Application of Bayesian MCMC to Reserving Noisy Data

Presented 11/09/2022 at Casualty Actuarial Annual Meeting

By Michael R Larsen, FCAS, MAAA

Polling Questions

- How many of you have taken Exam 5?
- How many of you have taken the prior Exam 4/c?
- How many of you have taken MAS II?
- How many of you have fit a regression model?
- How many of you have used R ?
- How many of you have experience in reserving?

Agenda

- Outline goals
- Vocabulary
- Generating Data Sets
- Format for Case Studies
- Case Study I
  - Varying inflation and exposure volume
  - Change in Operations
- Case Study II
  - Zero Payments
  - High Severity

Goals for Session

Cover examples

- Normalizing the triangle
- Options for inflation
- Varying exposure size
- Options for change in operations
- Zero payments in the triangle
- High claim severity

What won't be covered.

- Underlying theory of Bayesian MCMC
- How to monitor Bayesian MCMC behavior
- Extensive review of coding for packages

Bayesian MCMC Vocabulary

Model Structure Vocabulary

- Population variable: variable that can have a non-Normal prior distribution with posterior distribution as weighted average of prior and data indicated estimate (generalization of conjugate prior Bayesian distributions like Gamma-Poisson or Beta-Binomial)
- Random/Group variable: variable that is restricted to a Normal prior with zero mean (average across group) and the variance values estimated from the data set (something like the least squares credibility approach)
- Distribution model: each parameter of the distribution has an equation in the model structure
- MCMC: Markov Chain Monte Carlo approach where the parameters are entries in the Markov probability model and a simulation based algorithm is used to iteratively estimate the parameters
- Posterior Distribution: Data set containing the simulated parameter values for each pass through the MCMC process after warmup
- Prior distribution: Analyst's prior opinion of the distribution for each parameter

Modeling Environment

Software Vocabulary

- STAN: Bayesian MCMC program that implements MCMC via an efficient algorithm (Hamiltonian)
- Brms: software that is a macro writer to generate STAN code and prepare the data set for STAN
- Tidyverse: version of R code to simplify general data manipulation
- Tidybayes: version of R code to simplify Bayesian MCMC data manipulation
- Ggplot2: version of R code to assist in creating graphs
- Rstudio: integrated data environment to assist in creating and running models

Generating Data Sets

- Data sets created by simulation
    - Selected reported counts at 12 months development for exposure base
    - Poisson payment count distribution with rate varying by development year
    - Lognormal incremental payment amount with mu and sigma varying by development year
- Loss triangle description
    - Accident Years 2000 to 2021
    - Development Years 1 to 22
    - Incremental payments
- Loss Cost Trend
    - Inflation with lognormal assumption
    - Coverage adjustment with lognormal assumption multiplied with inflation

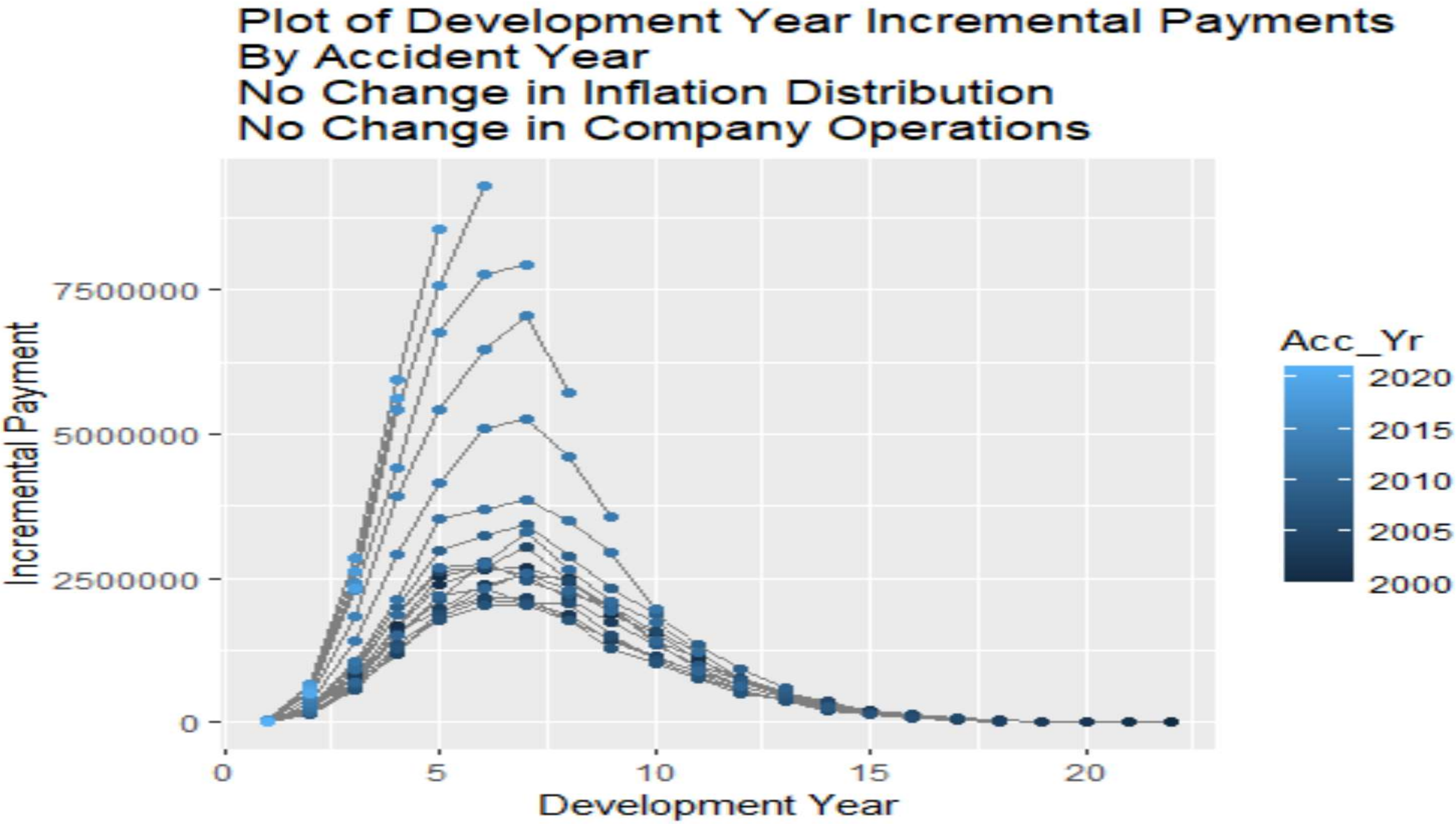
Format for Case Studies

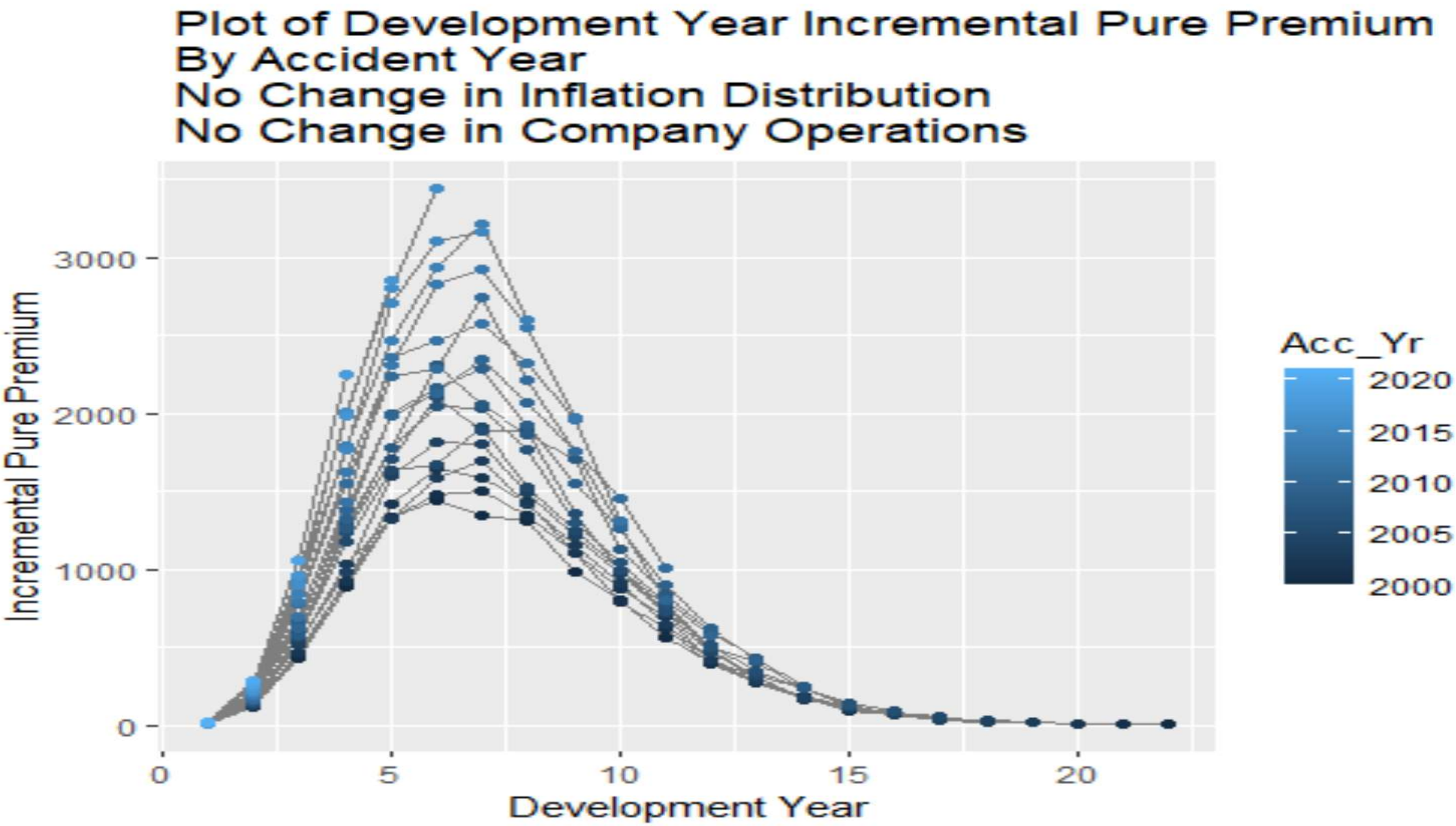
- Common sequence
  - Exploratory data analysis graphs
  - Model description
  - Model results
- Comments on modeling choices and results

Justification for Normalization using Case I Data Set

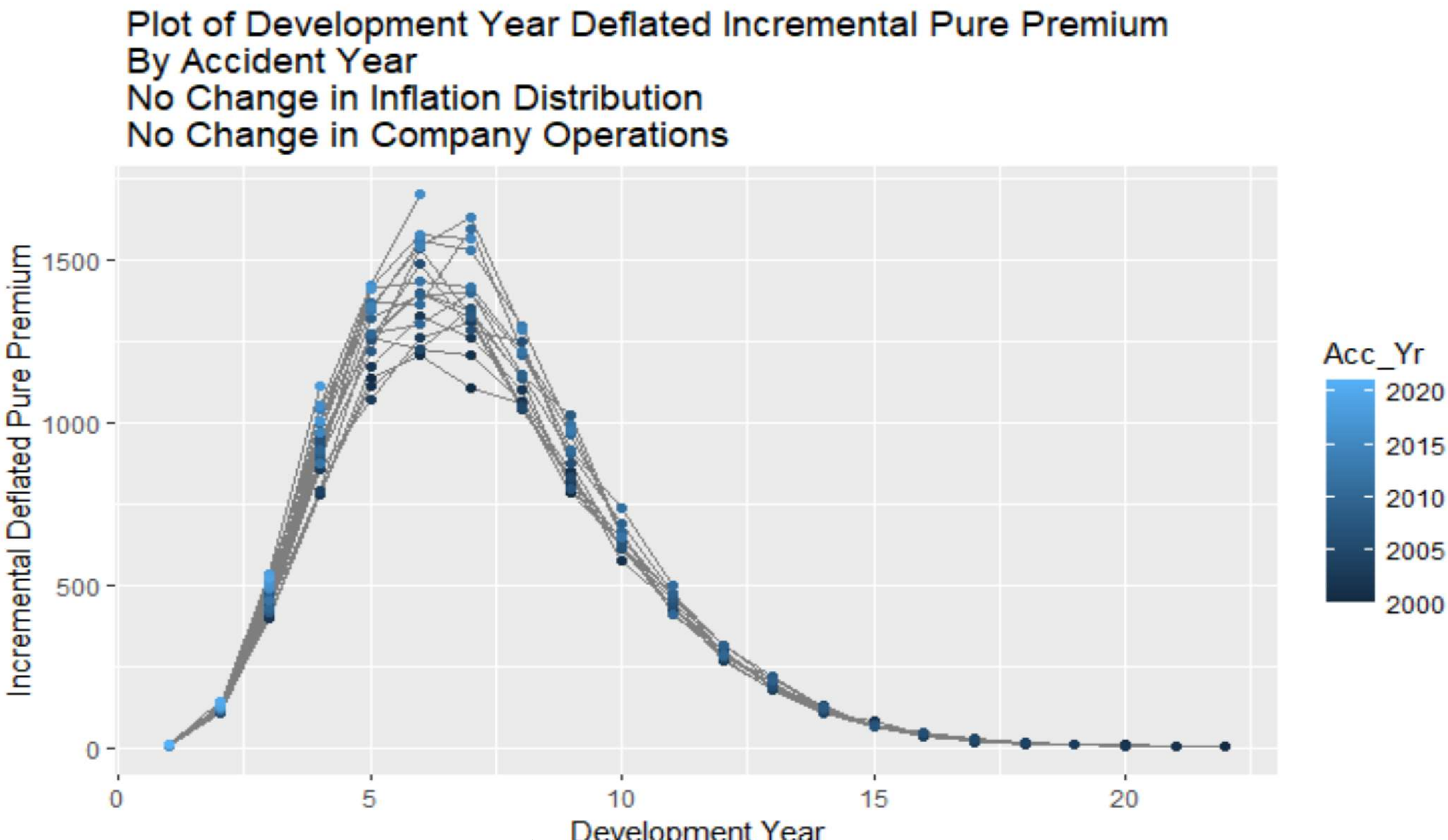
Normalization Translation to Natural Log

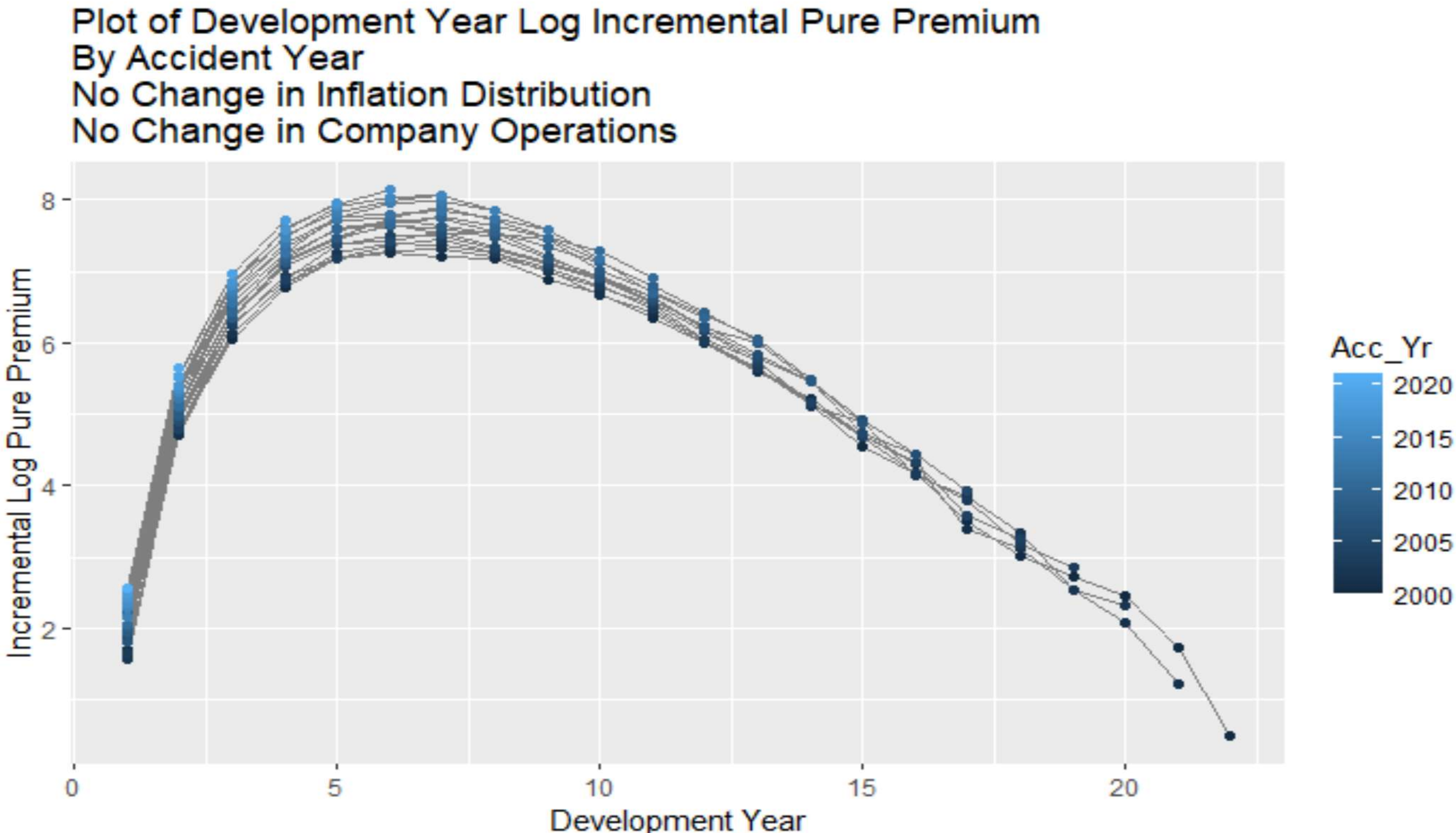
- Mechanics
  - Divide payments by accident year exposure
  - Divide payments by some inflation index
  - Take Natural log
- Justification
  - Payment amounts comparable across accident years
  - Diminishes correlation effect
  - Diminishes effect of large differences in amounts
- Illustration of effects
  - Graphs to illustrate benefits follow


Plot of Development Year Incremental Payments
By Accident Year
No Change in Inflation Distribution
No Change in Company Operations
Incremental Payment
7500000
5000000
2500000
0
0
5
10
15
20
Development Year
Acc_Yr
2020
2015
2010
2005
2000

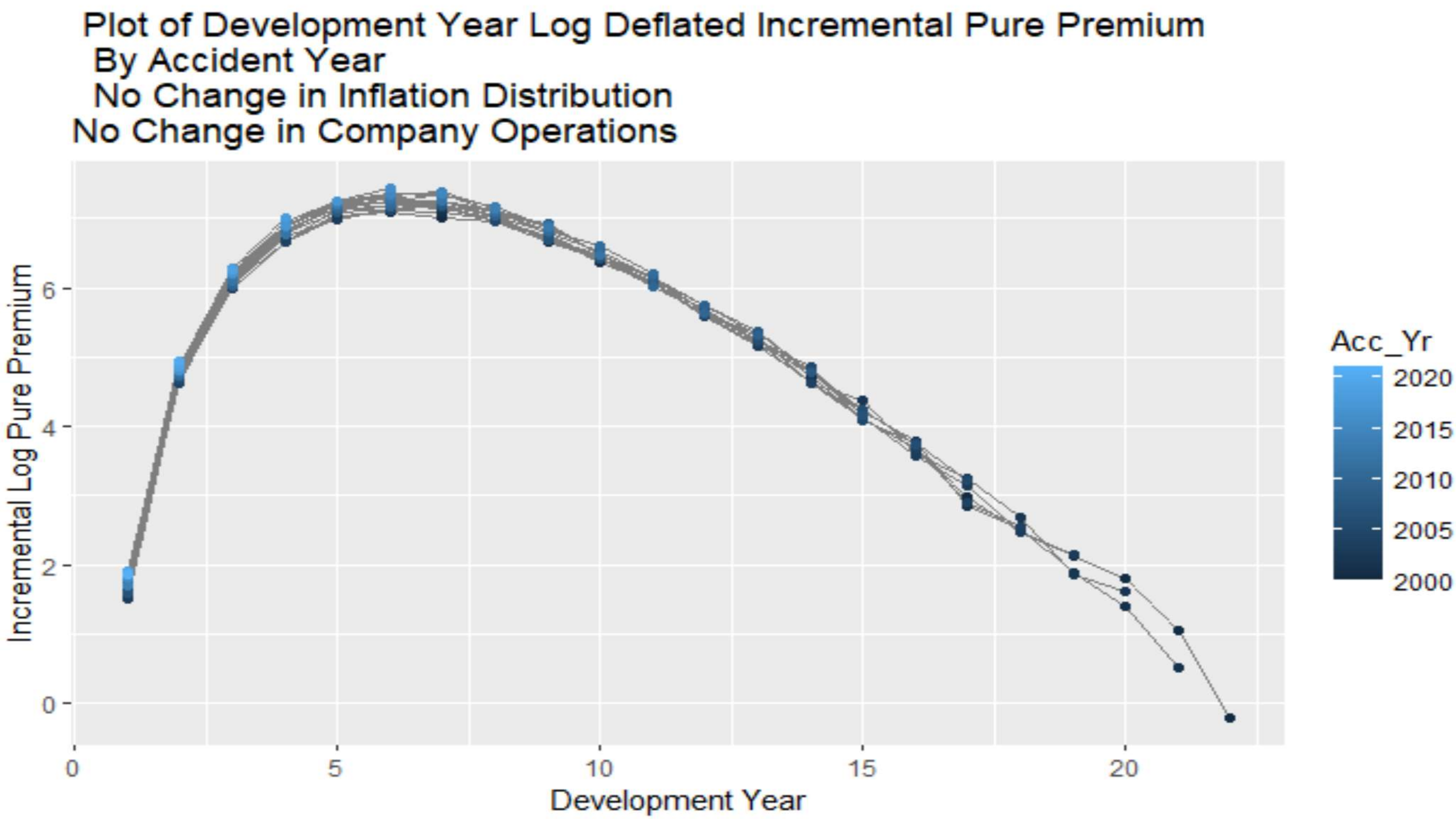


Plot of Development Year Incremental Pure Premium
By Accident Year
No Change in Inflation Distribution
No Change in Company Operations
Incremental Pure Premium
3000
2000
1000
0
Development Year
0
5
10
15
20
Acc_Yr
2020
2015
2010
2005
2000

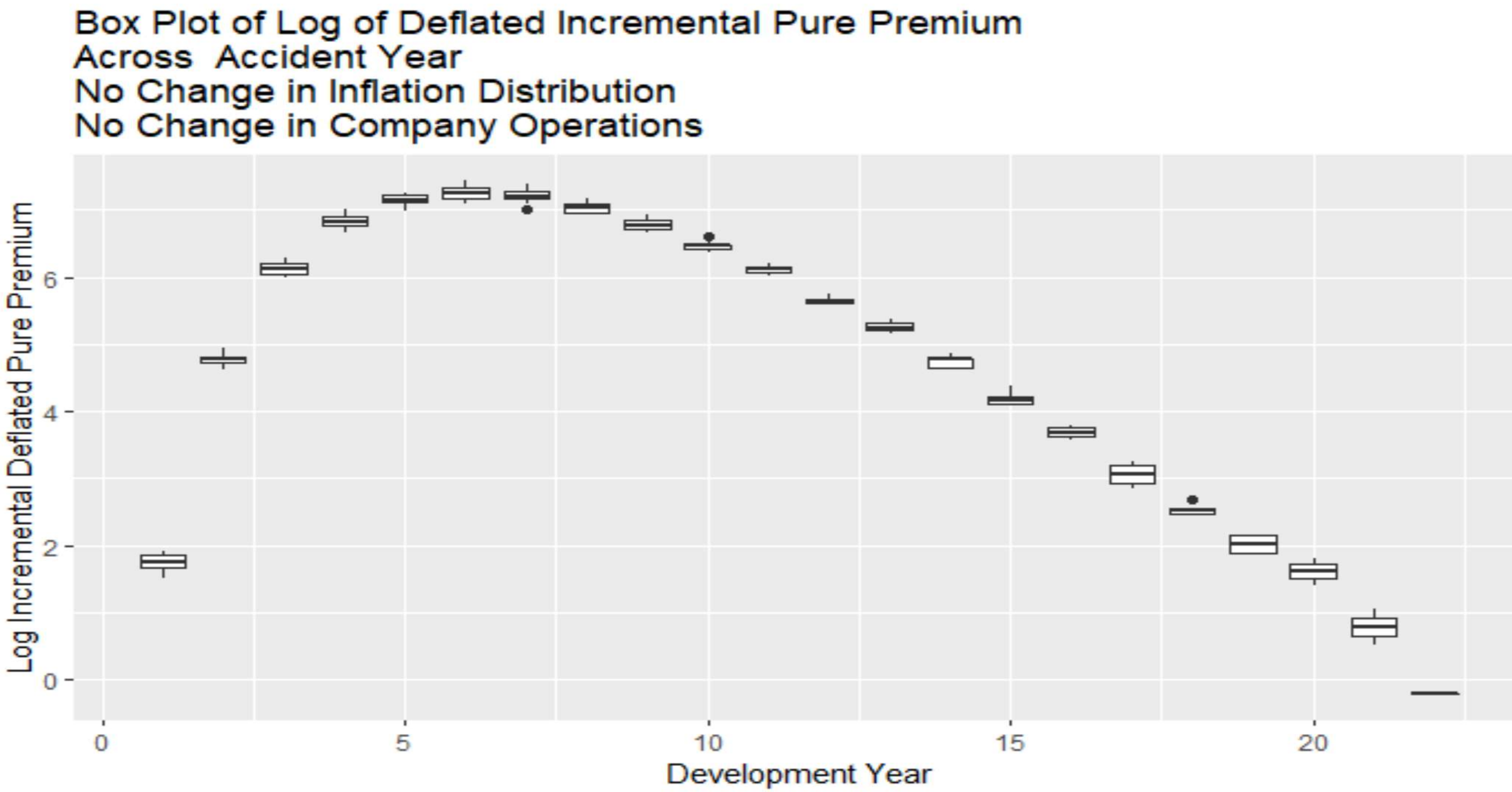

Plot of Development Year Deflated Incremental Pure Premium
By Accident Year
No Change in Inflation Distribution
No Change in Company Operations
Incremental Deflated Pure Premium
1500
1000
500
0
0
5
10
15
20
Development Year
Acc_Yr
2020
2015
2010
2005
2000

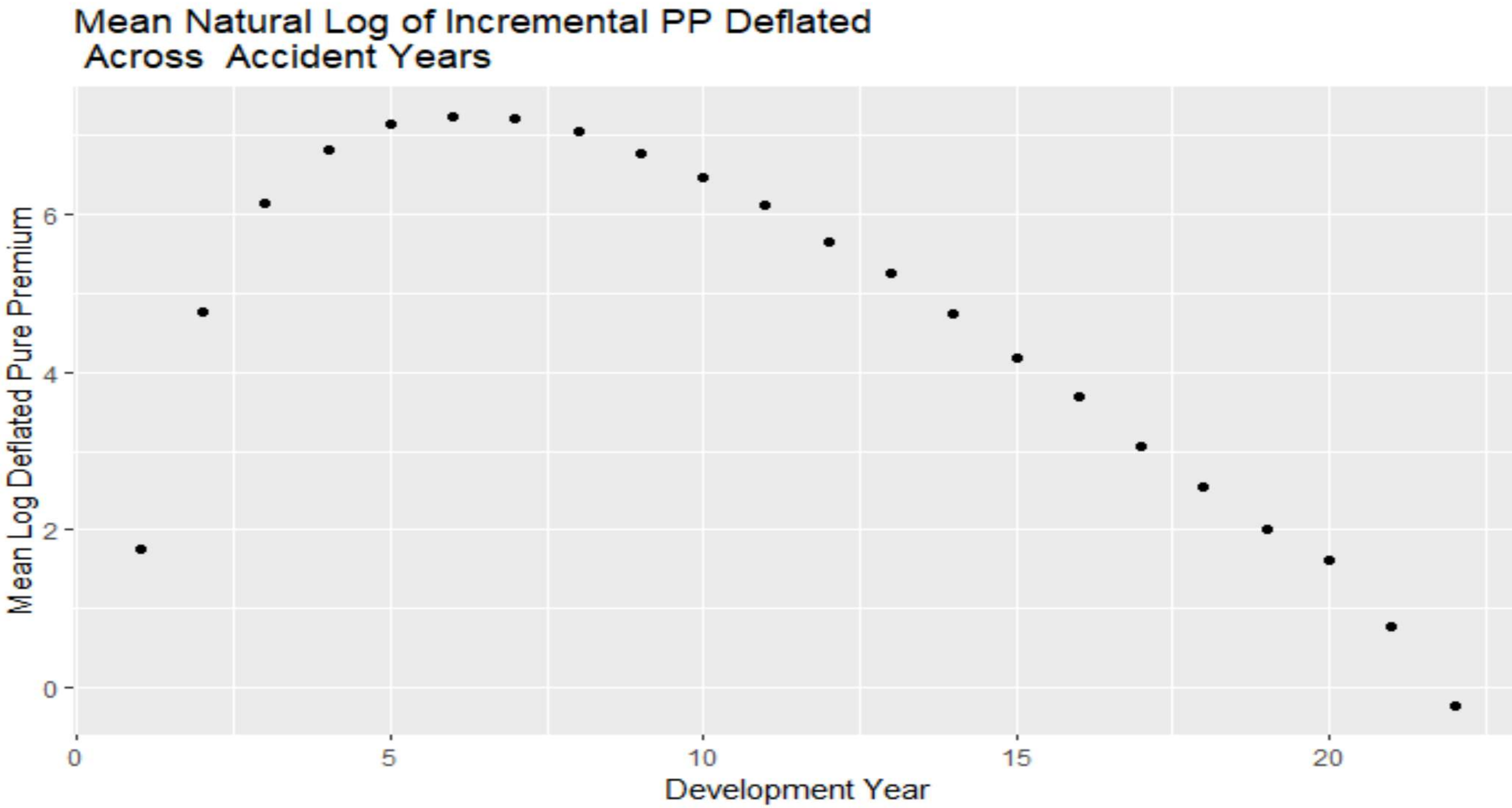


Plot of Development Year Log Incremental Pure Premium
By Accident Year
No Change in Inflation Distribution
No Change in Company Operations
Incremental Log Pure Premium
Development Year
Acc_Yr
2020
2015
2010
2005
2000
0
5
10
15
20
2
4
6
8

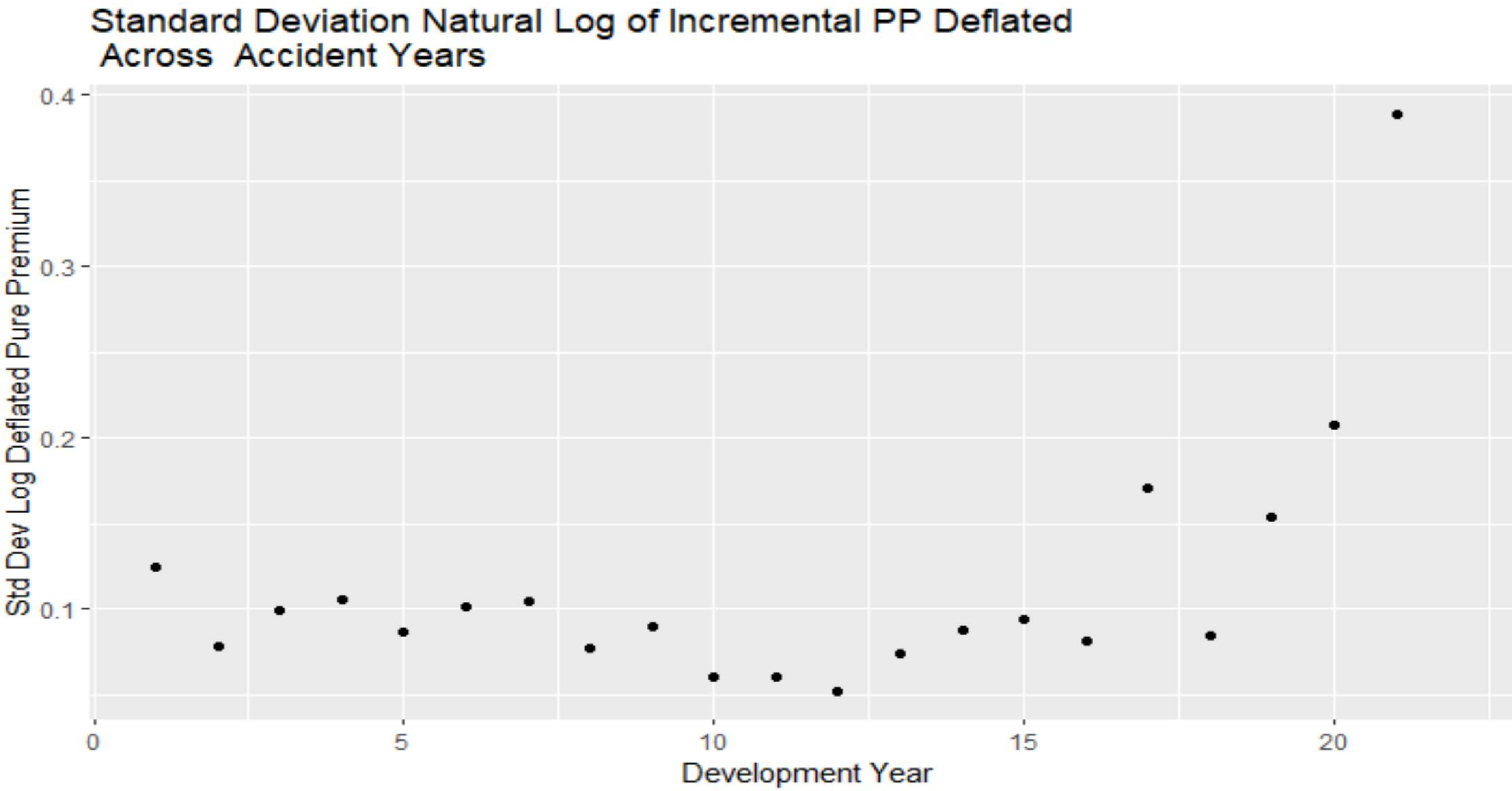

Plot of Development Year Log Deflated Incremental Pure Premium
By Accident Year
No Change in Inflation Distribution
No Change in Company Operations
Incremental Log Pure Premium
Development Year
Acc_Yr
2020
2015
2010
2005
2000
0
2
4
6
0
5
10
15
20

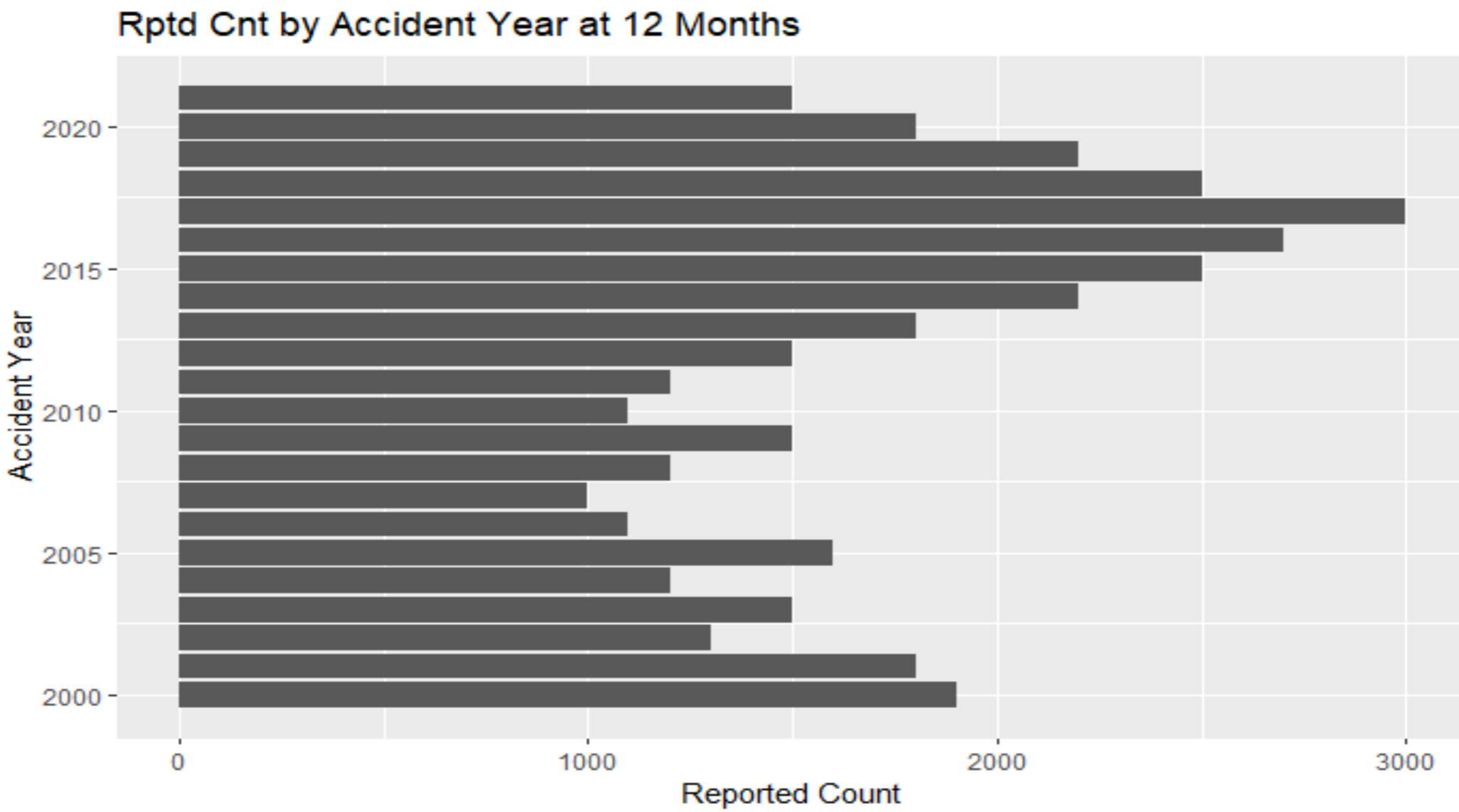
Case I:
Simplest Model: No Change in Operations and No Change in Inflation Distribution


Box Plot of Log of Deflated Incremental Pure Premium
Across Accident Year
No Change in Inflation Distribution
No Change in Company Operations
Log Incremental Deflated Pure Premium
0
2
4
6
0
5
10
15
20
Development Year



Mean Natural Log of Incremental PP Deflated
Across Accident Years
Mean Log Deflated Pure Premium
0
2
4
6
0
5
10
15
20
Development Year


Standard Deviation Natural Log of Incremental PP Deflated
Across Accident Years
Std Dev Log Deflated Pure Premium
0.4
0.3
0.2
0.1
0
5
10
15
20
Development Year


Rptd Cnt by Accident Year at 12 Months
Accident Year
2020
2015
2010
2005
2000
0
1000
2000
3000
Reported Count

Brms Structure

Brms structure

- Start with Regression Type formula:
  - `response | aterms ~ pterms + (gterms | group)`
- Add Other Modeling features:
  - Prior Distributions
  - Correlation structures
  - Variance modeling & other terms

Brms components

- Regression formula terms:
  - Response: dependent variable
  - Aterms: adjustments to dependent variable (exposure or censoring)
  - Pterms: GLM type betas for population
  - Group: variables to apply least squares credibility to in regression
- Other modeling features
  - Prior distribution: source to credibility weight against data set estimates
  - Correlation & Variance: options to model complex covariance structures

BRMS Code for Time as Trend Model Priors

```
ln_prior_operation_1 <- c(prior(normal(-3,.5),class=b, coef=Intercept),
                   prior(normal(1,1),class=b, coef=Dev_Yr_2_Factor2),
                   prior(normal(2,1),class=b, coef=Dev_Yr_2_Factor3),
                   prior(normal(.03,.01),class=b, coef=Cal_Yr_Time),
                   prior(normal(1,1),class=b, coef=Dev_Yr_6_Cap),
                   prior(normal(-.5,.5),class=b, coef=Dev_Yr_6_Cap_Sqrd),
                   prior(normal(-.5,.5),class=b, coef=Dev_Yr_8_Spline_Sqrd),
                   prior(normal(.5,.5),class=b, coef=Dev_Yr_8_Spline),
                   prior(normal(.4,.5),class=b, coef=Intercept,dpar=sigma),
                   prior(normal(0,.5),class=b, coef=Dev_Yr_1_Factor2,dpar=sigma),
                   prior(normal(.5,1),class=b, coef=Dev_Yr_13_Spline,dpar=sigma))
```

BRMS Code for Time as Trend Model

```
lognormal_pp_operation_I_1 <-brm(bf(Trended_Incr_PP ~ 0 +Intercept +
                  Dev_Yr_6_Cap+
                  Dev_Yr_6_Cap_Sqrd+
                  Dev_Yr_2_Factor +
                  Dev_Yr_8_Spline +
                  Dev_Yr_8_Spline_Sqrd +
                  Cal_Yr_Time ,
                 sigma ~0 + Intercept +
                  Dev_Yr_1_Factor+
                  Dev_Yr_13_Spline +
                  (1||Acc_Yr)), This introduces group effect for accident year
                seed = 8603529,
                data = Train_Triangle_All_Operation_I,
                family =lognormal(),
                prior =ln_prior_operation_1)
```

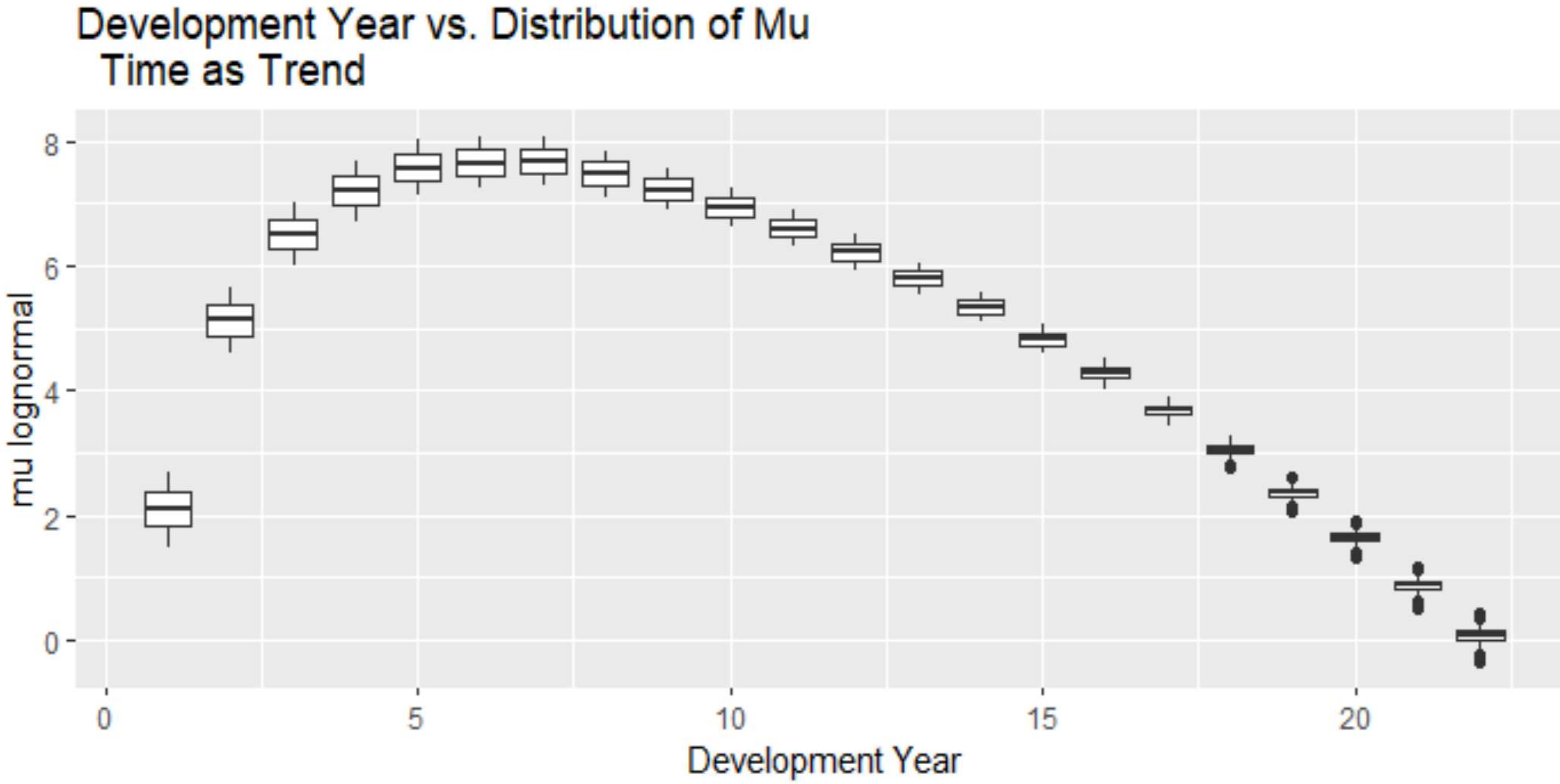
BRMS Code for Inflation Plus Correction Trend Model Priors

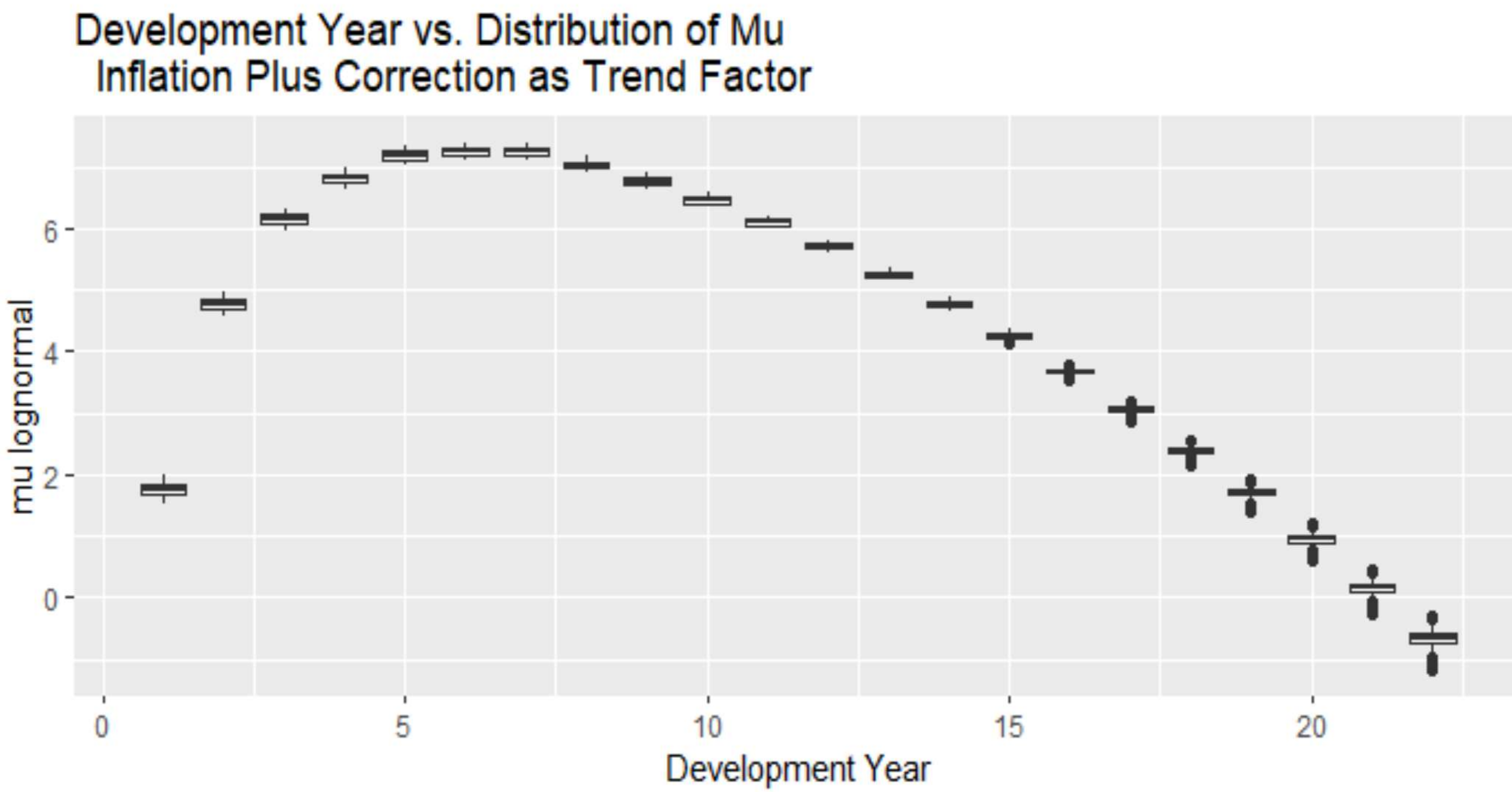
```
ln_prior_operation_3 <- c(prior(normal(-3,.5),class=b, coef=Intercept),
                prior(normal(1,1),class=b, coef=Dev_Yr_2_Factor2),
                prior(normal(2,1),class=b, coef=Dev_Yr_2_Factor3),
                prior(normal(.01,.01),class=b, coef=Cal_Yr_Time),
                prior(normal(1,1),class=b, coef=Dev_Yr_6_Cap),
                prior(normal(-.5,.5),class=b, coef=Dev_Yr_6_Cap_Sqrd),
                prior(normal(-.5,.5),class=b, coef=Dev_Yr_8_Spline_Sqrd),
                prior(normal(.5,.5),class=b, coef=Dev_Yr_8_Spline),
                prior(normal(.4,.5),class=b, coef=Intercept,dpar=sigma),
                prior(normal(0,.5),class=b, coef=Dev_Yr_1_Factor2,dpar=sigma),
                prior(normal(.5,1),class=b, coef=Dev_Yr_13_Spline,dpar=sigma))
```

Use blue for mu variables in examples

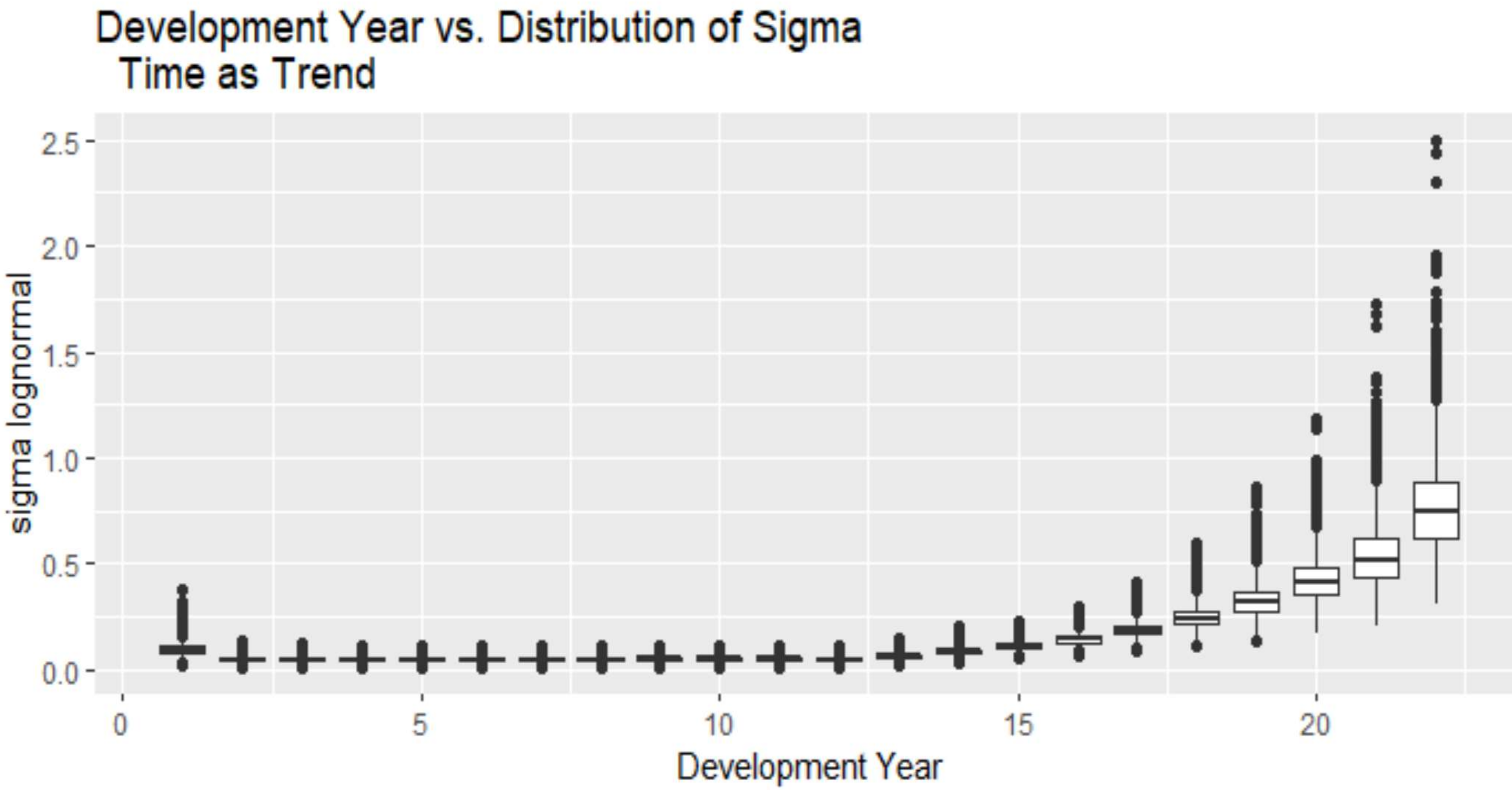
BRMS Code for Inflation Plus Correction as Trend Model

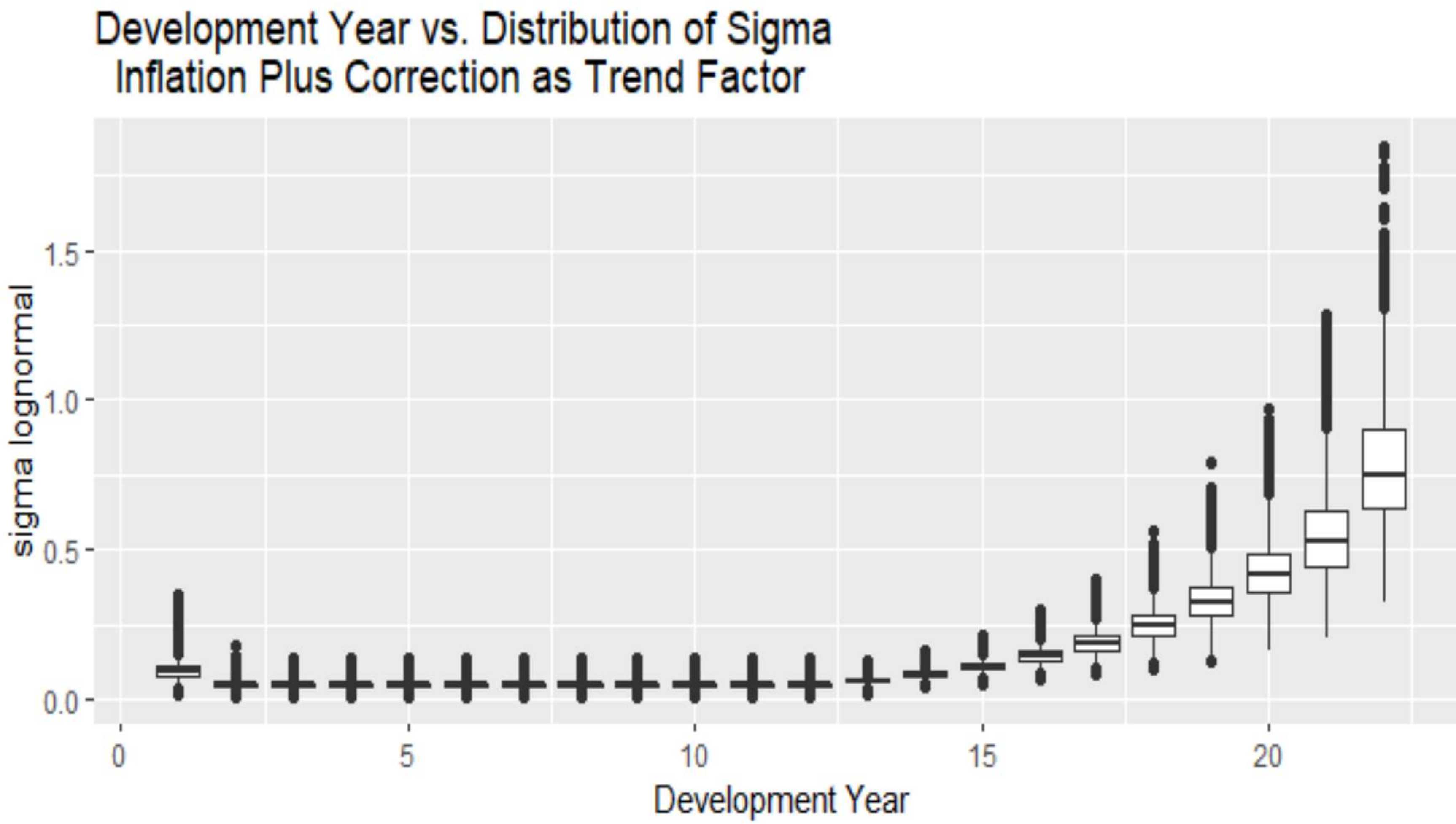
```
lognormal_pp_operation_I_3 <-brm(bf(Trended_Incr_PP_Def ~ 0 +Intercept +
                  Dev_Yr_6_Cap+
                  Dev_Yr_6_Cap_Sqrd+
                  Dev_Yr_2_Factor +
                  Dev_Yr_8_Spline +
                  Dev_Yr_8_Spline_Sqrd +
                  Cal_Yr_Time ,
                 sigma ~0 + Intercept +
                  Dev_Yr_1_Factor+
                  Dev_Yr_13_Spline +
                  (1||Acc_Yr)),
                seed = 8603529,
                data = Train_Triangle_All_Operation_I,
                family =lognormal(),
                prior =ln_prior_operation_3)
```

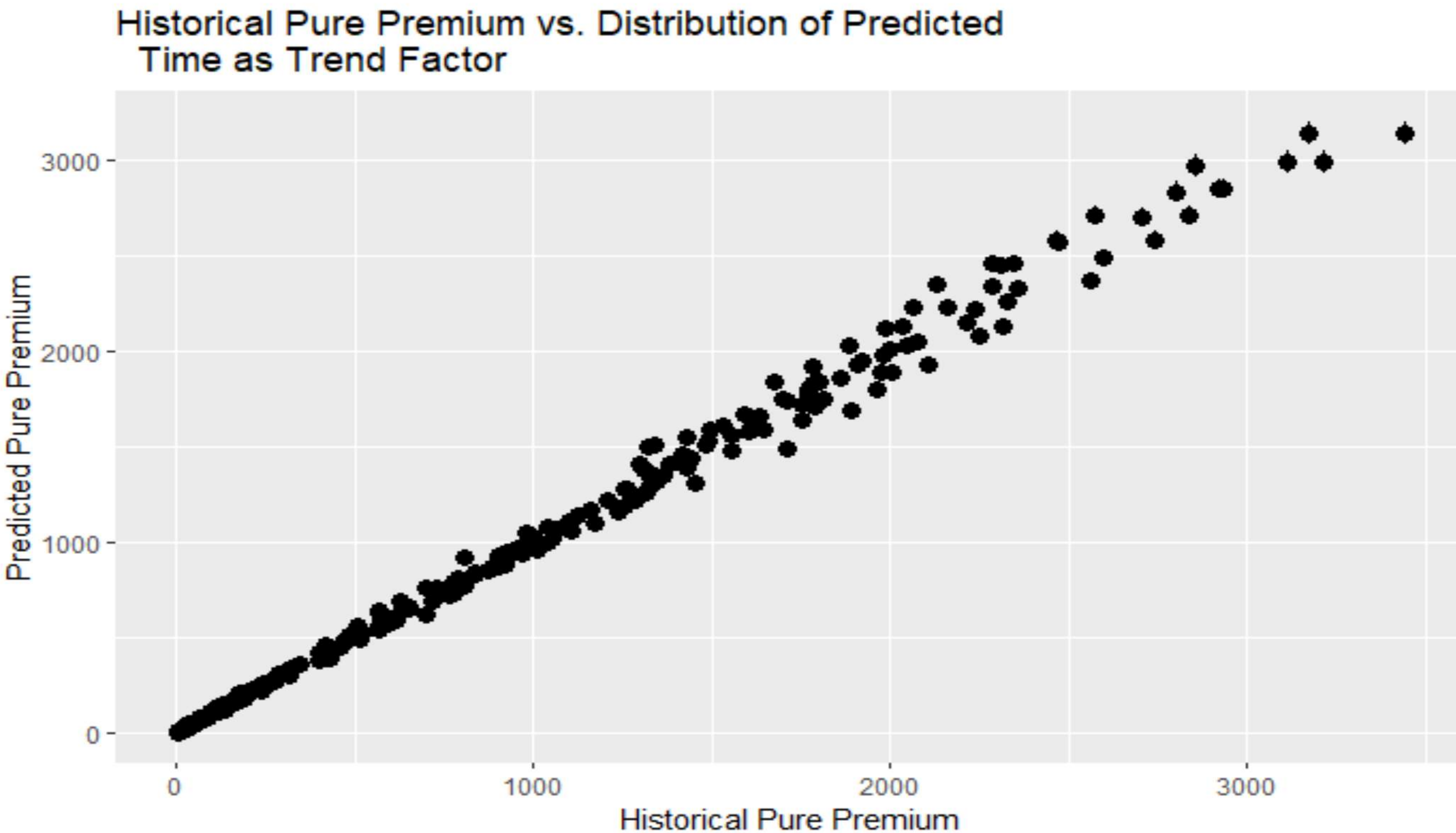

Development Year vs. Distribution of Mu
Time as Trend
mu lognormal
8
6
4
2
0
0
5
10
15
20
Development Year

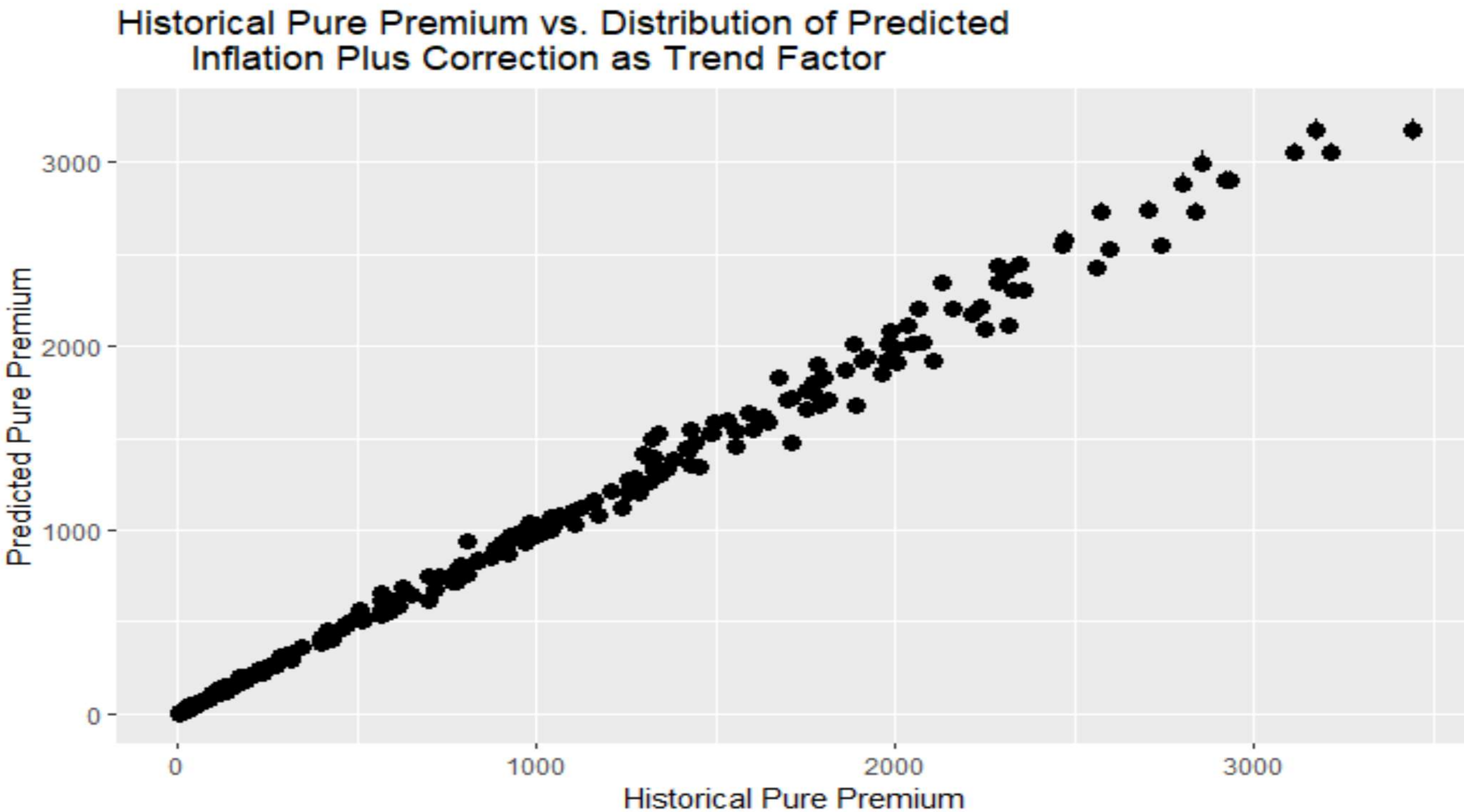


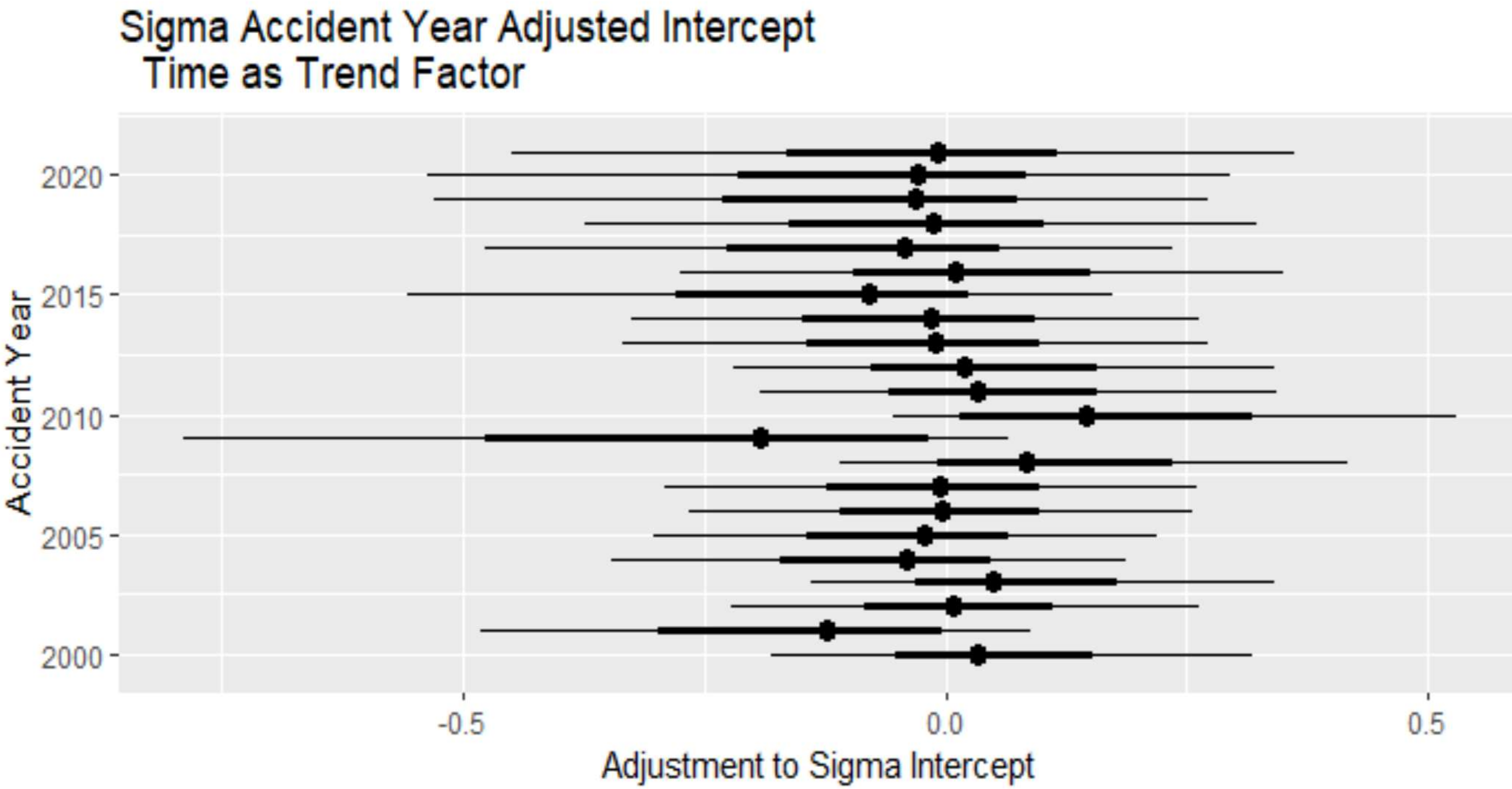
Development Year vs. Distribution of Mu
Inflation Plus Correction as Trend Factor
mu lognormal
0
2
4
6
0
5
10
15
20
Development Year

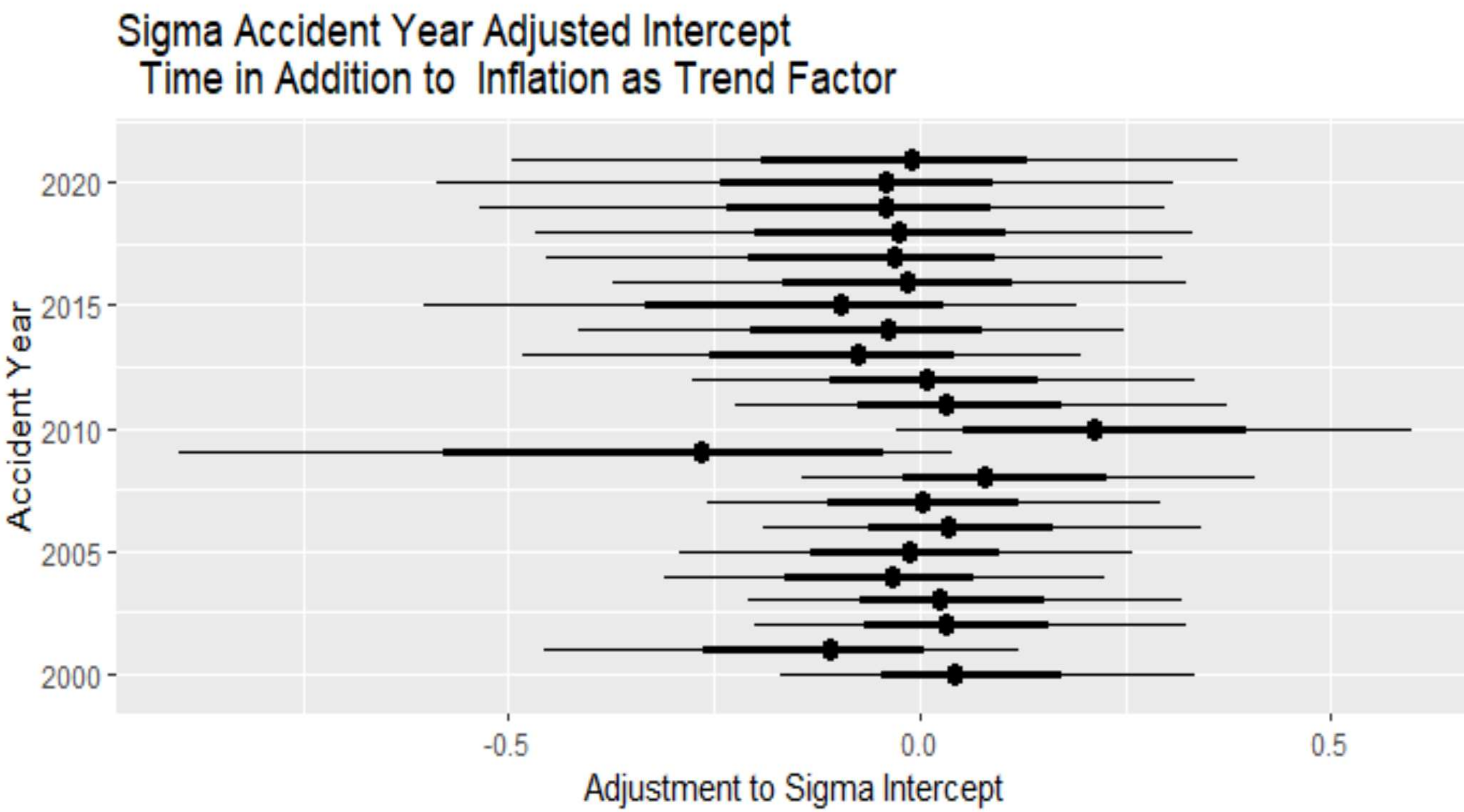

Development Year vs. Distribution of Sigma
Time as Trend
sigma lognormal
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
Development Year


Development Year vs. Distribution of Sigma
Inflation Plus Correction as Trend Factor
sigma lognormal
0.0
0.5
1.0
1.5
0
5
10
15
20
Development Year

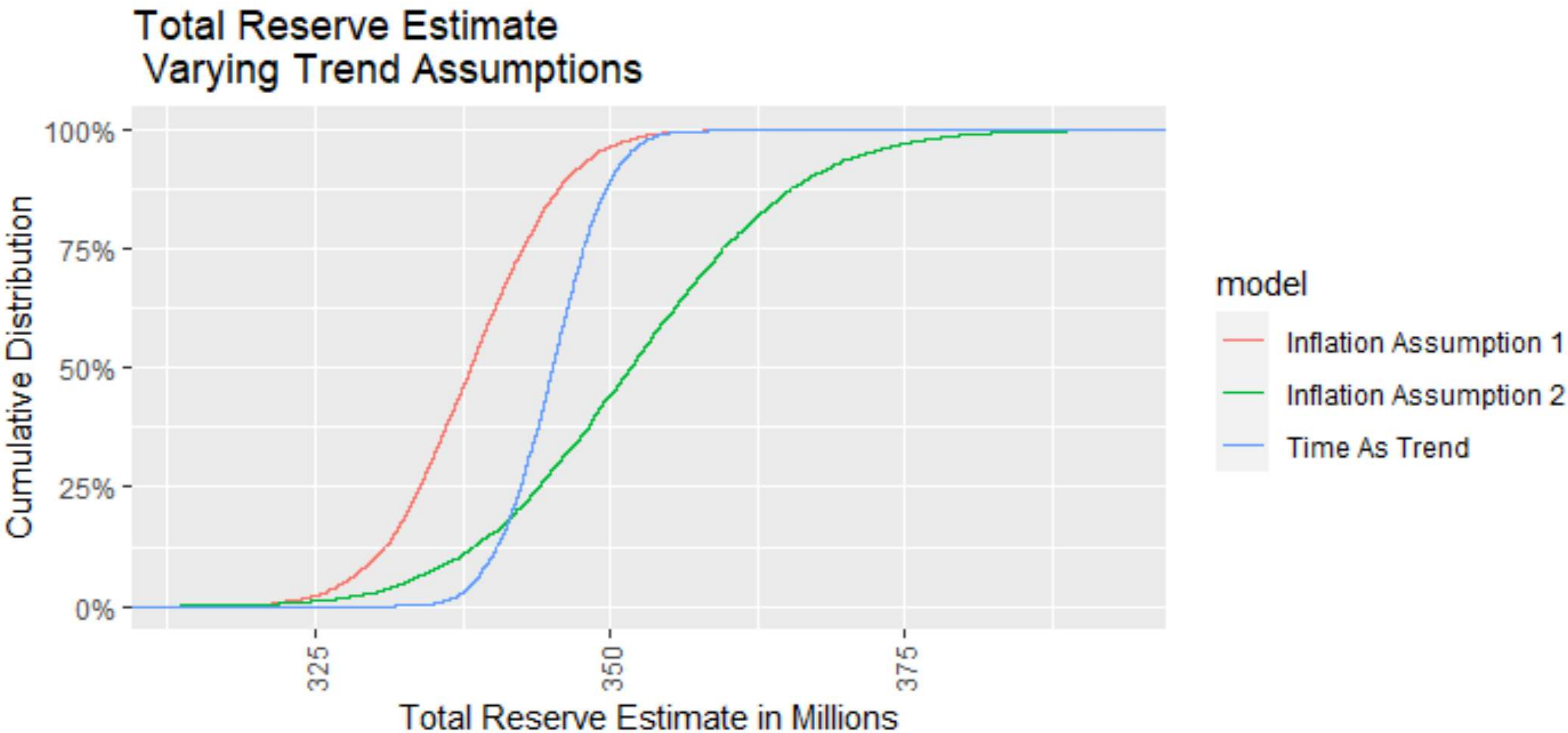

Historical Pure Premium vs. Distribution of Predicted
Time as Trend Factor
Predicted Pure Premium
3000
2000
1000
0
0
1000
2000
3000
Historical Pure Premium


Historical Pure Premium vs. Distribution of Predicted
Inflation Plus Correction as Trend Factor
Predicted Pure Premium
3000
2000
1000
0
0
1000
2000
3000
Historical Pure Premium


Sigma Accident Year Adjusted Intercept
Time as Trend Factor
Accident Year
2020
2015
2010
2005
2000
-0.5
0.0
0.5
Adjustment to Sigma Intercept



Sigma Accident Year Adjusted Intercept
Time in Addition to Inflation as Trend Factor
Accident Year
2020
2015
2010
2005
2000
-0.5
0.0
0.5
Adjustment to Sigma Intercept


Total Reserve Estimate
Varying Trend Assumptions
Cumulative Distribution
100%
75%
50%
25%
0%
325
350
375
Total Reserve Estimate in Millions
model
Inflation Assumption 1
Inflation Assumption 2
Time As Trend

Comments on Case I with Constant Inflation Assumption & No Change in Operation

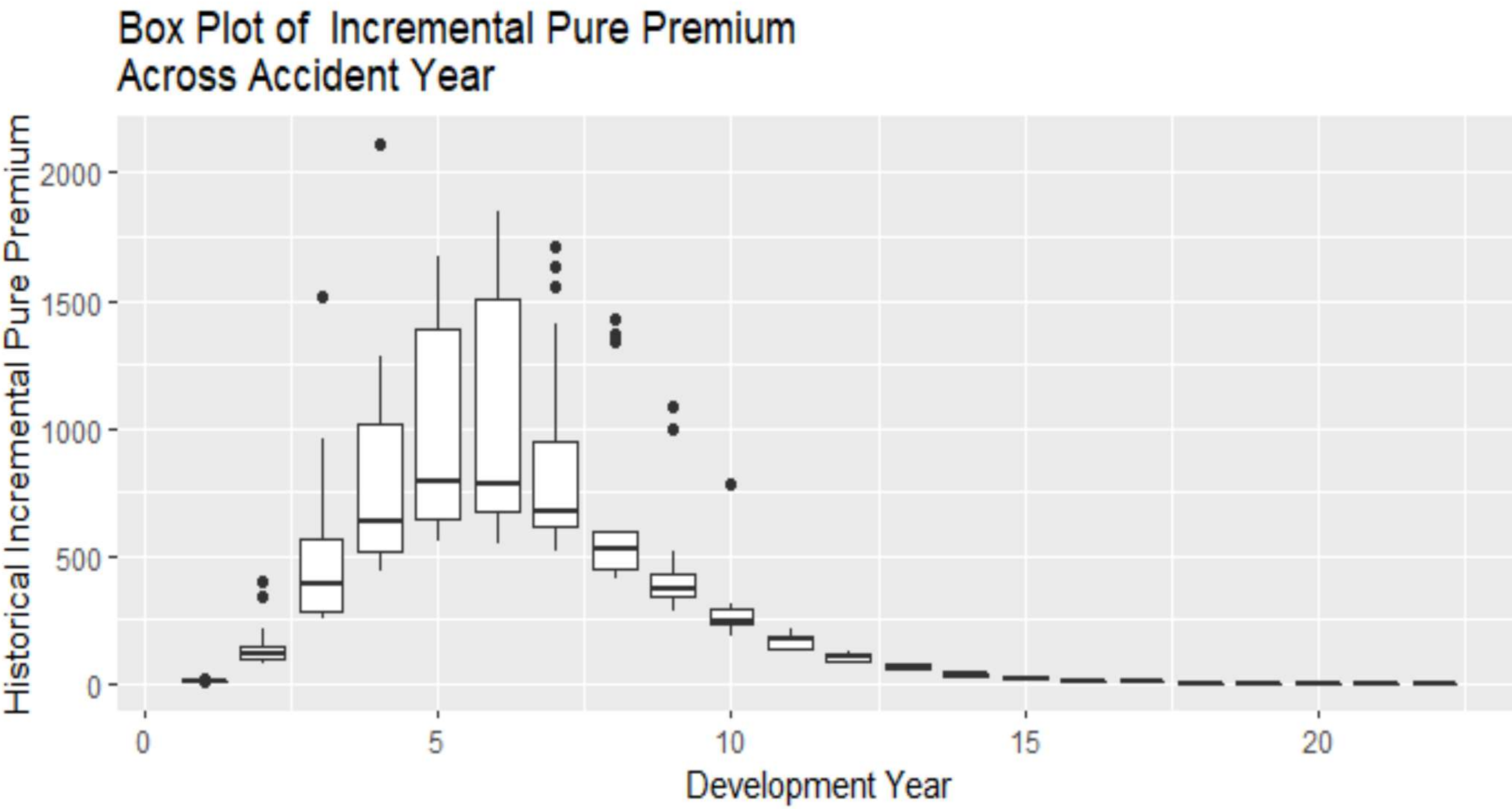
- Deflating the data to model let's one choose future inflation assumptions
- Varying inflation assumptions let's one explicitly account for changes in economy
- If one can assume no change in inflation and no change in operation time as trend factor gives similar fit
- Will retain the deflating option going forward in more complex examples

Case I With Change in Inflation , Change in Claim Operation & Change in Undewriting

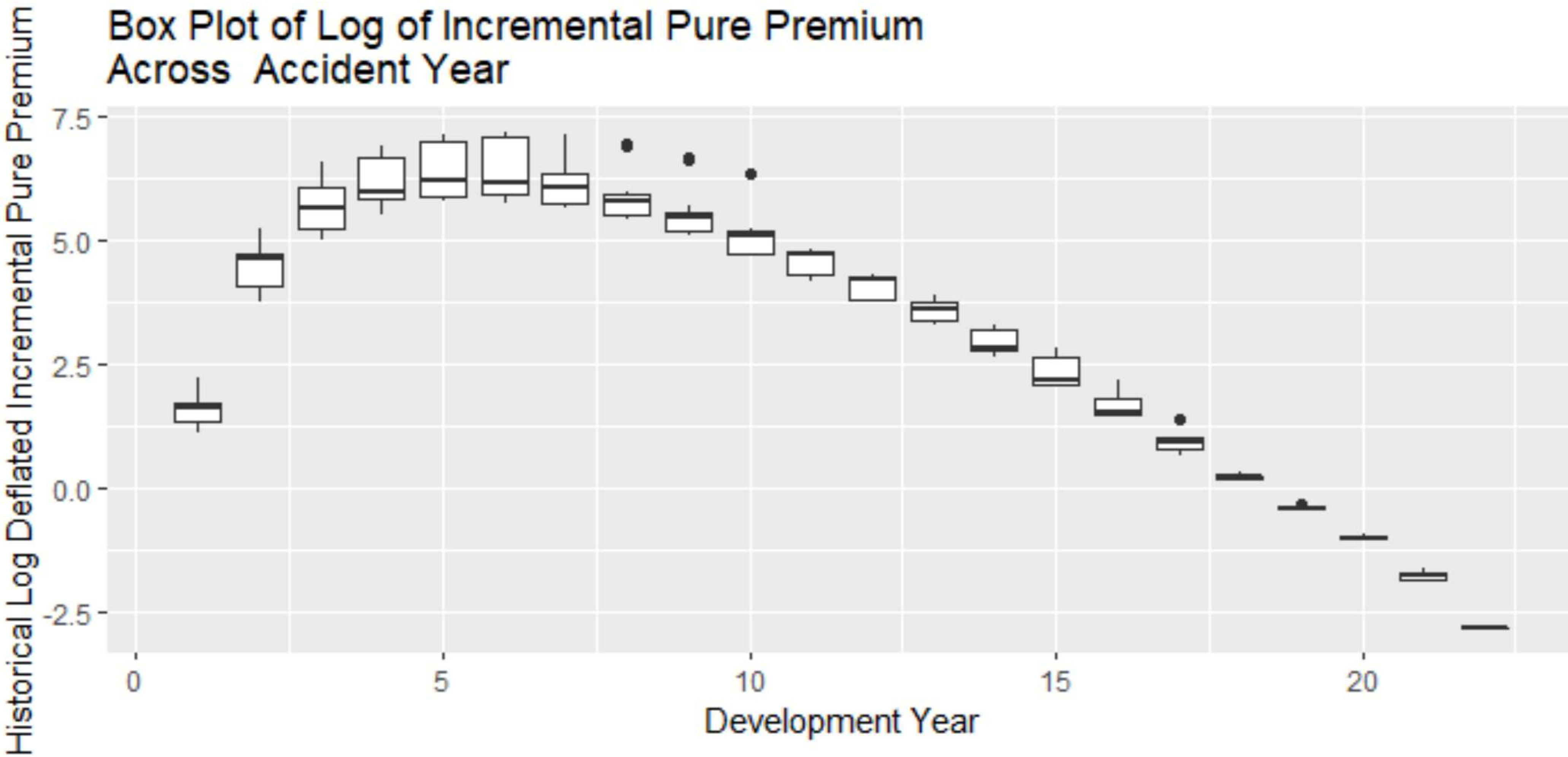
Case I with Operation & Inflation Change

- Data Set Changes
  - Introduce Calendar Year Shift in simulation parameters
    - Calendar Year Less than 2010
    - Calendar Year Less than 2017
  - Introduce Accident Year shift in simulation Parameters
    - Accident Year Less than 2018
    - Accident Year Equal to 2018
    - Accident Year Greater than2018
  - Introduce inflation change
    - Calendar year 2019 shifts to
    - Lognormal: mu .07 & sigma .02

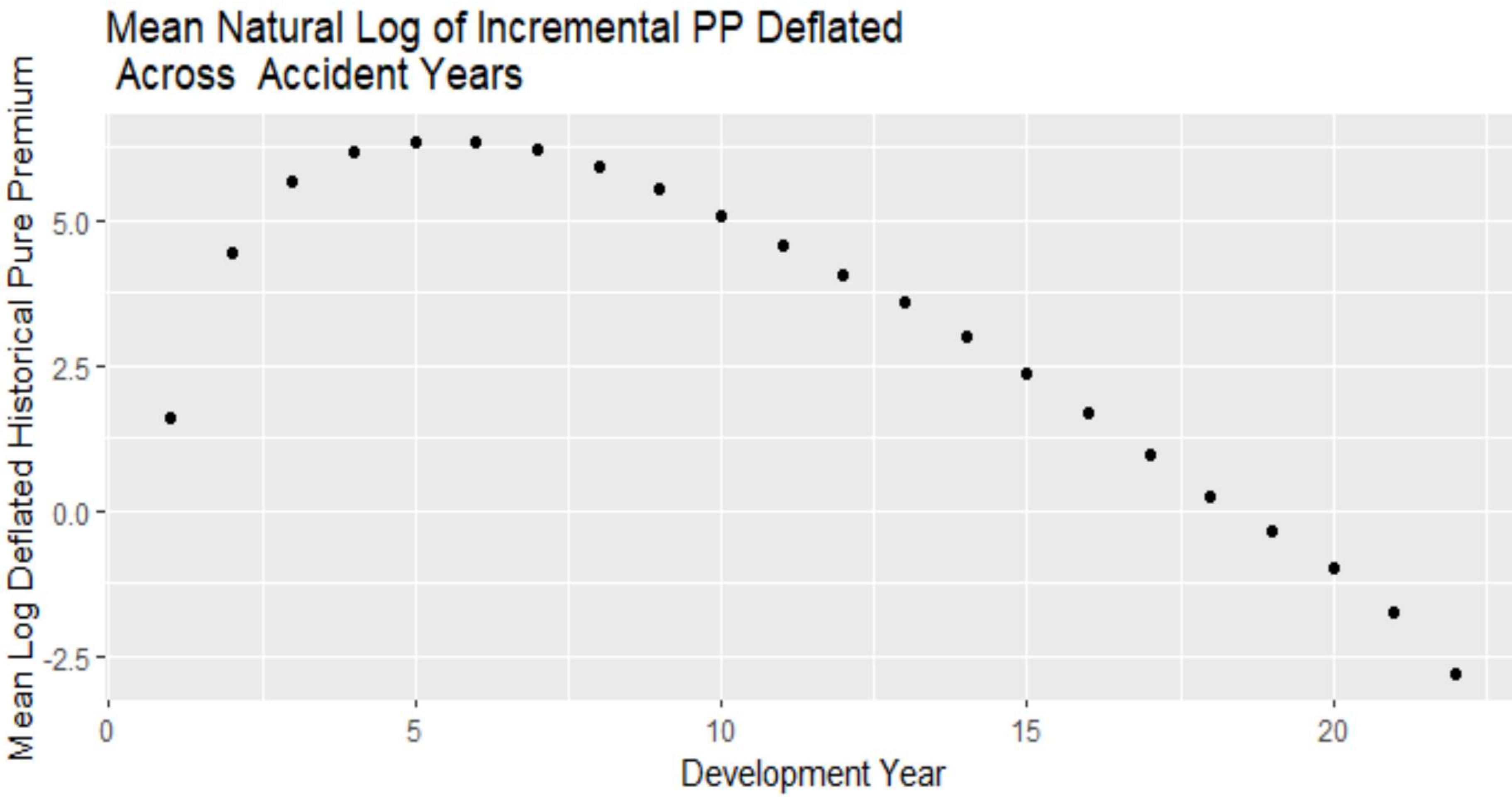
- Model Changes
  - Introduce Calendar Year group categories variable
  - Introduce Accident Year group categories variable
  - Inflation is handled by the effect of deflating data
    - Question of what inflation assumption to use going forward for reserve projections?
    - Look at different assumptions

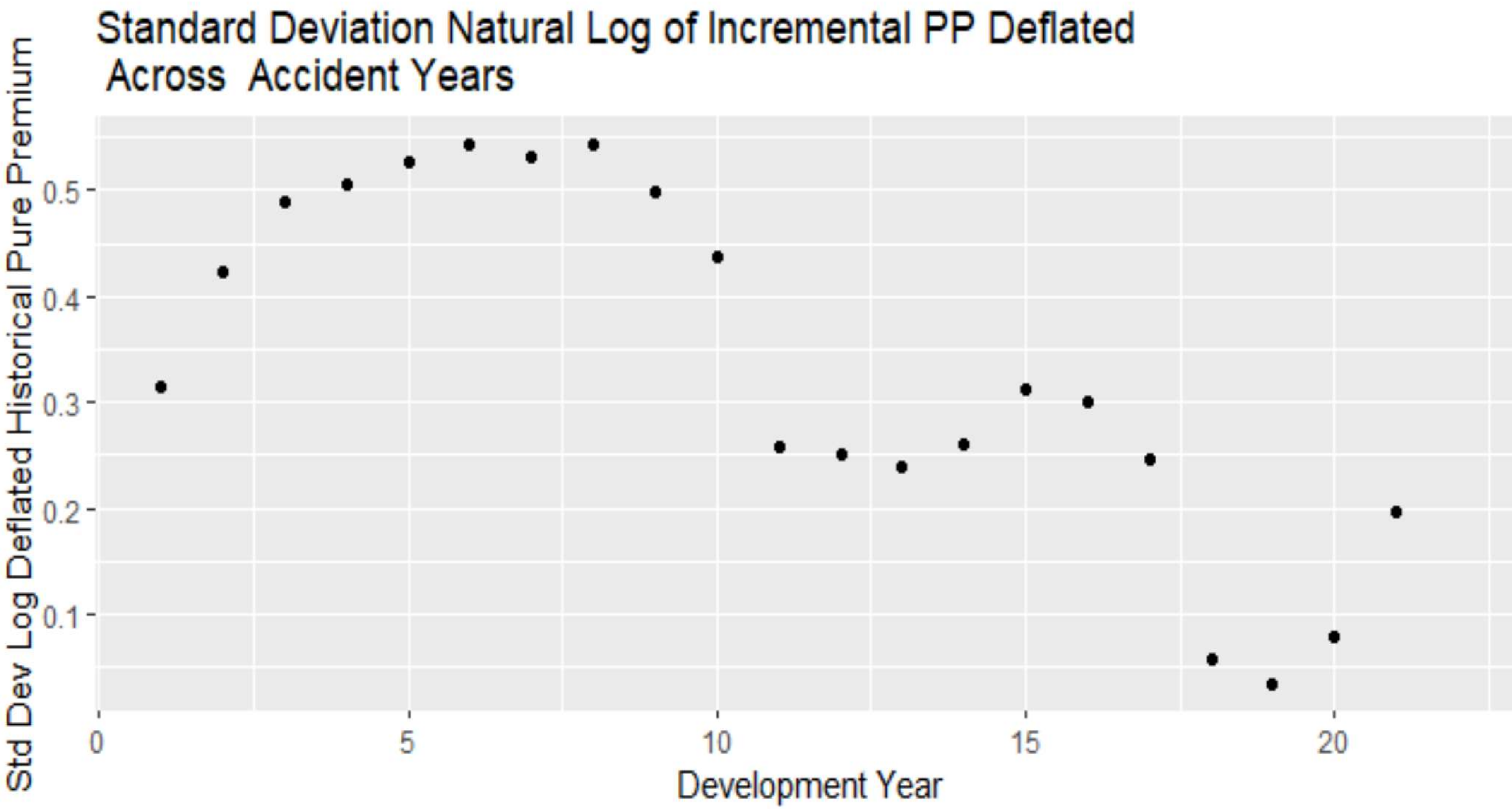


Box Plot of Incremental Pure Premium
Across Accident Year
Historical Incremental Pure Premium
2000
1500
1000
500
0
0
5
10
15
20
Development Year

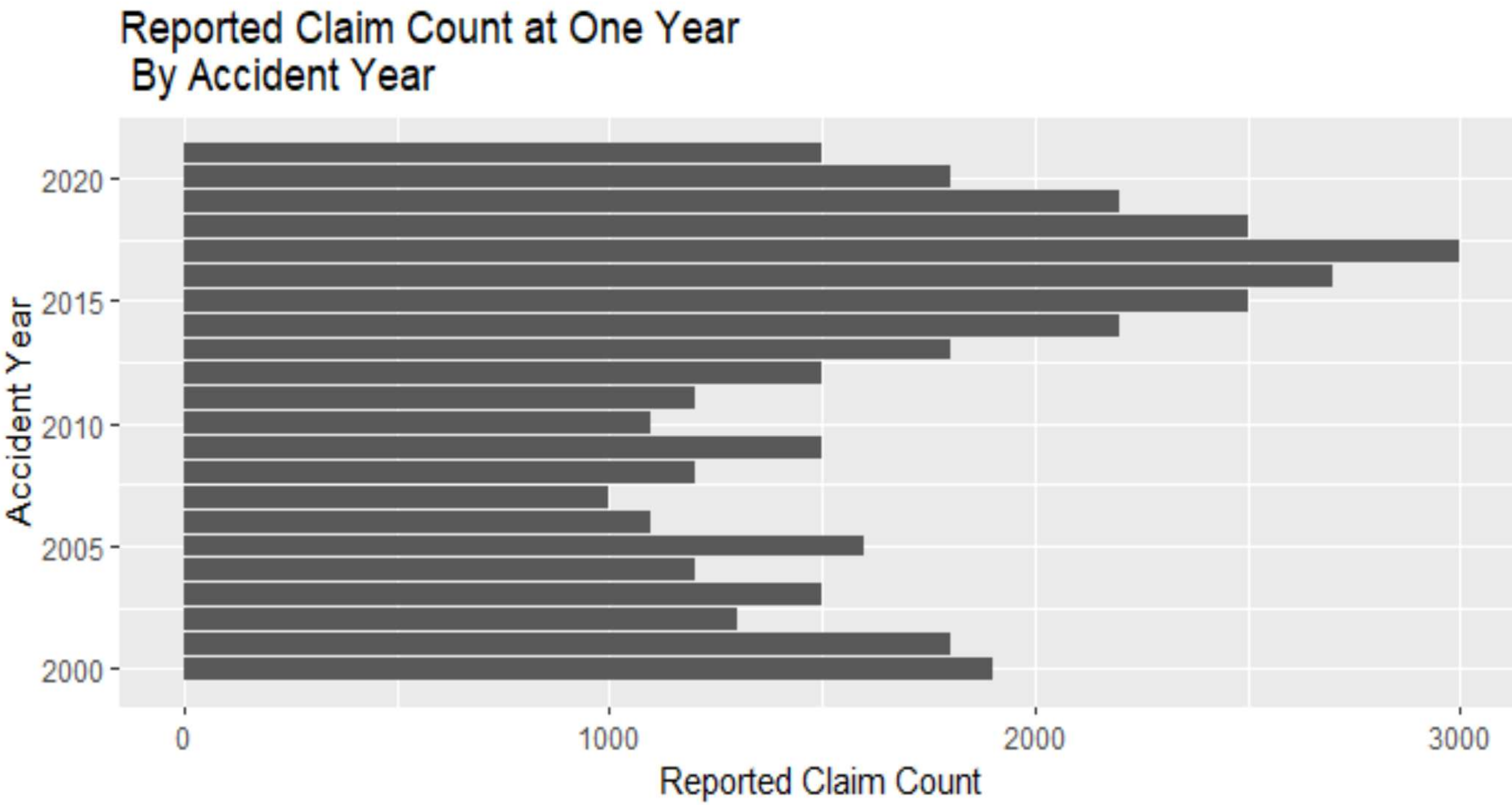


Box Plot of Log of Incremental Pure Premium
Across Accident Year
Historical Log Deflated Incremental Pure Premium
7.5
5.0
2.5
0.0
-2.5
0
5
10
15
20
Development Year


Mean Natural Log of Incremental PP Deflated
Across Accident Years
Mean Log Deflated Historical Pure Premium
5.0
2.5
0.0
-2.5
0
5
10
15
20
Development Year



Standard Deviation Natural Log of Incremental PP Deflated
Across Accident Years
Std Dev Log Deflated Historical Pure Premium
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20
Development Year


Reported Claim Count at One Year
By Accident Year
Accident Year
2020
2015
2010
2005
2000
0
1000
2000
3000
Reported Claim Count

BRMS Code for No Operation Change Model Priors

```
ln_prior_operation_1 <- c(prior(normal(-3,.5),class=b, coef=Intercept),
                    prior(normal(1,1),class=b, coef=Dev_Yr_2_Factor2),
                    prior(normal(2,1),class=b, coef=Dev_Yr_2_Factor3),
                    prior(normal(.01,.01),class=b, coef=Cal_Yr_Time),
                    prior(normal(1,1),class=b, coef=Dev_Yr_6_Cap),
                    prior(normal(-.5,.5),class=b, coef=Dev_Yr_6_Cap_Sqrd),
                    prior(normal(-.5,.5),class=b, coef=Dev_Yr_8_Spline_Sqrd),
                    prior(normal(.5,.5),class=b, coef=Dev_Yr_8_Spline),
                    prior(normal(-0,.5),class=b, coef=Dev_Yr_12_Spline_Sqrd),
                    prior(normal(0,.5),class=b, coef=Dev_Yr_12_Spline),
                    prior(normal(.4,.5),class=b, coef=Intercept,dpar=sigma),
                    prior(normal(0,.5),class=b, coef=Dev_Yr_1_Factor2,dpar=sigma),
                    prior(normal(-.5,1),class=b, coef=Ln_Dev_Yr,dpar=sigma),
                    prior(normal(.5,1),class=b, coef=Dev_Yr_13_Spline,dpar=sigma))
```

BRMS Code for No Operation Change Model

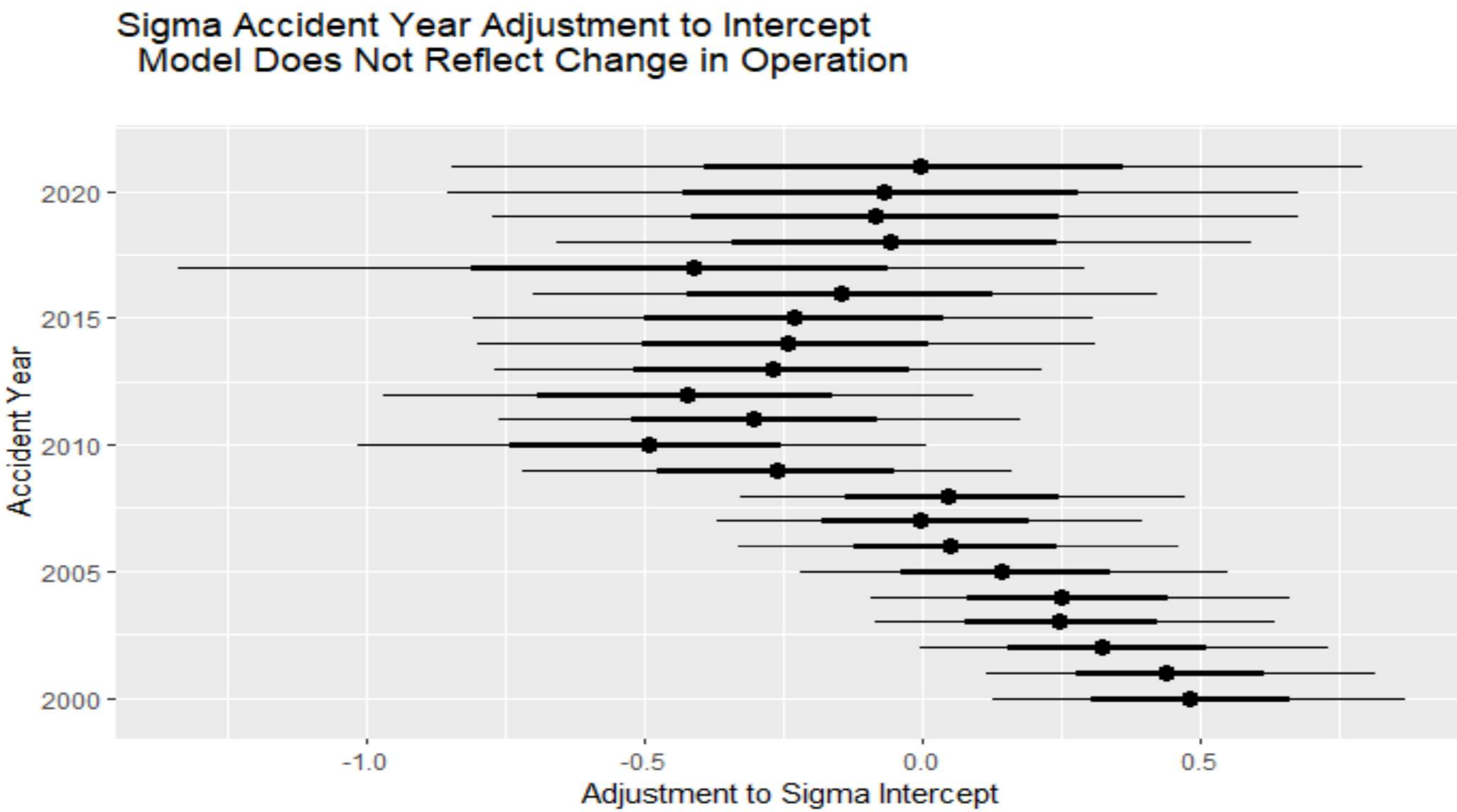
```
lognormal_pp_operation_II_1 <-brm(bf(Trended_Incr_PP_Def ~ 0 +Intercept +
                                   Dev_Yr_6_Cap+
                                   Dev_Yr_6_Cap_Sqrd+
                                   Dev_Yr_2_Factor +
                                   Dev_Yr_8_Spline +
                                   Dev_Yr_8_Spline_Sqrd +
                                    Dev_Yr_12_Spline +
                                    Dev_Yr_12_Spline_Sqrd +
                                   Cal_Yr_Time +
                                    (1||Acc_Yr),
                                 sigma ~0 + Intercept +
                                   Dev_Yr_1_Factor+
                                   Dev_Yr_13_Spline +
                                   Ln_Dev_Yr +
                                   (1||Acc_Yr)),
                               seed = 8603529,
                               data = Train_Triangle_All_Operation_II,
                               family =lognormal(),
                               prior =ln_prior_operation_1)
```

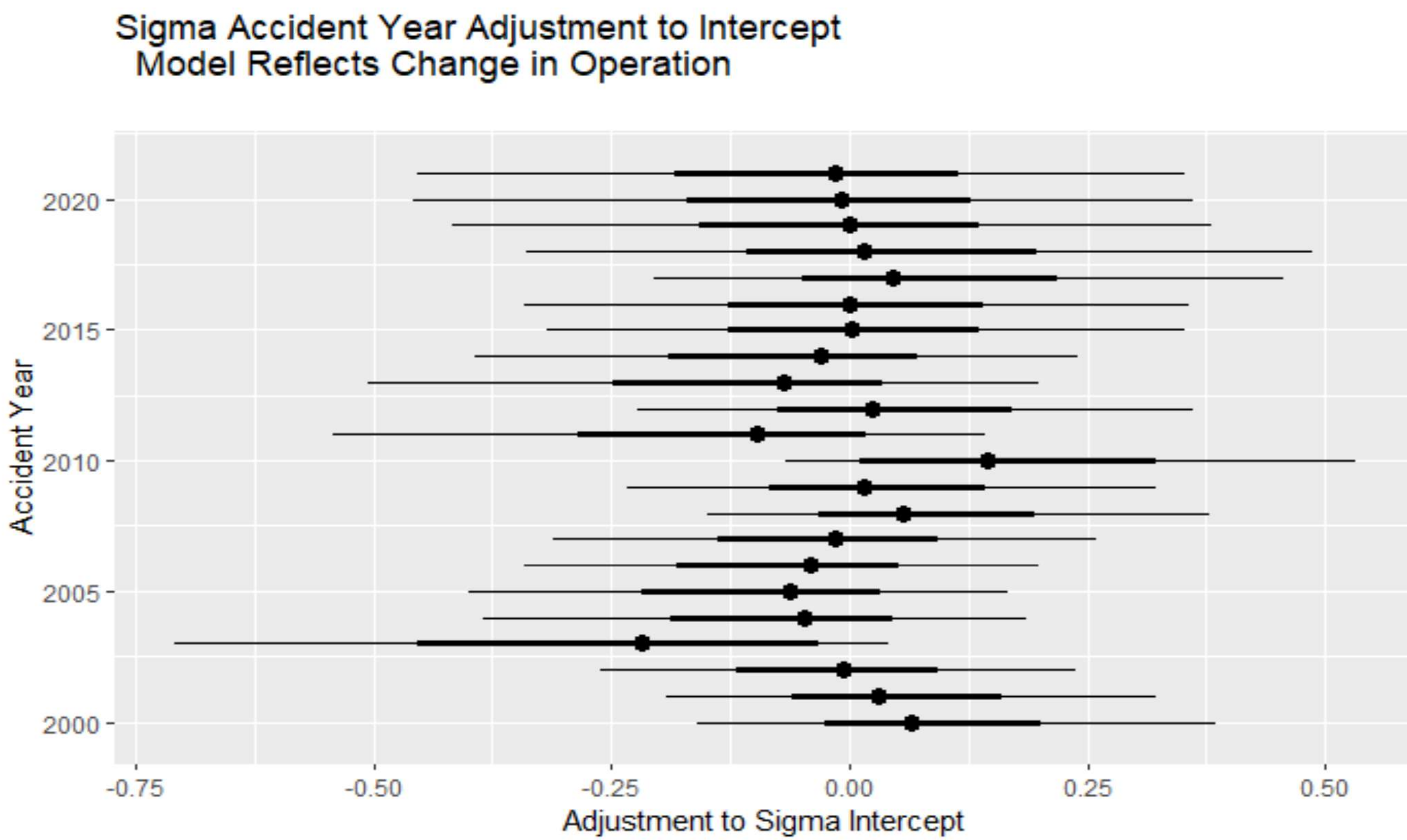
BRMS Code With Operation Change Model Priors

```
ln_prior_operation_1 <- c(prior(normal(-3,.5),class=b, coef=Intercept),
                          prior(normal(1,1),class=b, coef=Dev_Yr_2_Factor2),
                          prior(normal(2,1),class=b, coef=Dev_Yr_2_Factor3),
                          prior(normal(.01,.01),class=b, coef=Cal_Yr_Time),
                          prior(normal(1,1),class=b, coef=Dev_Yr_6_Cap),
                          prior(normal(-.5,.5),class=b, coef=Dev_Yr_6_Cap_Sqrd),
                          prior(normal(-.5,.5),class=b, coef=Dev_Yr_8_Spline_Sqrd),
                          prior(normal(.5,.5),class=b, coef=Dev_Yr_8_Spline),
                          prior(normal(-0,.5),class=b, coef=Dev_Yr_12_Spline_Sqrd),
                          prior(normal(0,.5),class=b, coef=Dev_Yr_12_Spline),
                          prior(normal(.4,.5),class=b, coef=Intercept,dpar=sigma),
                          prior(normal(0,.5),class=b, coef=Dev_Yr_1_Factor2,dpar=sigma),
                          prior(normal(-.5,1),class=b, coef=Ln_Dev_Yr,dpar=sigma),
                          prior(normal(.5,1),class=b, coef=Dev_Yr_13_Spline,dpar=sigma))
```

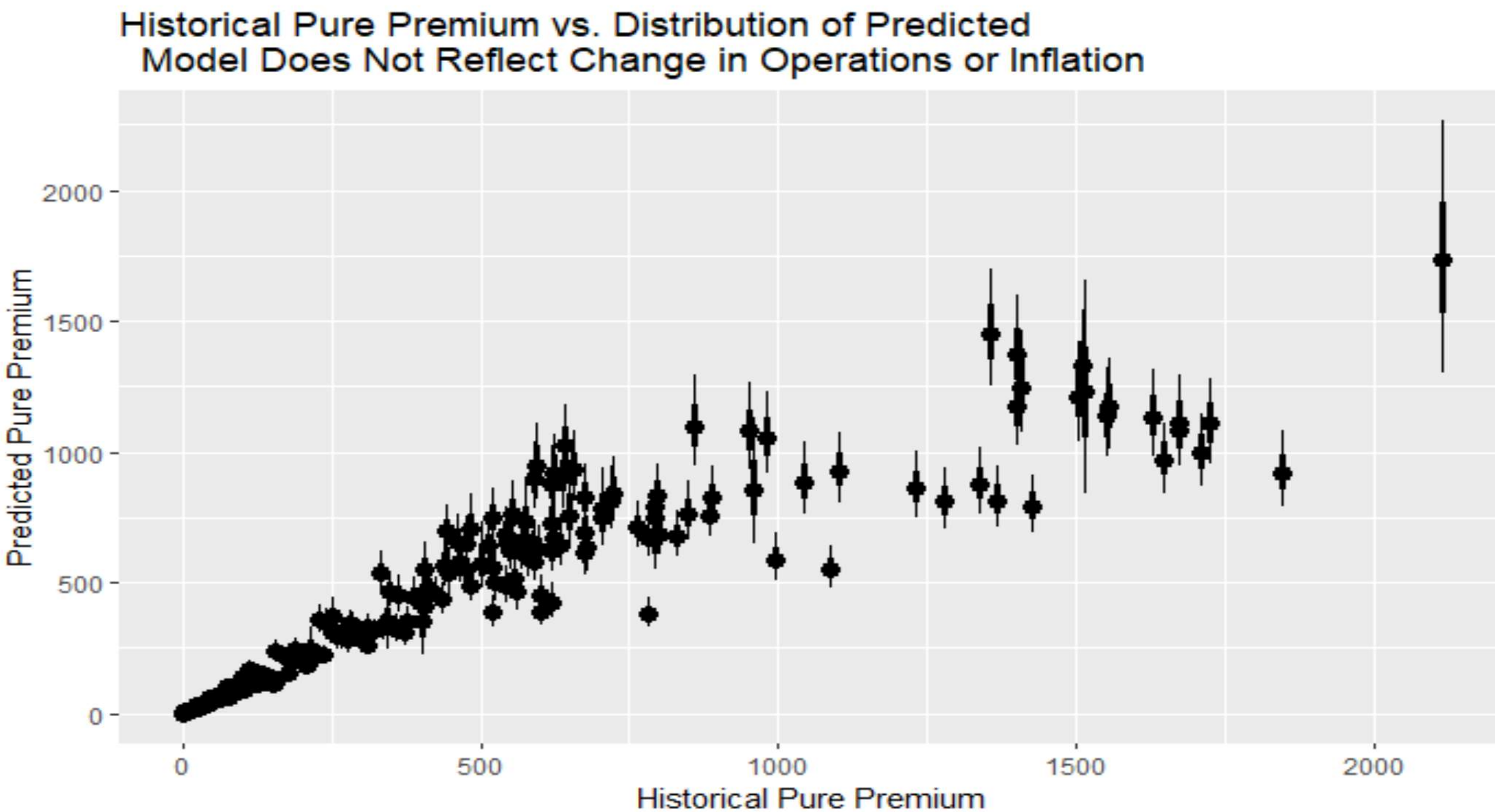
BRMS Code for With Operation Change Model

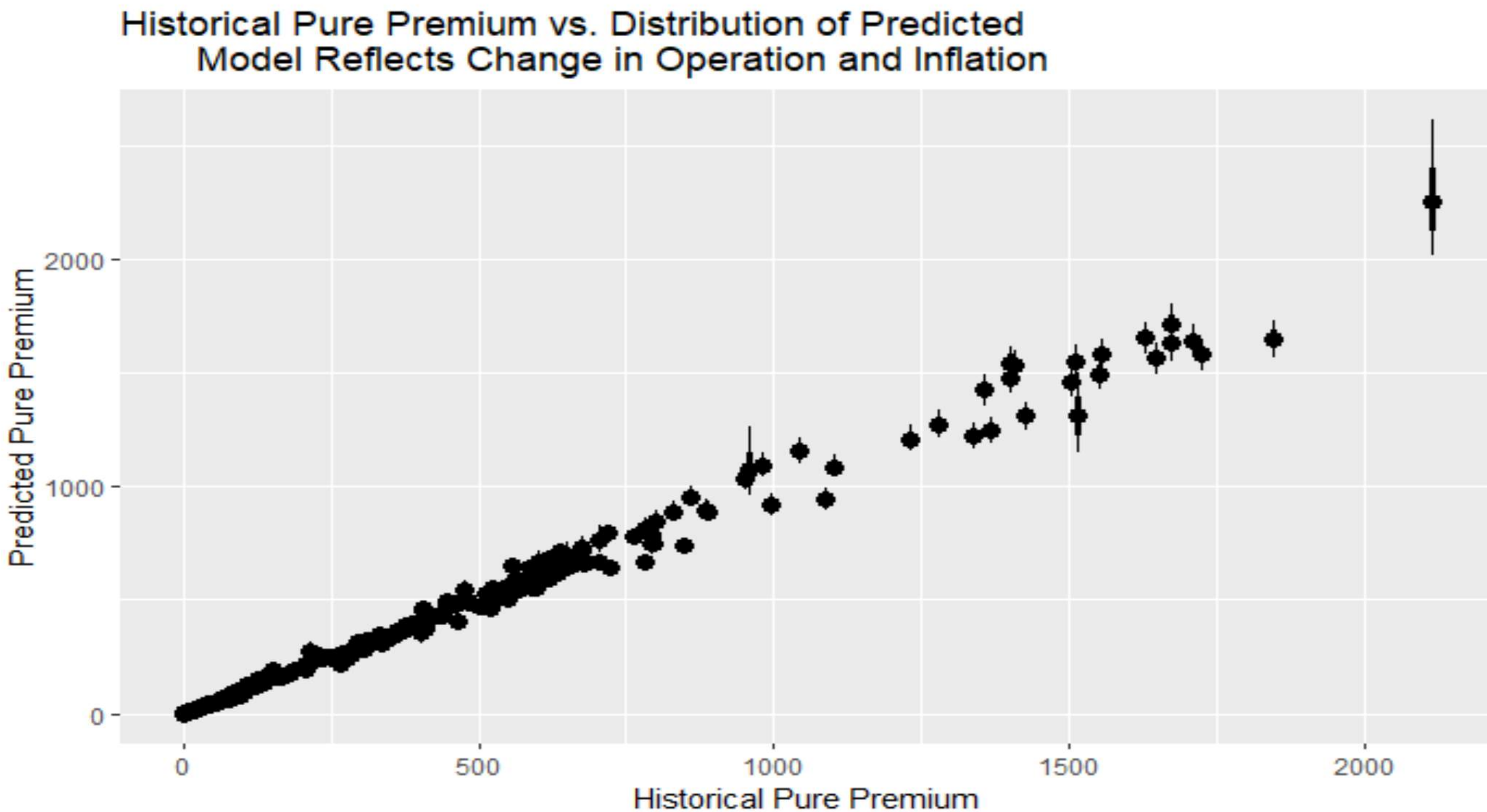
```
lognormal_pp_operation_II_4 <-brm(bf(Trended_Incr_PP_Def ~ 0 +Intercept +
                    Dev_Yr_6_Cap+
                    Dev_Yr_6_Cap_Sqrd+
                    Dev_Yr_2_Factor +
                    Dev_Yr_8_Spline +
                    Dev_Yr_8_Spline_Sqrd +
                    Dev_Yr_12_Spline +
                    Dev_Yr_12_Spline_Sqrd +
                    Cal_Yr_Time +
                    (1||Acc_Yr)+
                    Acc_Yr_Grp_2+Accident Year Group Categorical Variable
                    Cal_Yr_Grp_3,Calendar Year Group Categorical Variable
                   sigma ~0 + Intercept +
                    Dev_Yr_1_Factor+
                    Dev_Yr_13_Spline +
                    Ln_Dev_Yr +
                    (1||Acc_Yr)),
                 seed = 8603529,
                 data = Train_Triangle_All_Operation_II,
                 family =lognormal(),
                 prior =ln_prior_operation_4)
```


Sigma Accident Year Adjustment to Intercept
Model Does Not Reflect Change in Operation
Accident Year
2020
2015
2010
2005
2000
-1.0
-0.5
0.0
0.5
Adjustment to Sigma Intercept



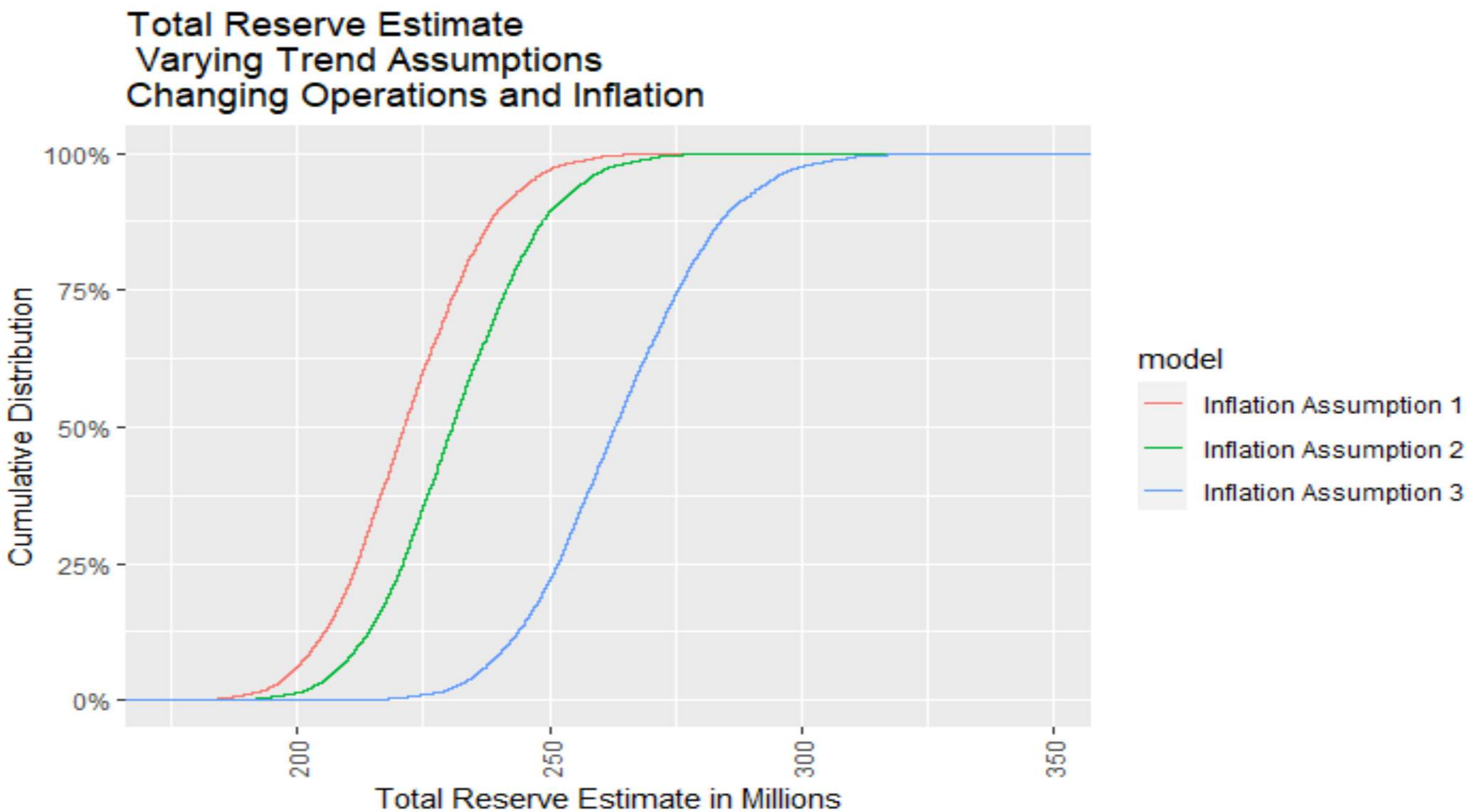
Sigma Accident Year Adjustment to Intercept
Model Reflects Change in Operation
2020
2015
2010
2005
2000
Accident Year
-0.75
-0.50
-0.25
0.00
0.25
0.50
Adjustment to Sigma Intercept


Historical Pure Premium vs. Distribution of Predicted
Model Does Not Reflect Change in Operations or Inflation
Predicted Pure Premium
2000
1500
1000
500
0
0
500
1000
1500
2000
Historical Pure Premium


Historical Pure Premium vs. Distribution of Predicted
Model Reflects Change in Operation and Inflation
Predicted Pure Premium
Historical Pure Premium
0
1000
2000
0
500
1000
1500
2000

Future Inflation Modeling Options

- Inflation Scenario 1
  - Mu = .03
  - Sigma =.02
- Inflation Scenario 2
  - Mu = .04
  - Sigma = .02
- Inflation Scenario 3
  - Mu = .07
  - Sigma = .02


Total Reserve Estimate
Varying Trend Assumptions
Changing Operations and Inflation
100%
75%
50%
25%
0%
Cumulative Distribution
200
250
300
350
Total Reserve Estimate in Millions
model
Inflation Assumption 1
Inflation Assumption 2
Inflation Assumption 3

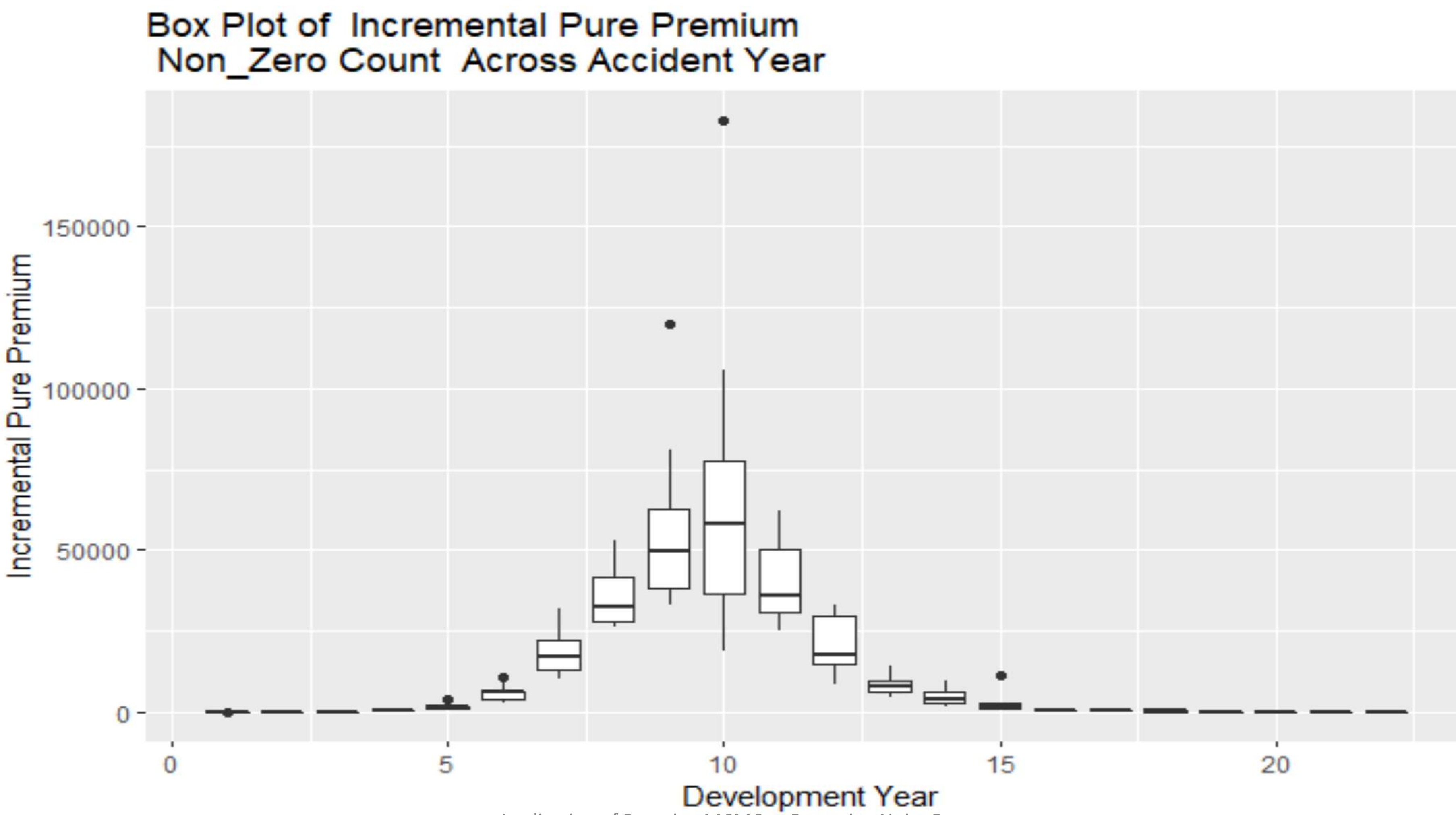
Comments on Case I with Changing Operation & Inflation

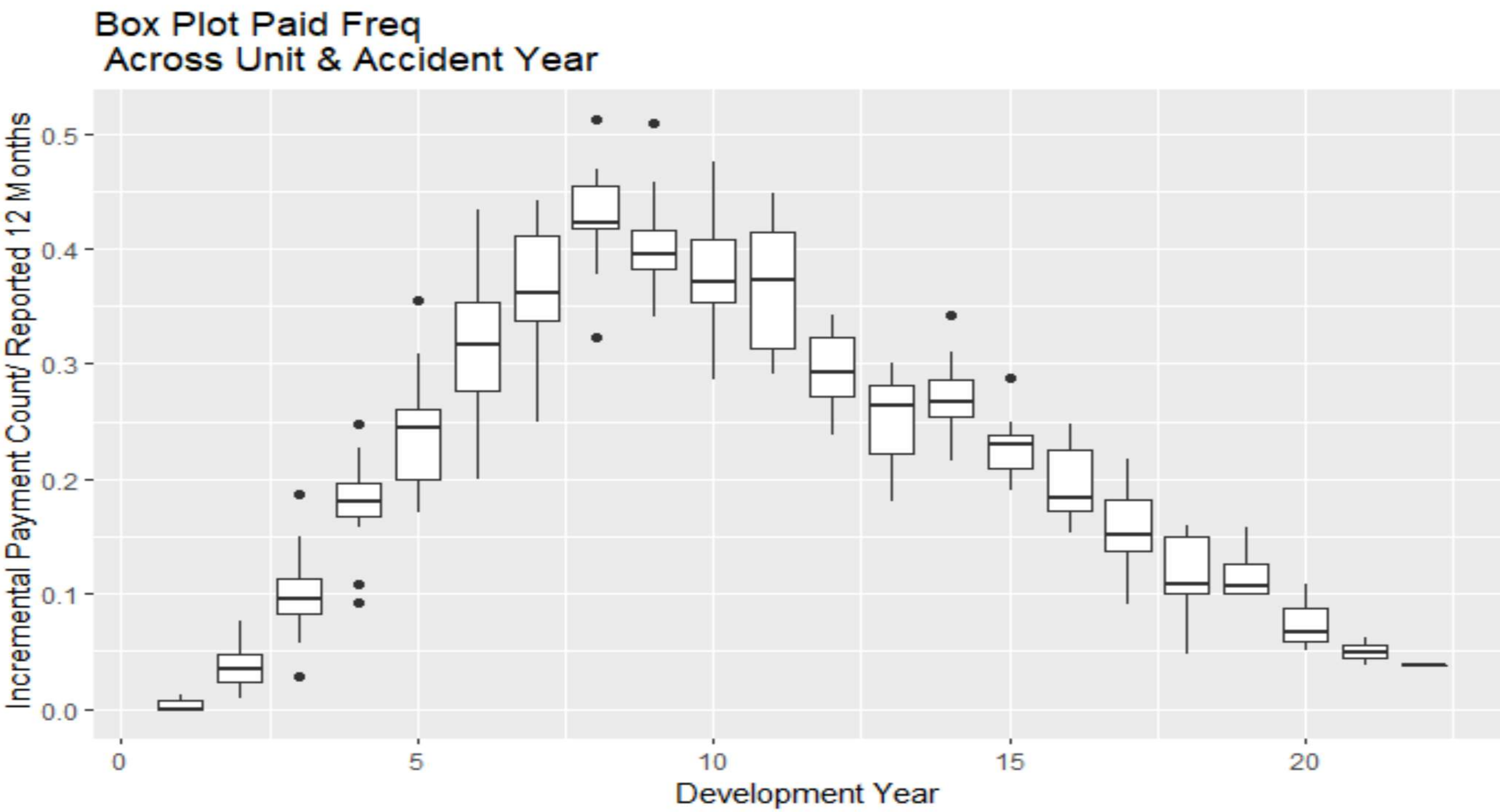
- Categorical variable can recognize shift in operations & improve fit
- No need for another calendar year group starting in 2019 due to inflation shift since deflating data picks that up.
- Assumptions on future economic environment are critical in setting reserves in changing enviroment

Case II: Zero Payments & High Severity (Lognormal)

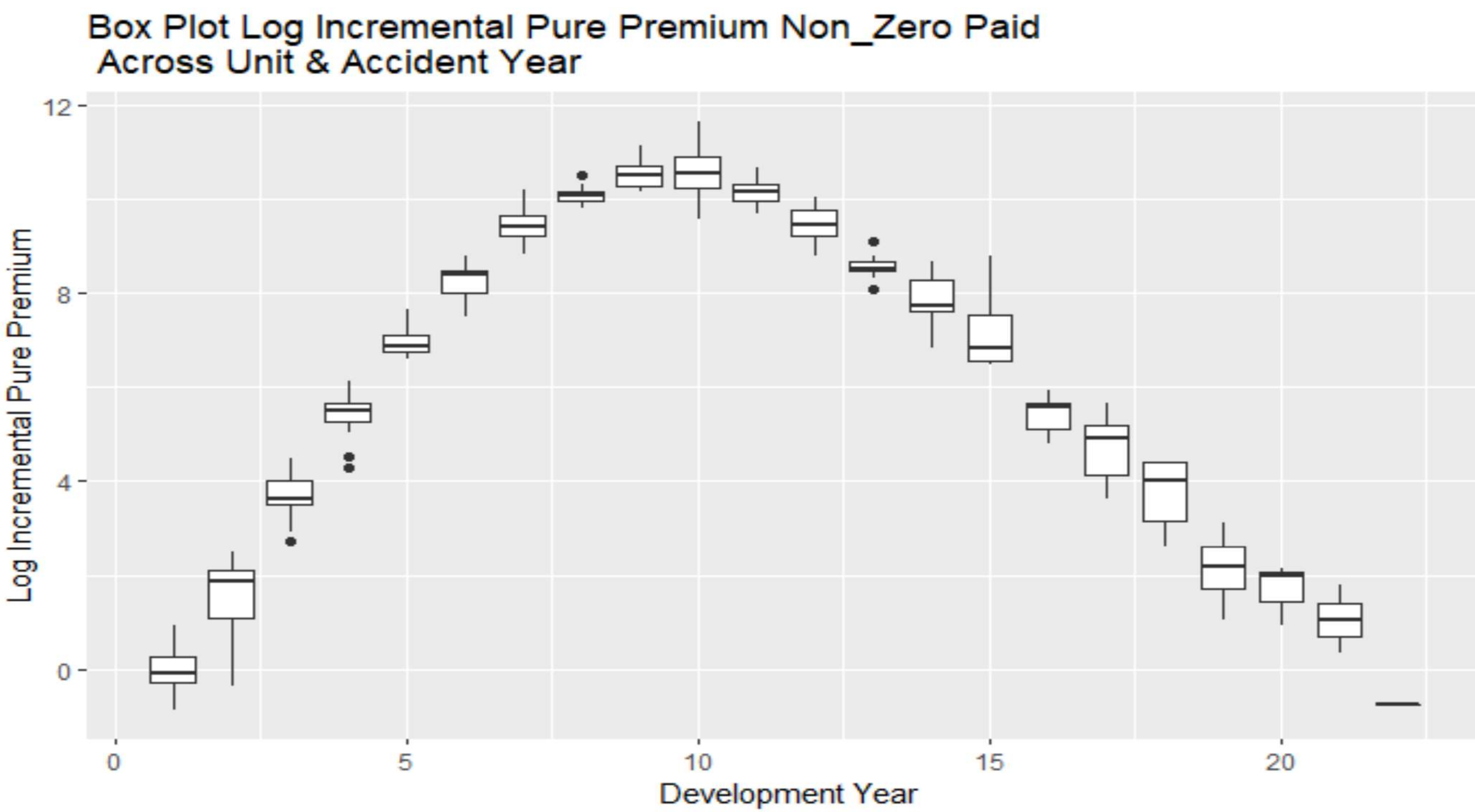
Case II Model Design Comments

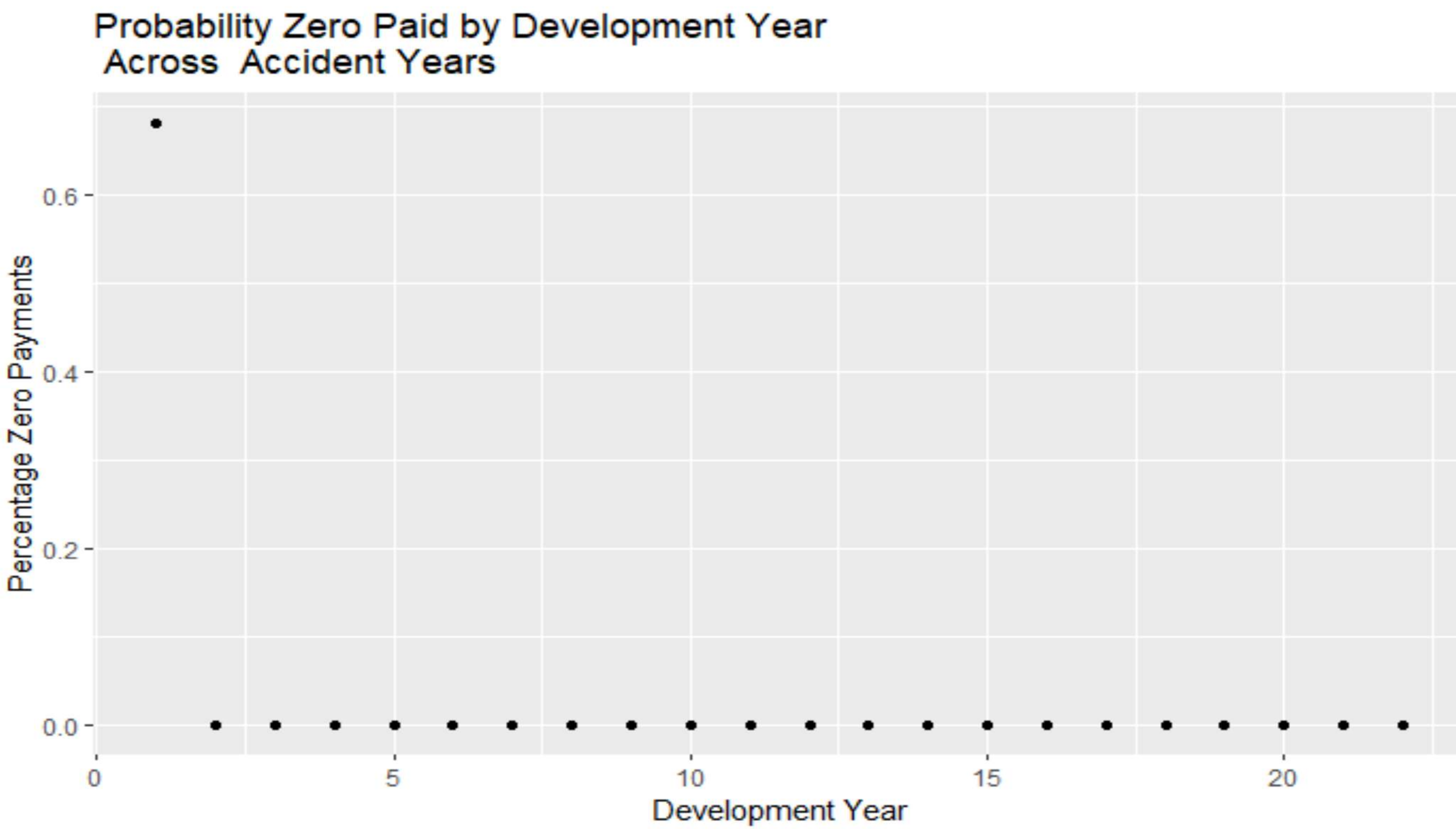
- Illustrates use of hurdle models to account for zero payments
- The density of a hurdle distribution can be specified as follows. If $x=0$ set $f(x)=\theta$. Else set $f(x)=(1-\theta)*g(x)/(1-G(0))$ where $g(x)$ and $G(x)$ are the density and distribution function of the non-hurdle part, respectively.
- This example combines
  - Binomial model for probability of zero payment
  - Lognormal for claim amounts

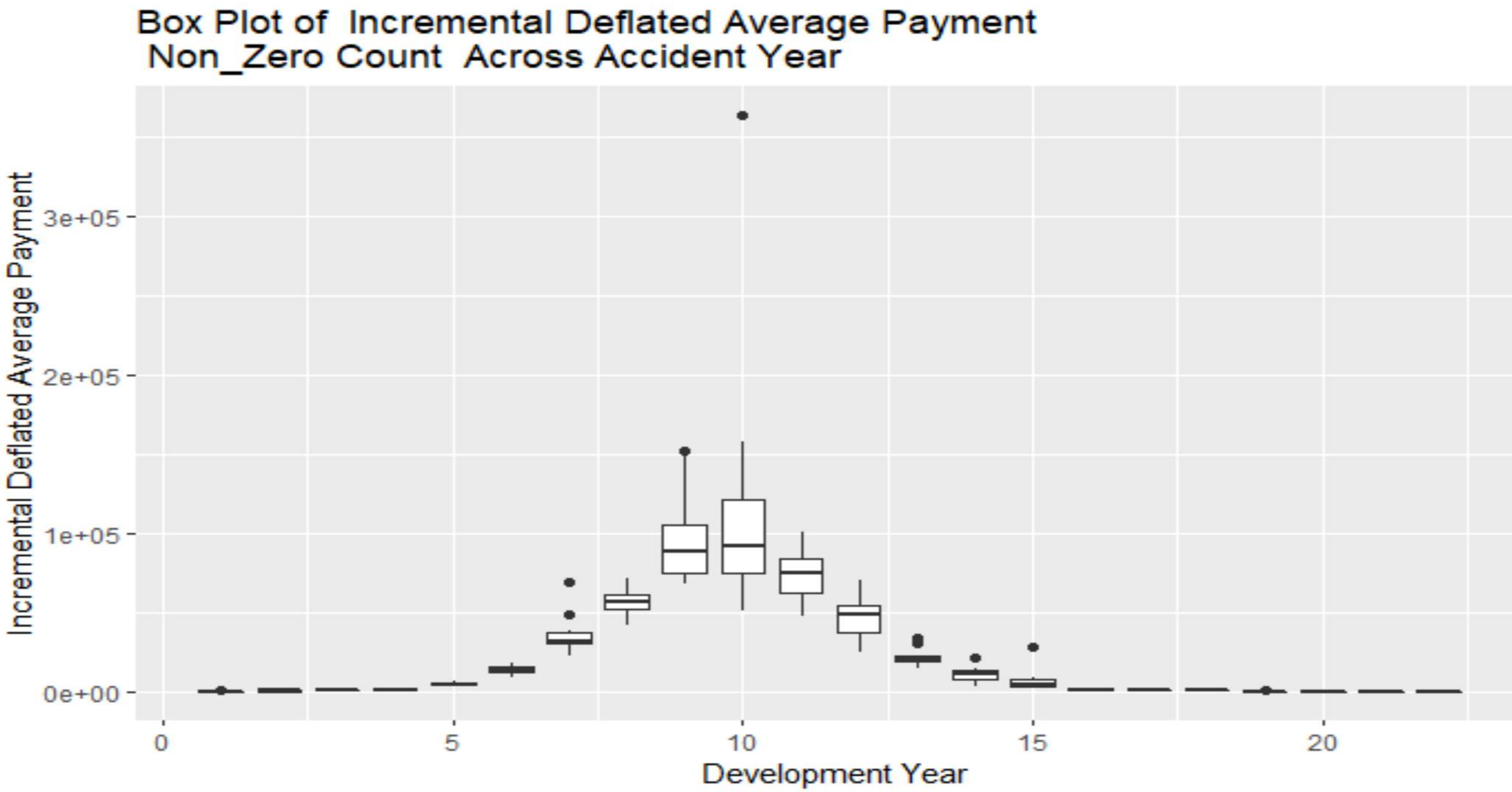

Box Plot of Incremental Pure Premium
Non_Zero Count Across Accident Year
Incremental Pure Premium
150000
100000
50000
0
0
5
10
15
20
Development Year

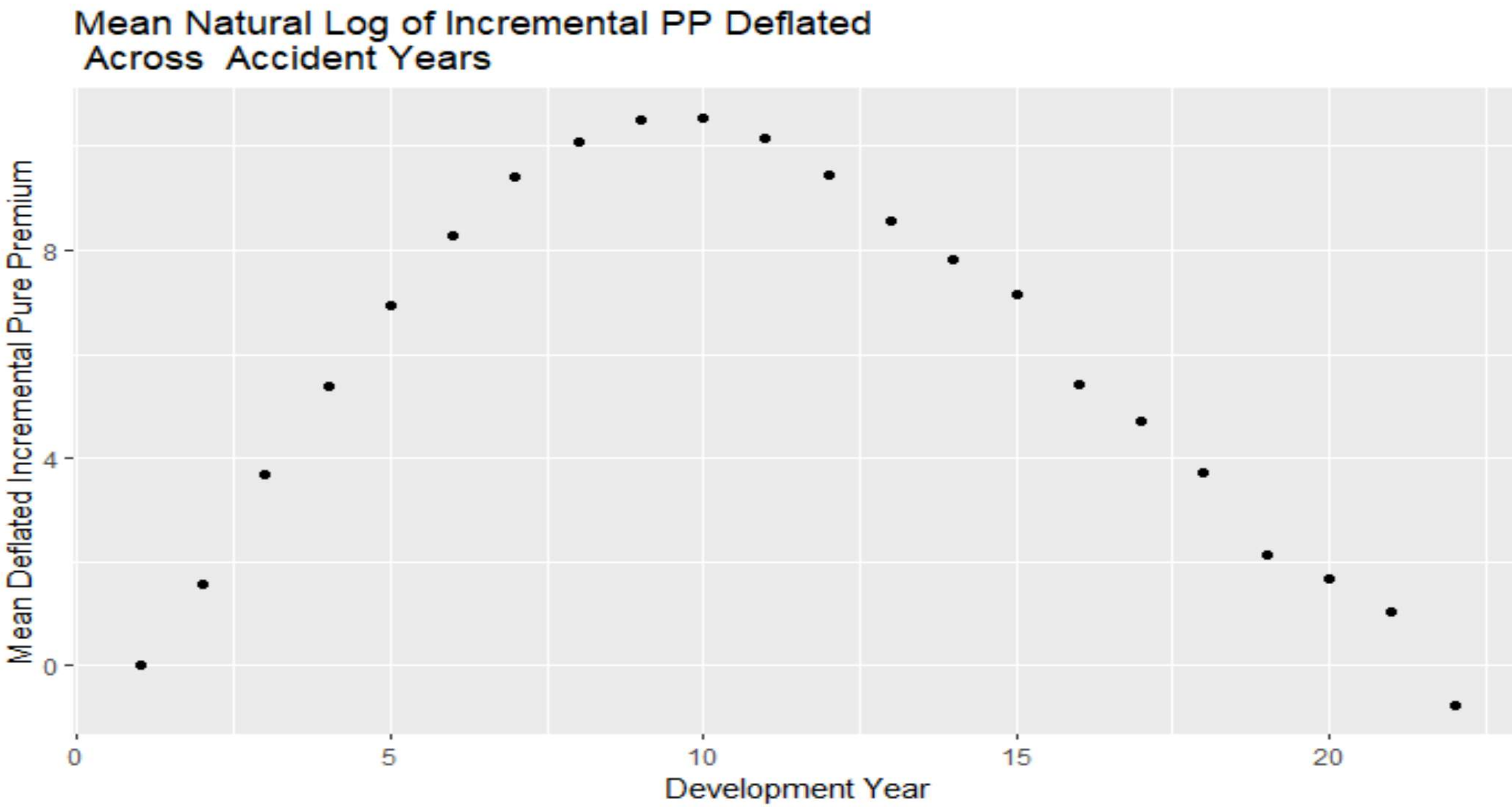


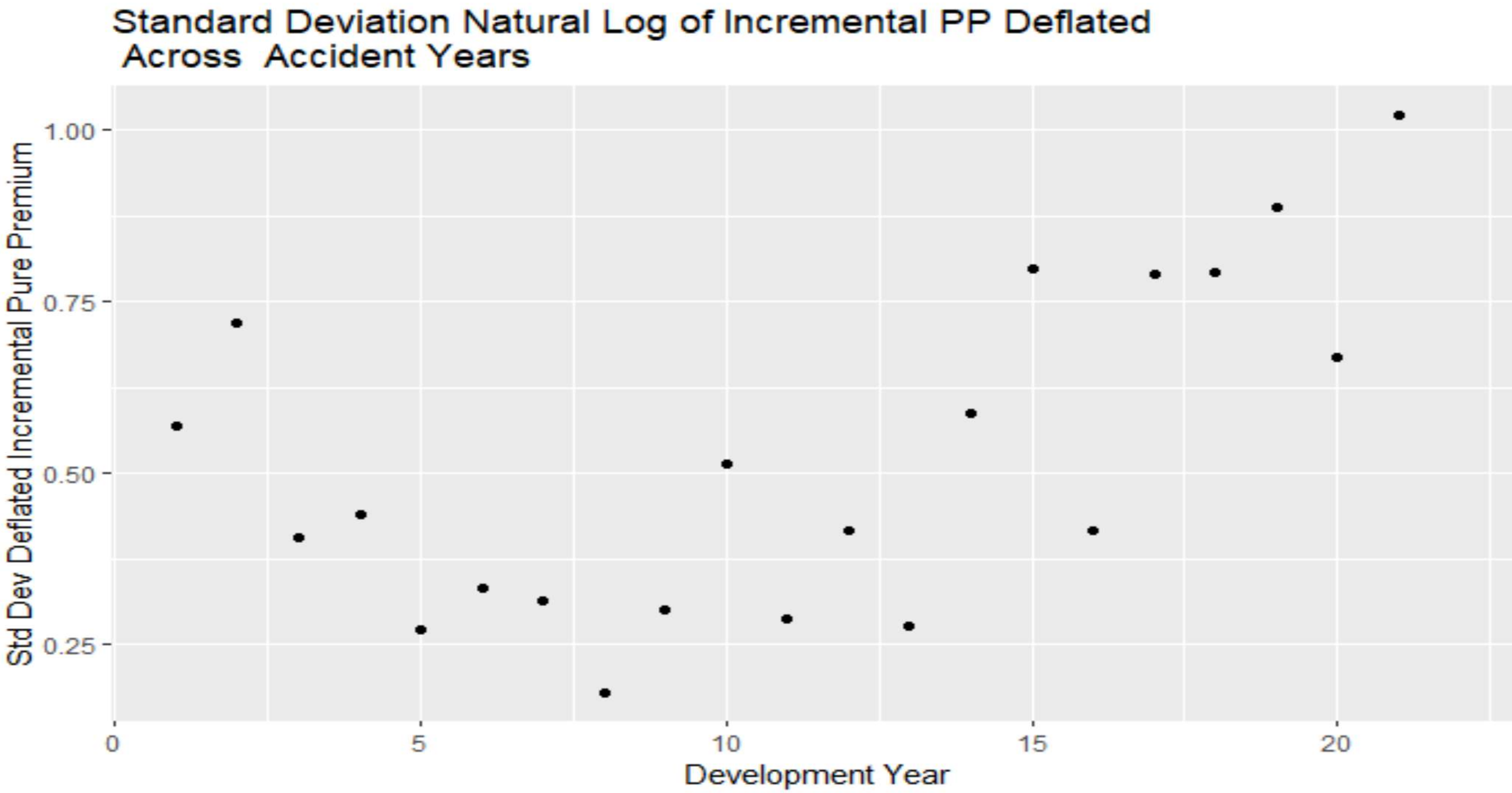
Box Plot Paid Freq
Across Unit & Accident Year
Incremental Payment Count/ Reported 12 Months
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20
Development Year


Box Plot Log Incremental Pure Premium Non_Zero Paid
Across Unit & Accident Year
Log Incremental Pure Premium
Development Year
0
4
8
12
0
5
10
15
20


Probability Zero Paid by Development Year
Across Accident Years
Percentage Zero Payments
0.0
0.2
0.4
0.6
0
5
10
15
20
Development Year


Box Plot of Incremental Deflated Average Payment
Non_Zero Count Across Accident Year
Incremental Deflated Average Payment
3e+05
2e+05
1e+05
0e+00
0
5
10
15
20
Development Year


Mean Natural Log of Incremental PP Deflated
Across Accident Years
Mean Deflated Incremental Pure Premium
8
4
0
0
5
10
15
20
Development Year


Standard Deviation Natural Log of Incremental PP Deflated
Across Accident Years
Std Dev Deflated Incremental Pure Premium
1.00
0.75
0.50
0.25
0
5
10
15
20
Development Year

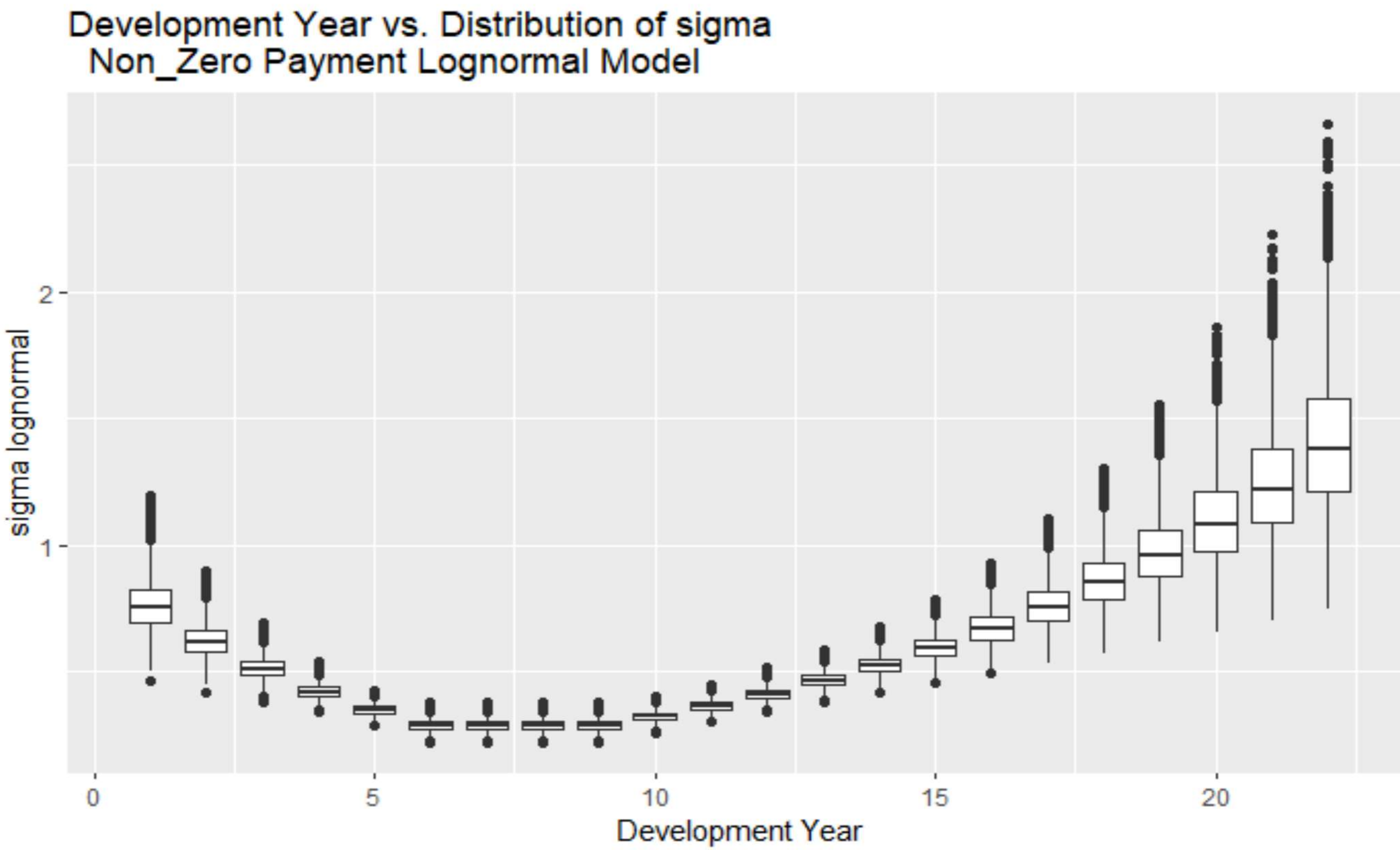
Bayesian Non_Zero Payment Priors

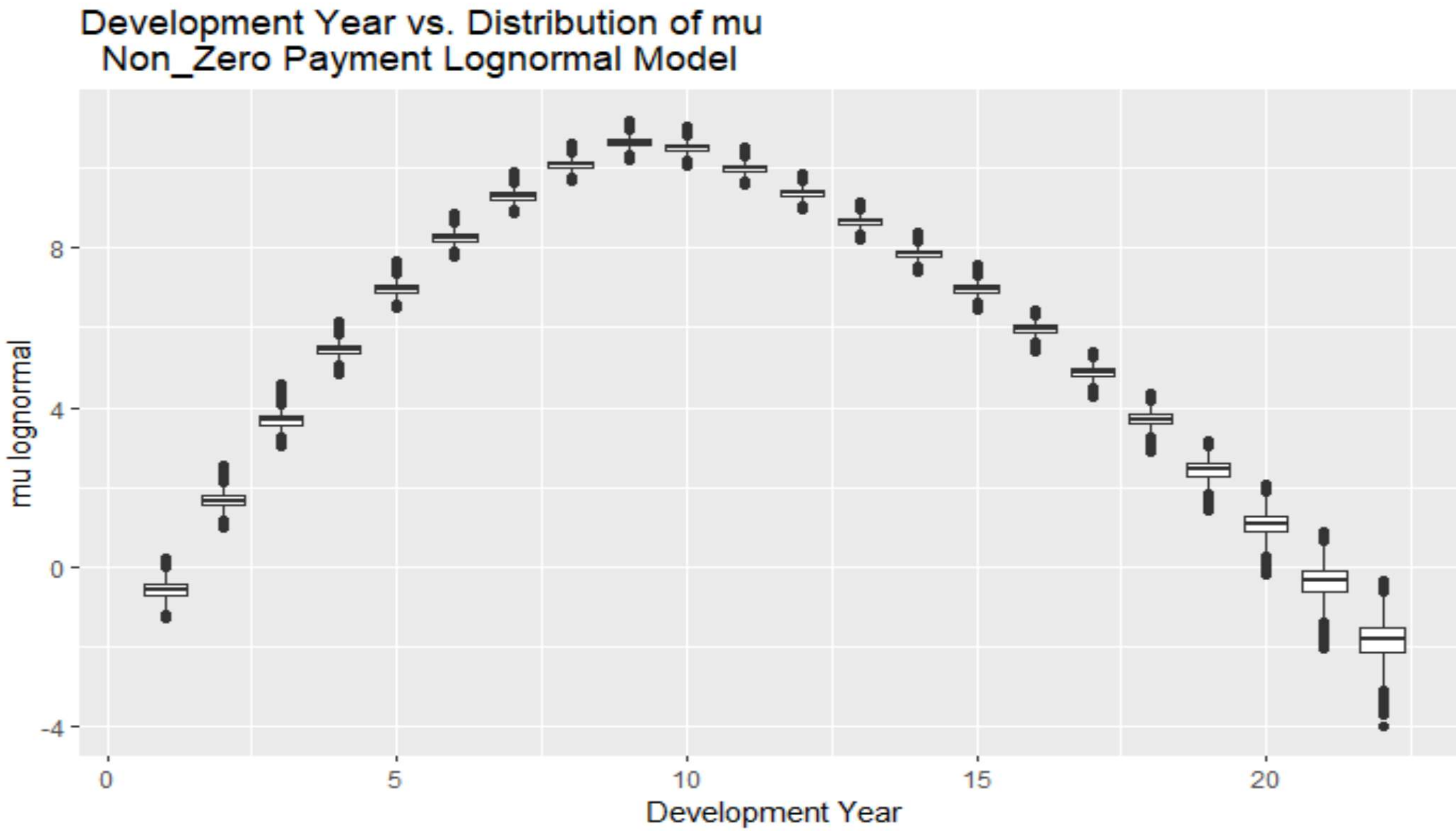
```
ln_prior_non_zero_2 <- c(prior(student_t(3,-4,.5),class=b, coef=Intercept),
                    prior(student_t(3,.03,.01),class=b, coef=Cal_Yr_Time),
                    prior(student_t(3,1,.5),class=b, coef=Dev_Yr_10_Cap),
                    prior(student_t(3,-.5,.5),class=b, coef=Dev_Yr_10_Cap_Sqrd),
                    prior(student_t(3,0,.5),class=b, coef=Dev_Yr_10_Spline),
                    prior(student_t(3,0,.5),class=b, coef=Dev_Yr_10_Spline_Sqrd),
                    prior(student_t(3,.1,.5),class=b, coef=Dev_Yr_6_Cap,dpar=sigma),
                    prior(student_t(3,.1,.5),class=b, coef=Dev_Yr_10_Spline,dpar=sigma))
```

Note use of student_t on prior to deal with outliers

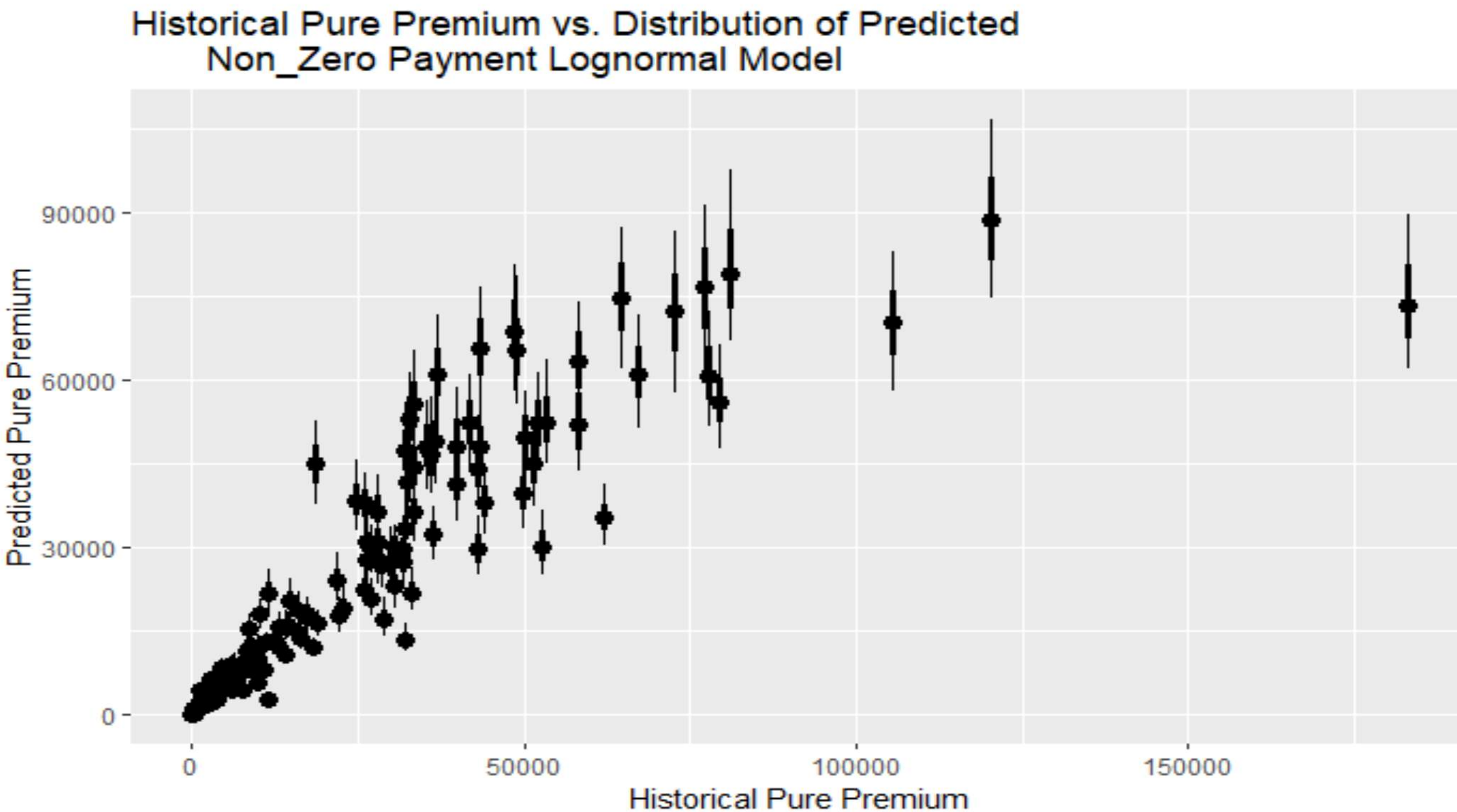
BRMS Model for Non_Zero Payments Lognormal Case

```
lognormal_pp_non_zero_2 <-brm(bf(Trended_Incr_PP_Def ~ 0 +Intercept +
Dev_Yr_10_Cap+
                       Dev_Yr_10_Cap_Sqrd+
                       Dev_Yr_10_Spline +
                       Dev_Yr_10_Spline_Sqrd +

                       Cal_Yr_Time +(1||Acc_Yr),
                      sigma ~
                       Dev_Yr_6_Cap +
                       Dev_Yr_10_Spline ),

                    data = Train_Triangle_Non_Zero_Count,
                    family =lognormal(),
                    prior =ln_prior_non_zero_2)
```


Development Year vs. Distribution of sigma
Non_Zero Payment Lognormal Model
sigma lognormal
2
1
0
5
10
15
20
Development Year



Development Year vs. Distribution of mu
Non_Zero Payment Lognormal Model
mu lognormal
8
4
0
-4
0
5
10
15
20
Development Year


Historical Pure Premium vs. Distribution of Predicted
Non_Zero Payment Lognormal Model
Predicted Pure Premium
90000
60000
30000
0
0
50000
100000
150000
Historical Pure Premium

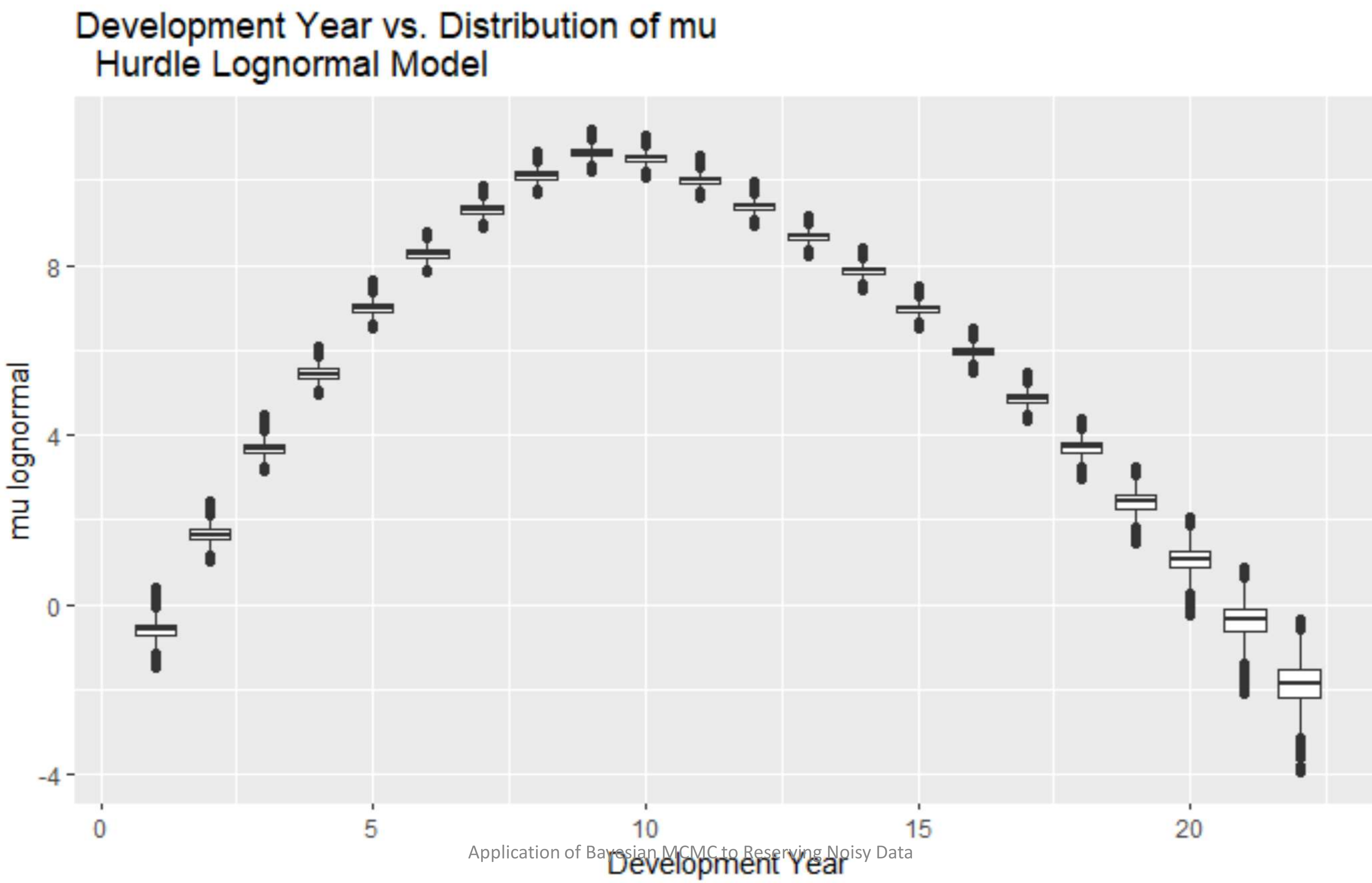
Bayesian Lognormal Hurdle ModelPriors

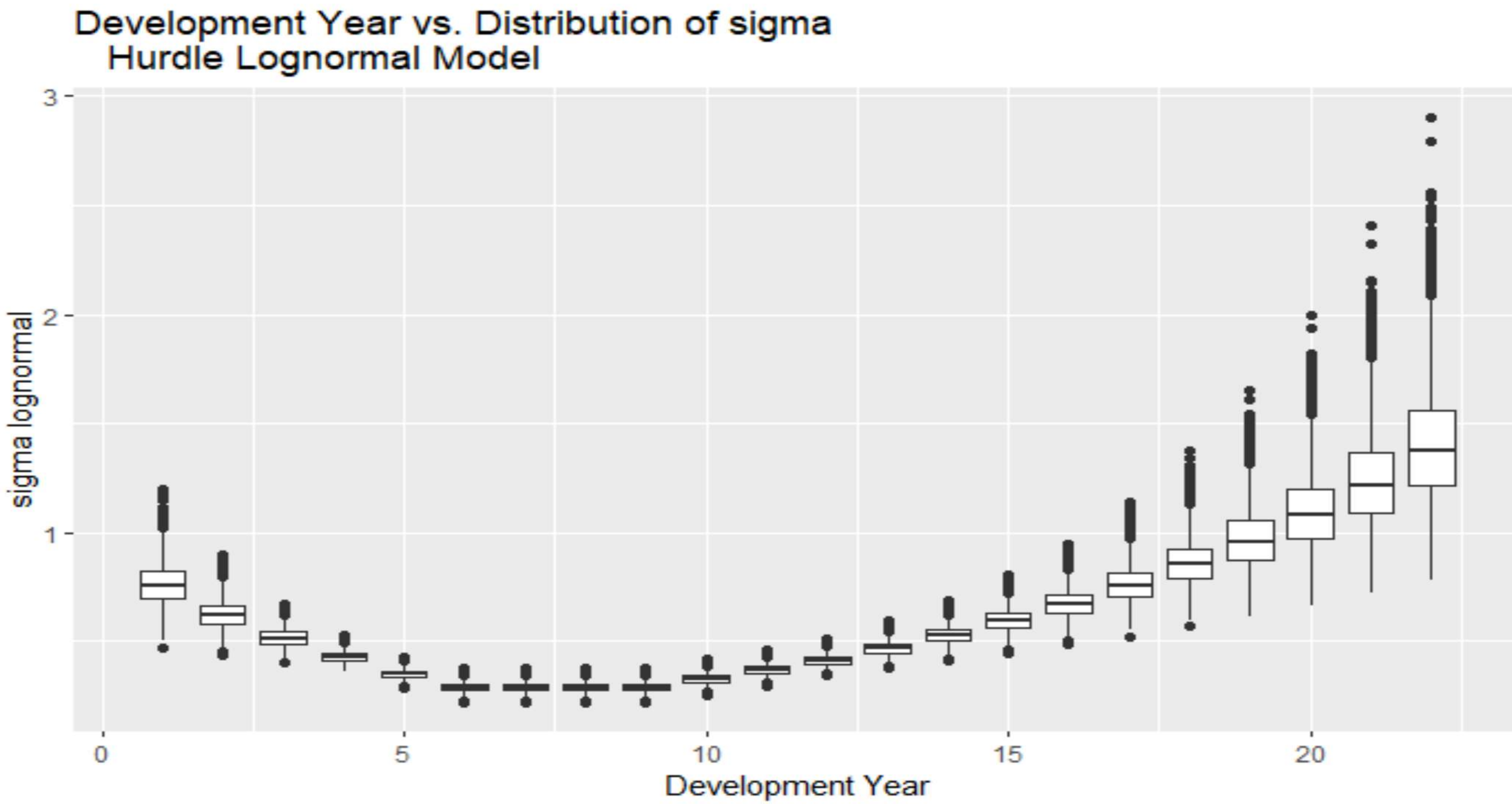
```
ln_hurdle_1_prior <- c(prior(student_t(3,-4,.5),class=b, coef=Intercept),
        prior(student_t(3,.03,.01),class=b, coef=Cal_Yr_Time),
        prior(student_t(3,1,.5),class=b, coef=Dev_Yr_10_Cap),
        prior(student_t(3,-.5,.5),class=b, coef=Dev_Yr_10_Cap_Sqrd),
        prior(student_t(3,0,.5),class=b, coef=Dev_Yr_10_Spline),
        prior(student_t(3,0,.5),class=b, coef=Dev_Yr_10_Spline_Sqrd),
        prior(student_t(3,.1,.5),class=b, coef=Dev_Yr_6_Cap,dpar=sigma),
        prior(student_t(3,.1,.5),class=b, coef=Dev_Yr_10_Spline,dpar=sigma),
       prior(student_t(3,-.5,.05),class=b, coef=Dev_Yr_1_Factor2,dpar=hu))
```

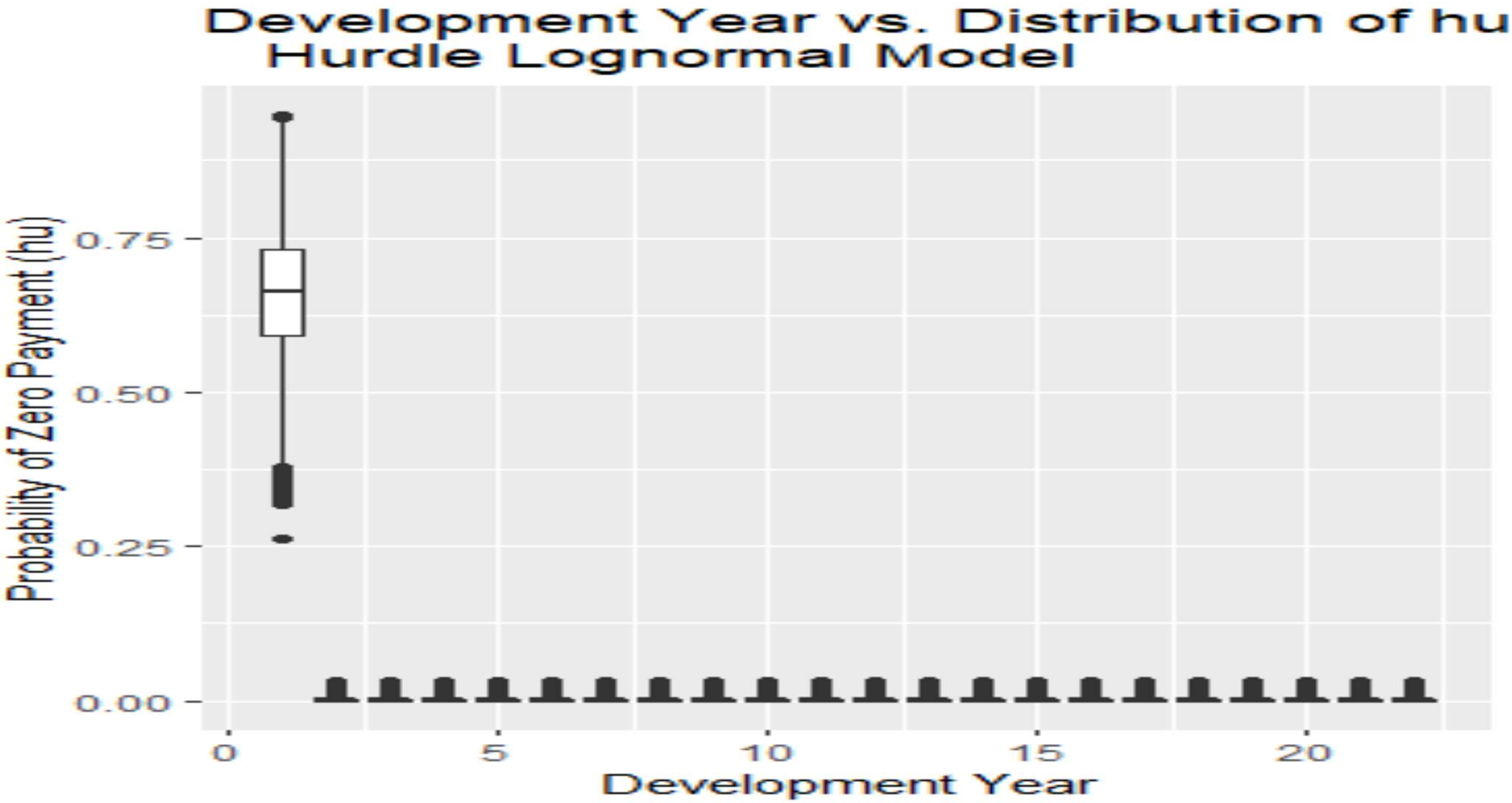
Note that hu is the binomial parameter

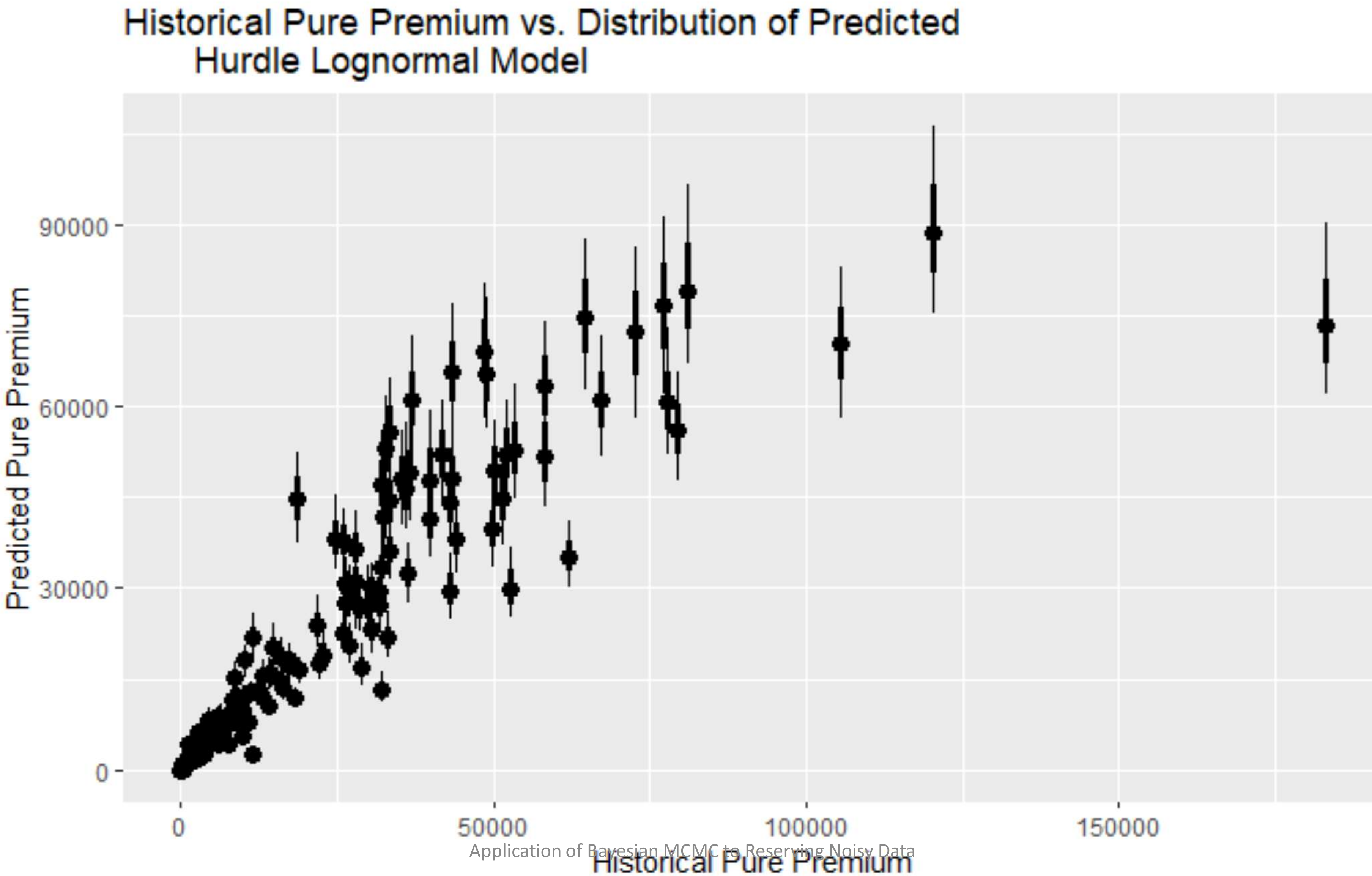
BRMS Code for Lognormal Hurdle Model

```
hurdle_lognormal_pp_1 <-brm(bf(Trended_Incr_PP_Def ~ 0 +Intercept + Dev_Yr_10_Cap+
                                Dev_Yr_10_Cap_Sqrd+
                                Dev_Yr_10_Spline +
                                Dev_Yr_10_Spline_Sqrd +
                                Cal_Yr_Time +(1||Acc_Yr),
                               sigma ~
                                Dev_Yr_6_Cap +
                                Dev_Yr_10_Spline ,
                               hu~0 + Intercept +Dev_Yr_1_Factor ),
                               data = Train_Triangle_All,
                               family =hurdle_lognormal(),
                               prior =ln_hurdle_1_prior)
```


Development Year vs. Distribution of mu
Hurdle Lognormal Model
mu lognormal
8
4
0
-4
0
5
10
15
20
Development Year


Development Year vs. Distribution of sigma
Hurdle Lognormal Model
sigma lognormal
Development Year
0
5
10
15
20
1
2
3


Development Year vs. Distribution of hu
Hurdle Lognormal Model
Probability of Zero Payment (hu)
0.75
0.50
0.25
0.00
0
5
10
15
20
Development Year


Historical Pure Premium vs. Distribution of Predicted
Hurdle Lognormal Model
Predicted Pure Premium
90000
60000
30000
0
0
50000
100000
150000
Historical Pure Premium

Comments on Case II

- Hurdle models allow one to combine probability of zero payment & non-zero payment estimates in one model
- Using the student_t prior is one approach to dealing with outliers
- Sigma's pattern of a slow decrease as the accident year ages for the first few development years then increasing as the accident year ages can be a common feature of low frequency – high severity business
- One could use other continuous distributions like Gamma and recognize the change in parameters:
  - Gamma would use "shape" rather than "sigma" in conjunction with the mean estimate as an example

Conclusion

- One can explicitly model for varying effect of inflation
- Inflation assumptions on reserve indications can be illustrated
- One can explicitly model effect of company operation changes
- The Bayesian MCMC product of a posterior distribution for results lends itself to management presentation
  - Recognize future estimates have some uncertainty
  - Uncertainty can be quantified
  - Helps with discussion on where to set published reserve estimates

Appendix

Additional Resources

- Brms
  - brms: An R Package for Bayesian Multilevel Models using Stan, Paul-Christian Burkner, Journal of Statistical Software 2017
  - Advanced Bayesian Multilevel Modeling with the R Package brms by Paul-Christian Bürkner, The R Journal Vol. 10/1, July 2018
- STAN
  - https://mc-stan.org/
- Rstudio
  - https://rstudio.com/
- Tidyverse
  - Rstudio help
  - R for Data Science: Import, Tidy, Transform, Visualize, and Model Data 1st Edition, Hadley Wickham, Garret Grolemund
- Bayesian MCMC textbooks
  - Statistical Rethinking: A Bayesian Course with Examples in R and STAN (Chapman & Hall/CRC Texts in Statistical Science) 2nd Edition, by Richard McElreath
  - Bayesian Data Analysis (Chapman & Hall/CRC Texts in Statistical Science) 3rd Edition, Gelman et.al.
- Tidybayes
  - https://cran.r-project.org/web/packages/tidybayes

Code for Creating Explanatory Variables Page 1

```
Cal_Yr_Time = Cal_Yr - 2000,
      Ln_Dev_Yr = log(Dev_Yr),
      Inv_Dev_Yr = 1/Dev_Yr,
      Dev_Yr_6_Cap = if_else(Dev_Yr < 6, Dev_Yr,
                        as.integer(6) ),
      Dev_Yr_10_Cap = if_else(Dev_Yr < 10, Dev_Yr,
                        as.integer(10) ),
      Dev_Yr_6_Cap_Ln =log(Dev_Yr_6_Cap),
      Dev_Yr_12_Cap = if_else(Dev_Yr < 12, Dev_Yr,
                        as.integer(12) ),
      Dev_Yr_13_Cap = if_else(Dev_Yr < 13, Dev_Yr,
                        as.integer(13) ),
```

Code for Creating Explanatory Variables Page 2

```
Dev_Yr_6_Spline=if_else(Dev_Yr < 6,0,
                     Dev_Yr-5),
    Dev_Yr_6_Spline_Sqrd = Dev_Yr_6_Spline *
Dev_Yr_6_Spline,
    Dev_Yr_8_Spline=if_else(Dev_Yr < 8,0,
                     Dev_Yr-7),
    Dev_Yr_8_Spline_Sqrd = Dev_Yr_8_Spline *
Dev_Yr_8_Spline,
    Dev_Yr_6_Cap_Sqrd = Dev_Yr_6_Cap * Dev_Yr_6_Cap ,
    Dev_Yr_12_Cap_Sqrd = Dev_Yr_12_Cap * Dev_Yr_12_Cap
,
    Dev_Yr_Sqrd = Dev_Yr * Dev_Yr ,
    Dev_Yr_Cbd = Dev_Yr * Dev_Yr * Dev_Yr,
    Dev_Yr_Ln = log(Dev_Yr),
    Dev_Yr_Inv = 1/Dev_Yr,
```

Code for Creating Explanatory Variables Page 3

```
Dev_Yr_1_Factor = as.factor(case_when(
     Dev_Yr == 1 ~ 1,
     Dev_Yr >1 ~2
    )),
    Dev_Yr_3_Factor = as.factor(case_when(
     Dev_Yr == 1 ~ 1,
     Dev_Yr == 2 ~2,
     Dev_Yr == 3 ~3,
     Dev_Yr > 3 ~4
    )),
    Dev_Yr_2_Factor = as.factor(case_when(
     Dev_Yr == 1 ~ 1,
     Dev_Yr == 2 ~2,
     Dev_Yr > 2 ~3
    )),
```

Code for Creating Explanatory Variables Page 4

```
Dev_Yr_Factor =as.factor(Dev_Yr),
        Dev_Yr_10_Spline =if_else(Dev_Yr < 10,0,
                                Dev_Yr - 9),
        Dev_Yr_12_Spline =if_else(Dev_Yr < 12,0,
                                Dev_Yr - 11),
        Dev_Yr_13_Spline =if_else(Dev_Yr < 13,0,
                                Dev_Yr - 12),
        Dev_Yr_10_Spline_Cap_15 = case_when(
          Dev_Yr < 10 ~0,
          Dev_Yr < 16 ~Dev_Yr-9,
          Dev_Yr >15 ~15),
        Dev_Yr_6_Spline_Ln =if_else(Dev_Yr < 6,0,
                                log(Dev_Yr_6_Spline)),
        Dev_Yr_10_Spline_Ln =if_else(Dev_Yr < 10,0,
                                log(Dev_Yr_10_Spline)),
        Dev_Yr_10_Spline_Ln =if_else(Dev_Yr < 10,0,
                                log(Dev_Yr_10_Spline)),
```

Code for Creating Explanatory Variables Page 5

```
Dev_Yr_12_Spline_Ln =if_else(Dev_Yr < 12,0,
                         log(Dev_Yr_12_Spline)),
      Dev_Yr_12_Spline_Sqrd = Dev_Yr_12_Spline *
Dev_Yr_12_Spline,
      Dev_Yr_15_Spline =if_else(Dev_Yr < 15,0,
                        Dev_Yr - 14),
      Dev_Yr_15_Spline_Ln =if_else(Dev_Yr < 15,0,
                         log(Dev_Yr_15_Spline)),
      Cal_Yr_Grp_3 =as.factor(case_when(
       Cal_Yr < 2010 ~ "LT_2010",
       Cal_Yr < 2017 ~"2010_To_2016",
       Cal_Yr > 2016 ~"GE_2017"
      )),
Acc_Yr_Grp_2 = as.factor(if_else( Acc_Yr < 2018,
                       "LT_2018",
                       "GE_2018")))
```