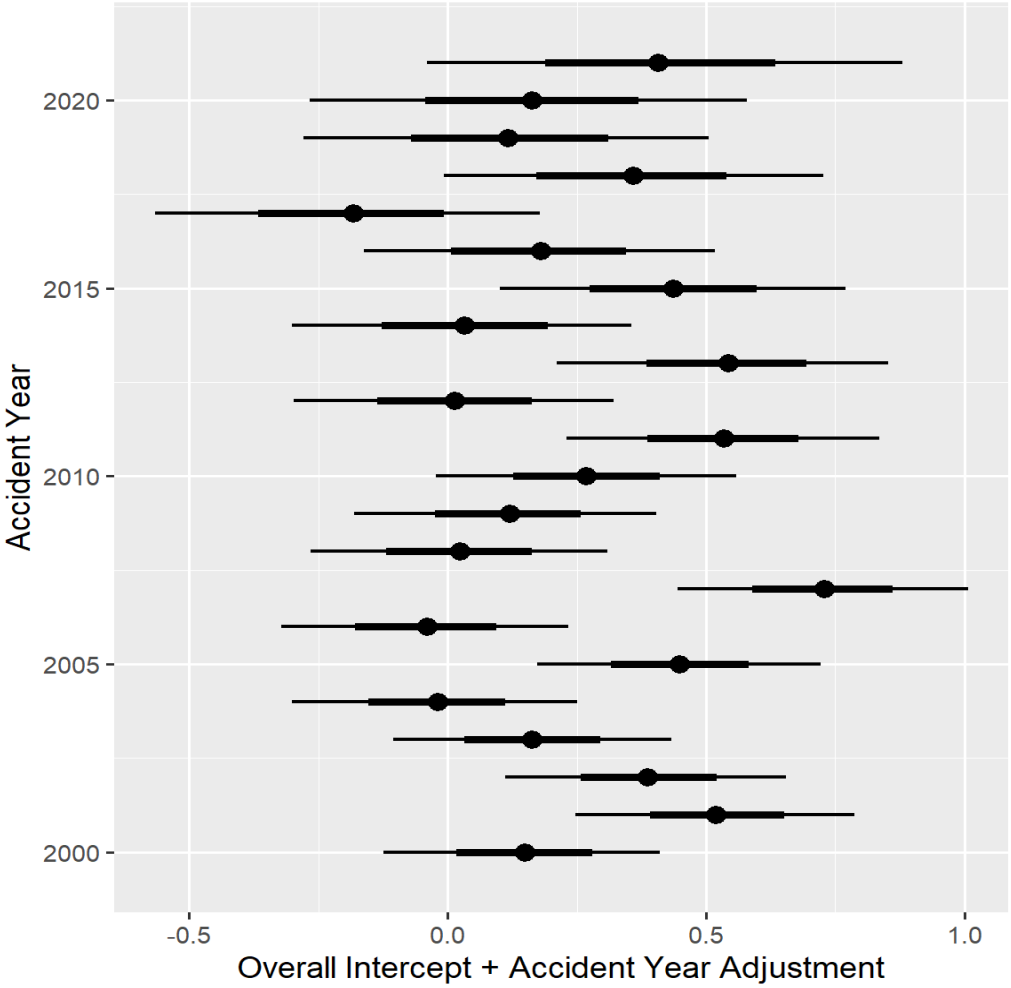
Additionally, comparing the length of the confidence intervals on the observed versus predicted graphs for Cases 1 and 2 reveals that the confidence intervals tend to become larger as the amount of information in the loss triangle decreases, even with tightening the standard deviation for selected coefficients. Later in the paper, comparisons between the cases will directly demonstrate how reducing information in the loss triangle affects the confidence intervals for μ and σ .

Case 3 With Selected Prior Distribution

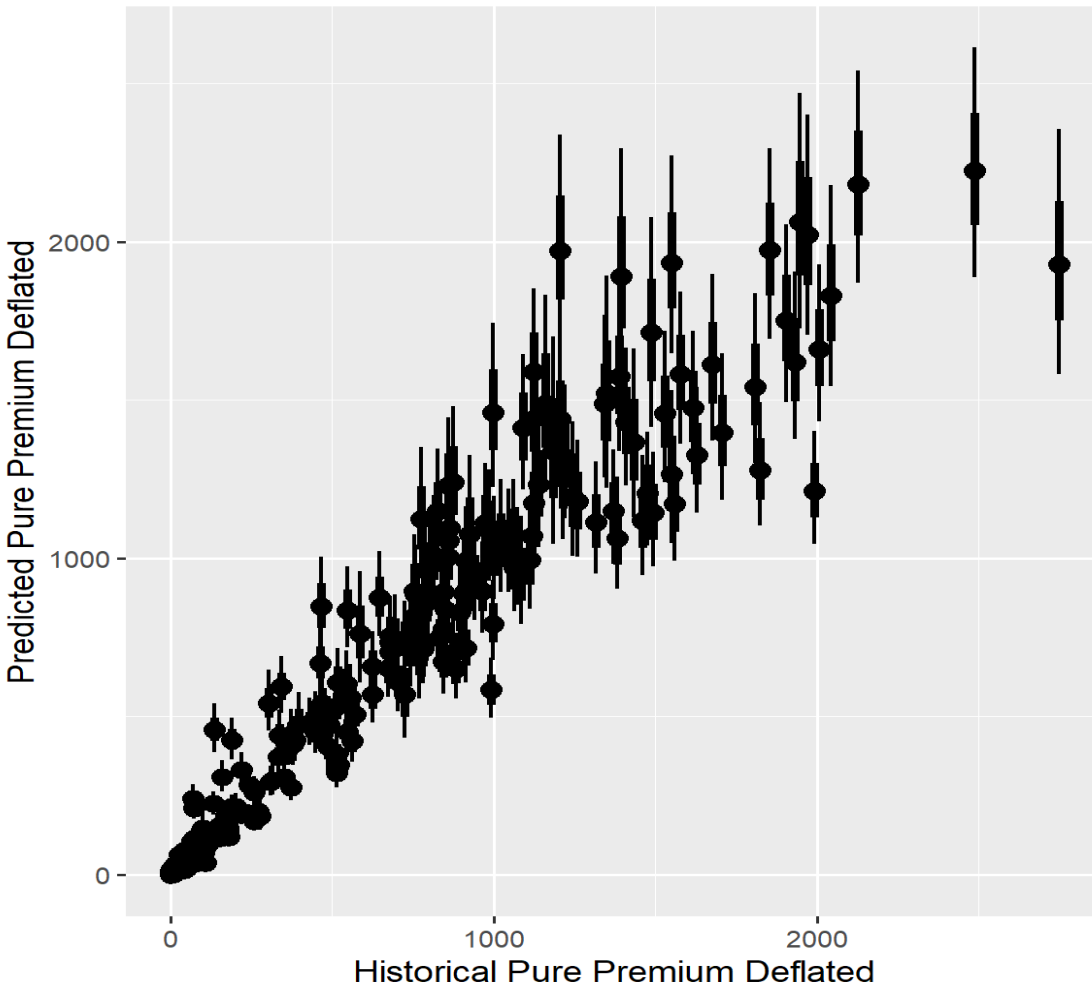
Coefficients	Estimate	Std_Err	LB_CI_95	UB_CI_95
Intercept	0.23	0.134	-0.041	0.49
Dev_Yr_2_Factor2	1.712	0.101	1.512	1.915
Dev_Yr_2_Factor3	2.13	0.15	1.837	2.424
Dev_Yr_6_Cap	1.681	0.112	1.462	1.903
Dev_Yr_6_Cap_Sqrd	-0.153	0.013	-0.178	-0.128
Dev_Yr_6_Spline	-0.112	0.023	-0.157	-0.068
Dev_Yr_6_Spline_Sqrd	-0.026	0.004	-0.033	-0.019
Cal_Yr_Time	0.019	0.007	0.006	0.032
Dev_Yr_10_Spline	-0.097	0.043	-0.179	-0.013
sigma_Intercept	-1.328	0.062	-1.452	-1.21
sigma_Dev_Yr_10_Spline_Ln	0.58	0.076	0.491	0.746

Prior_Distribution	Variable_Class	Coefficient	Group_Indicator	Distribution_Parameter	Source
	b				default
normal(0.02, 0.01)	b	Cal_Yr_Time			user
normal(-0.15, 0.05)	b	Dev_Yr_10_Spline			user
normal(1, 0.25)	b	Dev_Yr_2_Factor2			user
normal(2, 0.25)	b	Dev_Yr_2_Factor3			user
normal(1.5, 0.2)	b	Dev_Yr_6_Cap			user
normal(-0.1, 0.1)	b	Dev_Yr_6_Cap_Sqrd			user
normal(-0.1, 0.05)	b	Dev_Yr_6_Spline			user
normal(-0.1, 0.05)	b	Dev_Yr_6_Spline_Sqrd			user
normal(0, 0.25)	b	Intercept			user
	b			sigma	default
student_t(2, 0.5, 0.01)	b	Dev_Yr_10_Spline_Ln		sigma	user
normal(-1.5, 0.25)	b	Intercept		sigma	user
student_t(3, 0, 2.5)	sd				default
	sd		Acc_Yr		default
	sd	Intercept	Acc_Yr		default

Median fitted Adjusted Intercept by Accident Year
Deflated Incremental Pure Premium Model
Case 3 With Selected Prior Distributions



Historical Pure Premium Deflated vs. Distribution of Predicted Case 3 With Selected Prior Distributions



IV.2 Case 3 With Flat Prior

Dropping from 500 to 100 reported claims starts to affect the reliability of the coefficients when a flat prior distribution is used as the complement of credibility. The variable `Dev_Yr_10_Spline` was dropped, since its 95% confidence interval included zero. Dropping `Dev_Yr_10_Spline` tended to stabilize the coefficients for other variables. One can see that the coefficient for additional loss cost trend factor variable, `Ca_Yr_Time`, now has a confidence interval that border line includes zero, which was not the case for the model for this case that included selected prior distribution parameters.

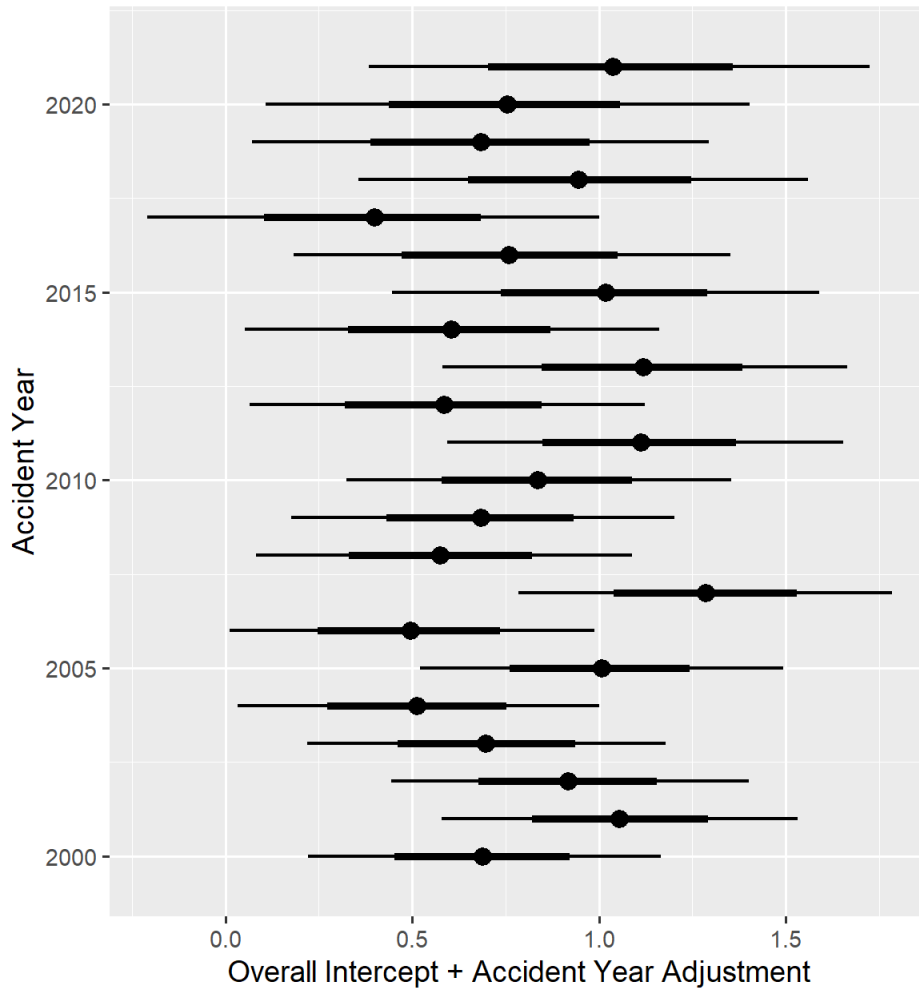
If you look at the graph comparing the observed deflated pure premium to the predicted, the length of the confidence interval bars has also started to grow as well. Using the model without some form of guard rails is becoming questionable.

Case 3 Without Selected Prior Distribution

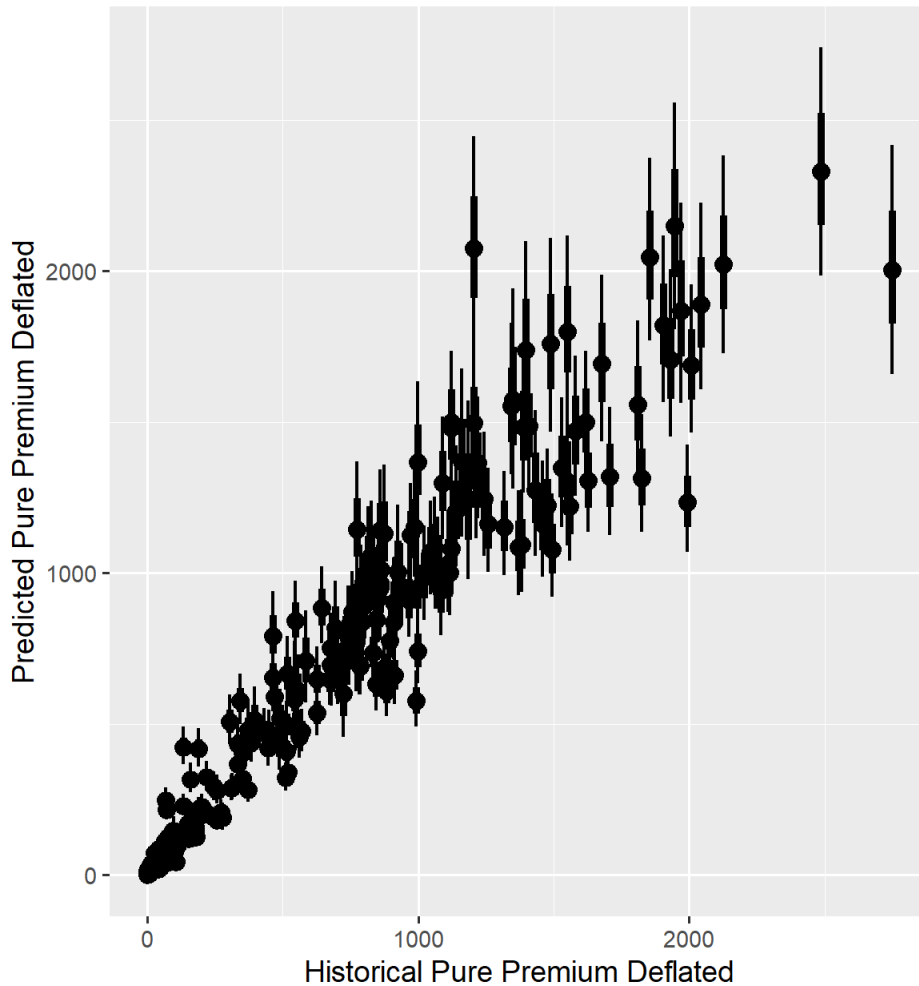
Coefficients	Estimate	Std_Err	LB_CI_95	UB_CI_95
Intercept	0.809	0.261	0.299	1.324
Dev_Yr_2_Factor2	2.204	0.19	1.84	2.579
Dev_Yr_2_Factor3	2.956	0.334	2.316	3.612
Dev_Yr_6_Cap	1.06	0.257	0.552	1.553
Dev_Yr_6_Cap_Sqrd	-0.085	0.028	-0.139	-0.03
Dev_Yr_6_Spline	-0.114	0.026	-0.165	-0.062
Dev_Yr_6_Spline_Sqrd	-0.031	0.003	-0.037	-0.026
Cal_Yr_Time	0.015	0.011	-0.006	0.036
sigma_Intercept	-1.403	0.054	-1.505	-1.291
sigma_Dev_Yr_10_Spline_Ln	0.704	0.062	0.585	0.83

Prior_Distribution	Variable_Class	Coefficient	Group_Indicator	Distribution_Parameter	Source
	b				default
	b	Cal_Yr_Time			default
	b	Dev_Yr_2_Factor2			default
	b	Dev_Yr_2_Factor3			default
	b	Dev_Yr_6_Cap			default
	b	Dev_Yr_6_Cap_Sqrd			default
	b	Dev_Yr_6_Spline			default
	b	Dev_Yr_6_Spline_Sqrd			default
	b	Intercept			default
	b			sigma	default
	b	Dev_Yr_10_Spline_Ln		sigma	default
	b	Intercept		sigma	default
student_t(3, 0, 2.5)	sd				default
	sd		Acc_Yr		default
	sd	Intercept	Acc_Yr		default

Median fitted Adjusted Intercept by Accident Year
 Deflated Incremental Pure Premium Model
 Case 3 Without Selected Prior Distributions



Historical Pure Premium Deflated vs. Distribution of Predicted Case 3 Without Selected Prior Distributions



IV.3 Case 3 Comparison of Reserve Distribution With and Without Selected Priors

Dropping down to 100 exposures in Case 3 from 500 exposures in Case 2 tended to boost the parameter estimates for sigma as the accident year ages.

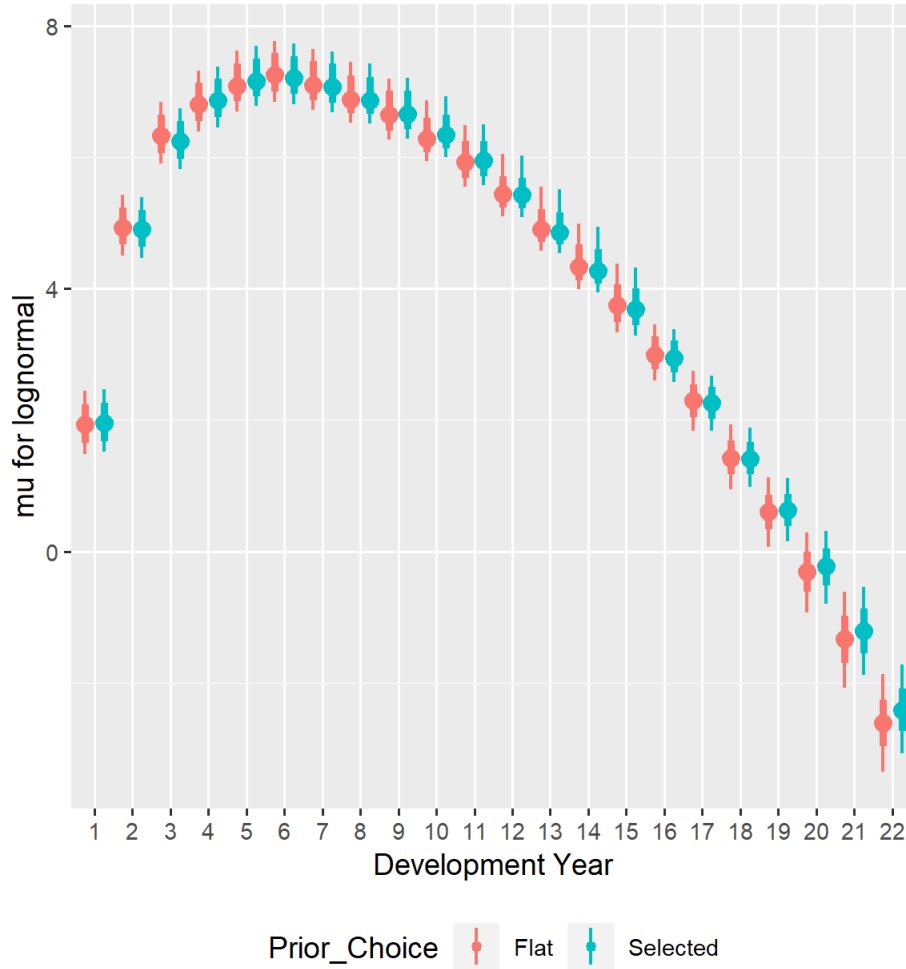
In practice, results from a larger sample could be used to inform prior distribution selections for all of the variables in a smaller sample. Using tighter standard deviation figures in the prior distribution for selected coefficients can compensate for limited information in the loss triangle. However, it is important to recognize that with a smaller amount of information, an increase in the variability of forecast results could be a realistic

expectation. Then too, a given data set may just behave differently, and some caution is warranted in dialing down the standard deviation to avoid overriding information in the data set.

Estimated mu vs. Development Year

Observed Data

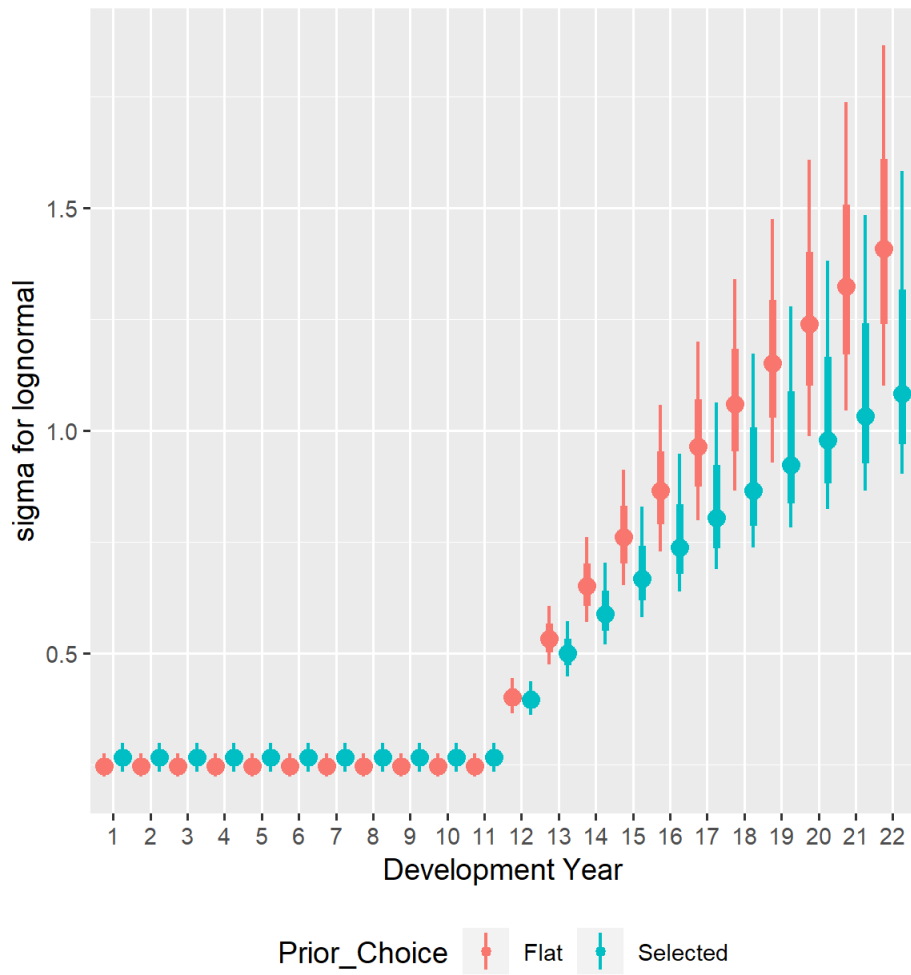
Case 3 by Prior Selection



Estimated sigma vs. Development Year

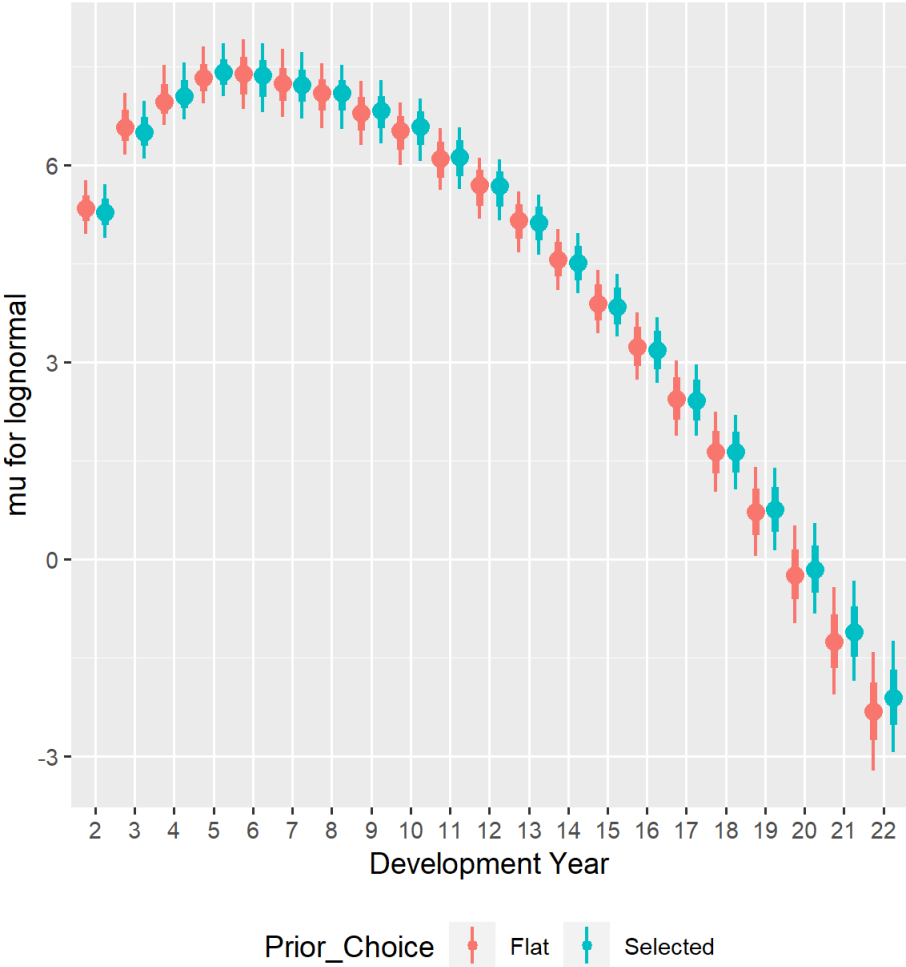
Observed Data

Case 3 by Prior Selection

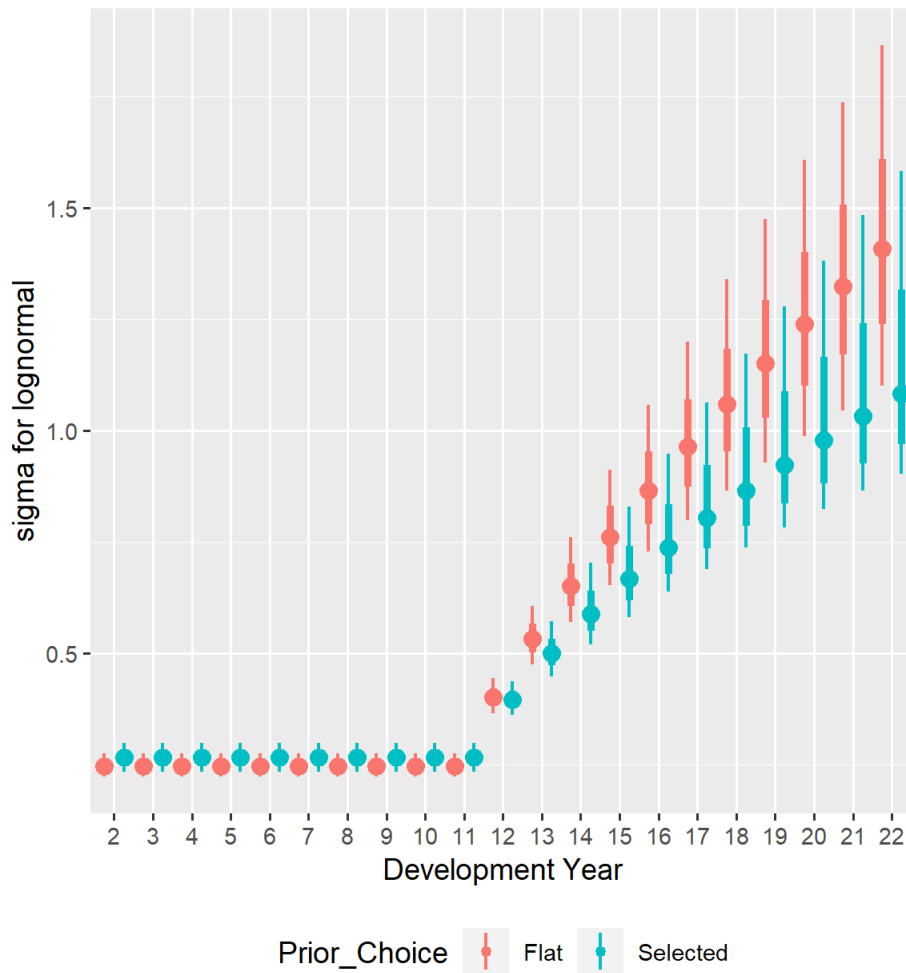


Forecast mu vs. Development Year Future Accident & Development Years

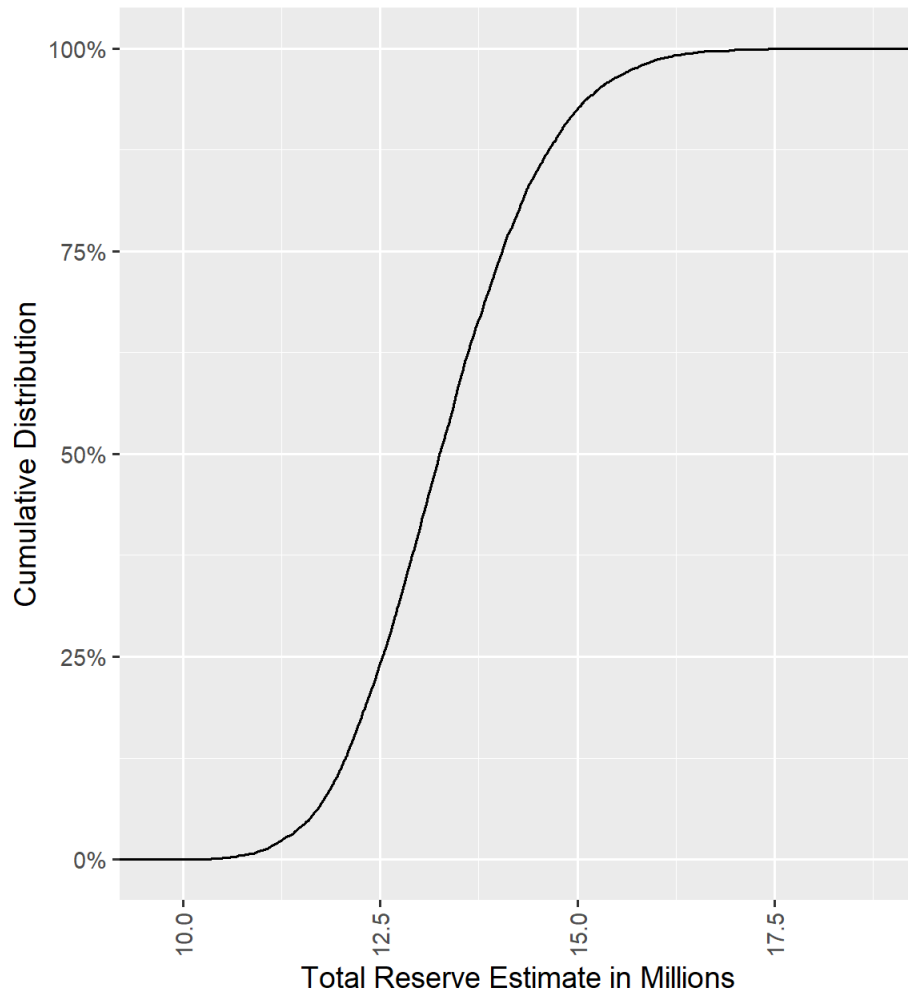
Case 3 by Prior Selection



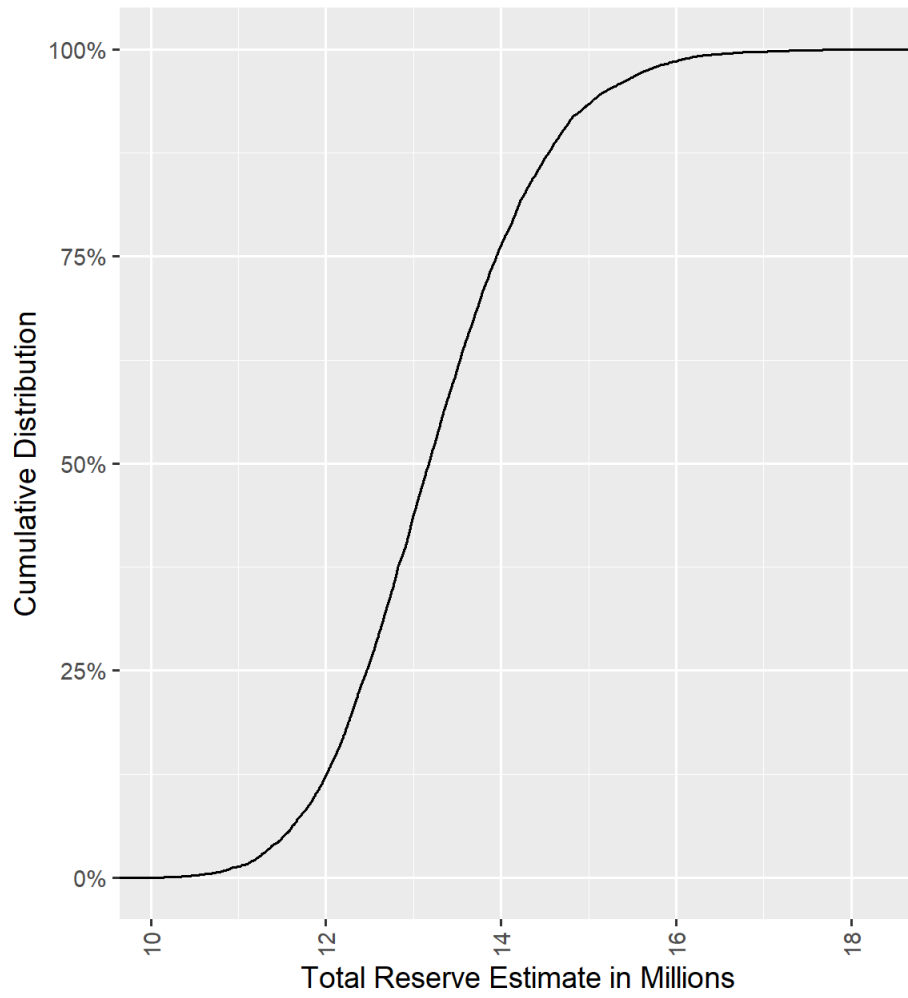
Forecast sigma vs. Development Year Future Accident & Development Years Case 3 by Prior Selection



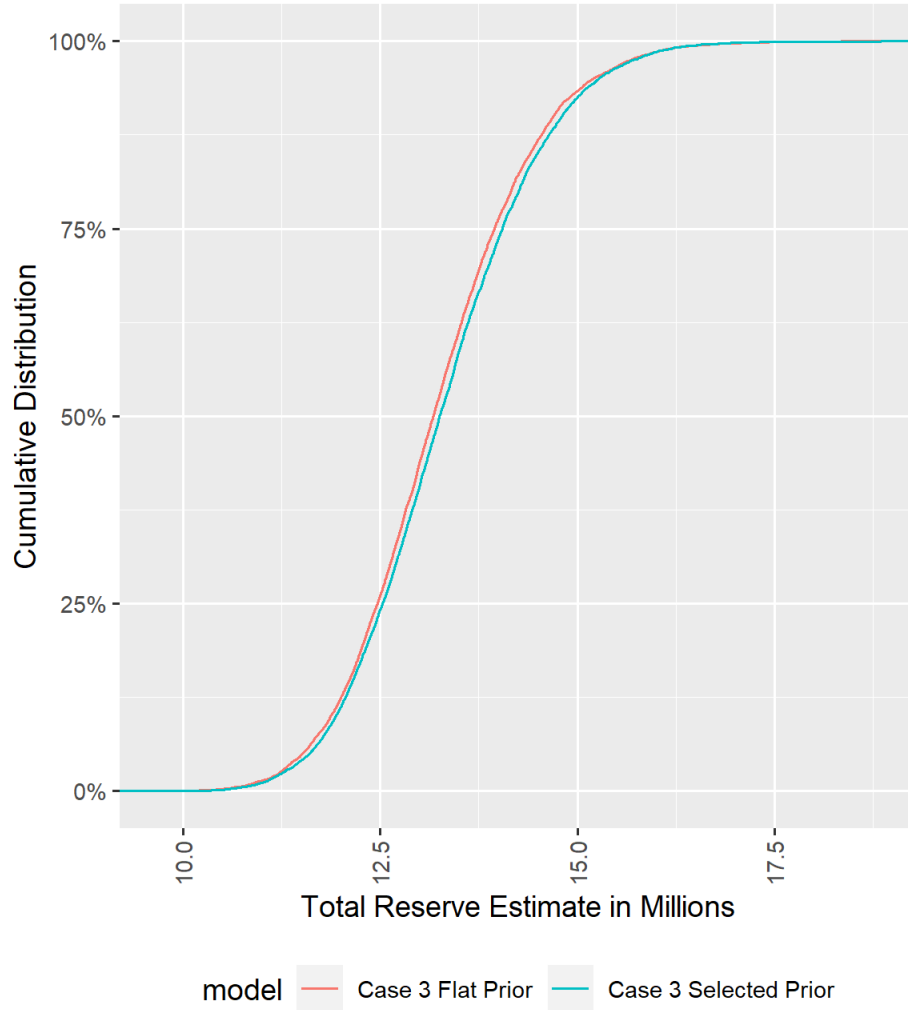
Total Reserve Estimate
Case 3 With Selected Prior Distribution
Simulated Future Inflation



Total Reserve Estimate
Case 3 Without Selected Prior Distribution
Simulated Future Inflation

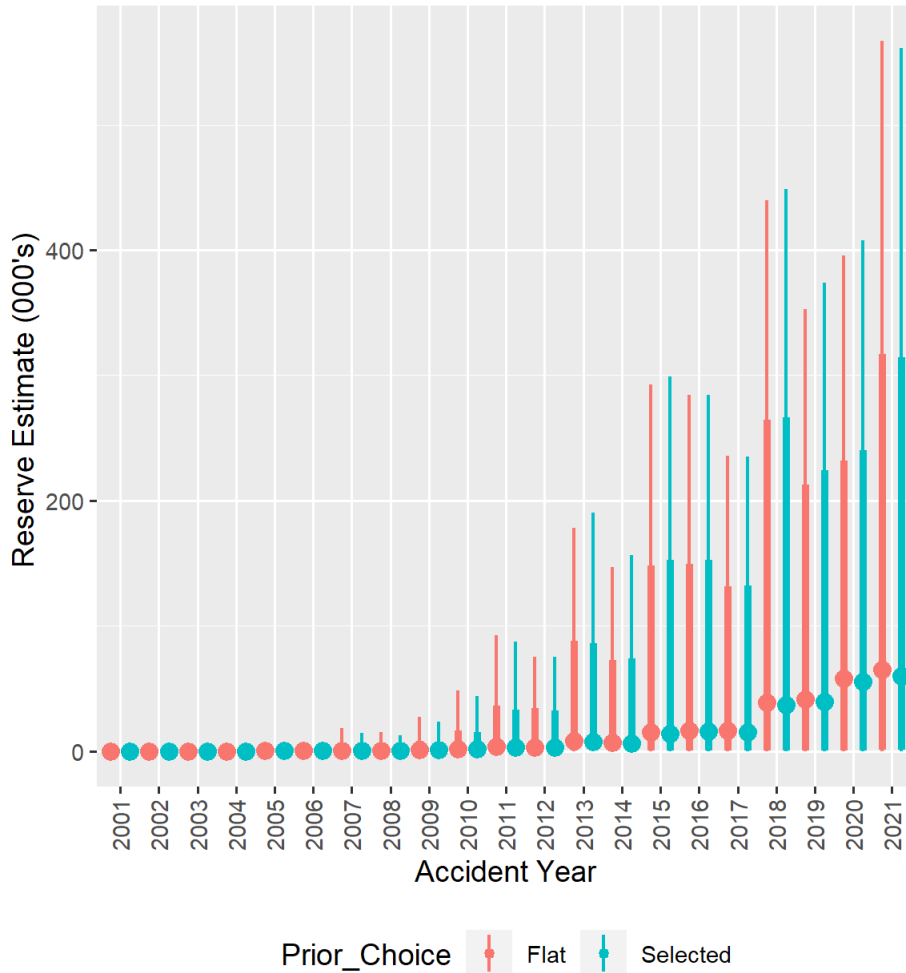


Total Reserve Estimate Case 3 Models With and Without Selected Prior



Accident Year Ultimate Reserve Estimates

Case 3 by Prior Selection



V. Case 4 Modeling Results

Case 4 has 22 accident years (2000 through 2021) with a uniform reported claim count of 50 reported claims at 12 months for each accident year and a maximum of 22 years of development time.

Case 4 is included to further explore how reducing the exposure amounts within accident years impacts the value of using selected prior distributions as the complement of credibility. This continuation of the experiment will help illustrate the effect of decreasing information in the loss triangle on the model's accuracy and reliability.

V.1 Case 4 Results With Prior Distribution Selected

The prior distribution for Case 4 was based on the results from Case 1 due to the lower exposure levels in Case 4. In this particular situation, it was necessary to set one of the parameters for sigma, specifically Dev_Yr_10_Spline_Ln, to a constant of 0.57 which is the mean value from the Case 1 results. This adjustment was required because the information for sigma beyond ten years of development was too limited to model any change in sigma successfully with this reduced exposure (50 reported claims per accident year).

Given the relatively small number of exposures for Case 4, the prior distribution was adjusted to reflect the results from Case 1, with the standard deviation for the coefficients in the prior distribution set to a smaller number. This adjustment places less weight on the results from the data and more on the actuary's prior estimates of the behavior of the development curve.

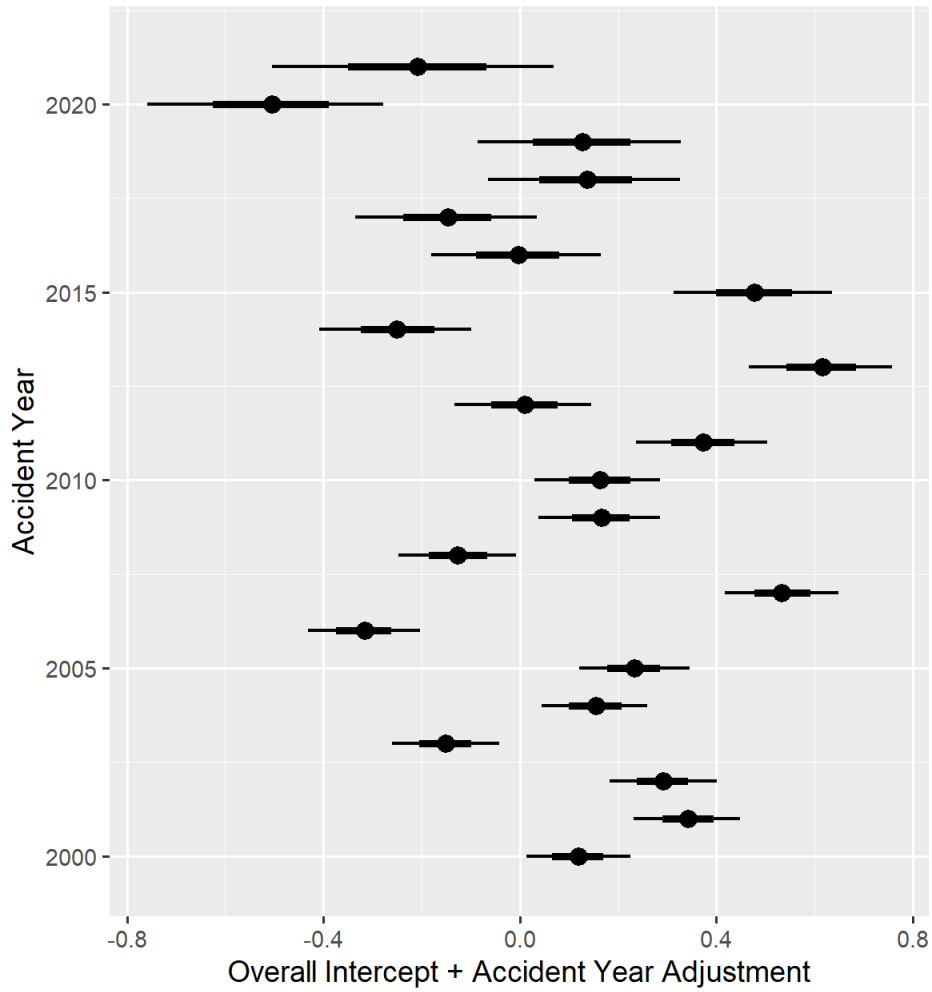
Please note that even though the reported claim count dropped from 100 in Case 3 to 50 in Case 4, the standard errors around the modeled parameters decreased with the use of tighter standard errors in the selected prior distributions. This made it practical to retain the Dev_Yr_10_Spline variable in the regression equation for mu.

Case 4 With Selected Prior Distribution

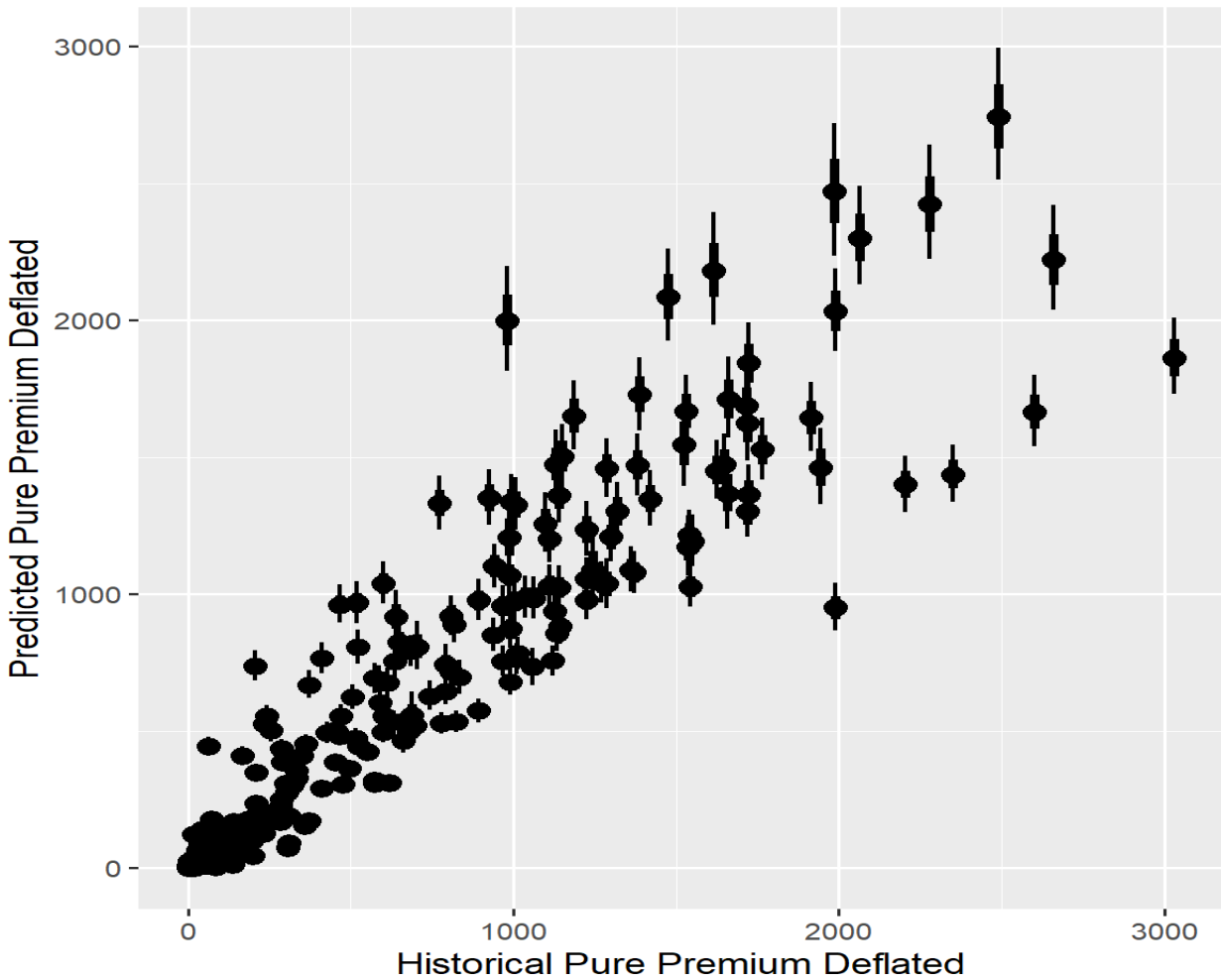
Coefficients	Estimate	Std_Err	LB_CI_95	UB_CI_95
Intercept	0.235	0.045	0.148	0.324
Dev_Yr_2_Factor2	1.794	0.031	1.734	1.853
Dev_Yr_2_Factor3	2.212	0.038	2.138	2.286
Dev_Yr_6_Cap	1.63	0.032	1.568	1.692
Dev_Yr_6_Cap_Sqrd	-0.138	0.004	-0.146	-0.131
Dev_Yr_6_Spline	-0.203	0.007	-0.217	-0.188
Dev_Yr_6_Spline_Sqrd	-0.024	0.001	-0.026	-0.023
Cal_Yr_Time	0.016	0.004	0.008	0.024
Dev_Yr_10_Spline	-0.146	0.005	-0.156	-0.136
sigma_Intercept	-2.068	0.007	-2.082	-2.054
sigma_Dev_Yr_10_Spline_Ln	0.57	0	0.57	0.57

Prior_Distribution	Variable_Class	Coefficient	Group_Indicator	Distribution_Parameter	Source
	b				default
normal(0.02, 0.005)	b	Cal_Yr_Time			user
normal(-0.15, 0.005)	b	Dev_Yr_10_Spline			user
normal(1.7, 0.05)	b	Dev_Yr_2_Factor2			user
normal(2.2, 0.05)	b	Dev_Yr_2_Factor3			user
normal(1.6, 0.05)	b	Dev_Yr_6_Cap			user
normal(-0.15, 0.05)	b	Dev_Yr_6_Cap_Sqrd			user
normal(-0.22, 0.01)	b	Dev_Yr_6_Spline			user
normal(-0.01, 0.005)	b	Dev_Yr_6_Spline_Sqrd			user
normal(0.3, 0.05)	b	Intercept			user
	b			sigma	default
constant(0.57)	b	Dev_Yr_10_Spline_Ln		sigma	user
normal(-2.5, 0.01)	b	Intercept		sigma	user
student_t(3, 0, 2.5)	sd				default
	sd		Acc_Yr		default
	sd	Intercept	Acc_Yr		default

Median fitted Adjusted Intercept by Accident Year
Deflated Incremental Pure Premium Model
Case 4 With Selected Prior Distributions



Historical Pure Premium Deflated vs. Distribution of Predicted Case 4 With Selected Prior Distributions



V.2 Case 4 Results With Flat Prior

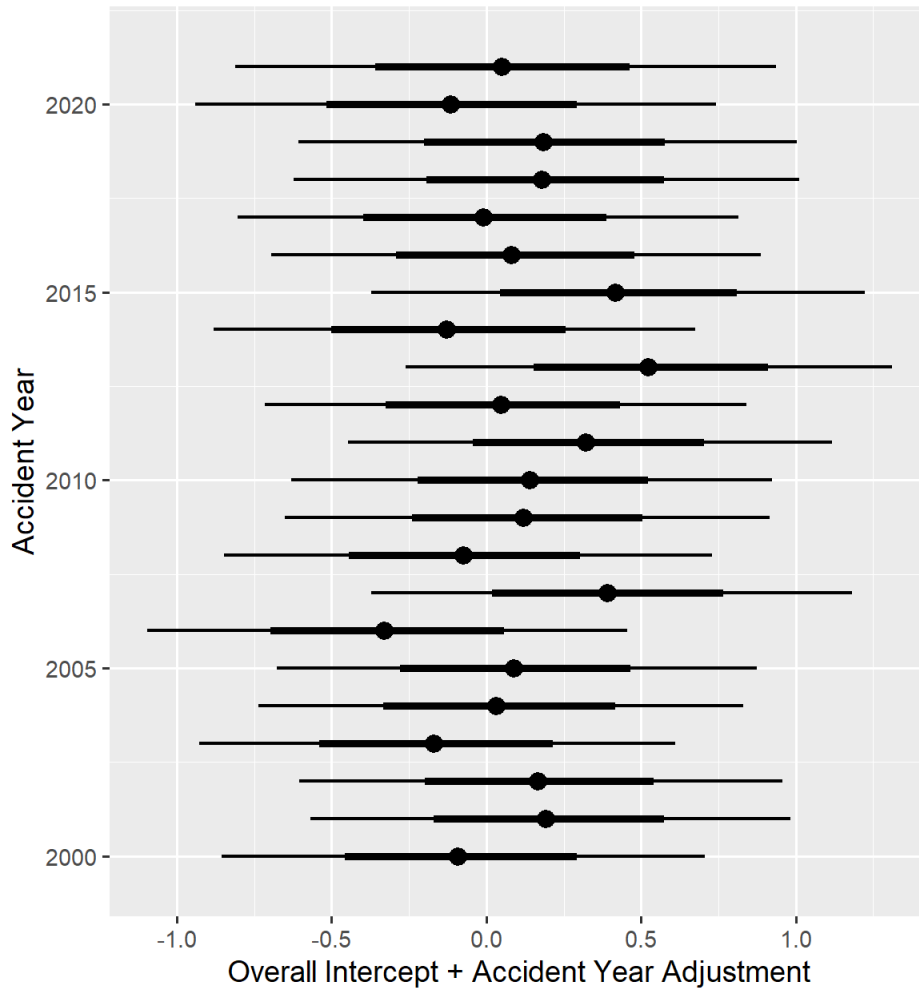
It was necessary to drop the Cal_Yr_Time and Dev_Yr_10_Spline variable from the model to obtain sensible coefficient results. The remaining coefficients now exhibit higher standard errors compared to earlier models, and the coefficient estimate for sigma has increased notably. Unfortunately, this adjustment has resulted in the graph comparing actual to modeled deflated pure premium values being distorted to such an extent that it is no longer informative.

Case 4 Without Selected Prior Distribution

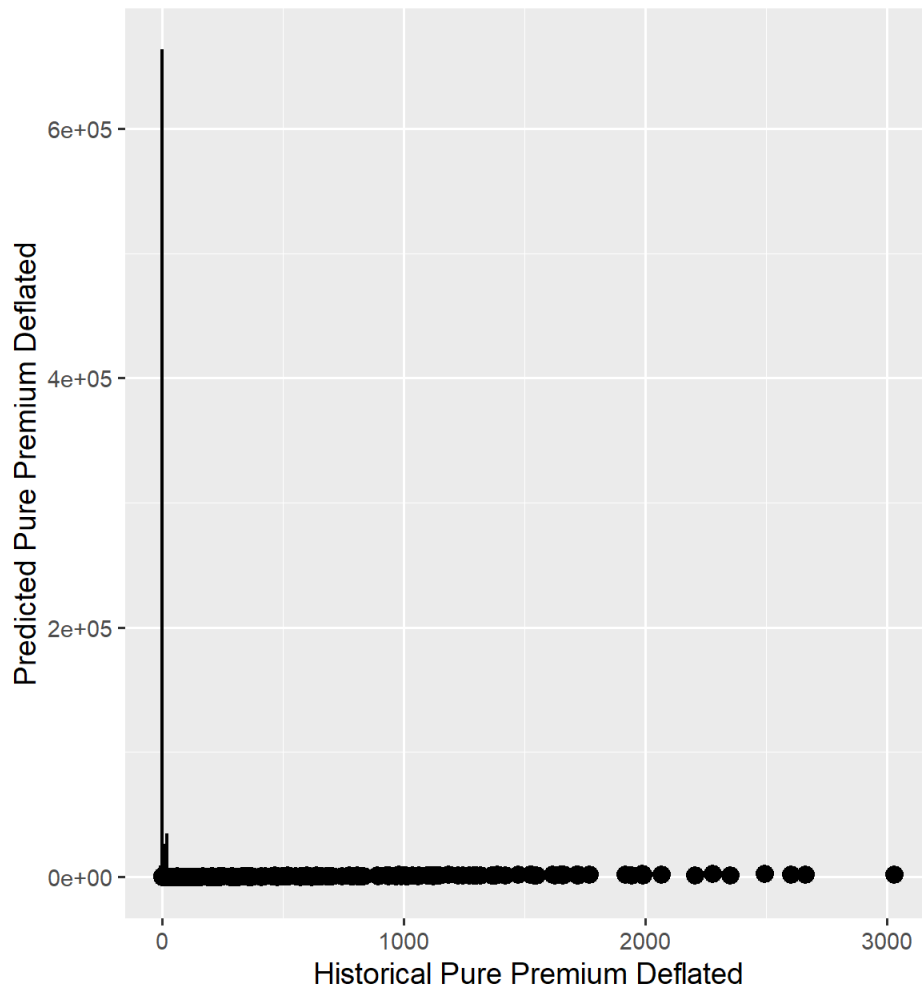
Coefficients	Estimate	Std_Err	LB_CI_95	UB_CI_95
Intercept	0.097	0.382	-0.645	0.854
Dev_Yr_2_Factor2	1.807	0.306	1.212	2.414
Dev_Yr_2_Factor3	2.164	0.533	1.126	3.216
Dev_Yr_6_Cap	1.754	0.413	0.927	2.547
Dev_Yr_6_Cap_Sqrd	-0.152	0.045	-0.241	-0.062
Dev_Yr_6_Spline	-0.147	0.048	-0.241	-0.052
Dev_Yr_6_Spline_Sqrd	-0.036	0.006	-0.049	-0.024
sigma_Intercept	-0.864	0.054	-0.969	-0.753
sigma_Dev_Yr_10_Spline_Ln	0.96	0.058	0.847	1.076

Prior_Distribution	Variable_Class	Coefficient	Group_Indicator	Distribution_Parameter	Source
	b				default
	b	Dev_Yr_2_Factor2			default
	b	Dev_Yr_2_Factor3			default
	b	Dev_Yr_6_Cap			default
	b	Dev_Yr_6_Cap_Sqrd			default
	b	Dev_Yr_6_Spline			default
	b	Dev_Yr_6_Spline_Sqrd			default
	b	Intercept			default
	b			sigma	default
	b	Dev_Yr_10_Spline_Ln		sigma	default
	b	Intercept		sigma	default
student_t(3, 0, 2.5)	sd				default
	sd		Acc_Yr		default
	sd	Intercept	Acc_Yr		default

Median fitted Adjusted Intercept by Accident Year
Deflated Incremental Pure Premium Model
Case 4 Without Selected Prior Distributions



Historical Pure Premium Deflated vs. Distribution of Predicted Case 4 Without Selected Prior Distributions



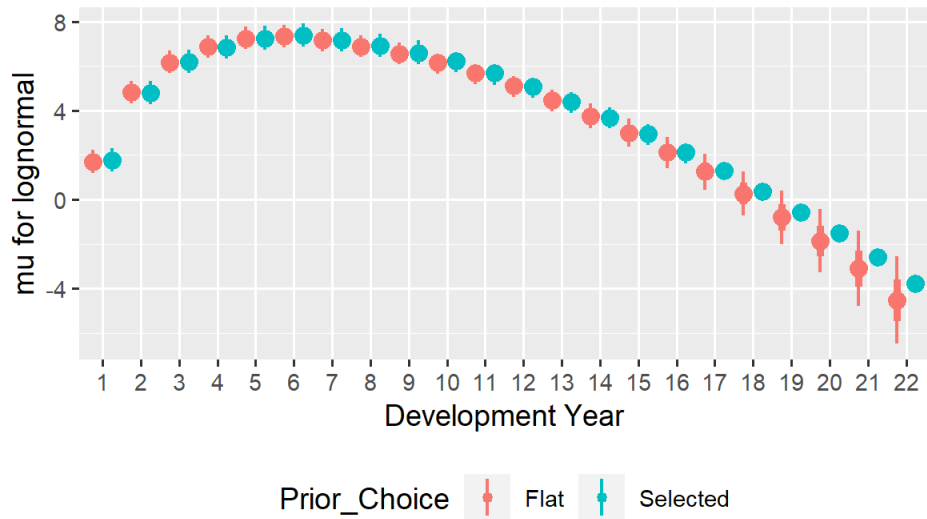
V.3 Case 4 Comparison of Results With and Without Prior Distribution Selected

The reserve estimates for Case 4 without selected priors can become excessively large due to a significant increase in the sigma value in the tail of the accident year development period. Additionally, as the exposure amount decreases, the variance around the parameter estimates increases, leading to instability in the estimates.

For Case 4, the graph overlaying the distribution with a uniform prior and selected prior distribution values is not informative because the values for the flat prior distribution became excessively large in some cases. However, it was retained to maintain consistency in the sequence of information displayed across Cases 1 to 5

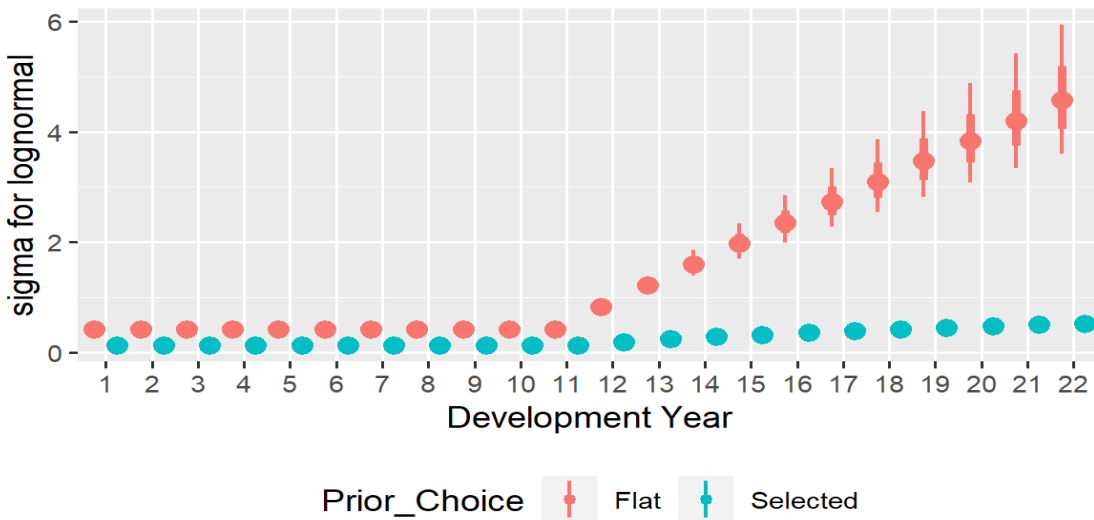
Estimated mu vs. Development Year Observed Data

Case 4 by Prior Selection



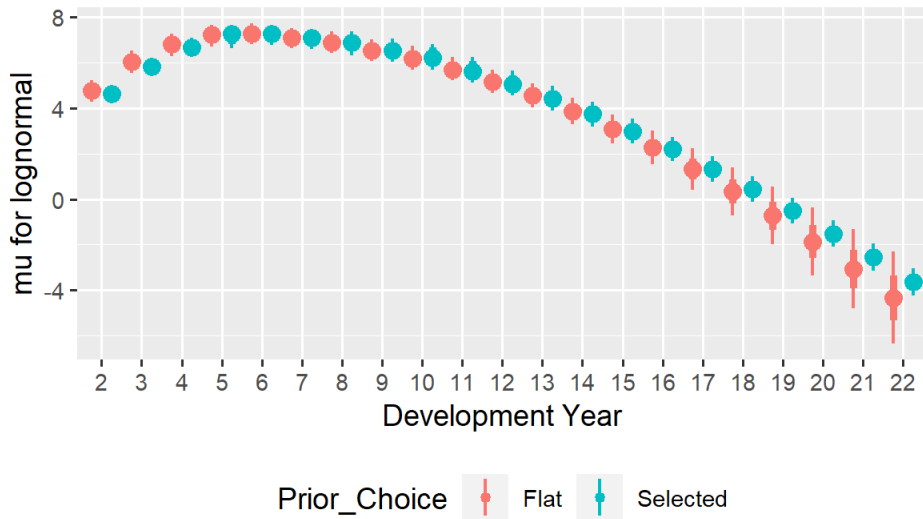
Estimated sigma vs. Development Year Observed Data

Case 4 by Prior Selection



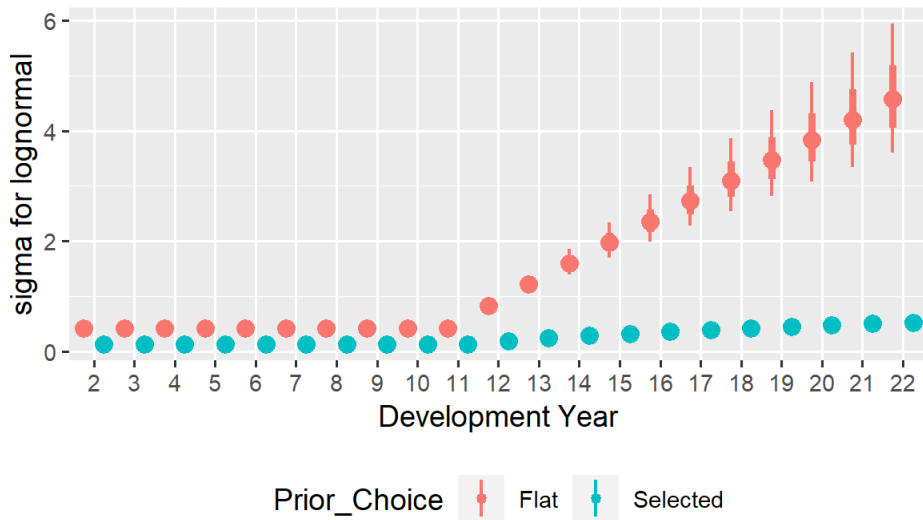
Forecast mu vs. Development Year Future Accident & Development Years

Case 4 by Prior Selection

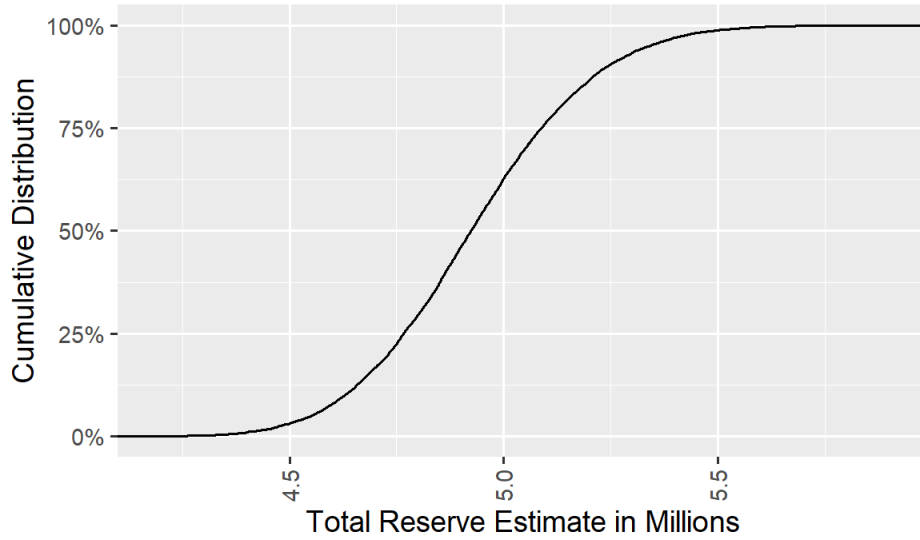


Forecast sigma vs. Development Year Future Accident & Development Years

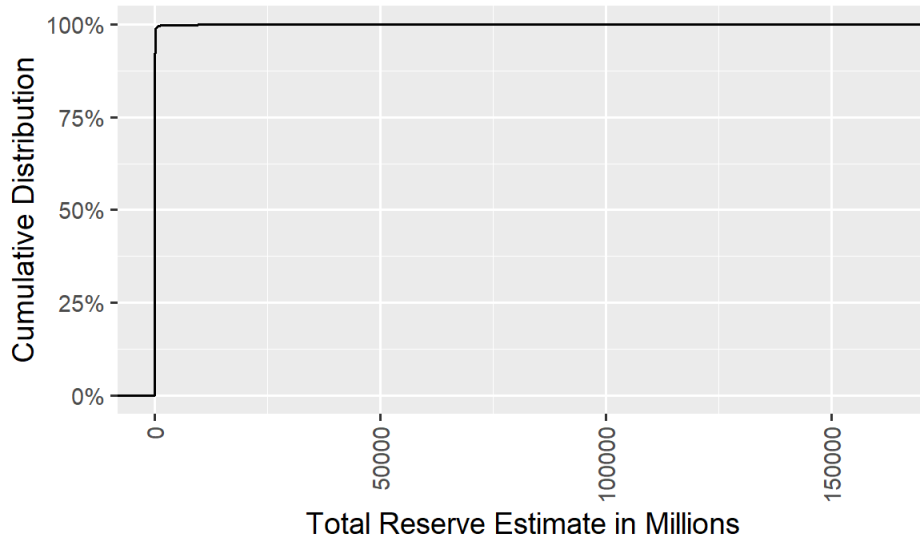
Case 4 by Prior Selection



Total Reserve Estimate
Case 4 With Selected Prior Distribution
Simulated Future Inflation



Total Reserve Estimate
Case 4 Without Selected Prior Distribution
Simulated Future Inflation



Accident Year Ultimate Reserve Estimates

Case 4 by Prior Selection



VI. Case 5 Modeling Results

Case 5 has 11 accident years (2011 through 2021) with a uniform reported claim count of 100 reported claims at 12 months for each accident year and we assume that incremental losses will continue out to a maximum of 22 years of development. We only have 11 years of development history though for the oldest accident year.

VI.1 Case 5 Model Results for Selected Prior Distribution

In this case, a lack of development information beyond 10 years necessitates supplementing the model with the parameter estimates from a larger data set, as demonstrated in Case 1, to obtain plausible reserve estimates. Specifically, due to the absence of development information after ten years, the distribution around the estimated parameters for the period beyond six years of development was constrained to be very small within the prior selection.

By comparing the parameter results from the summary below to the prior distributions used, it is evident that incorporating the information from the loss triangle did shift the

mean estimate, even with limited data for coefficients related to early development periods. This demonstrates that when a prior distribution is used as a compliment of credibility in a Bayesian MCMC model, the information within the loss triangle can still impact the results. The prior distribution effectively places guardrails around the estimation process for posterior distribution parameters for coefficients related to later development years.

The graph showing the historical normalized incremental losses (pure premium deflated) compared to the modeled results, along with residuals graphs, indicates that the model provides a reasonable fit to the observations in the loss triangle.

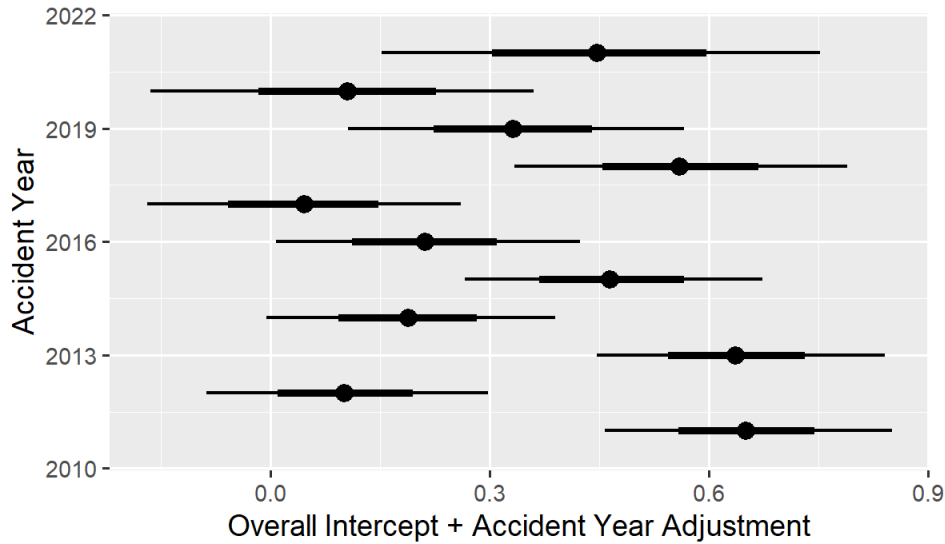
Case 5 With Selected Prior Distribution

Coefficients	Estimate	Std_Err	LB_CI_95	UB_CI_95
Intercept	0.311	0.048	0.216	0.404
Dev_Yr_2_Factor2	1.77	0.106	1.564	1.977
Dev_Yr_2_Factor3	2.392	0.162	2.077	2.714
Dev_Yr_6_Cap	1.504	0.112	1.284	1.721
Dev_Yr_6_Cap_Sqrd	-0.133	0.012	-0.157	-0.109
Dev_Yr_6_Spline	-0.215	0.009	-0.234	-0.197
Dev_Yr_6_Spline_Sqrd	-0.005	0.004	-0.012	0.002
Cal_Yr_Time	0.022	0.004	0.014	0.031
Dev_Yr_10_Spline	-0.15	0.005	-0.159	-0.14
sigma_Intercept	-1.903	0.056	-2.011	-1.793
sigma_Dev_Yr_10_Spline_Ln	0.57	0.019	0.53	0.607

Prior_Distribution	Variable_Class	Coefficient	Group_Indicator	Distribution_Parameter	Source
	b				default
normal(0.02, 0.005)	b	Cal_Yr_Time			user
normal(-0.15, 0.005)	b	Dev_Yr_10_Spline			user
normal(1.7, 0.5)	b	Dev_Yr_2_Factor2			user
normal(2.2, 0.5)	b	Dev_Yr_2_Factor3			user
normal(1.6, 0.5)	b	Dev_Yr_6_Cap			user
normal(-0.15, 0.5)	b	Dev_Yr_6_Cap_Sqrd			user
normal(-0.22, 0.01)	b	Dev_Yr_6_Spline			user
normal(-0.01, 0.005)	b	Dev_Yr_6_Spline_Sqrd			user
normal(0.3, 0.05)	b	Intercept			user
	b			sigma	default
student_t(2, 0.57, 0.01)	b	Dev_Yr_10_Spline_Ln		sigma	user
normal(-2.5, 0.1)	b	Intercept		sigma	user
student_t(3, 0, 2.5)	sd				default
	sd		Acc_Yr		default
	sd	Intercept	Acc_Yr		default

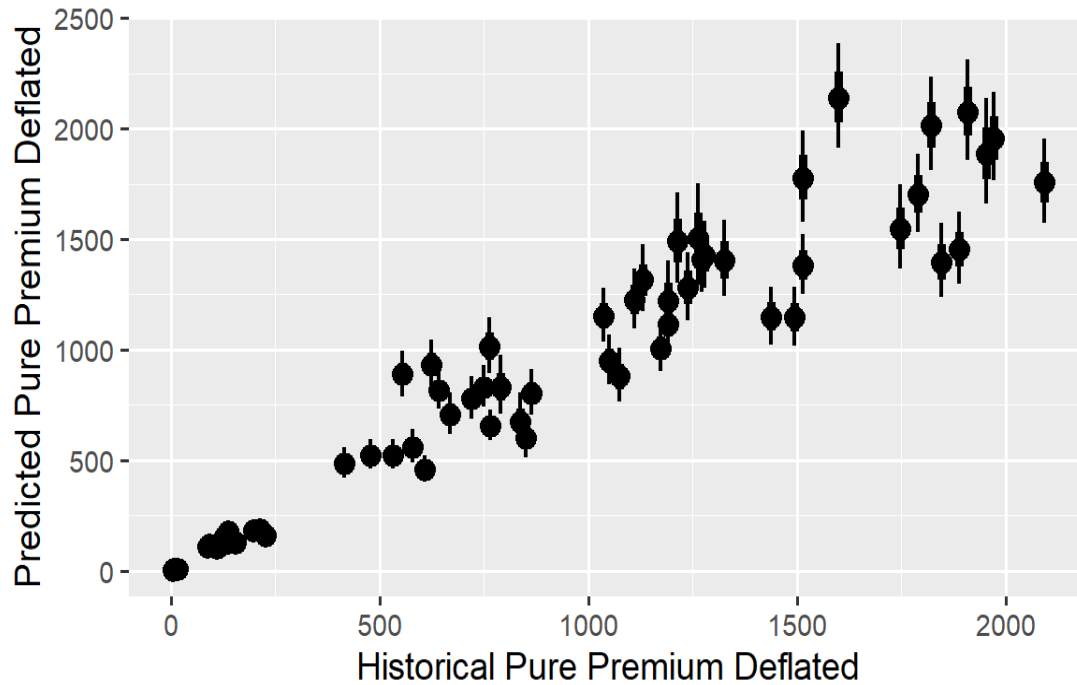
Median fitted Adjusted Intercept by Accident Year Deflated Incremental Pure Premium Model

Case 5 With Selected Prior Distributions



Historical Pure Premium Deflated vs. Distribution of Predicted

Case 5 With Selected Prior Distributions



VI.2 Case 5 With Flat Prior

Due to the lack of information beyond ten years of development, it was necessary to exclude the parameters for the ten-year development spline explanatory variables for both μ and σ .

If you examine the confidence interval results below, you'll notice that for development years beyond six, the confidence interval includes zero. This suggests that using those variables becomes questionable. This poses a challenge because forecasting incremental payments out to 22 years of development is essential for squaring the triangle.

On a positive note, the fit for the observed data is reasonable, indicating that choosing to tighten the standard errors for a subset of the selected prior distributions could be an effective approach. You may have enough information for early development periods to allow the results from the data set to influence the fit, but you can tighten up the standard deviation for later time periods where there is limited or no data. That approach allows one to utilize the data at hand to the extent it is reliable and supplement the lack of data with outside knowledge (actuarial experience) to obtain a plausible reserve distribution forecast.

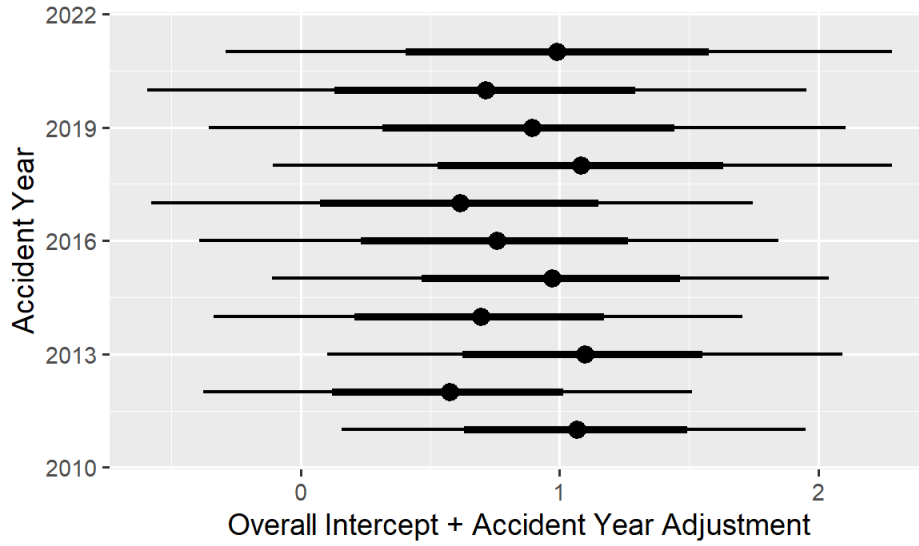
Case 5 Without Selected Prior Distribution

Coefficients	Estimate	Std_Err	LB_CI_95	UB_CI_95
Intercept	0.855	0.553	-0.244	1.927
Dev_Yr_2_Factor2	1.951	0.26	1.434	2.458
Dev_Yr_2_Factor3	2.751	0.452	1.844	3.642
Dev_Yr_6_Cap	1.239	0.361	0.538	1.959
Dev_Yr_6_Cap_Sqrd	-0.104	0.04	-0.183	-0.026
Dev_Yr_6_Spline	-0.183	0.103	-0.384	0.016
Dev_Yr_6_Spline_Sqrd	-0.007	0.022	-0.05	0.036
Cal_Yr_Time	0.005	0.03	-0.053	0.066
sigma_Intercept	-1.515	0.102	-1.708	-1.31

Prior_Distribution	Variable_Class	Coefficient	Group_Indicator	Distribution_Parameter	Source
	b				default
	b	Cal_Yr_Time			default
	b	Dev_Yr_2_Factor2			default
	b	Dev_Yr_2_Factor3			default
	b	Dev_Yr_6_Cap			default
	b	Dev_Yr_6_Cap_Sqrd			default
	b	Dev_Yr_6_Spline			default
	b	Dev_Yr_6_Spline_Sqrd			default
	b	Intercept			default
	b			sigma	default
	b	Intercept		sigma	default
student_t(3, 0, 2.5)	sd				default
	sd		Acc_Yr		default
	sd	Intercept	Acc_Yr		default

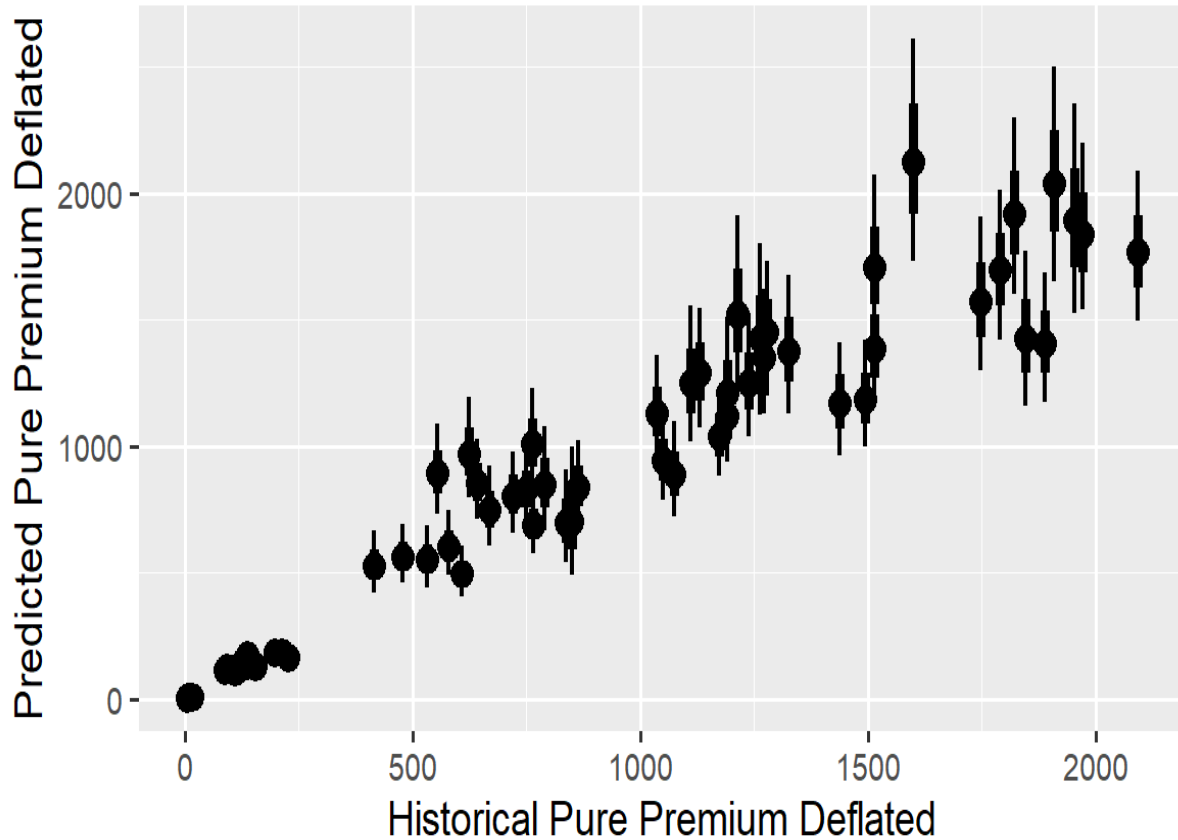
Median fitted Adjusted Intercept by Accident Year Deflated Incremental Pure Premium Model

Case 5 Without Selected Prior Distributions



Historical Pure Premium Deflated vs. Distribution of Predicted

Case 5 Without Selected Prior Distributions

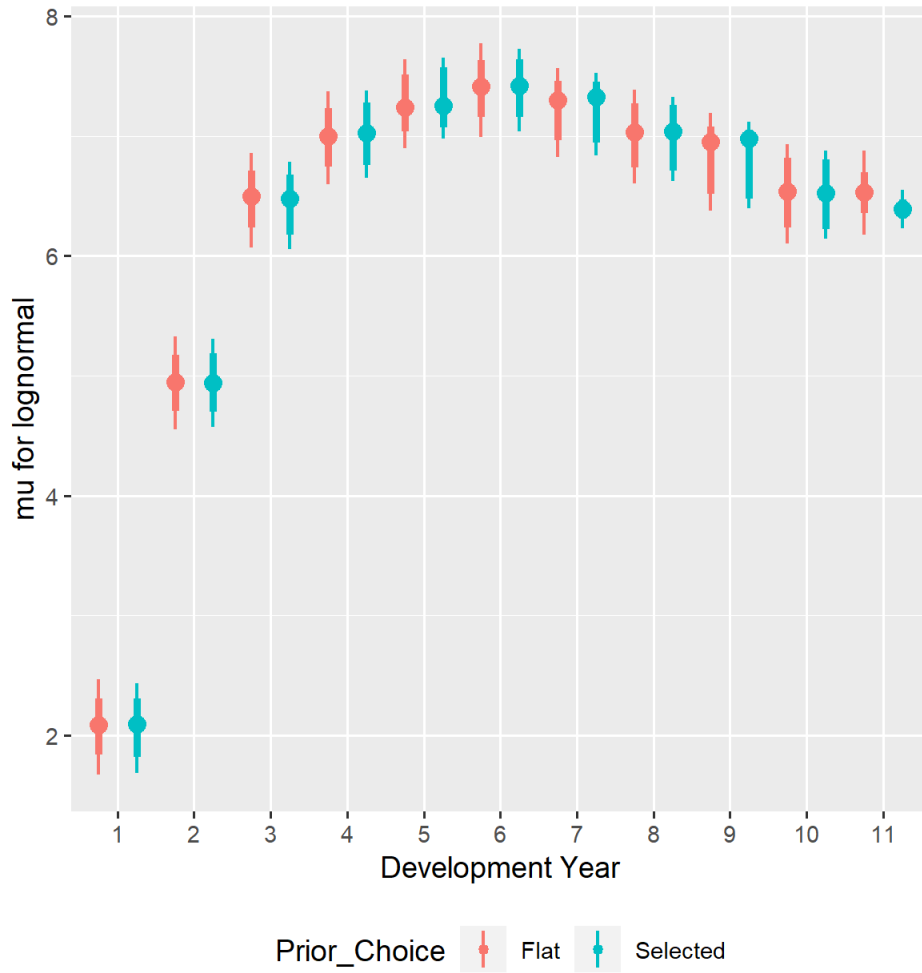


VI.3 Case 5 Comparison of With and Without Selected Prior Distributions

With no information for development years beyond ten years, reserve estimates become unstable when projecting incremental payments from eleven years to twenty-two years without a selected prior distribution. Additionally, sigma tends to increase gradually after ten years of development, but this trend cannot be accurately reflected in modeling without the ability to select prior distributions for the explanatory variables beyond ten years of development.

Estimated mu vs. Development Year Observed Data

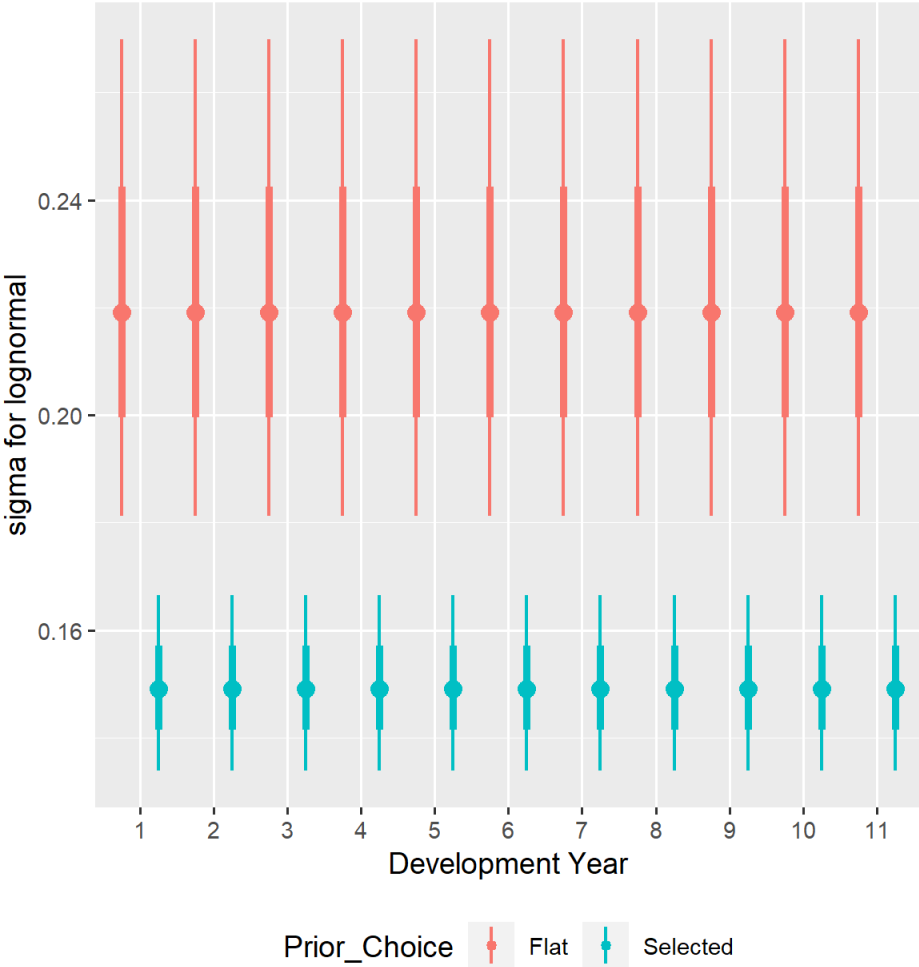
Case 5 by Prior Selection



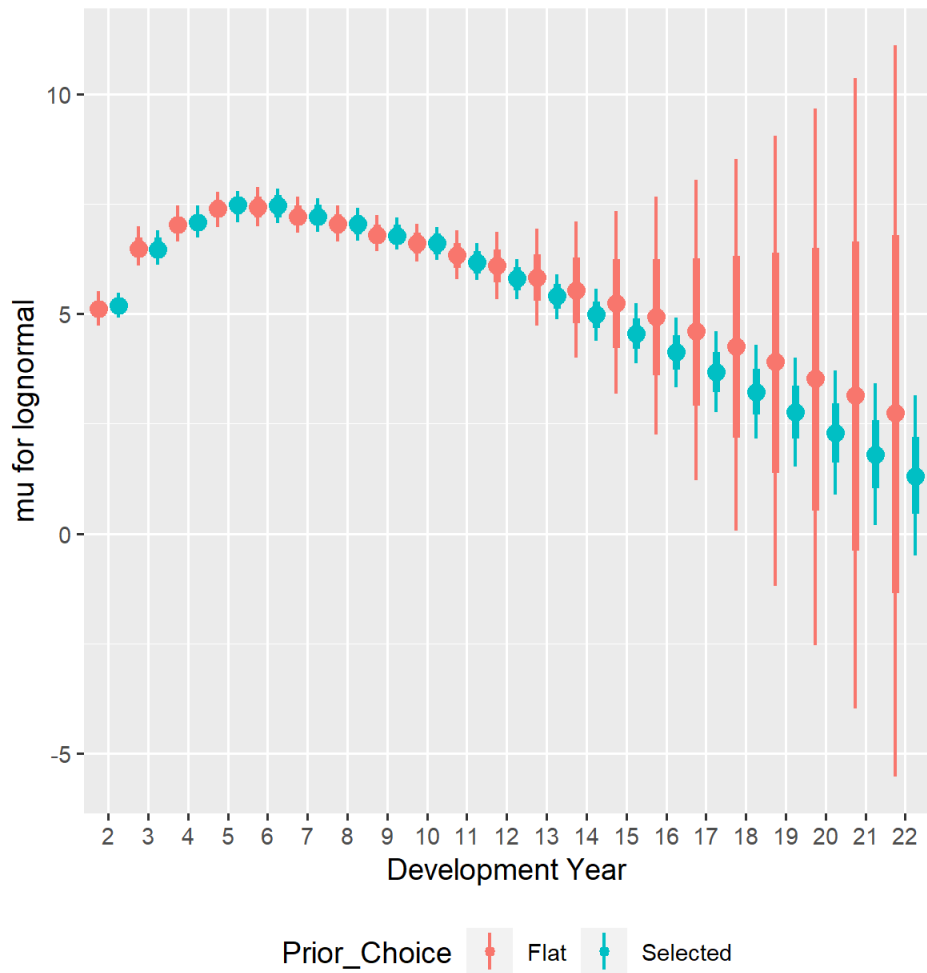
Estimated sigma vs. Development Year

Observed Data

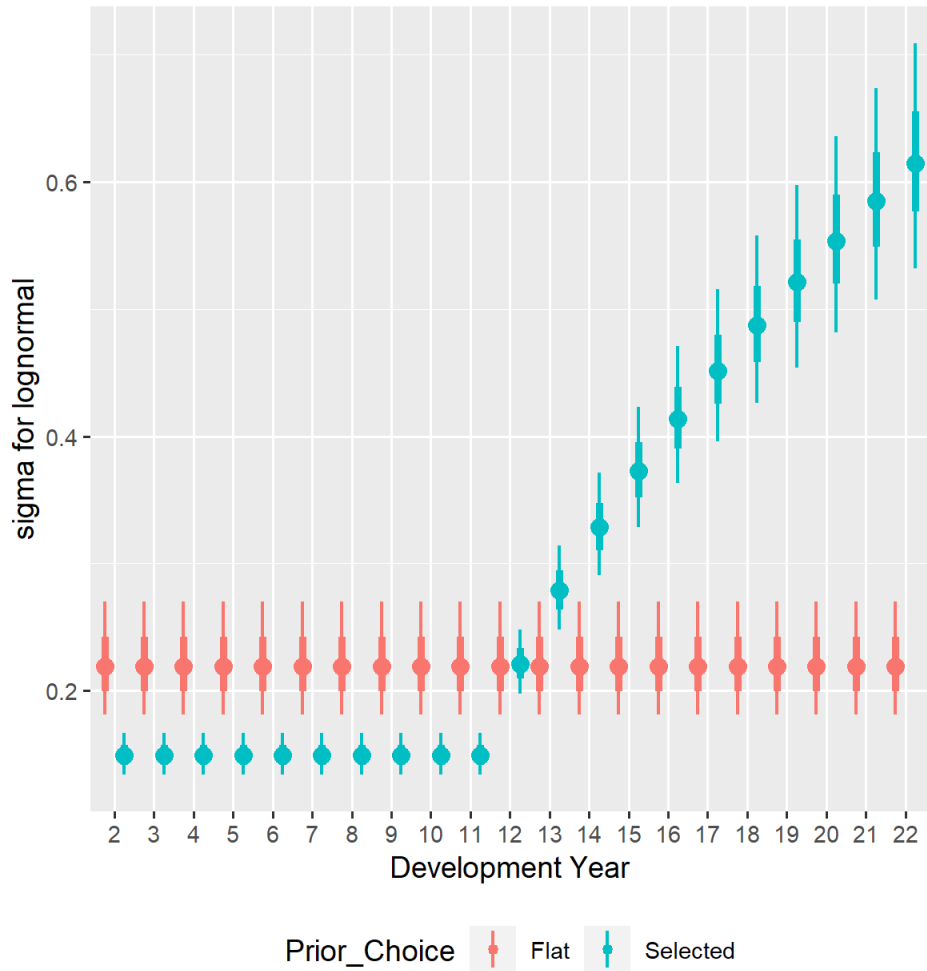
Case 5 by Prior Selection



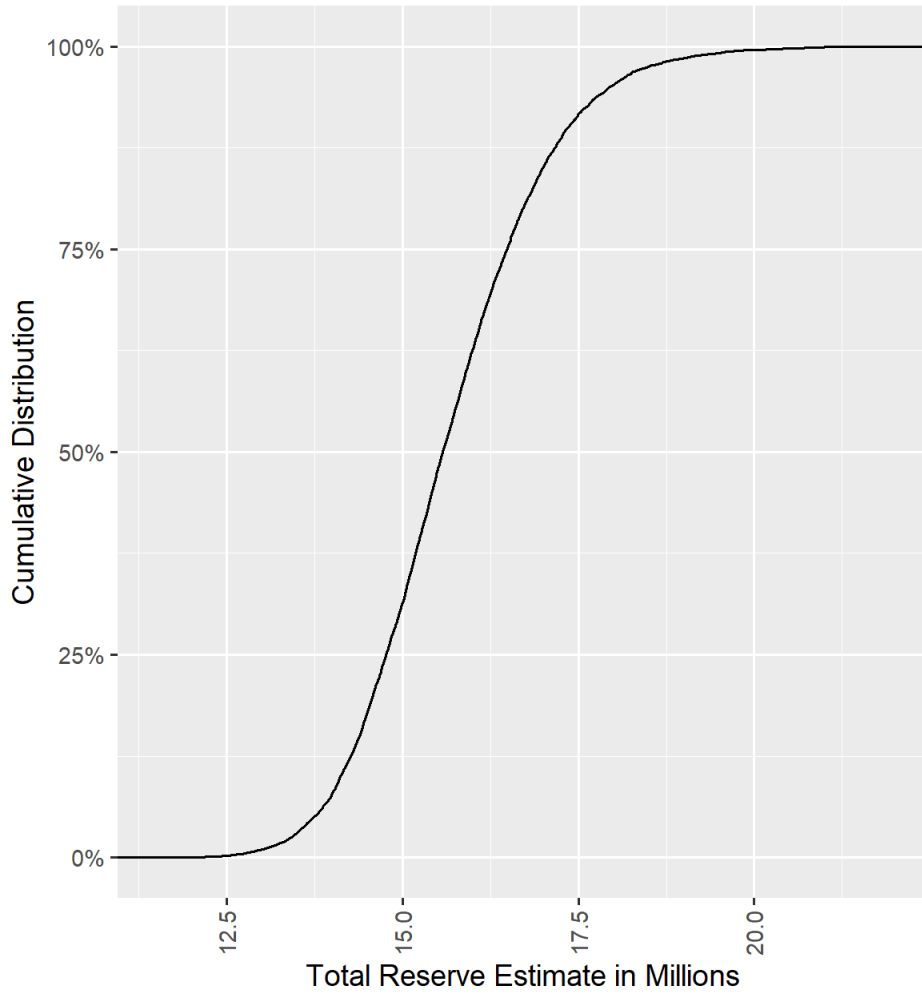
Forecast mu vs. Development Year Future Accident & Development Years Case 5 by Prior Selection



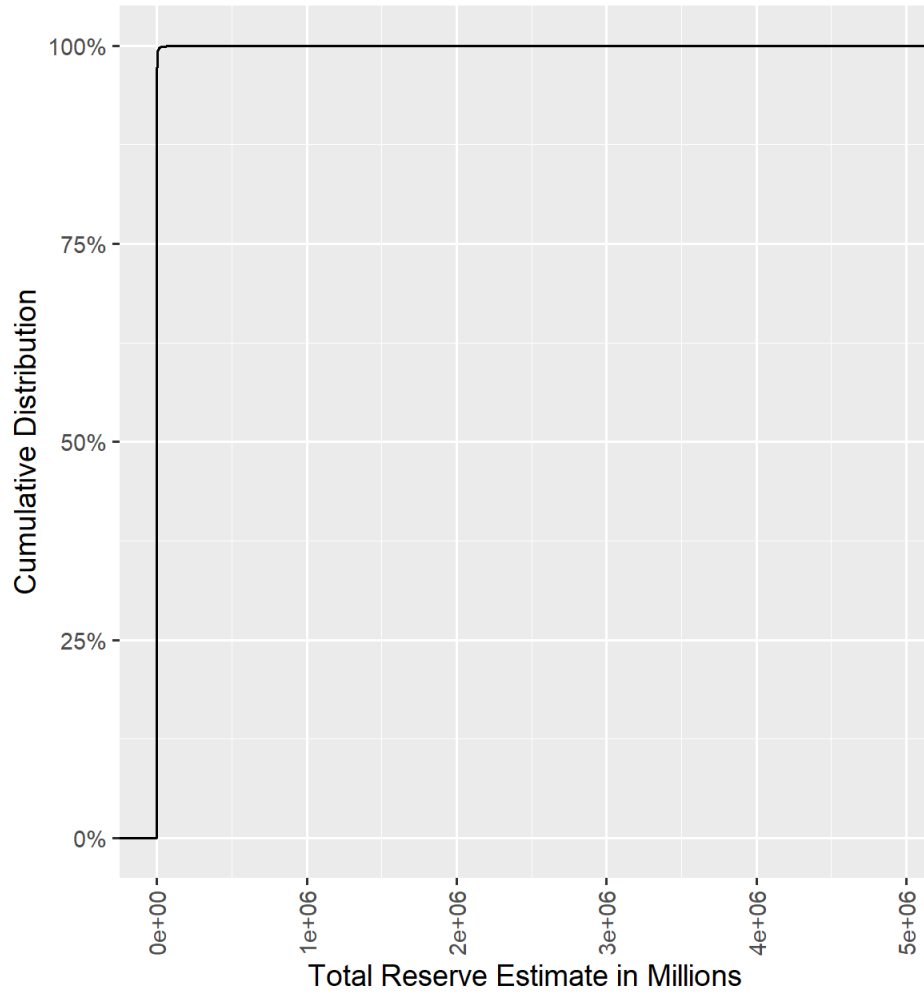
Forecast sigma vs. Development Year
Future Accident & Development Years
Case 5 by Prior Selection



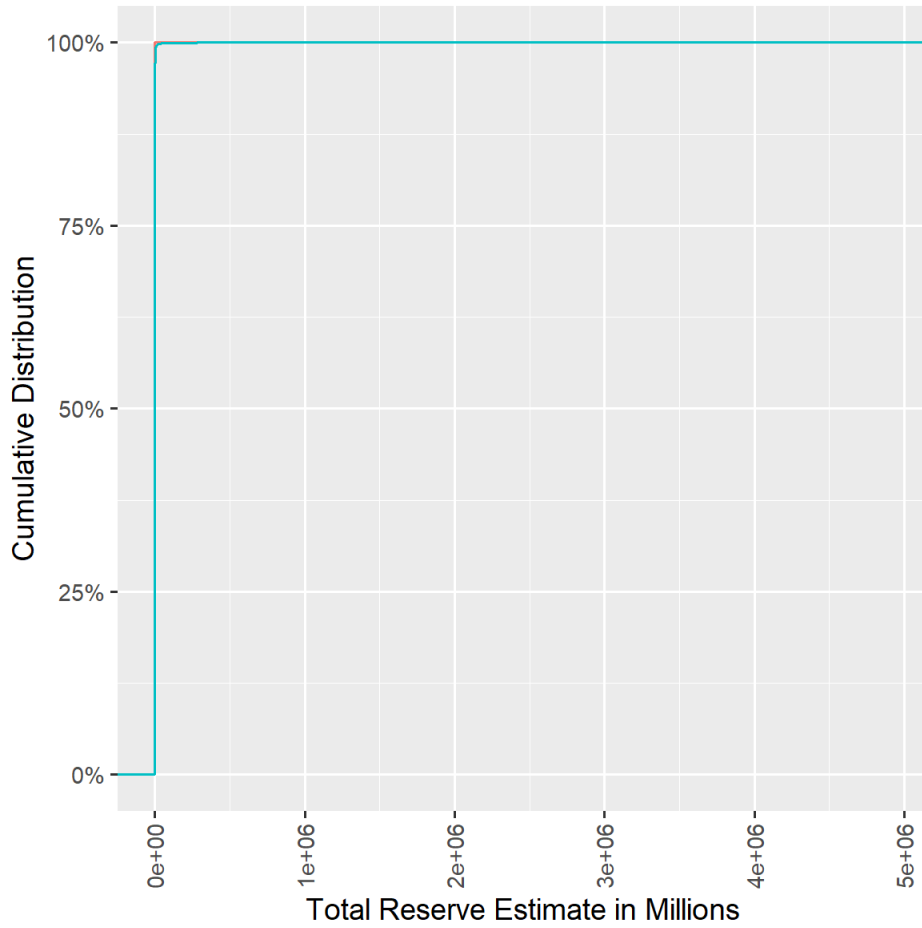
Total Reserve Estimate
Case 5 With Selected Prior Distribution
Simulated Future Inflation



Total Reserve Estimate
Case 5 With Selected Prior Distribution
Simulated Future Inflation



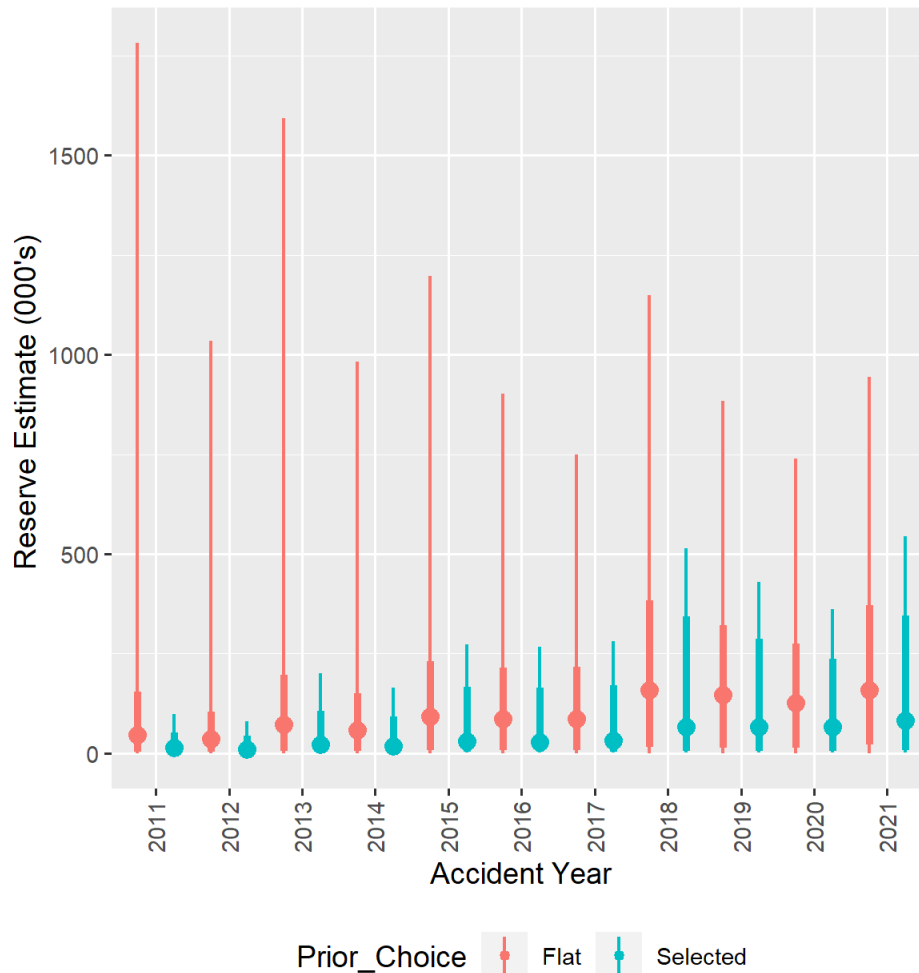
Total Reserve Estimate Case 5 Models With and Without Selected Prior



model Case 5 With Selected Prior Distribution Case 5 Without Selected Prior

Accident Year Ultimate Reserve Estimates

Case 5 by Prior Selection

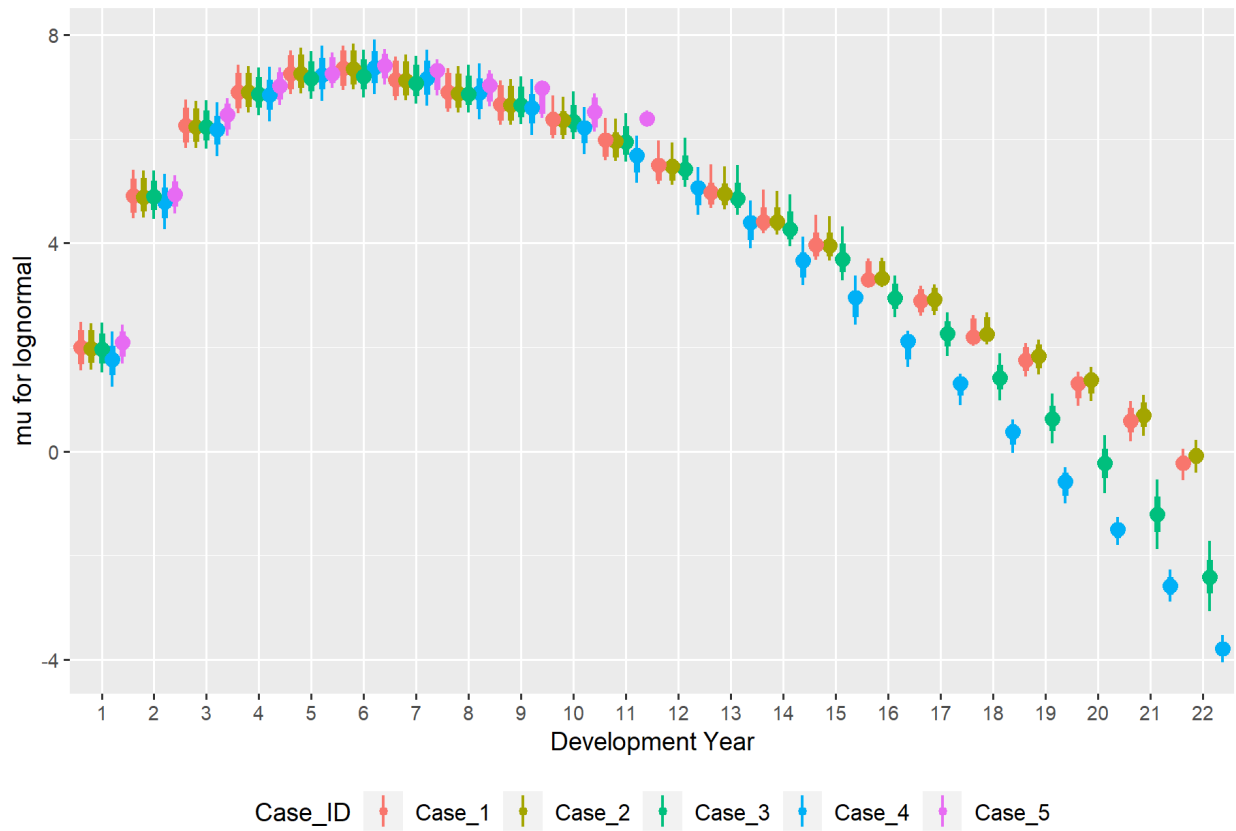


VII. Comparison Across Cases

VII.1 Comparison Across Cases for Selected Prior Parameters

Tightening the standard error parameter in the prior distributions for Cases 3, 4 and 5 resulted in minimal changes to the spread for the μ parameter estimates, even with less information. The sigma estimate for Case 3 is noticeably different and higher than those for Cases 1 and 2, which may realistically reflect the reduced information, while still producing plausible reserve forecasts. In Case 4, the sigma estimate is out of pattern with the rest but was set at a constant for the beyond ten years development period to ensure the MCMC routine converged to a plausible set of answers.

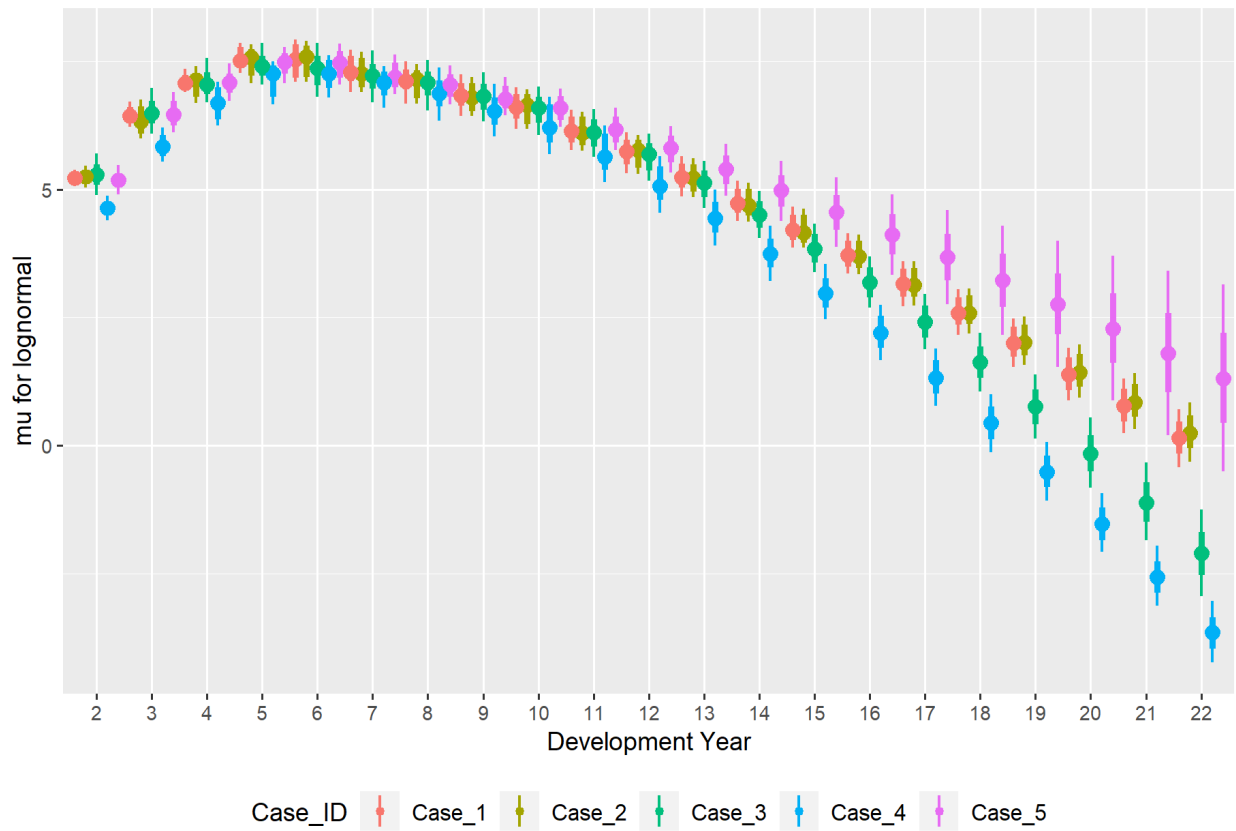
Estimated mu vs. Development Year
Observed Data
Selected Prior Parameter by Case Number



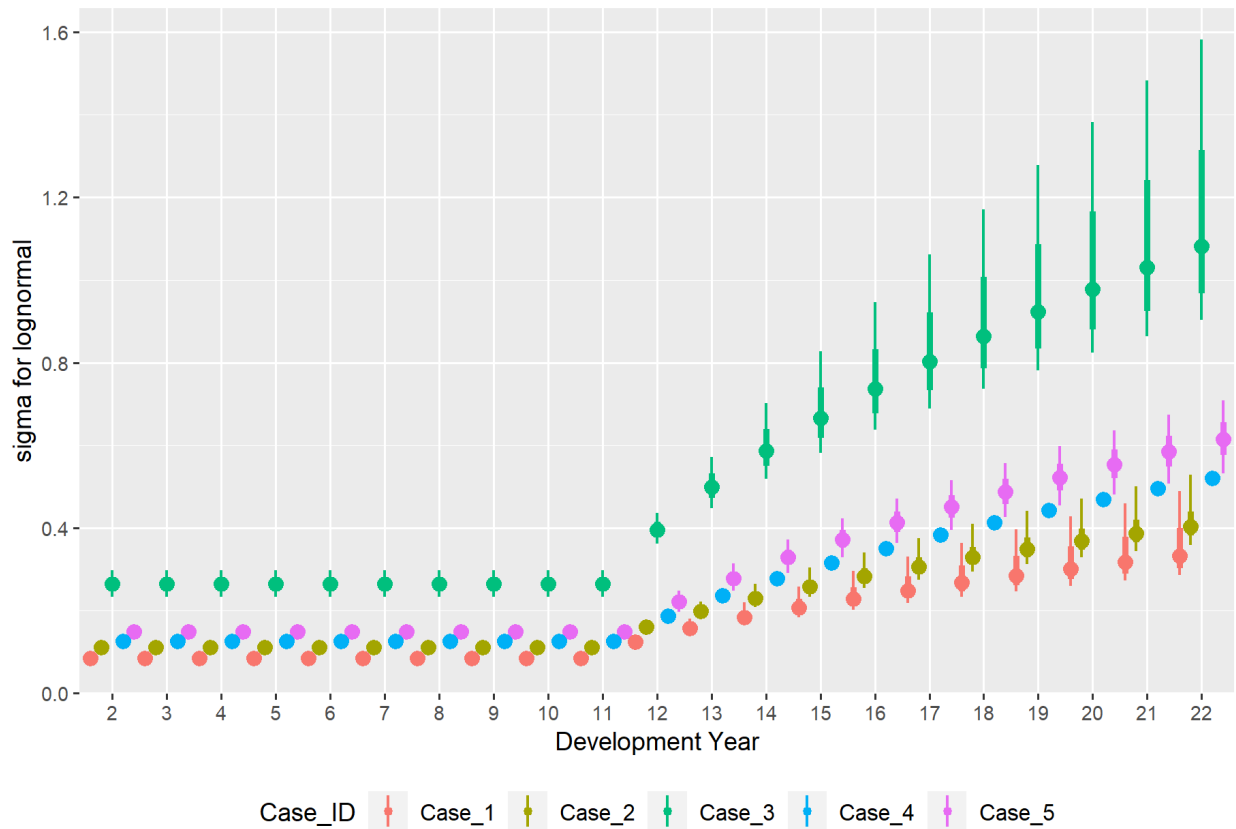
Estimated sigma vs. Development Year
Observed Data
Selected Prior Parameter by Case Number



Forecast mu vs. Development Year
Future Accident & Development Years
Selected Prior Parameter by Case Number



Forecast sigma vs. Development Year
 Future Accident & Development Years
 Selected Prior Parameter by Case Number

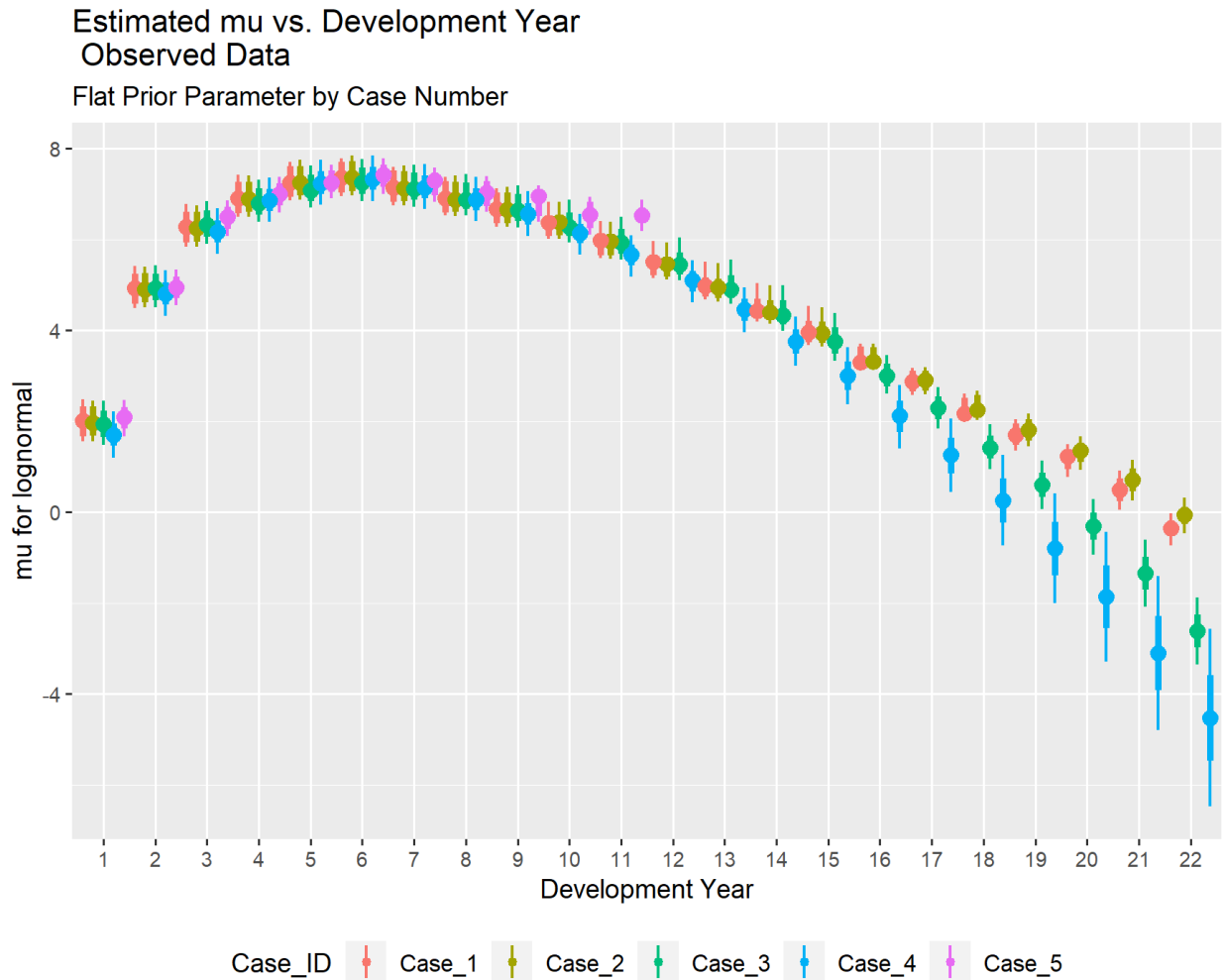


VII.2 Comparison Across Cases for Flat Prior Parameters

The variability in the forecast for mu and sigma exhibited more pronounced changes in the length of the confidence intervals as the information in the loss triangles varied, especially when compared to the results using selected prior distribution parameters. It is natural and reasonable to expect an increase in variability in the loss reserve forecast as the information in the loss triangle diminishes. The degree to which an actuary chooses to tighten the standard error range in a set of selected prior distributions requires professional judgement. Extensive experience in a given line may provide confidence in how the mean value should vary over time, but there remains an inherent increase in volatility as the sample size decreases.

The results for Case 5 clearly demonstrate the necessity of selecting prior distribution parameters to compensate for the lack of development history beyond eleven years when forecasting future time periods. The confidence intervals for mu beyond ten years are much larger than those for Case 3, and the forecast for mu does not decline at the same

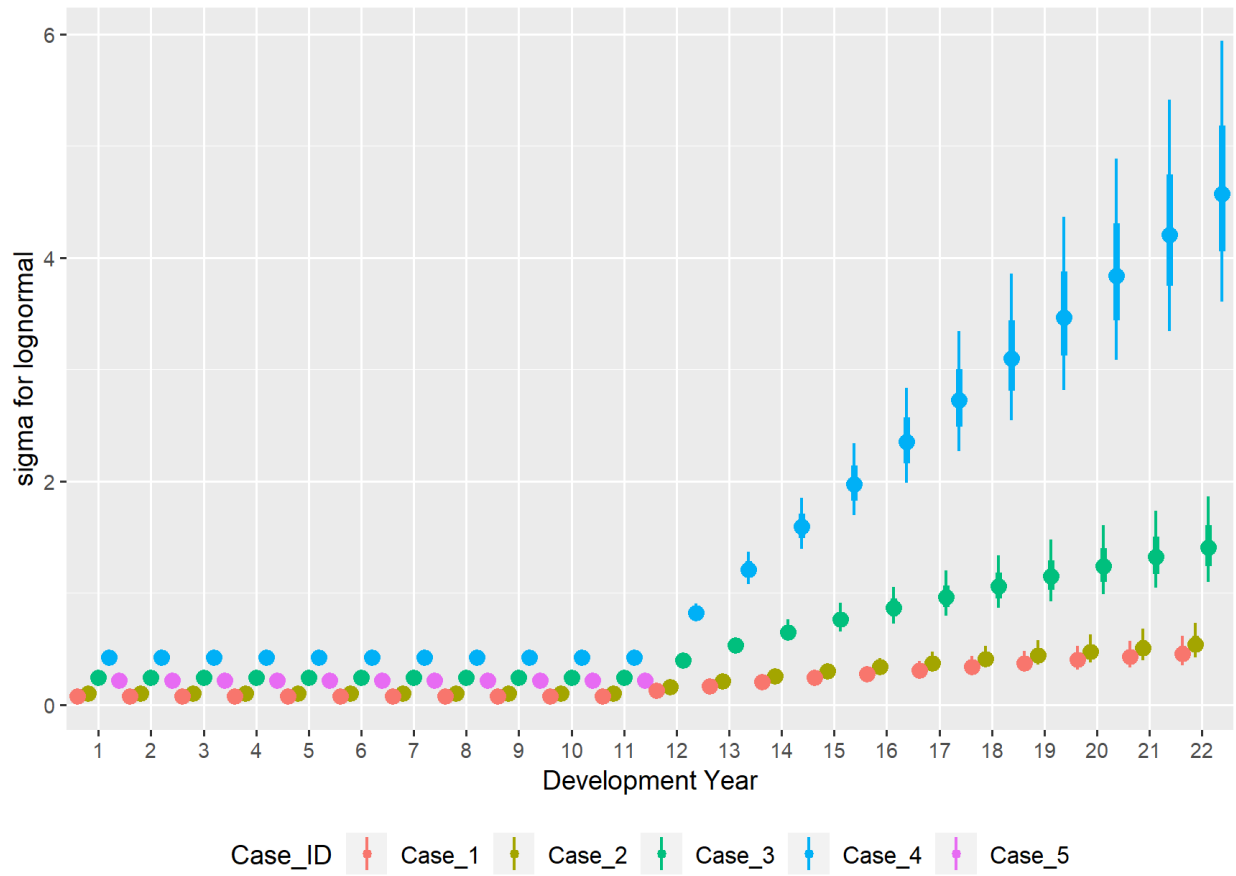
rate as seen in data sets that include experience for that time period. The assumption that sigma should remain flat as an accident year ages is not sensible, given the behavior observed in the data sets (and likely to be a true statement in general). This case generalizes the notion that some form of a selected prior distribution is essential when observations are not available for the time period being forecast.



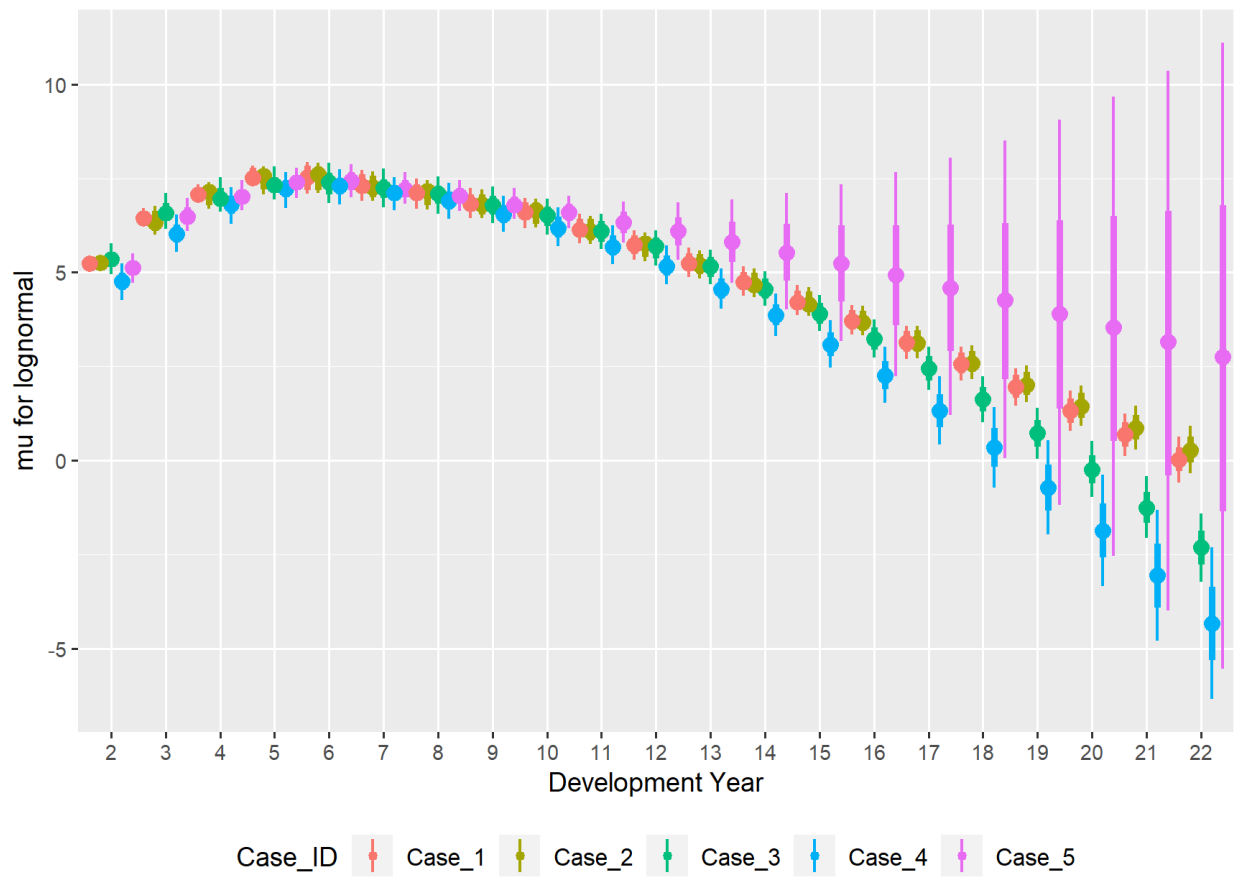
Estimated sigma vs. Development Year

Observed Data

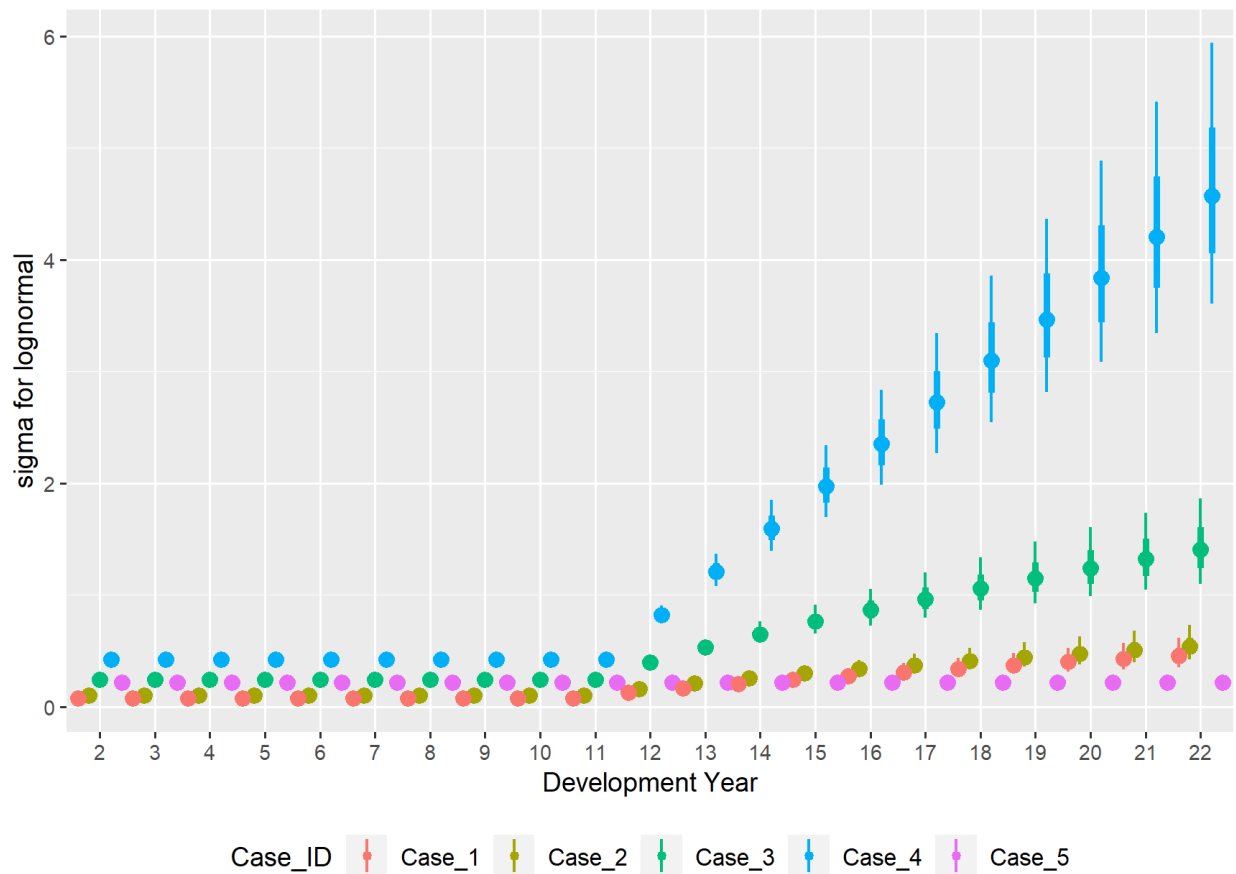
Flat Prior Parameter by Case Number



Forecast mu vs. Development Year
Future Accident & Development Years
Flat Prior Parameter by Case Number



Forecast sigma vs. Development Year
 Future Accident & Development Years
 Flat Prior Parameter by Case Number



VIII. Conclusion

The examples in this paper demonstrate how the Bayesian MCMC modeling environment can give an actuary the ability to supplement limited information in a sparse data set with experience gained from doing reserve analyses and general insurance knowledge. If you have observed development curves for a given line over time, you have an idea of what a plausible development curve should look like and how the stability in the development pattern varies as an accident year ages. Similarly, if you have worked with a given line of business for a while, you are likely to believe that the annual loss cost trend on top of a general inflation index is some small positive number rather than believing it is uniformly distributed over the real number line.

Using the terminology from the CAS exams, the prior distribution feature in a Bayesian MCMC analysis allows the actuary to effectively establish the compliment of credibility (or guardrails) for the population variable coefficients in a regression equation within the Bayesian MCMC environment. Variables that fall into the group category use the overall mean results in the data set across the group as the compliment of credibility. Note that

while the examples in this paper used the group variables to introduce a form of tempered interaction with the intercept by group member, one could also apply this technique to show an interaction effect with continuous variables that describe the smooth curve for a group variable like a business unit.

The paper is centered around examples of applying Bayesian MCMC to reserve modeling problems. There are many good textbooks on Bayesian MCMC and numerous articles and websites that provide a wealth of information. Providing a few examples that specifically address how to apply those techniques to loss reserving with sparse data is intended to help actuaries apply Bayesian MCMC concepts in practice.

The code that produced these examples is available to anyone interested in exploring the steps involved in building the models and generating the result summaries. To maintain a reasonable length, the paper does not delve into the mechanics of creating the model and organizing the results to evaluate the model. However, accessing the files, executing the code, and experimenting with different models and data sets exhibiting various behavior characteristics can be highly beneficial.

The paper also demonstrates that using simulation allows one to experiment with this technique and understand how different adjustments to the prior distributions affect the end results. There are various modeling forms available within the Bayesian MCMC environment, and using simulated data sets creates a safe sandbox for trying out those options.

The practical application of Bayesian MCMC became feasible after development of advances in software and algorithms between 2015 to 2020. Before this period, the time required to run a Bayesian MCMC, organize the information, and evaluate the results often made it impractical. However, the introduction of STAN, which uses the Hamiltonian algorithm to guide the MCMC search process, significantly reduced the time to run a Bayesian MCMC model constraint. The software and algorithms for organizing and evaluating the results of Bayesian MCMC modeling exercises also saw substantial improvements during this time.

While the software and algorithms available to run a Bayesian MCMC analysis have improved rapidly over recent years, it still requires an investment of time to learn how the technique works and how to operate the software. My hope is that this paper will provide some motivation to make that investment.

Appendix A

The data sets were created by simulation using the following steps:

1. The reported claim count was set at a single figure to be applied for all accident years within a given case.
2. Incremental paid claim counts were simulated by accident year and development year by inverting the Poisson distribution where lambda for the Poisson was set using a formula prior to calling the Poisson distribution. The formula for lambda has identical parameters for the development curve by development year for each accident year as it develops over time with minor differences in the intercept which varied by accident year.
3. The incremental severity figures came from inverting the lognormal distribution. The mu value for the lognormal distribution was set by a formula where the development by year curve for mu was identical across accident years, but there was a slight variation in the intercept by accident year. Sigma was set by a formula which created a value that varied by development year with the same curve applied across all accident years.
4. The incremental severity recorded by accident year and development year is the mean from the severity figures created calling a function to invert the lognormal with the number of times that function was called equal to the number of incremental paid counts generated by inverting the Poisson distribution.
5. The incremental paid amount prior to the effect of inflation and the additional loss cost trend was calculated as paid count times the average severity.
6. Overall inflation and loss cost trends were generated by inverting the Lognormal distribution for each calendar year included in the data set, calculating the geometric mean of the year end points to look at inflation as being centered at the midpoint of the calendar year then accumulating that effect on a compound basis.
7. The inflation and additional loss cost trend were linked to the accident year/development year table of incremental payments by calendar year and the trended incremental payment was calculated as the product of the accumulated inflation and loss cost trend effect and the incremental loss payments.
8. The normalized incremental payment (pseudo pure premium) that is the dependent variable that was modeled was calculated by dividing the trended incremental payment by the reported claim count and accumulated inflation and loss cost trend figures. Please note that many of the graphs refer to the pure premium as calculated at this step.
9. There were various transformations of development time created in the data development step that were used in the reserve modeling work.

Appendix B

For each of the reserve modeling cases in this paper, there are four steps that have to be completed to arrive at a forecast distribution. That forecast distribution is summarized in graphs and tables and is not shown in the paper.

1. Simulate the future inflation index results in the form of a data set that can be joined to the deflated pure premium posterior distribution.
2. Join the forecast data set for future development periods for each accident year to the model parameter estimates to create the forecast for normalized incremental payments (deflated pure premium) in the form of a posterior distribution data set.
3. Join the data sets created in steps 1 and 2 above by calendar year.
4. Multiply the forecast normalized incremental values (deflated pure premium) by the inflation forecasts and then multiply that result by the reported claim count at 12 months for each accident year.

Appendix C

The following software packages were used in creating this paper:

- Rstudio is the integrated development environment that was used.
- The brms macro writer for STAN was used to form the model structure then create STAN code. STAN is a powerful software package, but the coding requirements to create a model to run in STAN are detailed and brms simplified the coding required to produce models.
- The software which ran the Bayesian MCMC models is STAN.
- The tidyverse package was used to build the simulated data sets as well as organize the data to build graphs using ggplot2.
- The tidybayes package was used to forecast the future incremental losses as well as created modeled losses for the observed experience period for the loss triangles by joining the model object from the model constructed within brms and the loss triangles. There are also some specialized data manipulation packages used to organize the results and create graphs suited for Bayesian MCMC.
- Rmarkdown was used to organize the sequence in which code chunks for R are run to produce the model results plus add a few comments to help someone trying to use those program.
- Rtools links the brms macro writer to STAN from within Rstudio.

Bibliography

Brms

Bürkner, Paul-Christian. “brms: An R Package for Bayesian Multilevel Models Using Stan.” *Journal of Statistical Software* 80, no. 1 (January 1, 2017). <https://doi.org/10.18637/jss.v080.i01>.

Bürkner, Paul-Christian. “Advanced Bayesian Multilevel Modeling With the R Package Brms.” *The R Journal* 10, no. 1 (January 1, 2018): 395. <https://doi.org/10.32614/rj-2018-017>.

STAN

<https://mc-stan.org/>

The Stan Forums (mc-stan.org)

Rstudio

<https://rstudio.com/>

<https://community.rstudio.com/>

Tidyverse

Rstudio help

<http://tidyverser.org>

Wickham, Hadley, and Garrett Golemund. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*, 2016. https://openlibrary.org/books/OL26836898M/R_for_Data_Science_Import_Tidy_Transform_Visualize_and_Model_Data.

Background on Bayesian MCMC

Betancourt, Michael (2018). "A Conceptual Introduction to Hamiltonian Monte Carlo." [arXiv:1701.02434](https://arxiv.org/abs/1701.02434).

Ford, P., “MCMC Algorithms,” CAS Study Note, Version 0.7, November 2019.

Information Theory Applied to Bayesian MCMC

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC.” *Statistics and Computing* 27, no. 5 (August 30, 2016): 1413–32. <https://doi.org/10.1007/s11222-016-9696-4>.

Yao, Yuling, Aki Vehtari, Daniel Simpson, Andrew Gelman, Bertrand Clarke, Meng Li, Peter Grünwald, et al. “Using Stacking to Average Bayesian Predictive

Distributions (with Discussion).” *Bayesian Analysis* 13, no. 3 (September 1, 2018). <https://doi.org/10.1214/17-ba1091>.

Gelman, Andrew, Jessica Hwang, and Aki Vehtari. “Understanding predictive information criteria for Bayesian models.” *Statistics and Computing* 24, no. 6 (August 20, 2013): 997–1016. <https://doi.org/10.1007/s11222-013-9416-2>.

Watanabe, Sumio. “Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory.” *Journal of Machine Learning Research*, December 1, 2010. <https://doi.org/10.5555/1756006.1953045>.

Bayesian MCMC textbooks

McElreath, Richard. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, 2015. <http://pubman.mpdl.mpg.de/pubman/item/escidoc:2183367>.

Andrew Gelman et al., *Bayesian Data Analysis*, Third Edition. CRC Press, 2013. http://books.google.ie/books?id=ZXL6AQAQBAJ&printsec=frontcover&dq=%E2%80%A2%09Bayesian+Data+Analysis+Chapman+Hall&hl=&cd=2&source=gbs_api.

Tidybayes

<https://cran.r-project.org/web/packages/tidybayes>

[Tidy Data and Geoms for Bayesian Models • tidybayes \(mjskay.github.io\)](#)

Credibility Theory

Tse, Yiu-Kuen. *Nonlife Actuarial Models: Theory, Methods and Evaluation*, 2nd ed. (International Series on Actuarial Science). Cambridge: Cambridge University Press, 2023.

Casualty Actuarial Society Monographs

Taylor, Greg and Gráinne McGuire. “Stochastic Loss Reserving Using Generalized Linear Models.” 2016. CAS Monograph No. 3. <https://www.casact.org/monograph/cas-monograph-no-3>.

Meyers, Glenn. “Stochastic Loss Reserving Using Bayesian MCMC Models,” 2nd ed. 2019. CAS Monograph No. 8. <https://www.casact.org/sites/default/files/2021-02/08-Meyers.pdf>.

Casualty Actuarial Society Papers

Zehnwirth, Ben, and Glenn Barnett. “Best Estimates for Reserves,” 2000. *Proceedings of the Casualty Actuarial Society*, vo. LXXXVII, no. 167, 245-321.

https://www.casact.org/sites/default/files/database/proceed_proceed00_00245.pdf.

Guszcza, James. "Hierarchical Growth Curve Models for Loss Reserving." Casualty Actuarial Society E-Forum, Fall 2008.

https://www.casact.org/sites/default/files/database/forum_08forum_7guszcza.pdf.