

Bayesian Hierarchical Risk Premium Modeling with Model Risk: Addressing Non-Differential Berkson Error

Minkun Kim¹, Marija Bezbradica¹, and Martin Crane¹

¹ADAPT Centre, School of Computing, Dublin City University, Ireland

In general insurance pricing, aligning losses with accurate premiums is crucial for insurance companies' competitiveness. Traditional actuarial models often face challenges like data heterogeneity and mismeasured covariates, leading to misspecification bias. This paper addresses these issues from a Bayesian perspective, exploring connections between Bayesian hierarchical modeling, partial pooling techniques, and Gustafson correction method for mismeasured covariates. We focus on Non-Differential Berkson (NDB) mismeasurement and propose an approach that corrects such errors without relying on gold standard data. We discover the unique prior knowledge regarding the variance of the NDB errors, and utilize it to adjust the biased parameter estimates built upon the NDB covariate. Using simulated datasets developed with varying error rate scenarios, we demonstrate the superiority of Bayesian methods in correcting parameter estimates. However, our modeling process highlights the challenge in accurately identifying the variance of NDB errors. This emphasizes the need for a thorough sensitivity analysis of the relationship between our prior knowledge of NDB error variance and varying error rate scenarios.

Keywords: Bayesian hierarchical model; heterogeneity; non-differential Berkson measurement error; aggregate insurance claim; risk premium; partial pooling; Gustafson correction

1 Introduction

In Non-life insurance practice, the risk premium is the portion of the premium that specifically covers the expected loss arising from claims (Boland, 2006). Therefore, accurately predicting aggregate claim amounts (also known as 'loss severity' in insurance terms) is critical for major actuarial decisions to prevent insolvency. In order to estimate the risk premium (or *aggregate claim amounts*), one significant tool that is often considered in a conventional regression framework (such as Generalized Linear Models (GLMs), Generalized Additive Models (GAMs), etc.) is the identification of

risk clustering, where certain risks or claims are observed to cluster together within specific contexts (Werner and Modlin, 2010). By recognizing and analyzing these clusters - geographical regions, demographic groups, types of policies, or latent classes, etc. - within the GLM framework, insurers can make data-driven adjustments to their premium structures, ensuring that pricing reflects the actual risk accurately. This tailored homogeneity, in turn, improves their overall financial performance (Ohlsson and Johansson, 2010).

In the GLM framework, risk clustering often relies on the covariates \mathbf{X} in the model (Werner and Modlin, 2010). While practitioners typically verify model assumptions such as distributional properties, residual patterns, and statistical significance, etc., they may often overlook the broader implications of the inclusion of covariates \mathbf{X} . Aggarwal et al. (2016) discuss the concept of covariate-based model risk in non-life pricing, highlighting that risks such as model misspecification error or flawed data inputs often stem from the inclusion of covariates \mathbf{X} in the model. This oversight can become particularly problematic when estimating the aggregate claim amount $S_h|\mathbf{X}$.

1.1 Research questions

Suppose we are dealing with a group policy where each policy h can generate multiple claims (Baranoff et al., 2006). For example, each policy h is taken out by a different company, such as an oil company, a car manufacturer, etc., resulting in $h = 1, \dots, H$ unique policies in the insurer's portfolio. Each company holds a single group policy h to protect its multiple assets. Let the individual claim amount Y_{hi} , $i = 1, 2, \dots, N_h(t)$ associated with $N_h(t)$ different claims during a policy period t for a single policy h is log-normally distributed (i.e., $\ln Y_{hi} \sim \mathbf{N}(\mu(t), \sigma(t)^2)$). Assuming the policy period t is fixed and short, as is typical in a non-life insurance context, we can ignore t for simplicity. Under this assumption, the claim count N_h for policy h remains limited and follows a negative binomial distribution (Ohlsson and Johansson, 2010). The individual claim amount Y_{hi} , as well as the mean claim amount \bar{Y}_h for a policy h follows a log-normal distribution, assuming the claim count is not large enough for the Central Limit Theorem to apply. Each policy, treated as an individual observation, has a unique N_h value because it is linked to a distinct group of insured assets requiring insurance protection. Hence, for each policy h , the aggregate claim amounts S_h received by an insurer can be defined as a log-normal convolution: $S_h = \sum_{i=1}^{N_h} Y_{hi} = Y_{h1} + Y_{h2} + \dots + Y_{hN_h} = N_h \bar{Y}_h$. In order to determine the future value of the expected aggregate claim amounts $E[S_h]$ for a policy h , the traditional risk modeling principle applies the *frequency-severity* approach (Kaas et al., 2008), which involves modeling the frequency and severity of claims (the number of claims and the size of each claim) separately. Resting on the assumption that the summands Y_{hi} are mutually independent and identically distributed (i.i.d.) to maintain homogeneity, the expected aggregate claim amount $E[S_h] = E[Y_{h1}] + E[Y_{h2}] + \dots + E[Y_{hN_h}]$ for a policy h can be simplified as:

$$E[S_h] = E[N_h] \times E[\bar{Y}_h] \text{ for a policy } h \text{ by the Frequency-Severity approach.} \quad (1)$$

Now, let $\mathbf{X} = \{\mathbf{X}^F, \mathbf{X}^S\}$ represent the covariate matrix that is statistically significant to understand N_h , Y_{hi} , \bar{Y}_h , and S_h . The inclusion of covariates \mathbf{X} introduces new, potentially unknown structures into the data space of N_h , Y_{hi} , \bar{Y}_h , and S_h , altering the underlying distributional properties of the individual summand $\bar{Y}_h|\mathbf{X}_h^S$ (Neuhaus and McCulloch, 2006). As a result, their convolutions $E[S_h|\mathbf{X}]$ may be re-organized into hierarchical structures where observations at one level can be grouped into others. To be

specific, based on the group policy assumption, we can assume that $Y_{h1}, Y_{h2}, \dots, Y_{hN_h}$ are influenced by the same covariate vector $\mathbf{X}_h^S = \{x_h^S, z_h^S\}$, and thus the inclusion of \mathbf{X} still leads to $E[S_h|\mathbf{X}] = E[N_h|\mathbf{X}_h^F] \times E[\tilde{Y}_h|\mathbf{X}_h^S]$, as shown in Equation (1). However, if each claim amount Y_{hi} is influenced by individually different covariate vector $\mathbf{X}_{hi}^S = \{x_{hi}^S, z_{hi}^S\}$, $i = 1, \dots, N_h$, the summands - $Y_{h1}|\mathbf{X}_{h1}^S, Y_{h2}|\mathbf{X}_{h2}^S, \dots, Y_{hN_h}|\mathbf{X}_{hN_h}^S$ - may remain independent, but the assumption of the identically distributed observations (homogeneity) is no longer valid (since the distribution of each summand may have different parameter values). This heterogeneity introduced by \mathbf{X} renders $E[S_h|\mathbf{X}] \neq E[N_h|\mathbf{X}_h^F] \times E[Y_{hi}|\mathbf{X}_{hi}^S]$, making these log-normal convolutions analytically intractable under the traditional risk modeling principle (Kaas et al., 2008).

Alongside inherent heterogeneity, actuaries frequently encounter model risks linked to poor data quality, especially when covariates are affected by excessive noise. Various types of noise or measurement errors exist (which will be discussed in the next section); however, we focus specifically on Non-Differential Berkson (NDB) mismeasurement due to its unique impact on the modeling process and its simplicity in application. The inclusion of NDB covariates not only compromises the accuracy of the model but also amplifies existing heterogeneity, potentially leading to biased estimates and erroneous conclusions (Grace et al., 2021). These covariate-based model risks introduce new challenges, emphasizing the need for a deeper understanding of their effects and the development of advanced modeling techniques to address them. In this regard, we aim to tackle two fundamental covariate-based model risks by posing the following research questions:

- **RQ1.** If an additional unobservable heterogeneity is introduced by the inclusion of covariates \mathbf{X} , how to capture this and maintain homogeneity in risk clustering?
- **RQ2.** If an additional estimation bias results from the use of the mismeasured covariates (with Non-Differential Berkson (NDB) error in particular), what is the best way to mitigate this model risk?

1.2 Our contribution and paper outline

This paper is dedicated to the development of a novel strategy for modeling the conditional aggregate claim amount $S_h|\mathbf{X}$ by dealing with the covariate-based model risks: heterogeneity (RQ1) and NDB covariate (RQ2). We center our attention on the hierarchical GLM and Gustafson correction, with the aim of establishing connections between them and integrating them within the Bayesian parametric framework. This framework is especially relevant to actuarial risk management, as it facilitates the calculation of posterior credibility intervals to account for all sources of uncertainty in estimating the risk premium (Cairns, 2000). The primary contribution of this study is the novel elicitation of prior knowledge about the unknown variance of the NDB covariate conditional on the true covariate $\mathbf{x}^*|\mathbf{x}$. Understanding $\mathbf{x}^*|\mathbf{x}$ is central to implementing the Gustafson correction with a Gibbs sampler and facilitates its application in mitigating the model risk tied to the NDB covariate \mathbf{x}^* . To our knowledge, no previous studies have tackled the estimation of risk premium while accounting for NDB errors in the covariate within the hierarchical GLM framework.

The paper is organized as follows: Section 2 provides background information on the problem setting and discusses traditional approaches to the research questions RQ1 and RQ2. Section 3 explores our modeling methods for aggregate claim amounts under the frequency-severity principle, considering two scenarios: the complete covariate case and the NDB covariate case. This section details the components

of Bayesian inference within the hierarchical GLM framework and the development of Gibbs samplers, introducing our novel Gustafson correction for addressing NDB covariates. Section 4 presents empirical results and validates our approach using real insurance data. The effectiveness of our Bayesian hierarchical GLM incorporating the Gustafson correction technique is then compared to classical risk premium models using Simulation Extrapolation (SIMEX). The paper concludes with Section 5.

2 Background and Conventional Approaches to RQ1 and RQ2

This section provides a brief background on the problem setting and explores various conventional methods related to each research question.

2.1 Heterogeneity in risk clusters

The key objective in classical regression-based risk premium modeling is to accurately describe the stochastic relationships between the insured claim amounts and independent, homogeneous risk clusters. Homogeneity within a risk cluster reflects the similar characteristics of policyholders, which aids in managing the inherent variability observed in claim data. Common approaches in this domain include Generalized Linear Models (GLMs), Generalized Additive Models (GAMs), and Multivariate Adaptive Regression Splines (MARS), among others (Ohlsson and Johansson, 2010).

GLMs serve as a fundamental framework for analyzing premiums and risk clusters, owing to their clarity and theoretical robustness. These models effectively tackle heteroscedasticity through the application of Weighted Least Squares (WLS) and accommodate variability across different risk levels by incorporating a class-specific term in the linear predictor. Nonetheless, GLMs operate under the assumption that all heterogeneity stems from known covariates, which may lead to the neglect of unobserved factors and non-linear interactions that significantly contribute to heterogeneity (Werner and Modlin, 2010). GAMs build upon GLMs by integrating smoothing functions, which facilitate the capturing non-linear relationships among risk factors. This approach enhances flexibility by employing a nonparametric modeling strategy that adjusts the functional form based on the underlying data. However, GAMs fall short in capturing variations arising from categorical covariates (Denuit and Lang, 2004). MARSs represent another sophisticated advancement, further refining GLMs by applying piecewise linear regressions across data intervals. This characteristic not only enhances computational efficiency compared to GAMs but also makes MARSs particularly adept at handling high-dimensional data challenges (Francis, 2003).

In the context of risk premium modeling with GLMs, GAMs, and MARSs, addressing heterogeneity necessitates the inclusion of an additional class-specific effect term within the linear predictor. This term aims to capture the latent heterogeneity inherent in each risk cluster, reflecting unique deviations from the cluster mean (Wuthrich, 2020). However, a notable limitation of these classical risk premium modeling frameworks is that once parameters are estimated, predictions for new risk premiums are typically made solely based on these parameters and the associated covariates, neglecting latent factors and uncertainty inherent in these estimates. Consequently, overlooking this uncertainty constrains the insurer's ability to make more informed decisions and effectively formulate their financial strategies (Parodi, 2023).

2.2 NDB errors in covariates

Before delving into the methods for correcting measurement errors in covariates, we first provide an overview of the various types of measurement error. Consider $\mathbf{X}^S = \{\mathbf{x}^S, \mathbf{z}^S\}$, where only \mathbf{x}^S , a continuous covariate, is subject to mismeasurement, while \mathbf{z}^S , a binary covariate, is complete. For simplicity, we omit the superscript s from the covariate matrix \mathbf{X}^{S*} and the continuous covariate vector \mathbf{x}^{S*} , referring to them as \mathbf{X}^* and \mathbf{x}^* respectively. The measurement error ϵ and the error-prone observed covariate \mathbf{x}^* are assumed to follow $\epsilon \sim \mathbf{N}(0, \sigma_\epsilon^2)$ and $\mathbf{x}^* \sim \mathbf{N}(\mathbf{x}, \mathfrak{T}^2)$. The approach to handling mismeasured covariates depends on the nature of ϵ . Now, the different types of measurement errors ϵ are outlined as follows:

- Additive vs Multiplicative (Fewell, 2007) : With *additive* error, the mismeasured covariate \mathbf{x}^* can be expressed as $\mathbf{x}^* = \mathbf{x} + \epsilon$, while the *multiplicative* error can be represented as $\mathbf{x}^* = \mathbf{x} \cdot \epsilon$.
- Differential vs Non-Differential (Romann, 2008) : The error ϵ can be classified as *differential* or *non-differential* based on its relationship with other variables. Differential error occurs when the mismeasured covariate \mathbf{x}^* is correlated with the outcome, represented as $\mathbf{x}^* \sim Y | \mathbf{z}, \mathbf{x}$. Conversely, non-differential error arises when the error ϵ in \mathbf{x}^* provides no additional information about the outcome beyond what is already contained in \mathbf{x} , leading to $\mathbf{x}^* \perp Y | \mathbf{x}, \mathbf{z}$. This implies that $f(Y|\mathbf{x}^*, \mathbf{x}, \mathbf{z}) = f(Y|\mathbf{x}, \mathbf{z})$ and $f(\mathbf{x}^*|Y, \mathbf{x}, \mathbf{z}) = f(\mathbf{x}^*|\mathbf{x})$.
- Classical vs Berkson (Carroll et al., 2006) : *Classical* error occurs when the error ϵ is independent of the true covariates, leading to $\epsilon \perp Y, \mathbf{x}, \mathbf{z}$, $E[\epsilon|\mathbf{x}^*, \mathbf{z}] = 0$ and $\mathfrak{T}^2 = V(\mathbf{x}) + V(\epsilon) > V(\mathbf{x})$. In contrast, if the error ϵ is independent of the observed covariates, but associated with other latent factors with multiple levels, *Berkson* error applies, where $\epsilon_j \perp Y, \mathbf{x}^*, \mathbf{z}$, resulting in $E[\epsilon_j|\mathbf{x}^*, \mathbf{z}] = 0$ and $\mathfrak{T}_j^2 = V(\mathbf{x}_j) + V(\epsilon_j) < V(\mathbf{x})$ for $j = 1, \dots, J$. This implies that the error ϵ can display varying levels of heteroscedasticity across different risk clusters due to the latent factors.

We focus on the case in which the mismeasured covariate \mathbf{x}^* arises from additive measurement error in a manner that is non-differential (blind to the outcome and other covariates) and Berkson (correlated to the latent factors), allowing us to address error structures that may vary across risk clusters. Throughout this paper, we refer to it as the ‘NDB covariate.’

Regression Calibration (RC) is a widely used method for addressing the NDB covariate in the regression framework. RC intuitively corrects errors by leveraging the relationship between the NDB covariate and true covariate, regressing on the expected true covariates given the NDB ones (Freedman et al., 2004). Carroll et al. (2006) suggest that RC remains valid with GLMs when unbiased reference measurements (i.e., gold standard data) are available, but obtaining such data has been found to be costly and often impractical. In addition, Skrandal and Kuha (2012) demonstrate that parameter estimates from RC can be inconsistent as the size of the error variances σ_ϵ^2 increases.

An alternative method, Simulation-Extrapolation (SIMEX), proposed by Cook and Stefanski (1994), corrects parameter estimates through simulation rather than recovering true covariate values. SIMEX assesses the impact of measurement error by introducing additional artificial noise into the NDB covariate and observing how this

affects parameter estimates. By modeling these changes as a function of the noise, optimal parameter estimates can be obtained by setting the noise level back to zero. Unlike RC, SIMEX does not require external reference such as gold standard data for error correction; however, the primary challenge of SIMEX lies in the risk of inaccurate extrapolation due to the complex relationship between the estimated parameters and the error variance σ_ϵ^2 (Oh et al., 2018). Although a simple quadratic curve is often used for extrapolation stability, Carroll et al. (2006) show that parameter estimates may remain inconsistent until the extrapolation curve perfectly captures this complex relationship, which remains as a persistent limitation.

3 Modeling Method for $S_h | \mathbf{X}^F, \mathbf{X}^S$

3.1 Introduction of partial pooling to address heterogeneity in risk clustering: RQ1

The partial pooling method, proposed by Gelman and Carlin (2013), tackles data heterogeneity among risk clusters by finding a middle ground between two perspectives: ‘complete pooling’, where a single global parameter overlooks cluster differences, and ‘no-pooling’, where separate parameters are created for each cluster. Partial pooling techniques are naturally integrated within the hierarchical GLM framework, which permits each parameter to be modeled independently. Specifically, within the hierarchical GLM framework, the cluster-specific intercept, or ‘varying intercept’ models both overall data and cluster-level data, flexibly moving between complete and no-pooling settings. For example, in a log-normal model of aggregate claims given by:

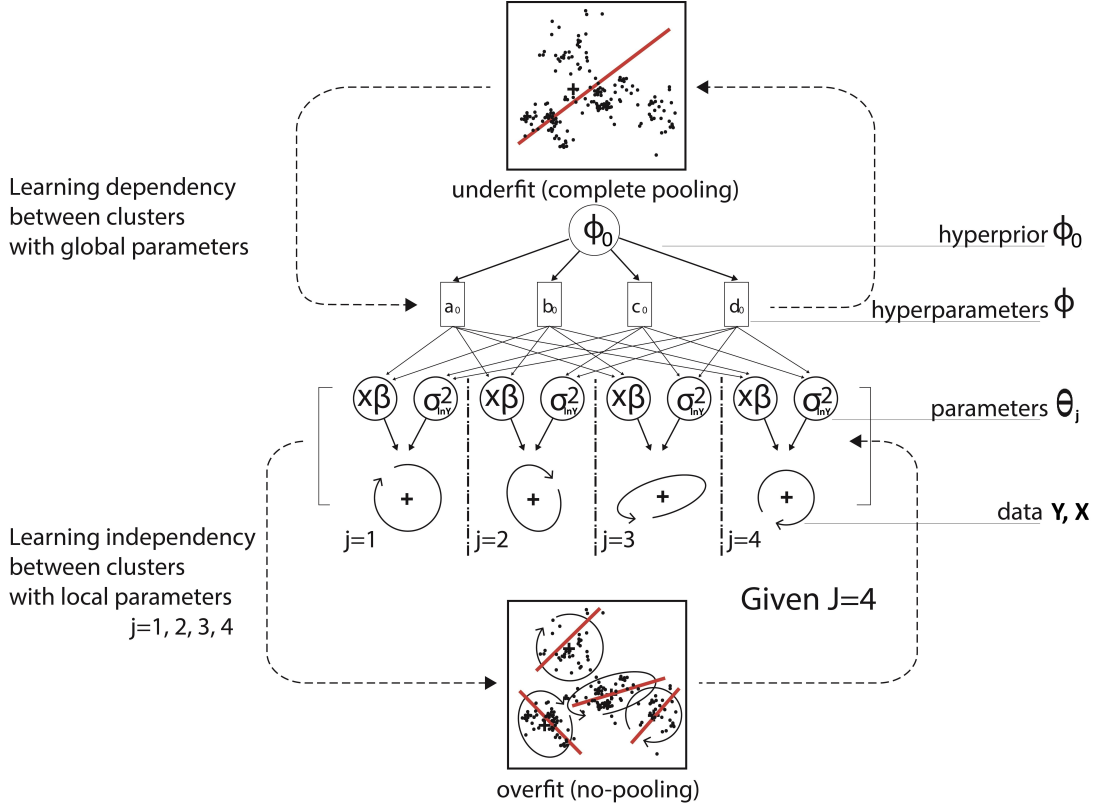
$$\begin{aligned} \ln Y_i &\sim \mathbf{N}\left(E[\ln Y_i | \mathbf{X}], \sigma_{\ln Y}^2\right), \text{ for } i = 1, \dots, N \\ E[\ln Y_i | \mathbf{X}] &= E[\alpha_{[j]}] + \mathbf{X}^T \boldsymbol{\beta}, \text{ for } j = 1, \dots, J \\ \alpha_{[j]} &\sim \mathbf{N}(\mu_\alpha, \sigma_\alpha^2), \end{aligned} \quad (2)$$

where $\boldsymbol{\beta}$ represents the regression parameters, and $\sigma_{\ln Y}^2$ is the log-scale variance of the outcome, the varying intercept $\alpha_{[j]}$ has its own normal density with parameters - $\mu_\alpha, \sigma_\alpha^2$ -, representing the cluster’s mean and variance. In no-pooling, $\alpha_{[j]}$ varies across clusters, while in complete pooling, $\alpha_{[j]}$ takes a single value for all data. The partial pooling approximates the value of the varying intercept by weighting the no-pooling estimate $\overline{\ln Y}_j - \bar{\mathbf{X}}_j^T \boldsymbol{\beta}_j$ and the complete pooling estimate μ_α as below (Gelman and Hill, 2007).

$$E[\alpha_{[j]}] \approx \frac{n_j / \sigma_{\ln Y_j}^2}{n_j / \sigma_{\ln Y_j}^2 + 1 / \sigma_\alpha^2} \cdot (\overline{\ln Y}_j - \bar{\mathbf{X}}_j^T \boldsymbol{\beta}_j) + \frac{1 / \sigma_\alpha^2}{n_j / \sigma_{\ln Y_j}^2 + 1 / \sigma_\alpha^2} \cdot \mu_\alpha \quad (3)$$

in which n_j is the sample size in the cluster j , $\sigma_{\ln Y_j}^2$ is the variance of the cluster j (within-cluster variance), and σ_α^2 is the total cluster variance. In interpreting Equation (3), when the cluster j has small samples (i.e., $n_j \rightarrow 0$) or the total cluster variance σ_α^2 is small and reliable (i.e., $\sigma_\alpha^2 \rightarrow 0$), the estimate $E[\alpha_{[j]}]$ shifts towards the global intercept μ_α . Otherwise, it leans towards the local cluster intercept $\overline{\ln Y}_j - \bar{\mathbf{X}}_j^T \boldsymbol{\beta}_j$.

Figure 1 gives a further illustration of how the partial pooling technique is integrated into the hierarchical GLM. Let the local cluster j ’s outcome parameter $\boldsymbol{\theta}_j = \{\boldsymbol{\beta}_j, \sigma_{\ln Y_j}^2\}$ for $j = 1, \dots, J$ be an independent sample from a population distribution with global



$$f[\ln Y_i | \mathbf{X}, \beta_{j=?}, \sigma_{\ln Y}^2_{j=?}]$$

Figure 1: As a Bayesian parametric example with $J = 4$ clusters, this diagram depicts a typical Bayesian hierarchical model with the Bias-Variance trade-off through the partial pooling. For prediction, the class membership j of each data point $\{Y_{hi}, \mathbf{X}_{hi}\}$ should be known beforehand.

parameters (hyperparameters) $\phi = \{a_0, b_0, c_0, d_0\}$, as shown in Figure 1. If ϕ is unknown and requires estimation, the uncertainties in both θ and ϕ would be evaluated using their joint posterior distribution given by:

$$p(\theta, \phi | \ln Y, \mathbf{X}) \propto f(\ln Y | \mathbf{X}, \theta, \phi) \cdot p(\theta, \phi) = \prod_{j=1}^J f(\ln Y_j | \mathbf{X}_j, \theta_j, \phi) \cdot p(\theta_j, \phi) \quad (4)$$

where the hyperparameters ϕ affect the outcome (or cluster) through the cluster parameter θ_j , as illustrated in Figure 1. To analyze the joint prior $p(\theta_j, \phi)$ in Equation (4), it is decomposed into the cluster parameter distribution $p(\theta_j | \phi)$ and the hyperprior $p(\phi)$. The posterior distributions are represented as $p(\theta_j | \phi, \ln Y_j, \mathbf{X}_j)$ for the cluster parameter and $p(\phi | \ln Y, \mathbf{X})$ for the hyperprior, which can then be evaluated as:

$$p(\theta | \phi, \ln Y, \mathbf{X}) = \prod_{j=1}^J p(\theta_j | \phi, \ln Y_j, \mathbf{X}_j) \quad (5)$$

$$p(\phi | \ln Y, \mathbf{X}) = \frac{p(\theta, \phi | \ln Y, \mathbf{X})}{p(\theta | \phi, \ln Y, \mathbf{X})} \quad (6)$$

where the parameter layer for θ_j is informed by the data points grouped by each cluster j in Equation (5) (this indicates no-pooling), while the hyperparameter layer for ϕ

communicates with the entire population without any clusters in Equation (6) (this indicates complete pooling). In Figure 1, complete pooling in the upper layer utilizes all data, which may overlook cluster-specific variations and result in underfitting, as hyperparameters ϕ are updated based on the entire dataset. In contrast, no pooling in the lower layer treats each cluster independently, risking overfitting, with parameters θ_j updated on a cluster-wise basis. The diagram in Figure 1 is translated into the analytic expression in Equations (4), (5), and (6). Note that the numerator in Equation (6) is just the joint posterior for θ, ϕ from Equation (4), and the denominator in Equation (6) is the posterior for $\theta|\phi$ in Equation (5). From Equations (4), (5), and (6), one can see that the cluster specific parameter θ_j and the global parameter ϕ can be drawn and estimated recursively, as per Figure 1. This balance enables optimal information sharing between clusters, allowing predictions to account for both cluster-level and individual-level variations simultaneously, thereby addressing heterogeneity across risk clusters effectively (Gelman and Carlin, 2013).

3.2 Introduction of Gustafson correction to address NDB covariate: RQ2

Assuming additive, non-differential Berkson errors (NDB) in a covariate, we examine the Bayesian framework to account for its cluster-wise error structure (i.e., heteroscedasticity) probabilistically. The strength of a Bayesian approach to address NDB covariate errors has been well-documented by Grace et al. (2021). From a Bayesian perspective, deviations from true values can be corrected through the incorporation of the prior knowledge that captures the relation between the unobservable true covariate \mathbf{x} and the observed NDB covariate \mathbf{x}^* . When the NDB error structure introduces cluster-wise heteroscedasticity, the framework specifies structural components to capture varying variances across risk clusters. This cluster-wise inference can also be achieved by leveraging the manageable joint product and Gibbs sampling (Gustafson, 2008).

The model component specification plays a crucial role in formulating the strategy, as it determines the parameterization when incorporating the necessary parameter knowledge. For the NDB covariate problem, Gustafson (2008) begins by specifying the full joint density of the relevant variables as

$$f(Y, \mathbf{x}^*, \mathbf{x}, \mathbf{z}) = f(Y|\mathbf{x}^*, \mathbf{x}, \mathbf{z}) \cdot f(\mathbf{x}^*|\mathbf{x}, \mathbf{z}) \cdot f(\mathbf{x}|\mathbf{z}) \cdot f(\mathbf{z}) \quad (7a)$$

$$f(Y, \mathbf{x}^*, \mathbf{x}|\mathbf{z}) = f(Y|\mathbf{x}^*, \mathbf{x}, \mathbf{z}) \cdot f(\mathbf{x}^*|\mathbf{x}, \mathbf{z}) \cdot f(\mathbf{x}|\mathbf{z}) \quad (7b)$$

in which the term for the precisely measured covariate \mathbf{z} is factored out for the sake of simplicity. Due to the assumption of the non-differential error, the conditional joint density in Equation (7b) can be further reduced to

$$f(Y, \mathbf{x}^*, \mathbf{x}|\mathbf{z}) = f(Y|\mathbf{x}^*, \mathbf{x}, \mathbf{z}) \cdot f(\mathbf{x}^*|\mathbf{x}, \mathbf{z}) \cdot f(\mathbf{x}|\mathbf{z}) \quad (8)$$

where the cancellations are based on the definition of NDB error as described in Section 2.2. This conditional joint density in Equation (8) is termed as the *complete joint* model (Gustafson, 2008). As the true covariate \mathbf{x} is not observable, the complete joint model is technically inaccessible or theoretical. However, in the construction of the complete joint model in Equation (8), three main components that make up the relationship between the true covariate \mathbf{x} and the observed covariate \mathbf{x}^* can be identified:

- outcome component $f(Y|\mathbf{x}, \mathbf{z})$
- linking (measurement) component $f(\mathbf{x}^*|\mathbf{x})$

- covariate (exposure) component $f(\mathbf{x}|\mathbf{z})$

In particular, the linking component incorporates the measurement error mechanism of $\mathbf{x}_j^* \sim \mathbf{N}(\mathbf{x}_j, \mathfrak{T}_j^2)$ into the analysis, allowing control over the simulation process (Grace et al., 2021). The linking (measurement) component term and \mathfrak{T}_j^2 also enable investigation of the NDB error's effect on the unknown risk clusters.

In contrast, the conditional joint model available in reality is referred to as the *incomplete joint* model (Gustafson, 2008), which can be represented as:

$$f(Y, \mathbf{x}^* | \mathbf{z}) = f(Y | \mathbf{x}^*, \mathbf{z}) \cdot f(\mathbf{x}^* | \mathbf{z}) \quad (9)$$

where the outcome term $f(Y | \mathbf{x}^*, \mathbf{z})$ and the exposure term $f(\mathbf{x}^* | \mathbf{z})$ are fully known. Note that the incomplete joint model in Equation (9) can be obtained from marginalizing the complete joint model in Equation (8) over the unobservable true covariate \mathbf{x} . For a concise comparison, refer to Table 1.

Table 1: Comparison of Complete and Incomplete Joint models

Complete Joint (unknown)	Incomplete Joint (known)
$\underbrace{f(Y \mathbf{x}, \mathbf{z})}_{\text{outcome}} \cdot \underbrace{f(\mathbf{x}^* \mathbf{x})}_{\text{measurement}} \cdot \underbrace{f(\mathbf{x} \mathbf{z})}_{\text{exposure}}$	$\underbrace{f(Y \mathbf{x}^*, \mathbf{z})}_{\text{outcome}} \cdot \underbrace{f(\mathbf{x}^* \mathbf{z})}_{\text{exposure}}$
$= f(Y, \mathbf{x}^*, \mathbf{x} \mathbf{z})$	$= f(Y, \mathbf{x}^* \mathbf{z})$

The implication is that when we align the two models by integrating out the unobservable true covariate \mathbf{x} from the complete joint model, the resulting equation reveals the connection between the parameters of the complete joint model and those of the incomplete joint model, as follows:

$$\int f(Y, \mathbf{x}^*, \mathbf{x} | \mathbf{z}) d\mathbf{x} = f(Y, \mathbf{x}^* | \mathbf{z}) \quad (10)$$

While explicitly solving the integral in Equation (10) can be difficult, comparing the parameterizations on both sides of the equation uncovers the link between the parameters of the model based on the NDB covariate \mathbf{x}^* and the true model using the actual covariate \mathbf{x} . Motivated by Romann (2008) and Grace et al. (2021), we analytically derived the solution to the integral in Equation (10), with the detailed derivation provided in Part 3 of the on-line supplementary file. Accordingly, a hybrid Gibbs sampler was developed using the relationship embedded in the parameterizations (highlighted in Part 2 of the on-line supplementary file) to mitigate the model risk associated with the NDB covariates.

3.3 Clustering $S_h | \mathbf{X}^F, \mathbf{X}^S$ with Complete Case Covariate: RQ1. Tackling heterogeneity

With accurately measured covariates, we have emphasized the necessity of maintaining homogeneity within each risk cluster to determine fair premiums. However, model risk—specifically the heterogeneity issue (RQ1)—emerges when covariates introduce variability, increasing the heterogeneity within these clusters. Assuming that the risk clusters $j = 1, \dots, J$ are already established, this section presents our foundational hierarchical GLM with varying coefficients, utilizing the partial pooling technique to ensure more homogeneous risk clusters.

Baseline GLM: For each policy $h = 1, \dots, H$, we suggest that the claim count N_h follows a negative binomial (NB) distribution with mean ζ_h and dispersion parameter ψ . This setting, along with the assumption that the individual claim amounts on a logarithmic scale $\ln Y_{hi}$, $i = 1, \dots, N_h$, are independent and normally distributed with mean μ_h and variance σ^2 , is grounded in Ohlsson and Johansson (2010). In short, we represent these two outcome models as follows:

$$N_h \sim \mathbf{NB}(\zeta_h, \psi) = \frac{\Gamma(N_h + \psi)}{N_h! \Gamma(\psi)} \left[\frac{\zeta_h}{\zeta_h + \psi} \right]^{N_h} \left[\frac{\psi}{\zeta_h + \psi} \right]^\psi \quad (11a)$$

$$\tilde{Y}_h \sim \mathbf{LogN}(\mu_h, \sigma^2) = \frac{1}{\tilde{Y}_h \sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} [\ln \tilde{Y}_h - \mu_h]^2\right) \quad (11b)$$

With the addition of covariates $\mathbf{X} = (\mathbf{X}^F : \{\mathbf{x}^F, \mathbf{z}^F\}, \mathbf{X}^S : \{\mathbf{x}^S, \mathbf{z}^S\})$, the covariate effects and the information for risk clusters $j = 1, \dots, J$ can be integrated into the outcome models via the expectation parameters ζ_h and μ_h as presented in Equation (11). Specifically, the covariates for claim count (frequency) and claim amount (severity) are represented by $\mathbf{X}^F : \{\mathbf{x}^F, \mathbf{z}^F\}$ and $\mathbf{X}^S : \{\mathbf{x}^S, \mathbf{z}^S\}$, respectively. This leads to the expectation parameters taking the form of GLMs:

$$\zeta_h = E[N_h] = E[E[N_h | \mathbf{X}^F \boldsymbol{\beta}^F + \epsilon_h^F]] = E[\exp(\mathbf{X}^F \boldsymbol{\beta}^F + \epsilon_h^F)] \approx e^{\mathbf{X}^F \boldsymbol{\beta}^F} \quad (12a)$$

$$e^{\mu_h + \frac{1}{2}\sigma^2} = E[\tilde{Y}_h] = E[E[\tilde{Y}_h | \mathbf{X}^S \boldsymbol{\beta}^S + \epsilon_h^S]] = E[\exp(\mathbf{X}^S \boldsymbol{\beta}^S + \frac{1}{2}\sigma^2 + \epsilon_h^S)] \approx e^{\mathbf{X}^S \boldsymbol{\beta}^S + \frac{1}{2}\sigma^2} \quad (12b)$$

where the residuals are normally distributed, $\epsilon_h^F \sim \mathbf{N}(0, \sigma_{\epsilon^F}^2)$, $\epsilon_h^S \sim \mathbf{N}(0, \sigma_{\epsilon^S}^2)$. Based on the frequency-severity principle, the conditional expected aggregate claim amount given $\mathbf{X} = \{\mathbf{X}^F, \mathbf{X}^S\}$ for a policy h is expressed as $E[S_h | \mathbf{X}] = E[N_h | \mathbf{X}^F] \times E[\tilde{Y}_h | \mathbf{X}^S]$. Consequently, the point estimate of the risk premium, considering the covariates \mathbf{X}^F and \mathbf{X}^S , is determined by

$$E[S_h | \mathbf{X}^F, \mathbf{X}^S] = \exp(\mathbf{X}^F \boldsymbol{\beta}^F + \mathbf{X}^S \boldsymbol{\beta}^S + \frac{1}{2}\sigma^2) \quad (13)$$

Hierarchical GLM with partial pooling: However, incorporating covariates \mathbf{X}^F and \mathbf{X}^S into the model may introduce unobserved risk factors, leading to increased heterogeneity within each risk cluster $j = 1, \dots, J$. To tackle this issue, we propose a hierarchical GLM that employs distinct model for each parameter, along with cluster-specific GLM coefficient vectors and dispersion parameters $\boldsymbol{\beta}_j^F, \psi_j, \boldsymbol{\beta}_j^S, \sigma_j^2$ across all policies $h = 1, \dots, H$. Consequently, Equation (13) with $j(h) \in \{1, \dots, J\}$ is redefined with the following prior selections:

$$E[S_{j(h)} | \mathbf{X}^F, \mathbf{X}^S] = \exp(\mathbf{X}^F \boldsymbol{\beta}_j^F + \mathbf{X}^S \boldsymbol{\beta}_j^S + \frac{1}{2}\sigma_j^2) \quad (14a)$$

$$\text{For } N_h \begin{cases} \boldsymbol{\beta}_j^F | \boldsymbol{\beta}_0^F, \Sigma_{\beta_0}^F \sim \mathbf{MVN}(\boldsymbol{\beta}_0^F, \Sigma_{\beta_0}^F) \\ \psi_j | u_0^F, v_0^F \sim \mathbf{Ga}(\frac{u_0^F}{2}, \frac{v_0^F}{2}) \end{cases} \quad (14b)$$

$$\text{For } \tilde{Y}_h \left\{ \begin{array}{l} \beta_j^S \mid \beta_0^S, \Sigma_{\beta_0}^S \sim \mathbf{MVN}(\beta_0^S, \Sigma_{\beta_0}^S) \\ \sigma_j^2 \mid u_0^S, v_0^S \sim \mathbf{InvGa}(\frac{u_0^S}{2}, \frac{v_0^S}{2}) \end{array} \right. \quad (14c)$$

$$\text{For } \mathbf{X}^F \left\{ \begin{array}{l} x_{j(h)}^F \sim \mathbf{N}(E[\mathbf{x}_j^F], \lambda_j^{2F}) \\ \lambda_j^{2F} \sim \mathbf{InvGa}(c_0^F, d_0^F) \\ z_{j(h)}^F \sim \mathbf{Bernoulli}(\pi_j^F) \\ \pi_j^F \sim \mathbf{Beta}(g_0^F, h_0^F) \end{array} \right. \quad (14d)$$

$$\text{For } \mathbf{X}^S \left\{ \begin{array}{l} x_{j(h)}^S \sim \mathbf{N}(E[\mathbf{x}_j^S], \lambda_j^{2S}) \\ \lambda_j^{2S} \sim \mathbf{InvGa}(c_0^S, d_0^S) \\ z_{j(h)}^S \sim \mathbf{Bernoulli}(\pi_j^S) \\ \pi_j^S \sim \mathbf{Beta}(g_0^S, h_0^S) \end{array} \right. \quad (14e)$$

Equation (14) presents a Bayesian hierarchical GLM in which the hyperparameters are dynamically updated based on the observed data. Gelman and Carlin (2013) and Winkelmann (2008) recommend a Multivariate Gaussian prior for β due to the Normality assumption, a Gamma prior for ψ , and an Inverse Gamma prior for σ^2 due to its positive nature and adjustability. In this way, the hierarchical structure in Equation (14) encompasses several layers, each influenced by the data differently to adjust the degree of sharing in pooling.

In Section 2, we previously discussed the partial pooling effect that can be attained by utilizing hierarchical structure within this GLM framework. In short, the hierarchical GLM in Equation (14) explores unique parameter values for each individual (saturated cohort) and each cluster (reduced cohort), pooling them across multiple clusters rather than averaging the parameter values based on the available information (Gelman and Carlin, 2013). To achieve this, the GLM coefficients and dispersion parameters - $\beta_j^F, \psi_j, \beta_j^S, \sigma_j^2$ - should vary by each cluster (i.e., no-pooling). At the same time, the corresponding hyperparameters $\beta_0^F, \Sigma_{\beta_0}^F, u_0^F, v_0^F, \beta_0^S, \Sigma_{\beta_0}^S, u_0^S, v_0^S$ should be updated by the entire data (i.e., complete pooling). The hyperpriors selected for these hyperparameters in this paper are listed below in Equations (15) to (18).

$$\left. \begin{array}{l} \beta_0^F \mid \underline{m}_0, \underline{\delta} \sim \mathbf{MVN}(\underline{m}_0, \frac{1}{\underline{\delta}} \Sigma_{\beta_0}^F) \\ \Sigma_{\beta_0}^F \mid \underline{q}_0, \underline{\Lambda} \sim \mathbf{IW}(\underline{q}_0, \underline{\Lambda}) \end{array} \right\} \text{for } \beta_j^F \quad (15)$$

$$\left. \begin{array}{l} u_0^F \mid \underline{\rho}_{u1}, \underline{\rho}_{u2} \propto \underline{\rho}_{u1}^{\left(\frac{u_0^F}{2}\right)^{-1}} / \Gamma\left(\frac{u_0^F}{2}\right)^{\underline{\rho}_{u2}} \\ v_0^F \mid \underline{\rho}_{v1}, \underline{\rho}_{v2} \sim \mathbf{Ga}(\underline{\rho}_{v1}, \underline{\rho}_{v2}) \end{array} \right\} \text{for } \psi_j \quad (16)$$

$$\left. \begin{array}{l} \beta_0^S \mid m_0, \delta \sim \mathbf{MVN}(m_0, \frac{1}{\delta} \Sigma_{\beta_0}^S) \\ \Sigma_{\beta_0}^S \mid q_0, \Lambda \sim \mathbf{IW}(q_0, \Lambda) \end{array} \right\} \text{for } \beta_j^S \quad (17)$$

$$\left. \begin{array}{l} u_0^S \mid \rho_{u1}, \rho_{u2} \propto \rho_{u1}^{\left(\frac{u_0^S}{2}\right)^{-1}} / \Gamma\left(\frac{u_0^S}{2}\right)^{\rho_{u2}} \\ v_0^S \mid \rho_{v1}, \rho_{v2} \sim \mathbf{Ga}(\rho_{v1}, \rho_{v2}) \end{array} \right\} \text{for } \sigma_j^2 \quad (18)$$

where: $1/\delta$ serves as a variance inflation factor; m_0 and Σ_{β_0} denote the mean vector and variance-covariance matrix of the GLM coefficients, respectively; q_0 and Λ are the degrees of freedom and the scale matrix of an Inverse Wishart hyperprior to

sample the variance-covariance matrix, respectively. As q_0 increases, its scale matrix Λ becomes smaller, and thus the variance-covariance matrix Σ_{β_0} becomes more influential (Kennedy and O'Hagan, 2001). ρ_{u1} and ρ_{u2} are the shape and rate parameters of a Gamma hyperprior. Such distributional knowledge discussed here can be taken into account when we simulate posterior parameter samples with the Gibbs sampler.

The selection of hyperpriors in Equations (15 to 18) is informed by distributions that are conjugate to the priors outlined in Equation (14), significantly streamlining the Bayesian updating process. For the hyperpriors of u_0^F and u_0^S specified in Equations (16) and (18), we utilize analytically derived kernels from Fink (1997) due to their conjugacy with the Gamma and Inverse Gamma distributions of ψ_j and σ_j^2 in Equations (14b) and (14c), which define their shape parameters. Consequently, the distributions of u_0^F and u_0^S will maintain their original kernel forms, even after incorporating updated values for ψ_j and σ_j^2 .

To break down Equations (14 to 18), we consider three distinct layers involved in the parameter inferences, as outlined below:

- Data layer: $N_1, \{Y_{1(1)}, \dots, Y_{1(N_1)}\}, \dots, N_H, \{Y_{H(1)}, \dots, Y_{H(N_H)}\} \mid \theta_j$
- Parameter layer: $\theta_j = \{\beta_j^F, \psi_j, \beta_j^S, \sigma_j^2 \mid \phi\}$
- Hyperparameter layer I and II: $\phi = \{\beta_0^F, \Sigma_{\beta_0}^F, u_0^F, v_0^F, \beta_0^S, \Sigma_{\beta_0}^S, u_0^S, v_0^S, \underline{m}_0, \underline{\delta}, \underline{q}_0, \underline{\Lambda}, \underline{\rho}_{u1}, \underline{\rho}_{u2}, \underline{\rho}_{v1}, \underline{\rho}_{v2}, m_0, \delta, q_0, \Lambda, \rho_{u1}, \rho_{u2}, \rho_{v1}, \rho_{v2}\}$

It is important to note that the inference for the main parameter level relies on cluster-specific data, while hyperparameter inference uses the entire dataset. However, assuming the hyperparameters $\underline{m}_0, \underline{\delta}, \underline{q}_0, \underline{\Lambda}, \underline{\rho}_{u1}, \underline{\rho}_{u2}, \underline{\rho}_{v1}, \underline{\rho}_{v2}, m_0, \delta, q_0, \Lambda, \rho_{u1}, \rho_{u2}, \rho_{v1}, \rho_{v2}$ are fixed makes their selection non-trivial. As a general guideline, flat hyperpriors - assigning equal probability to all potential parameter values - should be employed when specific knowledge about hyperparameter values is unavailable. Additional strategies for choosing hyperparameters can be found in the works of Fink (1997); Kennedy and O'Hagan (2001); Bousquet (2008), as well as their references.

As a Bayesian parametric framework, the hierarchical GLM necessitates defining the likelihood, prior, and hyperprior distributions at each tier of the hierarchy before deriving posterior estimates for the parameters. This is crucial because the calculation of marginal posterior means hinges on the joint distribution across all levels (Gelman and Carlin, 2013). The initial formulation of the cluster-specific joint probability (posterior \propto likelihood \times prior) based on Equation (14a) is as follows:

$$\prod_{h=1}^H f(N_{j(h)} \mid \beta_j^F, \psi_j) \prod_{i=1}^{N_j} f(Y_{j(hi)} \mid \beta_j^S, \sigma_j^2) p(\beta_j^F) p(\beta_j^S) p(\psi_j) p(\sigma_j^2) \quad \text{for cluster } j \quad (19)$$

However, within the complete hierarchical framework described in Equation (14), the baseline joint posterior for cluster j (to utilize in the Gibbs sampler) is expressed as:

$$\begin{aligned} & \prod_{h=1}^H f(N_{j(h)} \mid \beta_j^F, \psi_j) \prod_{i=1}^{N_j} f(Y_{j(hi)} \mid \beta_j^S, \sigma_j^2) && \Rightarrow \text{likelihood model} \\ & \times p(\beta_j^F \mid \beta_0^F, \Sigma_{\beta_0}^F) p(\beta_j^S \mid \beta_0^S, \Sigma_{\beta_0}^S) p(\psi_j \mid u_0^F, v_0^F) p(\sigma_j^2 \mid u_0^S, v_0^S) && \Rightarrow \text{prior model} \\ & \times p(\beta_0^F \mid \underline{m}_0, \underline{\delta}) p(\Sigma_{\beta_0}^F \mid \underline{q}_0, \underline{\Lambda}) p(\beta_0^S \mid m_0, \delta) p(\Sigma_{\beta_0}^S \mid q_0, \Lambda) && \Rightarrow \text{hyperprior model.I} \\ & \times p(u_0^F \mid \underline{\rho}_{u1}, \underline{\rho}_{u2}) p(v_0^F \mid \underline{\rho}_{v1}, \underline{\rho}_{v2}) p(u_0^S \mid \rho_{u1}, \rho_{u2}) p(v_0^S \mid \rho_{v1}, \rho_{v2}) && \Rightarrow \text{hyperprior model.II} \end{aligned} \quad (20)$$

As seen in Equation (20), we note that the hyperparameter layer remains unaffected by cluster membership j . To compute the expected aggregate claim amount $E[S_{j(h)}|\mathbf{X}]$ for policy h in cluster j from Equation (14a), we focus on calculating the marginal posterior means for the main parameters such as $E[\boldsymbol{\beta}_j^F|N_h]$, $E[\psi_j|N_h]$, $E[\boldsymbol{\beta}_j^S|Y_{h1}, Y_{h2}, \dots, Y_{hN_h}]$, and $E[\sigma_j^2|Y_{h(1)}, Y_{h(2)}, \dots, Y_{h(N_h)}]$. This can be also followed by constructing credibility intervals with the 5% level (given by the lower 2.5% and upper 2.5% of the posterior distribution for example) for Bayesian inference to estimate the likely range of true parameter values for the cluster j .

Posterior Computation for Bayesian Inference: The conjugate hyperpriors in Equations (15) to (18) yield closed-form full conditional posterior hyperpriors, and a Gibbs sampler, a Markov Chain Monte Carlo (MCMC) method applied to both negative binomial and log-normal GLMs, efficiently simulates the hyperparameters for posterior estimation using these hyperpriors. The analytically derived full conditional posterior hyperpriors (posterior hyperprior \propto prior \times hyperprior) utilized by the Gibbs sampler are shown below:

$$\left. \begin{aligned} \boldsymbol{\beta}_0^F | \underline{m}_0, \delta, N, \mathbf{X}^F, \Sigma_{\beta_0}^F &\sim \text{MVN}\left(\frac{\delta}{\delta+1}\underline{m}_0 + \frac{1}{\delta+1}\boldsymbol{\beta}^F, \frac{\Sigma_{\beta_0}^F}{\delta+1}\right) \\ \Sigma_{\beta_0}^F | \underline{q}_0, \underline{\Lambda}, N, \mathbf{X}^F, \boldsymbol{\beta}_0^F &\sim \text{IW}\left(\underline{q}_0 + 2, (\boldsymbol{\beta}_0^F - \boldsymbol{\beta}^F)(\boldsymbol{\beta}_0^F - \boldsymbol{\beta}^F)^T \right. \\ &\quad \left. + \delta(\boldsymbol{\beta}_0^F - \underline{m}_0)(\boldsymbol{\beta}_0^F - \underline{m}_0)^T + \underline{\Lambda}\right) \end{aligned} \right\} \text{for } \boldsymbol{\beta}^F \quad (21)$$

$$\left. \begin{aligned} u_0^F | \underline{\rho}_{u1}, \underline{\rho}_{u2}, N, \mathbf{X}^F, v_0^F &\propto (\psi_j \cdot \frac{v_0^F}{2} \cdot \underline{\rho}_{u1})^{(u_0^F/2)-1} / \Gamma(u_0^F/2)^{\underline{\rho}_{u2}+1} \\ v_0^F | \underline{\rho}_{v1}, \underline{\rho}_{v2}, N, \mathbf{X}^F, u_0^F &\sim \text{Ga}\left(\underline{\rho}_{v1} + \frac{u_{0j}^F}{2}, \underline{\rho}_{v2} + \psi_j\right) \end{aligned} \right\} \text{for } \psi_j \quad (22)$$

$$\left. \begin{aligned} \boldsymbol{\beta}_0^S | m_0, \delta, Y, \mathbf{X}^S, \Sigma_{\beta_0}^S &\sim \text{MVN}\left(\frac{\delta}{\delta+1}m_0 + \frac{1}{\delta+1}\boldsymbol{\beta}^S, \frac{\Sigma_{\beta_0}^S}{\delta+1}\right) \\ \Sigma_{\beta_0}^S | q_0, \Lambda, Y, \mathbf{X}^S, \boldsymbol{\beta}_0^S &\sim \text{IW}\left(q_0 + 2, (\boldsymbol{\beta}_0^S - \boldsymbol{\beta}^S)(\boldsymbol{\beta}_0^S - \boldsymbol{\beta}^S)^T \right. \\ &\quad \left. + \delta(\boldsymbol{\beta}_0^S - m_0)(\boldsymbol{\beta}_0^S - m_0)^T + \Lambda\right) \end{aligned} \right\} \text{for } \boldsymbol{\beta}^S \quad (23)$$

$$\left. \begin{aligned} u_0^S | \rho_{u1}, \rho_{u2}, Y, \mathbf{X}^S, v_0^S &\propto \left(\frac{1}{\sigma_j^2} \cdot \frac{v_0^S}{2} \cdot \rho_{u1}\right)^{(u_0^S/2)-1} / \Gamma(u_0^S/2)^{\rho_{u2}+1} \\ v_0^S | \rho_{v1}, \rho_{v2}, Y, \mathbf{X}^S, u_0^S &\sim \text{Ga}\left(\rho_{v1} + \frac{u_0^S}{2}, \rho_{v2} + \frac{1}{2\sigma_j^2}\right) \end{aligned} \right\} \text{for } \sigma_j^2 \quad (24)$$

Note that the expressions in Equations (21) to (24) are directly substituted into the joint conditional posterior of all parameters specified in Equation (20). Specifically, Equations (21) to (24) redefine the hyperprior terms $p(\boldsymbol{\beta}_0^F|\underline{m}_0, \delta)$, $p(\Sigma_{\beta_0}^F|\underline{q}_0, \underline{\Lambda})$, $p(\boldsymbol{\beta}_0^S|m_0, \delta)$, $p(\Sigma_{\beta_0}^S|q_0, \Lambda)$ and $p(u_0^F|\underline{\rho}_{u1}, \underline{\rho}_{u2})$, $p(v_0^F|\underline{\rho}_{v1}, \underline{\rho}_{v2})$, $p(u_0^S|\rho_{u1}, \rho_{u2})$, $p(v_0^S|\rho_{v1}, \rho_{v2})$ in Equation (20). For the main parameter term $p(\boldsymbol{\beta}_j^F|\boldsymbol{\beta}_0^F, \Sigma_{\beta_0}^F)$, $p(\boldsymbol{\beta}_j^S|\boldsymbol{\beta}_0^S, \Sigma_{\beta_0}^S)$, $p(\psi_j|u_0^F, v_0^F)$, $p(\sigma_j^2|u_0^S, v_0^S)$ in Equation (20); however, we employ the Metropolis-Hastings (MH) algorithm within the Gibbs sampler due to the lack of conjugate priors for the primary parameters $\boldsymbol{\beta}_j^F, \psi_j, \boldsymbol{\beta}_j^S, \sigma_j^2$ that align with our negative binomial and log-normal likelihoods.

Given the two outcome models - claim counts N_h and claim amounts \tilde{Y}_h , both Gibbs samplers can be executed in parallel to assess Equation (14a), assuming no covariate errors. Figure 2 illustrates the process of re-estimating the model parameters

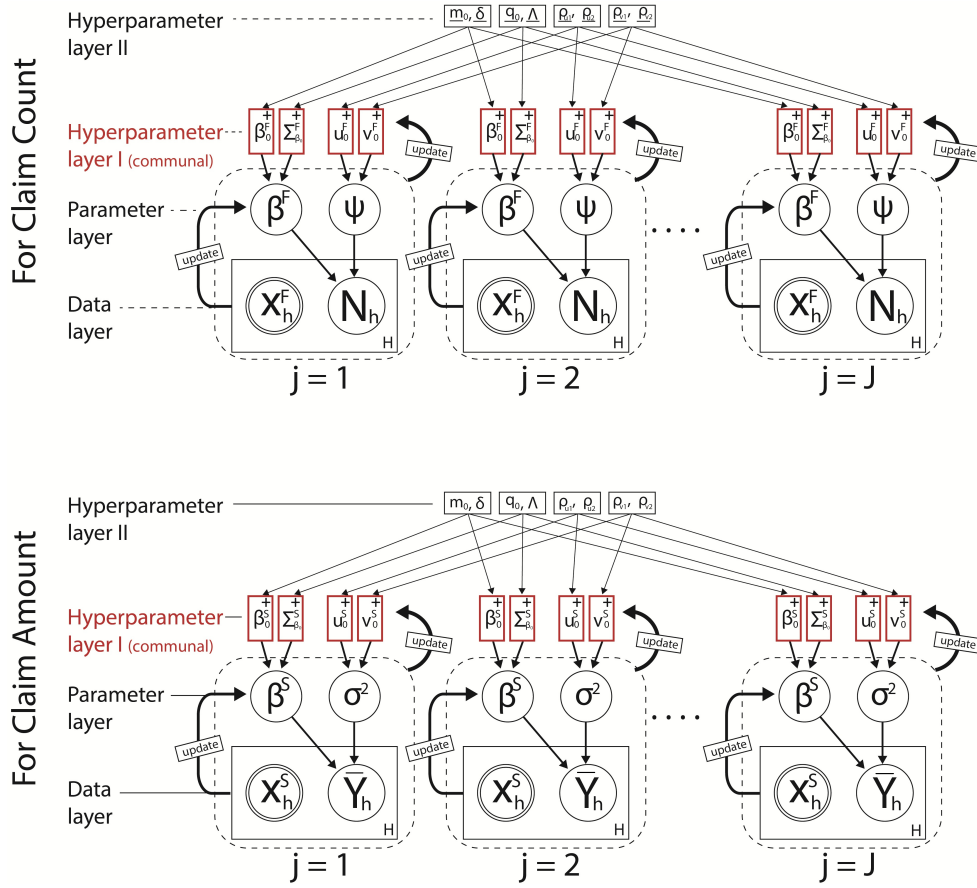


Figure 2: The acyclic graphical representation of the flows of the parameter updates in the hierarchical GLM. This is a snapshot for a single iteration ($M=1$).

$\beta_j^F, \psi_j, \beta_j^S, \sigma_j^2$ for each cluster j using the two distinct Gibbs samplers. Algorithm S2 in Part 2 of the on-line supplementary file provides a detailed description of the Gibbs sampler tailored for modeling claim amounts Y_h using the log-normal density. The approach for claim counts N_h based on the negative binomial distribution would be analogous and therefore omitted for brevity.

At the initialization of the Gibbs sampler, it is crucial to establish initial values for the hyperparameters $\phi : \{\beta_0^S, \Sigma_{\beta_0}^S, u_0^S, v_0^S, m_0, \delta, q_0, \Lambda, \rho_{u1}, \rho_{u2}, \rho_{v1}, \rho_{v2}, c_0^S, d_0^S, g_0^S, h_0^S\}$, which support the prior choices in Equations (14c) and (14e) for the log-normal GLM. Using these values, the initial parameters for the outcome $\theta^{(old)} : \{\beta_j^{S(old)}, \sigma_j^{2(old)}\}$ and covariate $w : \{\pi_j^S, \lambda_j^{2S}\}$ are determined. These parameters ultimately yield the values for the communal hyperparameters $\beta_0^{S+}, \Sigma_{\beta_0}^{S+}, u_0^{S+}, v_0^{S+}$, which guide both complete pooling and no pooling throughout the Gibbs sampling process. The two-stage Gibbs sampler for the log-normal model is executed as follows:

[Stage.1] Sampling with Complete Pooling

The Gibbs sampler first estimates the outcome parameters θ (without clustering) using a complete pooling approach. This aims to refine the communal hyperparameters $\beta_0^{S+}, \Sigma_{\beta_0}^{S+}, u_0^{S+}, v_0^{S+}$, as outlined in Equations (23) and (24). Given the absence of conjugate priors for the log-normal outcome, the Metropolis-Hastings (MH) algorithm is used, with the prior serving as its proposal distribution from where the candidate

values are drawn. Proposal samples $\theta^{(new)}$ are generated based on these communal hyperparameters, yielding $\theta^{(*)}$, which support the global mean $E[Y_{hi}|\mathbf{X}]$, while re-updating the communal hyperparameters $\beta_0^{S+}, \Sigma_{\beta_0}^{S+}, u_0^{S+}, v_0^{S+}$. See Algorithm S2 in Part 2 of the on-line supplementary file.

[Stage.2] Sampling without Pooling

In this stage, with the cluster membership j already established, the goal is to generate accurate parameter estimates $\theta^{(*)}$ for each cluster. These estimates are guided by the communal hyperparameters $\beta_0^{S+}, \Sigma_{\beta_0}^{S+}, u_0^{S+}, v_0^{S+}$ obtained from [Stage.1]. By using these communal hyperparameters, the Gibbs sampler optimizes $\theta_j^{(*)}$ for each risk cluster, minimizing within-cluster variability. The log-likelihood value can be computed after each iteration to monitor convergence. The implementation detail is provided in Algorithm S2 in Part 2 of the on-line supplementary file.

3.4 Clustering $S_h|\mathbf{X}^F, \mathbf{X}^S$ with NDB Case Covariate: RQ2.Tackling NDB errors

This section introduces a novel method for handling the NDB covariate (RQ2) based on the hierarchical GLM framework and partial pooling for risk premium modeling. Consistent with the parametric Bayesian principle (Hong and Martin, 2018), we assume that the risk clusters $j = 1, \dots, J$ have already been established. Given two covariate matrices $\mathbf{X}^F : \{\mathbf{x}^F, \mathbf{z}^F\}$ for the negative binomial outcome N_h and $\mathbf{X}^S : \{\mathbf{z}^S, \mathbf{x}^S\}$ for the log-normal outcome \tilde{Y}_h respectively, we consider the case where the continuous covariate \mathbf{x}^S contains mismeasurements, classified as Non-Differential Berkson (NDB) measurement errors. Since the covariates for the negative binomial outcome N_h are assumed complete, this section particularly focuses on the log-normal outcome \tilde{Y}_h and its covariate \mathbf{x}^S within the hierarchical framework. For simplicity, we will omit the superscript s and refer to \mathbf{x}^S as simply \mathbf{x} in the remainder of the discussion.

In Section 3.2, we briefly introduced Gustafson's error correction method to address model risk associated with the NDB covariate. The key idea is to estimate or sample the parameter values for the true covariate using a specialized joint model. This model incorporates prior knowledge of θ_j through the linking component $f(\mathbf{x}^*|\mathbf{x}, \theta_j)$, which captures the cluster-wise relationship between the true covariate \mathbf{x} and the NDB covariate \mathbf{x}^* (Grace et al., 2021). The main objective of this section is to accurately estimate the parameter values θ_j for the joint model.

To elaborate, the complete joint model that encompasses the outcome \tilde{Y}_h , the NDB covariate \mathbf{x}^* , the true covariate \mathbf{x} , and the additional covariate \mathbf{z} is expressed as per Equation (8) as:

$$f(\tilde{Y}_h, \mathbf{x}^*, \mathbf{x}|\mathbf{z}) = \underbrace{f(\tilde{Y}_h|\mathbf{x}^*, \mathbf{x}, \mathbf{z})}_{\text{outcome}} \times \underbrace{f(\mathbf{x}^*|\mathbf{x}, \mathbf{z})}_{\text{linking component}} \times \underbrace{f(\mathbf{x}|\mathbf{z})}_{\text{covariate}} \quad (25)$$

where the true covariate \mathbf{x} is included and is unknown in real life. As outlined in Section 3.2, the full joint distribution can be simplified to the form in Equation (25) because the NDB covariate \mathbf{x}^* is uncorrelated with any other variables except the true covariate \mathbf{x} itself. To integrate Gustafson's complete joint model for the NDB covariate from Equation (25) into the hierarchical GLM framework, we redefine the risk premium modeling structure along with the priors originally specified in Equation

(14) as below:

$$E[S_{j(h)}|\mathbf{X}^F, \mathbf{X}^{S*}] = \exp(\mathbf{X}^F \boldsymbol{\beta}_j^F + \mathbf{X}^{S*} \boldsymbol{\beta}_j^S + \frac{1}{2} \sigma_j^2) \quad (26a)$$

$$\text{For } N_h \left\{ \begin{array}{l} \boldsymbol{\beta}_j^F | \boldsymbol{\beta}_0^F, \Sigma_{\beta_0}^F \sim \mathbf{MVN}(\boldsymbol{\beta}_0^F, \Sigma_{\beta_0}^F) \\ \psi_j | u_0^F, v_0^F \sim \mathbf{Ga}(\frac{u_0^F}{2}, \frac{v_0^F}{2}) \end{array} \right. \quad (26b)$$

$$\text{For } \bar{Y}_h \left\{ \begin{array}{l} \boldsymbol{\beta}_j^S | \boldsymbol{\beta}_0^S, \Sigma_{\beta_0}^S \sim \mathbf{MVN}(\boldsymbol{\beta}_0^S, \Sigma_{\beta_0}^S) \\ \sigma_j^2 | u_0^S, v_0^S \sim \mathbf{InvGa}(\frac{u_0^S}{2}, \frac{v_0^S}{2}) \end{array} \right. \quad (26c)$$

$$\text{For } \mathbf{X}^F \left\{ \begin{array}{l} x_{j(h)}^F \sim \mathbf{N}(E[\mathbf{x}_j^F], \lambda_j^{2F}) \\ \lambda_j^{2F} \sim \mathbf{InvGa}(c_0^F, d_0^F) \\ z_{j(h)}^F \sim \mathbf{Bernoulli}(\pi_j^F) \\ \pi_j^F \sim \mathbf{Beta}(g_0^F, h_0^F) \end{array} \right. \quad (26d)$$

$$\text{For } \mathbf{X}^{S*} \left\{ \begin{array}{l} \mathbf{x}_{j(h)}^{S*} | \mathbf{x}_{j(h)}^S \sim \mathbf{N}(\mathbf{x}_{j(h)}^S, \boldsymbol{\Sigma}_j^2) \\ \boldsymbol{\Sigma}_j^2 \sim \text{undetermined} \\ \mathbf{x}_{j(h)}^S | z_{j(h)}^S \sim \mathbf{N}(\kappa_{j0} + \kappa_{j1} z_{j(h)}^S, \lambda_j^{2S}) \\ \kappa_j \sim \mathbf{MVN}(\tilde{\boldsymbol{\kappa}}, \lambda_j^{2S} \tilde{\Sigma}_{\boldsymbol{\kappa}}) \\ \lambda_j^{2S} \sim \mathbf{InvGa}(c_0^S, d_0^S) \\ z_{j(h)}^S \sim \mathbf{Bernoulli}(\pi_j^S) \\ \pi_j^S \sim \mathbf{Beta}(g_0^S, h_0^S) \end{array} \right. \quad (26e)$$

Within the hierarchical framework established in Equation (26), both models for $N_h|\mathbf{X}^F$ and $\bar{Y}_h|\mathbf{X}^{S*}$ are elaborated to compute $S_h|\mathbf{X}$. However, this section will focus solely on modeling the claim amount $\bar{Y}_h|\mathbf{X}^{S*}$ for demonstration purposes. This is based on the assumption that the claim amount model $\bar{Y}_h|\mathbf{X}^{S*}$ is solely influenced by the NDB covariate \mathbf{x}^* . For simplicity, we will denote the covariate matrix \mathbf{X}^{S*} and the continuous covariate vector \mathbf{x}^{S*} as \mathbf{X}^* and \mathbf{x}^* , respectively.

We will start by constructing the complete joint model outlined in Equation (25) to derive the model for the claim amount $\bar{Y}_h|\mathbf{X}^*$. This entails identifying previously defined 'linking component' (also known as measurement model, termed by Gustafson, 2008) that connects \mathbf{x}^* to \mathbf{x} within a hierarchical framework. We define the linking component under the assumption that \mathbf{x}^* is a normally distributed random variable.

$$f_N(x_h^*|x_h) = \frac{1}{\sqrt{2\pi\boldsymbol{\Sigma}_j^2}} \exp\left\{-\frac{(x_h^* - x_h)^2}{2\boldsymbol{\Sigma}_j^2}\right\} \quad (27)$$

where $x_h^*|x_h = x_h + \epsilon_j \sim \mathbf{N}(x_h, \boldsymbol{\Sigma}_j^2)$, $\boldsymbol{\Sigma}_j^2 : V(\mathbf{x}^*|\mathbf{x})$, $\epsilon_j \sim \mathbf{N}(0, \sigma_{j\epsilon}^2)$, and $\boldsymbol{\Sigma}_j^2 = \sigma_x^2 + \sigma_{j\epsilon}^2$. This suggests that the prior knowledge of the dispersion $\boldsymbol{\Sigma}_j^2$ in Equation (27) captures the cluster-specific traits of the NDB covariate, derived from the relationship between the NDB covariate \mathbf{x}^* and the true covariate \mathbf{x} . However, initially estimating $\boldsymbol{\Sigma}_j^2$ is challenging because both the true covariate variance σ_x^2 and the cluster-wise error variance $\sigma_{j\epsilon}^2$ are unknown.

To complete the joint model in Equation (25), we define the *outcome* and *covariate* components (referred to as ‘exposure model’ in Gustafson’s terminology as outlined below:

$$f_{\text{LogN}}(\bar{Y}_h | x_h, z_h) = \frac{1}{\bar{Y}_h \sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{1}{2} \left[\frac{\ln \bar{Y}_h - (\beta_{j0} + \beta_{j1}x_h + \beta_{j2}z_h)}{\sigma_j} \right]^2\right\} \quad (28a)$$

$$f_N(x_h | z_h) = \frac{1}{\sqrt{2\pi\lambda_j^2}} \exp\left\{-\frac{(x_h - \{\kappa_{j0} + \kappa_{j1}z_h\})^2}{2\lambda_j^2}\right\} \quad (28b)$$

where $\bar{Y}_h | x_h, z_h \sim \text{LogN}(\mathbf{X}_h \boldsymbol{\beta}_j, \sigma_j^2)$, $\sigma_j^2 : V(\bar{Y} | \mathbf{X})$, $x_h | z_h \sim \mathbf{N}(\kappa_{j0} + \kappa_{j1}z_h, \lambda_j^2)$, and $\lambda_j^2 : V(\mathbf{x} | \mathbf{z})$ as shown in Equations (26c) and (26e). Note that the three terms in Equations (27) and (28) are essential building blocks, as they must be multiplied to form the complete joint model specified in Equation (25). However, the two terms in Equation (28) are largely theoretical since the true covariate \mathbf{x} remains unknown. Instead, more practical terms available for implementation are:

$$f_{\text{LogN}}(\bar{Y}_h | x_h^*, z_h) = \frac{1}{\bar{Y}_h \sqrt{2\pi\hat{\sigma}_j^2}} \exp\left\{-\frac{1}{2} \left[\frac{\ln \bar{Y}_h - (\hat{\beta}_{j0} + \hat{\beta}_{j1}x_h^* + \hat{\beta}_{j2}z_h)}{\hat{\sigma}_j} \right]^2\right\} \quad (29a)$$

$$f_N(x_h^* | z_h) = \frac{1}{\sqrt{2\pi\hat{\lambda}_j^2}} \exp\left\{-\frac{(x_h^* - \{\hat{\kappa}_{j0} + \hat{\kappa}_{j1}z_h\})^2}{2\hat{\lambda}_j^2}\right\} \quad (29b)$$

where $\bar{Y}_h | x_h^*, z_h \sim \text{LogN}(\mathbf{X}^* \hat{\boldsymbol{\beta}}_j, \hat{\sigma}_j^2)$, $\hat{\sigma}_j^2 : V(\bar{Y} | \mathbf{X}^*)$, $x_h^* | z_h \sim \mathbf{N}(\hat{\kappa}_{j0} + \hat{\kappa}_{j1}z_h, \hat{\lambda}_j^2)$, and $\hat{\lambda}_j^2 : V(\mathbf{x}^* | \mathbf{z})$ (the notation $\hat{\cdot}$ is used to indicate that these parameters are derived from the covariate with NDB errors, prior to correction). By multiplying these two terms - the outcome model and the covariate model - in Equation (29), we obtain the *incomplete joint model* in Equation (30), as per Equation (9), which serves as the practical available solution:

$$f(\bar{Y}_h, x_h^* | z) = \frac{1}{\bar{Y}_h (2\pi) \hat{\sigma}_j \hat{\lambda}_j} \times \exp\left(-\frac{1}{2\hat{\sigma}_j^2} \left[(\ln \bar{Y}_h - \hat{\beta}_{j0} - \hat{\beta}_{j2}z_h) - \hat{\beta}_{j1}x_h^* \right]^2\right) \\ \times \exp\left(-\frac{1}{2\hat{\lambda}_j^2} \left[x_h^* - (\hat{\kappa}_{j0} + \hat{\kappa}_{j1}z_h) \right]^2\right) \quad (30)$$

However, this joint model has a notable limitation: it is developed without the true covariate \mathbf{x} . Fortunately, connecting the complete joint model in Equation (25) with the incomplete joint model in Equation (30) is a straightforward process. As noted in Section 3.2, we can obtain another incomplete model by marginalizing the complete joint model in Equation (25) over the true covariate \mathbf{x} using the following integral:

$$\int f(\bar{Y}_h, x_h^*, x_h | z_h) dx_h \\ = \int \frac{1}{\sigma_j \bar{Y}_h \sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left[\frac{\ln \bar{Y}_h - (\beta_{j0} + \beta_{j1}x_h + \beta_{j2}z_h)}{\sigma_j} \right]^2\right\} \\ \times \frac{1}{\sqrt{2\pi\boldsymbol{\Sigma}_j^2}} \exp\left\{-\frac{(x_h^* - x_h)^2}{2\boldsymbol{\Sigma}_j^2}\right\} \times \frac{1}{\sqrt{2\pi\lambda_j^2}} \exp\left\{-\frac{(x_h - \{\kappa_{j0} + \kappa_{j1}z_h\})^2}{2\lambda_j^2}\right\} dx_h \quad (31)$$

The evaluation process for the integral in Equation (31) is detailed in Part 3 of the on-line supplementary file, and the resulting solution is presented below.

$$\begin{aligned}
& \int f(\tilde{Y}_h, x_h^*, x_h | z_h) dx_h \\
&= \frac{1}{\tilde{Y}_h (2\pi) \sigma_j \mathfrak{T}_j \lambda_j} \left(\frac{\sigma_j^2 \lambda_j^2 \mathfrak{T}_j^2}{\beta_{j1}^2 \mathfrak{T}_j^2 \lambda_j^2 + \sigma_j^2 \lambda_j^2 + \sigma_j^2 \mathfrak{T}_j^2} \right)^{1/2} \\
& \times \exp \left(-\frac{1}{2} \left(\frac{\lambda_j^2 + \mathfrak{T}_j^2}{\beta_{j1}^2 \mathfrak{T}_j^2 \lambda_j^2 + \sigma_j^2 \lambda_j^2 + \sigma_j^2 \mathfrak{T}_j^2} \right) \left[(\ln \tilde{Y}_h - \beta_{j0} - \beta_{j2} z_h) - \frac{\beta_{j1} \left(\frac{x_h^*}{\mathfrak{T}_j} + \frac{\kappa_{j0} + \kappa_{j1} z_h}{\lambda_j^2} \right)}{\frac{1}{\mathfrak{T}_j} + \frac{1}{\lambda_j^2}} \right]^2 \right) \\
& \times \exp \left(-\frac{1}{2} \left(\frac{1}{\mathfrak{T}_j^2 + \lambda_j^2} \right) \left[(x_h^* - (\kappa_{j0} + \kappa_{j1} z_h))^2 \right] \right) = f(\tilde{Y}_h, x_h^* | z_h)
\end{aligned} \tag{32}$$

In Equation (32), the unobservable true covariate term \mathbf{x} is completely eliminated through this marginalization process. As a result, the integral solution in Equation (32), derived from the complete joint model, aligns with the incomplete joint model in Equation (30). The key point is that while these two solutions originate from different sources, they both describe the same incomplete joint model $f(\tilde{Y}_h, x_h^* | z_h)$ in a practical framework, without needing to evaluate or sample the true covariate \mathbf{x} . This allows for directly matching the parameters of the complete and incomplete joint models.

This relationship can be formalized through a system of equations. Since the parameters of the incomplete joint model $\hat{\beta}_{j0}, \hat{\beta}_{j1}, \hat{\beta}_{j2}, \hat{\sigma}_j^2, \hat{\lambda}_j^2, \hat{\kappa}_{j0}, \hat{\kappa}_{j1}$ (with the $\hat{\cdot}$ denoting the parameter estimates based on NDB error) in Equation (30) are accessible, the parameters of the marginalized complete joint model $\beta_{j0}, \beta_{j1}, \beta_{j2}, \sigma_j^2, \lambda_j^2, \kappa_{j0}, \kappa_{j1}$ in Equation (32) can be expressed in terms of these counterparts. Solving this system of equations could provide a guideline for adjusting parameter estimates in the presence of the NDB covariate \mathbf{x}^* (RQ2). The full derivation and detailed explanation of the system of equations are provided in Part 3 of the on-line supplementary file, and the resulting system of equations for the complete joint model parameters is shown below.

$$\lambda_j^2 = \hat{\lambda}_j^2 - \mathfrak{T}_j^2 \tag{33a}$$

$$\kappa_{j0} = \hat{\kappa}_{j0} \tag{33b}$$

$$\kappa_{j1} = \hat{\kappa}_{j1} \tag{33c}$$

$$\beta_{j1} = \frac{\hat{\beta}_{j1} \hat{\lambda}_j^2}{\hat{\lambda}_j^2 - \mathfrak{T}_j^2} \tag{33d}$$

$$\beta_{j0} = \hat{\beta}_{j0} - \frac{\hat{\beta}_{j1} \hat{\kappa}_{j0} \mathfrak{T}_j^2}{\hat{\lambda}_j^2 - \mathfrak{T}_j^2} \tag{33e}$$

$$\beta_{j2} = \hat{\beta}_{j2} - \frac{\hat{\beta}_{j1} \hat{\kappa}_{j1} \mathfrak{T}_j^2}{\hat{\lambda}_j^2 - \mathfrak{T}_j^2} \tag{33f}$$

$$\sigma_j^2 = \hat{\sigma}_j^2 - \frac{\beta_{j1}^2 \mathfrak{T}_j^2 (\hat{\lambda}_j^2 - \mathfrak{T}_j^2)}{\hat{\lambda}_j^2} \tag{33g}$$

Our novel prior knowledge on \mathfrak{T}_j^2 and scaling factor ζ : As shown in Equation (33), the adjustment of the parameter estimation in the incomplete joint model largely hinges

on the value of \mathfrak{I}_j^2 . However, as mentioned previously, estimating \mathfrak{I}_j^2 in Equation (27) is challenging until the relationship between the NDB covariate \mathbf{x}^* and the true covariate \mathbf{x} is clarified. The system of equations, particularly Equation (33a), now reveals that \mathfrak{I}_j^2 can be expressed as $\hat{\lambda}_j^2 - \lambda_j^2$. In other words, $V(\mathbf{x}^*|\mathbf{x}) = V(\mathbf{x}^*|\mathbf{z}) - V(\mathbf{x}|\mathbf{z})$. From this, several key insights (our own findings) can be drawn:

- (i) $\hat{\lambda}_j^2 : V(\mathbf{x}^*|\mathbf{z})$ is always greater than $\lambda_j^2 : V(\mathbf{x}|\mathbf{z})$ according to Equation (33a).
- (ii) Given that $x^*|z \sim \mathbf{N}(\hat{\kappa}_{j0} + \hat{\kappa}_{j1}z, \hat{\lambda}_j^2)$ and $x|z \sim \mathbf{N}(\kappa_{j0} + \kappa_{j1}z, \lambda_j^2)$, it appears that $\hat{\kappa}_{j0} = \kappa_{j0}$ and $\hat{\kappa}_{j1} = \kappa_{j1}$, hence $E[\mathbf{x}^*|\mathbf{z}] = E[\mathbf{x}|\mathbf{z}]$ according to Equation (33b) and (33c).
- (iii) Given (i) and (ii), it is safe to say that the variance of the true covariate can be a scalar multiple of the variance of the NDB covariate: $\lambda_j^2 = \zeta \times \hat{\lambda}_j^2$ where $0 < \zeta < 1$ is a *scaling factor*.

Equation (33a), along with the key insights (i), (ii), and (iii), establishes a crucial prior knowledge regarding the dispersion parameter of the linking component $\mathfrak{I}_j^2 : V(\mathbf{x}^*|\mathbf{x})$. Together, they highlight that

- $\mathfrak{I}_j^2 = \hat{\lambda}_j^2 - \lambda_j^2$ i.e., $V(\mathbf{x}^*|\mathbf{x}) = V(\mathbf{x}^*|\mathbf{z}) - V(\mathbf{x}|\mathbf{z})$ from Equation (33a).
- $\lambda_j^2 = \zeta \times \hat{\lambda}_j^2$ i.e., $V(\mathbf{x}|\mathbf{z}) = \zeta \times V(\mathbf{x}^*|\mathbf{z})$ from the findings (i),(ii),(iii).

Ultimately, the relationships described above can be expressed through the equation:

$$\mathfrak{I}_j^2 = (1 - \zeta)\hat{\lambda}_j^2 \text{ or } V(\mathbf{x}^*|\mathbf{x}) = (1 - \zeta) V(\mathbf{x}^*|\mathbf{z}) \quad (34)$$

which demonstrates that the dispersion parameter \mathfrak{I}_j^2 can be represented as a fraction $(1 - \zeta)$ of the variance $\hat{\lambda}_j^2 : V(\mathbf{x}^*|\mathbf{z})$. Thus, with an estimate for $\hat{\lambda}_j^2$, we can determine \mathfrak{I}_j^2 using the scaling factor $0 < \zeta < 1$. In other words, given the availability of both \mathbf{x}^* and \mathbf{z} , we can utilize $\hat{\lambda}_j^2 : V(\mathbf{x}^*|\mathbf{z})$ as a proxy for estimating $\mathfrak{I}_j^2 : V(\mathbf{x}^*|\mathbf{x})$. The scaling factor $0 < \zeta < 1$ reflects our confidence in the adequacy of the known covariate \mathbf{z} as a substitute for the unobservable true covariate \mathbf{x} . A higher value of ζ indicates that \mathfrak{I}_j^2 relies less on $V(\mathbf{x}^*|\mathbf{z})$. Conversely, a lower ζ enhances the dependency on the observed variance $V(\mathbf{x}^*|\mathbf{z})$ to inform our understanding of \mathfrak{I}_j^2 .

In this paper, we suggest utilizing the insights gained about \mathfrak{I}_j^2 as the prior knowledge for modeling the probability distribution of $\mathbf{x}^*|\mathbf{x}$, which elucidates the cluster-specific relationship between the NDB covariate \mathbf{x}^* and the true covariate \mathbf{x} . Additionally, we will implement a sensitivity analysis for ζ and assess how changes in this scaling factor impact the estimates of \mathfrak{I}_j^2 , in order to shed light on how the scaling factor ζ influences the effectiveness of the error correction approach. This, in turn, will facilitate the identification of the ideal value for \mathfrak{I}_j^2 to improve estimation results impacted by the model risk associated with the NDB covariate \mathbf{x}^* .

Gibbs sampler modification with the Gustafson correction: We propose implementing the following enhancements (additional steps) to integrate the resulting system of equations from Equation (33) into the Gibbs sampler for our hierarchical GLM development, as described in Algorithm S2 in Part 2 of the on-line supplementary file:

- (a) In line 12, assuming the NDB covariate value x_h^* in \mathbf{x}^* at observation h in the risk cluster j , we add a step to sample the posterior parameters for the linking component in Equation (27) and the covariate model in Equation (29b). This sets the stage for parameter adjustment using the system of equations outlined in line 18.

$$w_j : \begin{cases} \pi_j \sim \mathbf{Beta}(g_0 + \Sigma z_j, h_0 + n_j - \Sigma z_j) \\ \hat{\boldsymbol{\kappa}}_j \sim \mathbf{MVN}\left(\left[(\tilde{\Sigma}_k^{-1} + \mathbb{K}_1^T \mathbb{K}_1)^{-1}(\tilde{\Sigma}_k^{-1} \tilde{\boldsymbol{\kappa}} + \mathbb{K}_2)\right], \hat{\lambda}_j^2 \left[\tilde{\Sigma}_k^{-1} + \mathbb{K}_1^T \mathbb{K}_1\right]^{-1}\right) \\ \hat{\lambda}_j^2 \sim \mathbf{InvGa}\left(\frac{c_0 + n_j}{2}, \frac{1}{2}(d_0 + \Sigma(x_j - \hat{\kappa}_{0j} + \hat{\kappa}_{1j} z_j)^2)\right) \\ \mathfrak{I}_j^2 = (1 - \zeta) \hat{\lambda}_j^2 \end{cases} \quad (35)$$

The derivations of the posterior densities for the covariate model parameters - $\pi_j, \hat{\boldsymbol{\kappa}}_j, \hat{\lambda}_j^2$ - are thoroughly outlined in Part 1 of the on-line supplementary file. The scaling factor ζ is set by researchers based on the findings from the sensitivity analysis, ensuring it aligns with the anticipated error levels in the NDB covariate across various scenarios. A more in-depth exploration of this experiment will be presented in the following section.

- (b) In line 18, we utilize the system of equations from Equation (33) to refine the estimated outcome parameter values $\boldsymbol{\theta}_j^{(*)} : \{\beta_j^{(*)}, \sigma_j^{2(*)}\}$ by using the parameter samples from the incomplete joint model: $\hat{\beta}_{j0}, \hat{\beta}_{j1}, \hat{\beta}_{j2}, \hat{\sigma}_j, \hat{\lambda}_j, \hat{\kappa}_{j0}, \hat{\kappa}_{j1}$, along with the variance $\mathfrak{I}_j^2 = (1 - \zeta) \hat{\lambda}_j^2$ (where $0 < \zeta < 1$) of the linking component in Equation (27). It is important to note that the parameter samples obtained from the incomplete joint model during the Gibbs sampler must satisfy specific criteria outlined in Equation (33). For instance, as indicated in Equation (33a), $\hat{\lambda}_j^2$ values must always exceed λ_j^2 values. Additionally, according to Equation (33g), $\hat{\sigma}_j^2$ must be always greater than the value given by $\frac{\beta_{j1}^2 \mathfrak{I}_j^2 \lambda_j^2}{\lambda_j^2}$. These constraints ensure that the sampled parameters maintain valid relationships with the true parameters, allowing the Gibbs sampler to filter out any samples that do not conform to these requirements.

4 Empirical Study

4.1 Data: Local Government Property Insurance Fund

We evaluate our hierarchical GLM using an insurance dataset from the Wisconsin Local Government Property Insurance Fund (LGPIF)¹. Compiled by the actuarial research team at the University of Wisconsin, this dataset encompasses information on insurance coverage for $H = 1,679$ policies pertaining to various government building units across Wisconsin. This dataset presents unique challenges, particularly regarding unobservable heterogeneity (RQ1) in the log-normal outcome variable influenced by the NDB covariate (RQ2). This paper employs a frequency-severity approach to

¹ The Local Government Property Insurance Fund was established to provide insurance coverage to government properties not owned by the State of Wisconsin. The Fund made insurance available for local government property such as municipal buildings, schools, libraries and vehicles. For further details, refer to <https://sites.google.com/a/wisc.edu/local-government-property-insurance-fund>

risk premium modeling, incorporating four covariates: two for the claim count model - a binary covariate \mathbf{z}^F (*AC15*: 1 or 0) and a continuous covariate \mathbf{x}^F (*LnCoverage*) - and two for the claim amount model - also a binary covariate \mathbf{z}^S (*Fire5*: 1 or 0) and a continuous covariate \mathbf{x}^S (*lnDeduct*). The outcome variables - claim count and claim amount - are represented as N_h and \tilde{Y}_h respectively. Therefore, the structure of this dataset is outlined as follows:

$$\begin{aligned}
& \text{Year}_1, \quad \dots, \quad \text{Year}_y \\
\text{Policy } (h = a): & \quad \{(N_a, \mathbf{X}_a^F, Y_{a(1)}, \dots, Y_{a(N_a)}, \mathbf{X}_a^S), \dots, (N_a, \mathbf{X}_a^F, Y_{a(1)}, \dots, Y_{a(N_a)}, \mathbf{X}_a^S)\} \\
\text{Policy } (h = b): & \quad \{(N_b, \mathbf{X}_b^F, Y_{b(1)}, \dots, Y_{b(N_b)}, \mathbf{X}_b^S), \dots, (N_b, \mathbf{X}_b^F, Y_{b(1)}, \dots, Y_{b(N_b)}, \mathbf{X}_b^S)\} \\
& \quad \vdots \\
\text{Policy } (h = H): & \quad \{(N_H, \mathbf{X}_H^F, Y_{H(1)}, \dots, Y_{H(N_H)}, \mathbf{X}_H^S), \dots, (N_H, \mathbf{X}_H^F, Y_{H(1)}, \dots, Y_{H(N_H)}, \mathbf{X}_H^S)\}
\end{aligned}$$

The experiment concerns predicting the aggregate claim amount $E[S_h|\mathbf{X}]$ using the frequency-severity principle to establish risk premiums for specific policies. Predictions are categorized into six distinct entity types: *city*, *county*, *school*, *town*, *village*, and *miscellaneous*. These classifications represent the source of the insured property, effectively organizing the policies into six fixed risk clusters (i.e., $j = 1, \dots, 6$), representing six distinct entity types. This approach allows for the incorporation of varying risk characteristics associated with each entity type in the predictions. As discussed in the main text, the continuous covariate \mathbf{x}^S in the claim amount model is assumed to be influenced by NDB error.

4.2 Model Implementation

Using simulation data, we evaluate our hierarchical GLM-based approach for correcting NDB errors against the SIMEX error correction method discussed in Section 2.2. Both methodologies seek to simultaneously tackle two model risk issues - heterogeneity (RQ1) and NDB covariates (RQ2) - for risk premium development. As mentioned in Section 2.2, many existing error correction techniques rely on gold standard data (e.g., subsets of the true covariate) to estimate the true covariate by establishing relationships between mismeasured and true values. However, in practical applications, access to gold standard data is often limited or nonexistent. In this regard, this experiment aims to create methodologies that effectively address the NDB measurement errors without the need for gold standard data, leveraging the prior knowledge derived: $\mathfrak{T}_j^2 = (1 - \zeta)\hat{\lambda}_j^2$ from Section 3.4. We develop simulation data to explore the relationship between the scaling factor ζ and the severity of NDB errors, represented by the error rate R_{ϵ_x} , in the NDB covariate \mathbf{x}^* . This process hinges on two elements: a) gold-standard datasets with no errors, and b) datasets with controlled error rates in the NDB covariate \mathbf{x}^* . By comparing models from error-free data with those generated from datasets with varying NDB error levels, we assess the efficiency of Gustafson's correction approach, which is based on a hierarchical GLM, in relation to the error rate R_{ϵ_x} and the scaling factor $0 < \zeta < 1$. Our ultimate goal is to establish a guideline for selecting an optimal scaling factor ζ that accounts for the unknown variance $\mathfrak{T}_j^2 = (1 - \zeta)\hat{\lambda}_j^2$ when the error rate R_{ϵ_x} is known. Once established, this guideline could aid in estimating the true covariate when gold-standard data is unavailable.

Design of simulation data: We construct simulation datasets with different error rates R_{ϵ_x} , derived from real LGPIF data, while preserving the original LGPIF dataset as the gold standard. These simulation datasets are generated by deliberately introducing controlled NDB errors into the true covariate \mathbf{x}^S at different error rates R_{ϵ_x} . The method for creating NDB errors in the covariate \mathbf{x}^{S*} , leading to variations in R_{ϵ_x} , is summarized in Figure 3. As defined in Section 2.2, the NDB error $\epsilon_j \sim \mathbf{N}(0, \sigma_{j\epsilon}^2)$

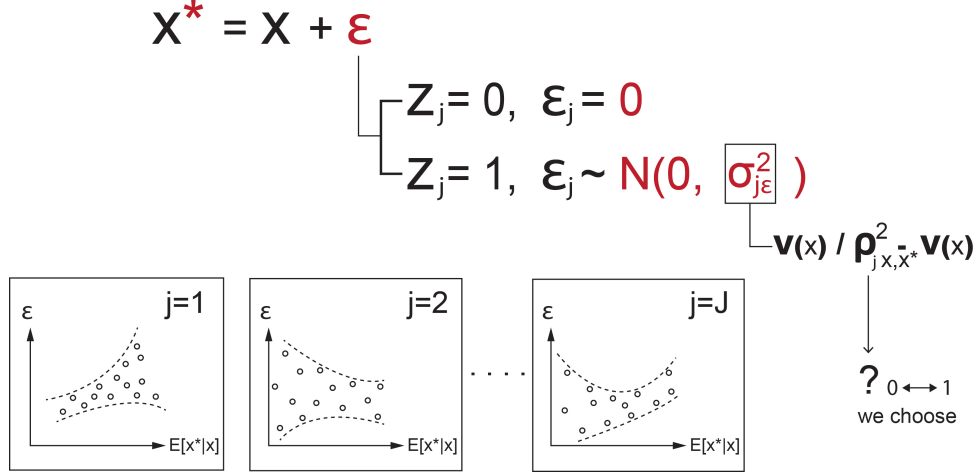


Figure 3: Design of Non-differential Berkson (NDB) Error in \mathbf{x}^* and the induced heteroscedasticity varying by cluster j

is characterized by its independence from the outcome and other covariates, while being correlated with the latent factors. However, as indicated in Equation (34), the Gustafson correction method suggests that the variance of the linking component $V(\mathbf{x}^*|\mathbf{x})$ can be approximated using the variance of the covariate component $V(\mathbf{x}^*|\mathbf{z})$. Hence, to replicate this setting, we connect the NDB covariate \mathbf{x}^{S*} with the binary covariate \mathbf{z}^S as shown in Equation (29b) by conditionally introducing the NDB errors to the true covariate \mathbf{x}^S only when $\mathbf{z}^S = 1$. Additionally, as proposed by Hoffmann et al. (2017), the cluster-specific error variance $\sigma_{j\epsilon}^2$ is modeled using the random correlation $-1 < \rho_{j(x, x^*)} < 1$ between the true covariate \mathbf{x}^S and the NDB covariate \mathbf{x}^{S*} , simulating the NDB error structure. This method utilizes the NDB noise generation technique introduced by Klau et al. (2021), and it is given by:

$$\sigma_{j\epsilon}^2 = \frac{V(\mathbf{x}^S)}{\rho_{j(x^S, x^{S*})}^2} - V(\mathbf{x}^S) \quad (36)$$

Therefore, the selection of cluster-specific random correlations $\rho_{1(x^S, x^{S*})}, \dots, \rho_{J(x^S, x^{S*})}$ can lead to variations in the error severity in the NDB covariate \mathbf{x}^{S*} . We generate three simulation datasets, based on the LGPIF dataset, corresponding to different error rate R_{ϵ_x} scenarios in \mathbf{x}^{S*} , with error rates set at 1%, 10%, and 40%². The definition of error rate that we use is:

$$R_{\epsilon_x} = \frac{\sum_{h=1}^H |x_h^{S*} - x_h^S|}{\sum_{h=1}^H x_h^S} : \begin{cases} 0.01 \text{ (1\% error rate in } \mathbf{x}^{S*} \text{ for dataset A.} \\ 0.1 \text{ (10\% error rate in } \mathbf{x}^{S*} \text{ for dataset B.} \\ 0.4 \text{ (40\% error rate in } \mathbf{x}^{S*} \text{ for dataset C.} \end{cases} \quad (37)$$

² We raise a question of whether an NDB covariate with an excessive error rate (e.g., $> 50\%$) is worth including at all, suggesting that it might be better to exclude such a covariate.

Candidate models: As noted earlier, our goal is to determine the optimal scaling factor ζ , which varies depending on the error rate R_{ϵ_x} - 1%, 10%, 40% - in the NDB covariate \mathbf{x}^{S^*} . The performance of the hierarchical GLM, with the corresponding optimal ζ for each error rate, is then compared to that of the SIMEX method. Figure 4 presents the development process for four distinct risk premium models, facilitating a comprehensive comparison between the Gustafson correction method and SIMEX within the Bayesian hierarchical GLM and traditional GLM frameworks. Each model,

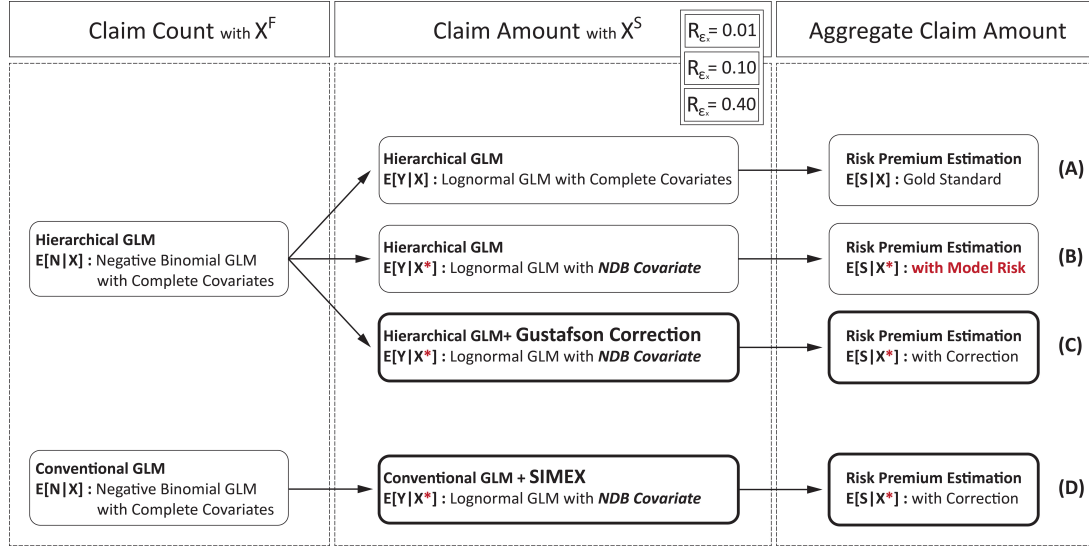


Figure 4: Four candidate models - (A) to (D) - for risk premium development. Specifically, Model(B),(C),(D) need to be thoroughly compared across various error rates R_{ϵ_x} - 1%, 10%, 40% - in the NDB covariate \mathbf{x}^* .

labeled (A) through (D), is designed to systematically assess the impact of different error rates in the NDB covariate \mathbf{x}^{S^*} . Model(A), built using the true covariates within the hierarchical GLM framework, serves as the gold standard and provides a performance benchmark. This benchmark is critical for evaluating the effectiveness of the correction methods. In contrast, Model(B) highlights the detrimental effect (model risk) of using the NDB covariate \mathbf{x}^{S^*} in the same hierarchical GLM framework. It is essential to compare this error-prone Model(B) with Model(C), which incorporates Gustafson's correction, and Model(D), which applies the SIMEX correction with conventional GLM. The purpose of this comparison is to assess how well these correction methods mitigate the negative effects of NDB errors in the covariate \mathbf{x}^{S^*} .

Choice of hyperparameters: To run the Gibbs sampler for the hierarchical models - Model(A),(B),(C) - in Figure 4, flat priors have been selected on the hyperparameters $\phi : \{m_0, \delta, q_0, \Lambda, \rho_{u1}, \rho_{u2}, \rho_{v1}, \rho_{v2}, c_0, d_0, g_0, h_0\}$ as outlined in Equations (21) through

(24) and Equation (26):

$$\begin{aligned}
\{\underline{m}_0 = \beta_{GLM}^F, m_0 = \beta_{GLM}^S, \underline{\delta} = 0.01, \delta = 0.01\} & \quad \text{for } \beta_0^F \text{ and } \beta_0^S \\
\{q_0 = p + 2, q_0 = p + 2, \underline{\Lambda} = \Sigma_{GLM}^F, \Lambda = \Sigma_{GLM}^S\} & \quad \text{for } \Sigma_{\beta_0}^F \text{ and } \Sigma_{\beta_0}^S \\
\{\underline{\rho}_{u1} = 0.125, \rho_{u1} = 0.125, \underline{\rho}_{u2} = 1.5, \rho_{u2} = 1.5\} & \quad \text{for } u_0^F \text{ and } u_0^S \\
\{\underline{\rho}_{v1} = 8, \rho_{v1} = 8, \underline{\rho}_{v2} = 1, \rho_{v2} = 1\} & \quad \text{for } v_0^F \text{ and } v_0^S \\
\{c_0^F = 0.5, c_0^S = 0.5, d_0^F = 0.5, d_0^S = 0.5\} & \quad \text{for } \lambda^{2F} \text{ and } \lambda^{2S} \\
\{g_0^F = 0.5, g_0^S = 0.5, h_0^F = 0.5, h_0^S = 0.5\} & \quad \text{for } \pi^F \text{ and } \pi^S
\end{aligned} \tag{38}$$

where $\beta_{GLM}^F, \beta_{GLM}^S$ denote the naïve GLM coefficient vectors, serving as initial values for the Gibbs sampler. The variable p represents the number of covariates in the model, while $\Sigma_{GLM}^F, \Sigma_{GLM}^S$ are the variance-covariance matrices obtained from the naïve GLM output. The values specified in Equation (38) have been meticulously chosen based on preliminary findings from a trial run of the Gibbs sampler with randomly assigned starting values. While using inappropriate initial values may lead to less efficient posterior samples, it can still yield some samples that shed light on the true posterior distribution behavior. By employing the Method of Moments technique (Pearson, 1936), we can gain insights into the nature of these posterior samples, guiding our selection of starting values to improve the Gibbs sampler's efficiency.

4.3 Model Validation

To assess the effectiveness of the proposed models in the following section, the methods outlined below will be utilized:

- **LPPD** : To assess predictive performance in the Bayesian context, we can employ the Log Pointwise Predictive Density (LPPD) (McElreath, 2018), which captures model uncertainty through the complete posterior distribution. LPPD is calculated by averaging the log-predictive likelihood for each data point across posterior samples $\theta_1, \dots, \theta_M$, as follows:

$$\text{LPPD}(\bar{Y}_1, \dots, \bar{Y}_H, \mathbf{X}_1, \dots, \mathbf{X}_H, \theta) = \sum_{h=1}^H \ln \left(\frac{1}{M} \sum_{m=1}^M \mathbf{L}(\theta_m; \bar{Y}_h, \mathbf{X}_h) \right) \tag{39}$$

The likelihood function takes values from 0 to 1 (since it is a probability function), resulting in LPPD values from $-\infty$ to 0. Multiplying LPPD by -2 yields values similar to Mean Squared Error (MSE), where a perfect fit equals zero and a poor fit results in a large positive value (Cousineau and Allan, 2015). Note that LPPD emphasizes predictive performance. This is because LPPD not only assesses how well the model fits the training data but also evaluates its generalization to new, unseen data by incorporating parameter uncertainty from the full posterior distribution (McElreath, 2018).

- **D_{KL}** : The Kullback-Leibler Divergence D_{KL} quantifies how one probability distribution differs from another by comparing their entropies $H[.]$ (Anderson and Burnham, 2004). Among competing models, the one with the minimized D_{KL} is preferred. For our predictive model $\mathbf{L}(\theta; \bar{Y}_h, \mathbf{X})$ and target model $\mathbf{P}(\theta^{true}; \bar{Y}_h, \mathbf{X})$,

$D_{KL}(\mathbf{P}, \mathbf{L}) = H[\mathbf{P}, \mathbf{L}] - H[\mathbf{P}]$ for our model can be computed as:

$$D_{KL} = - \sum_{h=1}^H \ln(\mathbf{L}(\boldsymbol{\theta}; \tilde{Y}_h, \mathbf{X}_h)) \cdot \mathbf{P}(\boldsymbol{\theta}^{true}; \tilde{Y}_h, \mathbf{X}_h) + \sum_{h=1}^H \ln(\mathbf{P}(\boldsymbol{\theta}^{true}; \tilde{Y}_h, \mathbf{X}_h)) \cdot \mathbf{P}(\boldsymbol{\theta}^{true}; \tilde{Y}_h, \mathbf{X}_h) \quad (40)$$

where $\sum_{h=1}^H \ln(\mathbf{L}(\boldsymbol{\theta}; \tilde{Y}_h, \mathbf{X}_h))$ is the LPPD discussed previously. We aim to identify which candidate model achieves a larger decrease in D_{KL} and the extent of that reduction.

- **SSPE + SAPE** : Prediction performance can be evaluated by measuring the discrepancy between predicted and observed values using the Sum of Square Error (SSE) criterion. Specifically, the Sum of Square Prediction Error (SSPE) and the Sum of Absolute Prediction Error (SAPE) can be used to assess different facets of accuracy. SSPE focuses on squared differences between predicted values $g(\mathbf{X}_h)$ and actual values \tilde{Y}_h , while SAPE calculates the absolute differences across all observations $h = 1, \dots, H$:

$$\text{SSPE: } \sum_{h=1}^H (g(\mathbf{X}_h) - \tilde{Y}_h)^2 \quad (41a)$$

$$\text{SAPE: } \sum_{h=1}^H |g(\mathbf{X}_h) - \tilde{Y}_h| \quad (41b)$$

SSPE and SAPE assess prediction performance differently. SSPE penalizes large deviations more heavily, while SAPE treats all deviations equally by focusing on absolute differences. Given the skewed nature of our outcome \tilde{Y}_h , where outliers may be common, SAPE is preferred. This approach considers each data point equally important, making it unnecessary to disproportionately penalize larger errors, particularly when we are concerned with potential outliers (Parodi, 2023).

- **CTE** : The final aspect of this validation process is assessing risk within the predictive distributions, specifically through the Conditional Tail Expectation (CTE), defined as follows:

$$\text{CTE}(q) = E[\tilde{Y}_h | \tilde{Y}_h > Q_q(\tilde{Y}_h)], \quad q \in (0, 1) \quad (42)$$

where $Q_q(\tilde{Y}_h)$ is the q th quantile of the predictive distribution. The CTE analyzes the tail behavior of predictive distributions to provide insights into expected aggregate losses (risk premium) under extreme conditions (Brazauskas et al., 2008). A lower CTE value indicates the model predicts less severe losses in extreme scenarios.

We have outlined several criteria for model validation, including: 1) LPPD, 2) D_{KL} , 3) SSPE, 4) SAPE, and 5) CTE. In the upcoming section, these metrics will be used to evaluate the performance and accuracy of our proposed models in predicting risk premiums.

4.4 Results with LGPIF ($H = 1,679$)

We constructed a training set with 1,276 records for modeling and a test set with 403 records for validation. As shown in Figure 4, Model(A) serves as the gold standard, Model(B) represents a flawed model that reflects real-world risks, Model(C) applies the Gustafson correction, and Model(D) utilizes the SIMEX correction within a traditional GLM framework. We systematically evaluated the performance of Models(B), (C), and (D) across three scenarios with varying error rates $R_{\epsilon_x} = 0.01, R_{\epsilon_x} = 0.1,$ and $R_{\epsilon_x} = 0.4$ in the NDB covariate \mathbf{x}^{S*} . Additionally, we compared the effects of the scaling factor $0 < \zeta < 1$ in Models(C) and (D) within each error rate scenario. For each hierarchical GLM model, Model(A), (B), (C) shown in Figure 4, we ran two independent Markov chains with $M = 60,000$ iterations of Gibbs sampling, following Algorithm S2 in Part 2 of the on-line supplementary file. Since Model(A) is the gold standard, the focus was on comparing Models(B), (C), and (D) across datasets with error rates $R_{\epsilon_x} = 0.01, R_{\epsilon_x} = 0.1,$ and $R_{\epsilon_x} = 0.4$. Model(C) employed the Gustafson correction with scaling factors $\zeta = \{0.1, \dots, 0.9\}$. Within each Gibbs sampling iteration in Model(A), (B), and (C), a Metropolis-Hastings (MH) technique was embedded to update the outcome parameters - $\beta_j^F, \psi_j, \beta_j^S, \sigma_j^2$ - as conjugate priors were not available. The initial 10,000 iterations were treated as burn-in and discarded. Convergence was confirmed using the Brooks-Gelman statistic (Brooks and Gelman, 1998), ensuring adequate mixing of the chains. For Model(D), a conventional GLM, parameters were estimated using the Maximum Likelihood Estimation (MLE) method.

The aggregate claim amount model $f(S_h | \mathbf{X}^F, \mathbf{X}^S, \zeta_h, \psi, \mu_h, \sigma^2)$, representing the risk premium function, is derived by integrating these two gold standard models through Monte Carlo simulation: the i) estimated claim count model $f(N_h | \mathbf{X}^F, \zeta_h, \psi)$ and the ii) estimated claim amount model $f(\tilde{Y}_h | \mathbf{X}^S, \mu_h, \sigma^2)$ (Detailed results for each gold standard model are provided in Part 4 of the on-line supplementary file). The integration process begins by simulating a claim count sample N_h for a policy h . This sample is drawn from the claim count model $f(N_h | \mathbf{X}^F, \zeta_h, \psi)$. Subsequently, for each drawn claim count N_h , a corresponding set of claim amount samples is generated. These samples are obtained from the claim amount model $f(\tilde{Y}_h | \mathbf{X}^S, \mu_h, \sigma^2)$. This iterative process is repeated numerous times, allowing for the construction of a comprehensive distribution of aggregate claims S_h for a policy h (Scollnik, 2001). The resulting predictive distribution

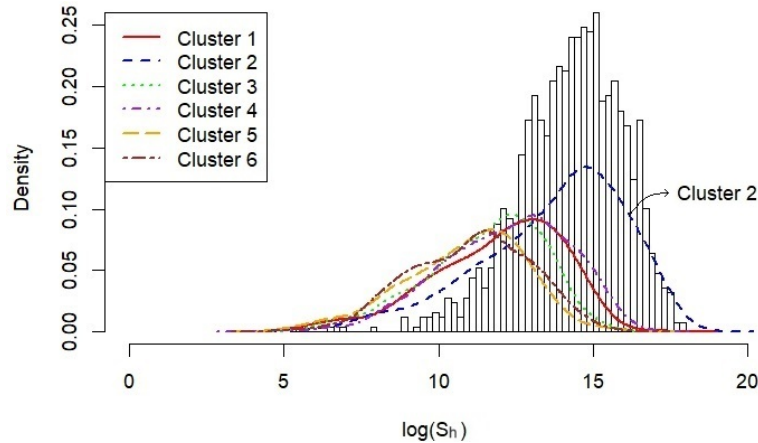


Figure 5: Model(A) Result V: A histogram of the overall expected aggregate claim amount on a log scale, overlaid with the individual cluster-wise distributions $\ln S_h | \mathbf{X}^F, \mathbf{X}^S$.

of the aggregate claim amount modeling is illustrated in Figure 5. This figure compares the cluster-wise distributions of the aggregate claim amount S_h on a log scale with the overall distribution of S_h . This comparison facilitates the evaluation of each cluster's risk profile and its impact on the total aggregate claim. Significantly, the distribution for 'Cluster 2' (shown as a dotted blue curve) closely resembles the shape of the overall distribution, indicating a substantial contribution to the overall risk premium.

In what follows, we will examine the results of Models (A), (B), (C), and (D), focusing exclusively on the claim amount component $\tilde{Y}_h | \mathbf{X}^{S*} \sim \mathbf{LogN}(\mu_h = \ln(E[\tilde{Y}_h | \mathbf{X}^{S*}]) - \frac{1}{2}\sigma_j^2, \sigma_j^2)$ (log-normal GLM). As outlined in Section 4.2, we account for scenarios where the covariates associated with the negative binomial outcome N_h are complete, while the continuous covariate \mathbf{X}^S for the claim amount outcome is subject to Non-Differential Berkson (NDB) measurement errors. Our focus will center on the log-normal outcome \tilde{Y}_h and its associated NDB covariate \mathbf{X}^{S*} within the hierarchical modeling framework.

Comparisons [Model(B) vs Model(C) vs Model(D)]: As illustrated in Figure 4, this experiment evaluates which candidate models, Models (C) and (D), yield results closest to the gold standard, Model (A). Specifically, we compare Model (C), utilizing the Gustafson correction, with Model (D), which employs the SIMEX correction. Models (A), (B), and (C) are developed within the Bayesian hierarchical GLM framework, while Model (D) is constructed in the conventional GLM framework, providing a frequentist perspective. Additionally, a sensitivity analysis for the claim amount model is conducted to determine the optimal scaling factor value ζ , leveraging the prior knowledge: $\mathfrak{T}_j^2 = (1 - \zeta)\hat{\lambda}_j^2$. Table 2 presents a com-

Table 2: Comparison of the scale parameter σ_j^2 estimates from the hierarchical log-normal GLMs in Model(A),(B), and (C) across risk clusters $j = 1, \dots, 6$

Model	Scale	Parameter estimates σ_j^2								
Model(A) Gold standard	$\sigma_{j=1}^2$	3.35 with 95% Credible Interval: $\{1.04 \leq \sigma_{j=1}^2 \leq 9.85\}$								
	$\sigma_{j=2}^2$	3.44 with 95% Credible Interval: $\{1.05 \leq \sigma_{j=2}^2 \leq 10.01\}$								
	$\sigma_{j=3}^2$	3.22 with 95% Credible Interval: $\{1.02 \leq \sigma_{j=3}^2 \leq 9.48\}$								
	$\sigma_{j=4}^2$	3.40 with 95% Credible Interval: $\{1.04 \leq \sigma_{j=4}^2 \leq 9.78\}$								
	$\sigma_{j=5}^2$	3.40 with 95% Credible Interval: $\{1.23 \leq \sigma_{j=5}^2 \leq 10.13\}$								
	$\sigma_{j=6}^2$	3.23 with 95% Credible Interval: $\{1.01 \leq \sigma_{j=6}^2 \leq 9.65\}$								
		Error rate R_{ϵ_x} in \mathbf{x}^* : 0.01			Error rate R_{ϵ_x} in \mathbf{x}^* : 0.10			Error rate R_{ϵ_x} in \mathbf{x}^* : 0.40		
Model(B) Before correction	$\sigma_{j=1}^2$	3.36			3.38			3.59		
	$\sigma_{j=2}^2$	3.44			3.51			3.65		
	$\sigma_{j=3}^2$	3.26			3.38			3.46		
	$\sigma_{j=4}^2$	3.43			3.58			3.67		
	$\sigma_{j=5}^2$	3.44			3.57			3.65		
	$\sigma_{j=6}^2$	3.28			3.45			3.51		
		$\zeta : 0.5$	$\zeta : 0.6$	$\zeta : 0.7$	$\zeta : 0.5$	$\zeta : 0.6$	$\zeta : 0.7$	$\zeta : 0.5$	$\zeta : 0.6$	$\zeta : 0.7$
Model(C) After correction	$\sigma_{j=1}^2$	3.28	3.30	3.33	3.22	3.32	3.56	3.36	3.34	3.31
	$\sigma_{j=2}^2$	3.37	3.38	3.41	3.29	3.46	3.48	3.46	3.45	3.33
	$\sigma_{j=3}^2$	3.19	3.22	3.34	3.26	3.26	3.57	3.24	3.23	3.21
	$\sigma_{j=4}^2$	3.27	3.31	3.48	3.31	3.35	3.61	3.41	3.38	3.36
	$\sigma_{j=5}^2$	3.10	3.29	3.49	3.25	3.35	3.57	3.42	3.38	3.35
	$\sigma_{j=6}^2$	3.11	3.29	3.33	3.39	3.32	3.57	3.26	3.25	3.19
LPPD ($\times 10^3$)		-16.79	-16.77	-16.78	-16.41	-16.39	-16.41	-16.19	-16.20	-16.21

parison of the marginal posterior means of the scale parameter σ_j^2 obtained from Model(A), (B), and (C). Each hierarchical GLM describes the log-normal outcome: $\bar{Y}_h|\mathbf{X}^S \sim \mathbf{LogN}(\mu_h = \ln(E[\bar{Y}_h|\mathbf{X}^S]) - \frac{1}{2}\sigma_j^2, \sigma_j^2)$. Additional comparisons of the marginal posterior means of the GLM coefficients β_j are provided in Tables 3 and 4, which also include results from the sensitivity analysis on the optimal scaling factor ζ .

We first examine Model(B), which incorporates the NDB covariate \mathbf{x}^{S*} , introducing model risk. Our investigation focuses on how the parameter estimates - the scale parameter σ^2 and the GLM coefficients β - vary with increasing error rates R_{ϵ_x} , reflecting different degrees of model risk. As seen in Tables 2, 3, and 4, estimates from Model(B) consistently show an upward trend in response to higher error rates ($R_{\epsilon_x} : 0.01 \rightarrow R_{\epsilon_x} : 0.40$). Both the scale parameter σ^2 and the GLM coefficients β_0, β_2 (intercept and binary covariate) inflate, while β_1 (linked to \mathbf{x}^{S*}) decreases across all clusters $j = 1, \dots, 6$. This pattern aligns with the expectation that as the error rate R_{ϵ_x} increases, the added noise results in greater variability in the outcomes, causing a proportional rise in the scale parameter σ^2 . The consistent directional shift in the coefficient β_1 , tied to the NDB covariate \mathbf{x}^{S*} , indicates that \mathbf{x}^{S*} follows a systematic trend characteristic of NDB error. Turning to Model(C), which applies the Gustafson

Table 3: Comparison of the GLM intercept β_0 estimates from the hierarchical log-normal GLMs (claim amount component) in Model(A),(B), and (C) across risk clusters $j = 1, \dots, 6$

Model	Intercept	Parameter estimates β_{j0}								
Model(A) Gold standard	$\beta_{0j=1}$	7.18 with 95% Credible Interval: $\{4.98 \leq \beta_{0j=1} \leq 9.32\}$								
	$\beta_{0j=2}$	7.69 with 95% Credible Interval: $\{5.50 \leq \beta_{0j=2} \leq 9.71\}$								
	$\beta_{0j=3}$	6.37 with 95% Credible Interval: $\{4.11 \leq \beta_{0j=3} \leq 8.86\}$								
	$\beta_{0j=4}$	7.36 with 95% Credible Interval: $\{3.83 \leq \beta_{0j=4} \leq 10.52\}$								
	$\beta_{0j=5}$	7.14 with 95% Credible Interval: $\{3.83 \leq \beta_{0j=5} \leq 10.32\}$								
	$\beta_{0j=6}$	7.36 with 95% Credible Interval: $\{4.87 \leq \beta_{0j=6} \leq 10.46\}$								
		Error rate R_{ϵ_x} in \mathbf{x}^* : 0.01			Error rate R_{ϵ_x} in \mathbf{x}^* : 0.10			Error rate R_{ϵ_x} in \mathbf{x}^* : 0.40		
Model(B) Before correction	$\beta_{0j=1}$	7.28			8.89			9.40		
	$\beta_{0j=2}$	7.72			9.14			9.51		
	$\beta_{0j=3}$	6.44			8.82			9.28		
	$\beta_{0j=4}$	7.33			9.28			9.52		
	$\beta_{0j=5}$	7.12			9.37			9.57		
	$\beta_{0j=6}$	7.46			9.11			9.72		
		$\zeta : 0.5$	$\zeta : 0.6$	$\zeta : 0.7$	$\zeta : 0.5$	$\zeta : 0.6$	$\zeta : 0.7$	$\zeta : 0.5$	$\zeta : 0.6$	$\zeta : 0.7$
Model(C) After correction	$\beta_{0j=1}$	6.08	7.24	7.29	6.86	7.28	8.03	7.21	7.09	6.81
	$\beta_{0j=2}$	7.35	8.07	8.21	7.37	8.12	8.43	7.56	7.39	7.24
	$\beta_{0j=3}$	5.20	6.07	6.21	5.93	6.19	7.25	6.35	6.11	5.87
	$\beta_{0j=4}$	6.40	7.33	7.87	6.91	7.53	8.16	7.39	7.14	6.73
	$\beta_{0j=5}$	6.52	6.95	7.20	6.23	6.94	7.28	7.18	7.15	7.11
	$\beta_{0j=6}$	6.20	7.40	7.48	6.87	7.27	7.67	7.39	7.15	6.42
LPPD ($\times 10^3$)		-16.79	-16.77	-16.78	-16.41	-16.39	-16.41	-16.19	-16.20	-16.21

correction technique, we observe that as the error rate R_{ϵ_x} increases, the parameter estimates from Model(C) progressively align with those of the gold standard, Model(A), with the values falling within the credible intervals established by Model(A). This alignment is a promising indicator of our correction method's effectiveness. Note that a key aspect of our error correction technique is the selection of an optimal scaling factor ζ . Table 2, 3 and 4 highlight that Model(C) performs best within the scaling factor range $0.5 \leq \zeta \leq 0.7$. This suggests that the relationship between the conditional error variance terms - $V(\mathbf{x}^*|\mathbf{x})$ and $V(\mathbf{x}^*|\mathbf{z})$ - is largely captured in this window of ζ , as reflected by the prior knowledge: $\mathfrak{T}_j^2 = (1 - \zeta)\hat{\lambda}_j^2$ in Equation (34).

Table 4: Comparison of the GLM slope β_{1j}, β_{2j} estimates from the hierarchical log-normal GLMs (claim amount component) in Model(A),(B), and (C) across risk clusters $j = 1, \dots, 6$.

Model	Slope	Parameter estimates $\beta_{j1}; \beta_{j2}$								
Model(A) Gold standard	$\beta_{1j=1}$	0.31 with 95% Credible Interval: $\{0.03 \leq \beta_{1j=1} \leq 0.62\}$								
	$\beta_{1j=2}$	0.24 with 95% Credible Interval: $\{-0.04 \leq \beta_{1j=2} \leq 0.59\}$								
	$\beta_{1j=3}$	0.43 with 95% Credible Interval: $\{0.09 \leq \beta_{1j=3} \leq 0.67\}$								
	$\beta_{1j=4}$	0.33 with 95% Credible Interval: $\{-0.10 \leq \beta_{1j=4} \leq 0.84\}$								
	$\beta_{1j=5}$	0.34 with 95% Credible Interval: $\{-0.12 \leq \beta_{1j=5} \leq 0.89\}$								
	$\beta_{1j=6}$	0.29 with 95% Credible Interval: $\{-0.08 \leq \beta_{1j=6} \leq 0.67\}$								
		Error rate R_{ϵ_x} in \mathbf{x}^* : 0.01			Error rate R_{ϵ_x} in \mathbf{x}^* : 0.10			Error rate R_{ϵ_x} in \mathbf{x}^* : 0.40		
Model(B) Before correction	$\beta_{1j=1}$	0.29			0.06			0.01		
	$\beta_{1j=2}$	0.24			0.04			0.01		
	$\beta_{1j=3}$	0.42			0.07			0.01		
	$\beta_{1j=4}$	0.32			0.07			0.01		
	$\beta_{1j=5}$	0.34			0.06			0.01		
	$\beta_{1j=6}$	0.28			0.06			0.01		
		$\zeta : 0.5$	$\zeta : 0.6$	$\zeta : 0.7$	$\zeta : 0.5$	$\zeta : 0.6$	$\zeta : 0.7$	$\zeta : 0.5$	$\zeta : 0.6$	$\zeta : 0.7$
Model(C) After correction	$\beta_{1j=1}$	0.56	0.38	0.21	0.38	0.36	0.27	0.34	0.38	0.42
	$\beta_{1j=2}$	0.38	0.26	0.11	0.34	0.27	0.20	0.21	0.33	0.38
	$\beta_{1j=3}$	0.83	0.56	0.32	0.52	0.48	0.39	0.43	0.47	0.51
	$\beta_{1j=4}$	0.47	0.32	0.28	0.41	0.35	0.26	0.37	0.42	0.48
	$\beta_{1j=5}$	0.81	0.41	0.29	0.34	0.31	0.28	0.36	0.39	0.45
	$\beta_{1j=6}$	0.55	0.40	0.31	0.31	0.26	0.19	0.31	0.34	0.38
Model(A) Gold standard	$\beta_{2j=1}$	0.24 with 95% Credible Interval: $\{-0.55 \leq \beta_{2j=1} \leq 1.01\}$								
	$\beta_{2j=2}$	0.21 with 95% Credible Interval: $\{-0.47 \leq \beta_{2j=2} \leq 0.82\}$								
	$\beta_{2j=3}$	0.12 with 95% Credible Interval: $\{-0.67 \leq \beta_{2j=3} \leq 0.94\}$								
	$\beta_{2j=4}$	0.08 with 95% Credible Interval: $\{-1.06 \leq \beta_{2j=4} \leq 0.59\}$								
	$\beta_{2j=5}$	-0.12 with 95% Credible Interval: $\{-1.05 \leq \beta_{2j=5} \leq 0.60\}$								
	$\beta_{2j=6}$	0.17 with 95% Credible Interval: $\{-0.63 \leq \beta_{2j=6} \leq 0.67\}$								
		Error rate R_{ϵ_x} in \mathbf{x}^* : 0.01			Error rate R_{ϵ_x} in \mathbf{x}^* : 0.10			Error rate R_{ϵ_x} in \mathbf{x}^* : 0.40		
Model(B) Before correction	$\beta_{2j=1}$	0.27			0.31			0.38		
	$\beta_{2j=2}$	0.24			0.34			0.40		
	$\beta_{2j=3}$	0.13			0.32			0.40		
	$\beta_{2j=4}$	0.09			0.17			0.22		
	$\beta_{2j=5}$	0.10			0.12			0.17		
	$\beta_{2j=6}$	0.12			0.23			0.36		
		$\zeta : 0.5$	$\zeta : 0.6$	$\zeta : 0.7$	$\zeta : 0.5$	$\zeta : 0.6$	$\zeta : 0.7$	$\zeta : 0.5$	$\zeta : 0.6$	$\zeta : 0.7$
Model(C) After correction	$\beta_{2j=1}$	0.17	0.26	0.29	0.19	0.28	0.31	0.25	0.23	0.19
	$\beta_{2j=2}$	0.23	0.27	0.31	0.18	0.23	0.28	0.24	0.21	0.18
	$\beta_{2j=3}$	0.07	0.11	0.19	0.07	0.12	0.17	0.19	0.17	0.15
	$\beta_{2j=4}$	0.01	0.03	0.11	0.11	0.13	0.18	0.10	0.07	0.02
	$\beta_{2j=5}$	-0.11	-0.14	0.16	-0.21	-0.15	-0.07	-0.16	-0.19	-0.21
	$\beta_{2j=6}$	0.12	0.19	0.21	0.08	0.17	0.21	0.16	0.12	0.09
LPPD ($\times 10^3$)		-16.79	-16.77	-16.78	-16.41	-16.39	-16.41	-16.19	-16.20	-16.21

The challenge, however, lies in identifying the optimal range of ζ values for a specific error rate R_{ϵ_x} in the covariate \mathbf{x}^* . Our findings from this LGPIF experiment show that when ζ deviates from the $0.5 \leq \zeta \leq 0.7$ range, Model(C)'s corrections become ineffective, often performing worse than the erroneous Model(B). This worsens as the error rate R_{ϵ_x} increases, which is expected, because a higher error rate implies greater bias, making it much harder for the model to correct. Interestingly, our experimental results suggest that the optimal ζ range can be effectively identified by evaluating the LPPD for each modeling result, as the LPPD reflects the degree of predictive performance. This reflects how closely the estimated parameter values align with the

gold standard. Therefore, by searching for the modeling results with the maximum LPPD, we can systematically pinpoint the optimal ζ , leading to enhanced correction performance without requiring access to gold standard data.

Our LGPIF experimental results are further analyzed in Figure 6 and Table 5 for the error rate $R_{\epsilon_x} = 0.01$, Figure 7 and Table 6 for the error rate $R_{\epsilon_x} = 0.10$, and Figure 8 and Table 7 for the error rate $R_{\epsilon_x} = 0.40$. For comparisons, Tables 5, 6, 7 present (i) LPPD, (ii) SSPE, (iii) SAPE, and (iv) Kullback-Leibler Divergence (D_{KL}) for the individual claim amount model $f(\bar{Y}_h|\mathbf{X}^S)$ as well as CTEs for the aggregate claim amount model $f(S_h|\mathbf{X}^F, \mathbf{X}^S)$ within each error rate scenario. The results from Model(D), traditional GLM with SIMEX correction, are also included. Figures 6,7,8 display histograms of the testing set across all clusters $j = 1, \dots, 6$, comparing the out-of-sample prediction curves from Model(C) (blue) against the gold standard Model(A) (red) and the erroneous Model(B) (dotted). The prediction curve for Model(D) is omitted due to its subpar performance.

As expected, Tables 5, 6, and 7 show that the gold standard model, Model(A), achieves the highest LPPD value of $-16,155.90$, while the naïve model with model risk, Model(B), consistently exhibits the lowest LPPD values across datasets with varying error rates. Note that the LPPD for the GLM-based SIMEX method is unavailable, as LPPD calculations require posterior densities. Applying Gustafson corrections in Model(C) produces LPPD values of $-16,770.85$ with $\zeta = 0.6$ for an error rate 0.01, $-16,370.44$ with $\zeta = 0.6$ for an error rate 0.10, and $-16,188.31$ with $\zeta = 0.5$ for an error rate 0.40, closely aligning with the LPPD of the error-free gold standard model, Model(A). This result aligns with other metrics, such as SSPE and SAPE, across different error rate scenarios. The erroneous model with model risk, Model(B), consistently exhibits the highest SSPE and SAPE values, reflecting its weaker predictive performance. In contrast, the Gustafson correction model, Model(C), achieves significantly lower SSPE and SAPE values, closely approximating those of the gold standard model, Model(A). Additionally, as the error rate increases, Model(C) increasingly outperforms the SIMEX method, Model(D), underscoring its robustness.

The effectiveness of the Gustafson correction technique is further substantiated by D_{KL} , which quantifies the distance between the estimated distribution produced by Model(B) and Model(C) and the target (gold standard) distribution, Model(A). A smaller D_{KL} value indicates a more accurate error correction. Notably, as shown in Tables 5, 6, and 7, within the range of $0.5 \leq \zeta \leq 0.7$, Model(C) exhibits a reduction in D_{KL} from 0.61 to 0.33 as the error rate increases from $R_{\epsilon_x} = 0.01$ to $R_{\epsilon_x} = 0.40$. This trend suggests that the Gustafson correction effectively mitigates the impact of the NDB covariate, thereby enhancing the model's fidelity to the true data distribution as error rates rise. It is important to note that Model(D), a GLM with SIMEX, is excluded from this analysis due to its frequentist nature, which does not yield the LPPD values necessary for this divergence assessment.

In the Conditional Tail Expectation (CTE) analysis, the predictive distribution generated by the Gustafson correction in Model(C) exhibits thicker tails compared to Model(D), with higher CTE values of $CTE_{95\%} = 269,656$ at $R_{\epsilon_x} = 0.01$, $CTE_{95\%} = 281,371$ at $R_{\epsilon_x} = 0.10$, $CTE_{95\%} = 278,099$ at $R_{\epsilon_x} = 0.40$. Notably, the CTE values of Model(C) do not demonstrate a consistent trend across the varying error rate scenarios. However, the higher CTE values in Model(C) suggest that while the Gustafson correction enhances the model's overall predictive accuracy, the underlying hierarchical GLM may effectively address extreme values or outliers. This indicates a potential improvement in the model's applicability in contexts where accurately

capturing outliers is essential.

Upon inspection of Figures 6, 7, and 8, it is clear that the Gustafson correction (Model(C), blue curve) effectively mitigates the model risk from the NDB covariate (Model(B), dotted curve) because the corrected hierarchical GLM curves (Model(C), blue) align closely with that of the gold standard curve (Model(A), red) across all clusters $j = 1, \dots, 6$, particularly in scenarios with higher error rates: $R_{\epsilon_x} = 0.40$ in Figure 8. However, a slight gap still persists between Models(A) and (C), indicating that the correction, while beneficial, has limitations.

In contrast, Model(B) demonstrates significant distortions, including extreme variations and multiple peaks that worsen with increasing error rates. Although both Models(B) and (C) are derived from the same LGPIF dataset, the Gustafson correction in Model(C) appears to effectively restore the properties of the original distribution, particularly pronounced at higher error rates such as $R_{\epsilon_x} = 0.40$. This improvement is observed with scaling factors in the range $0.5 \leq \zeta \leq 0.7$, suggesting that the relationship between $V(\mathbf{x}^*|\mathbf{x})$ and $V(\mathbf{x}^*|\mathbf{z})$ in the LGPIF dataset is characterized by this range of ζ . Consequently, further investigation may be warranted using different datasets where $V(\mathbf{x}^*|\mathbf{x})$ and $V(\mathbf{x}^*|\mathbf{z})$ have relationships defined by different range of scaling factors.

$\zeta = 0.6$ $R_{\epsilon_x} = 0.01$	Feature	Model(A): Gold standard	Model(B): with Model Risk	Model(C) Gustafson correction	Model(D): SIMEX correction
$f(\ln \bar{Y}_h \mathbf{X}^S)$	LPPD	-16,155.90	-17,437.59	-16,770.85	-
	SSPE	784.52	817.16	798.89	816.80
	SAPE	415.21	422.53	419.83	421.60
	D_{KL}	0.00	1.28	0.61	-
$f(\ln S_h \mathbf{X}^F, \mathbf{X}^S)$	CTE 10%	48,782.40	42,995.64	49,599.75	45,237.36
	CTE 50%	81,593.58	89,103.11	82,218.10	78,736.26
	CTE 90%	209,761.38	226,599.19	196,273.16	121,342.20
	CTE 95%	274,996.31	260,859.33	269,656.52	170,487.20

Table 5: Comparison of predictive performances among three Bayesian hierarchical GLMs—Model (A), (B), and (C)—and the GLM-based SIMEX, based on the LGPIF data with a covariate error rate of $R_{\epsilon_x} = 0.01$ and a scaling factor of $\zeta = 0.6$.

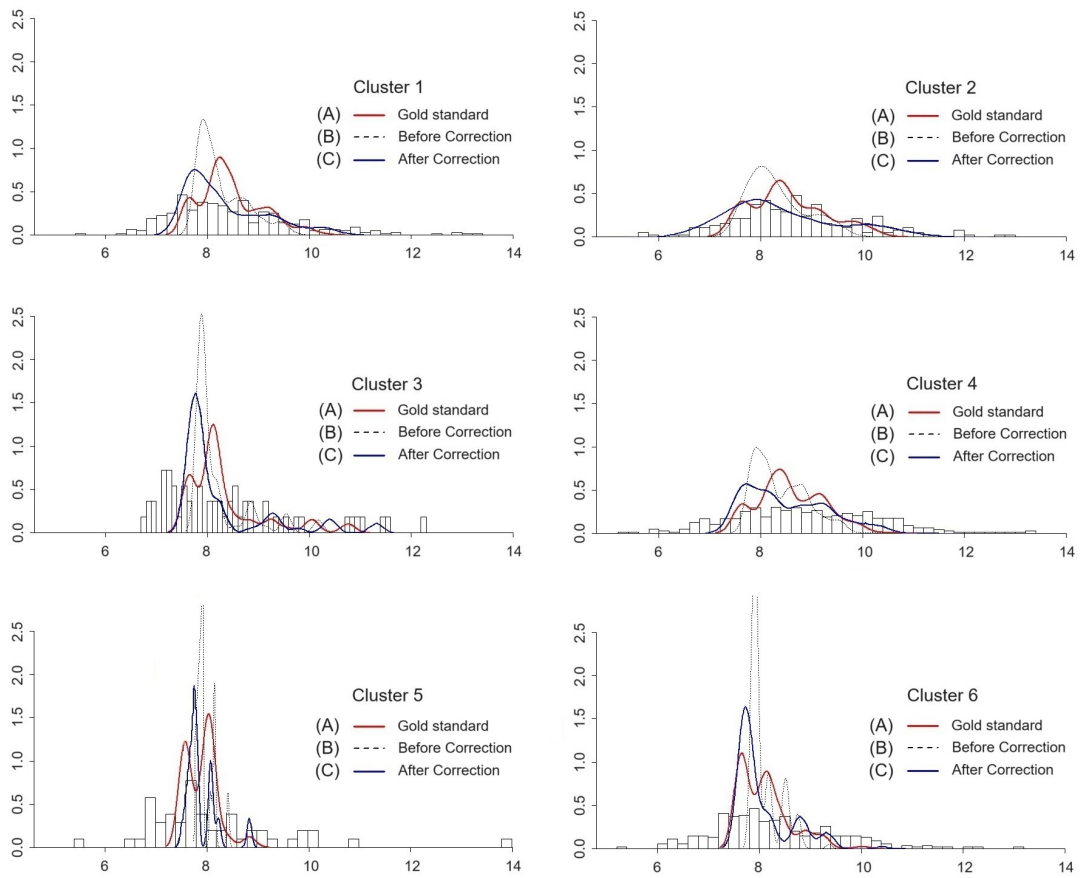


Figure 6: Curve alignment under the condition of ‘error rate $R_{\epsilon_x} = 0.01$ ’ with $\zeta = 0.6$: Cluster-wise histograms (for $j = 1, \dots, 6$) of the observed claim amount Y_h on a log scale and the out-of-sample predictive densities obtained from Model(A), (B), and (C)

$\zeta = 0.6$ $R_{\epsilon_x} = 0.10$	Feature	Model(A): Gold standard	Model(B): with Model Risk	Model(C) Gustafson correction	Model(D): SIMEX correction
$f(\ln \bar{Y}_h \mathbf{X}^S)$	LPPD	-16,155.90	-17,731.03	-16,390.44	-
	SSPE	784.52	840.02	795.07	892.72
	SAPE	415.21	430.22	418.68	439.92
	D_{KL}	0.00	1.57	0.41	-
$f(\ln S_h \mathbf{X}^F, \mathbf{X}^S)$	CTE 10%	48,782.40	50,222.22	49,764.81	45,846.51
	CTE 50%	81,593.58	84,118.93	83,595.88	73,558.45
	CTE 90%	209,761.38	233,257.58	221,359.16	129,851.81
	CTE 95%	274,996.31	295,325.47	281,371.20	164,324.93

Table 6: Comparison of predictive performances among three Bayesian hierarchical GLMs—Model (A), (B), and (C)—and the GLM-based SIMEX, based on the LGPIF data with a covariate error rate of $R_{\epsilon_x} = 0.10$ and a scaling factor of $\zeta = 0.6$.

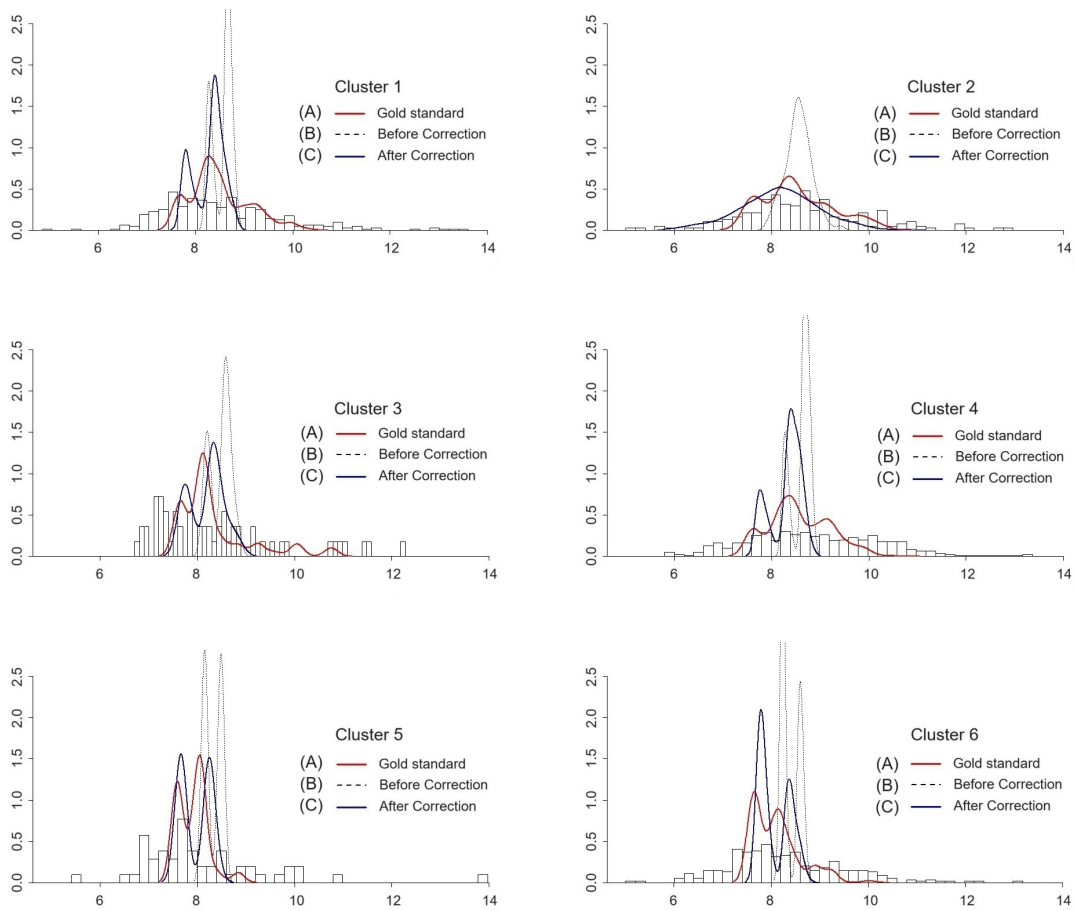


Figure 7: Curve alignment under the condition of ‘error rate $R_{\epsilon_x} = 0.10$ ’ with $\zeta = 0.6$: Cluster-wise histograms (for $j = 1, \dots, 6$) of the observed claim amount Y_h on a log scale and the out-of-sample predictive densities obtained from Model(A), (B), and (C)

$\zeta = 0.5$ $R_{\epsilon_x} = 0.40$	Feature	Model(A): Gold standard	Model(B): with Model Risk	Model(C) Gustafson correction	Model(D): SIMEX correction
$f(\ln \bar{Y}_h \mathbf{X}^S)$	LPPD	-16,155.90	-18,058.43	-16,188.31	-
	SSPE	784.52	861.44	793.50	954.04
	SAPE	415.21	437.73	417.36	532.72
	D_{KL}	0.00	1.94	0.33	-
$f(\ln S_h \mathbf{X}^F, \mathbf{X}^S)$	CTE 10%	48,782.40	54,671.42	49,155.18	54,716.98
	CTE 50%	81,593.58	89,824.04	84,038.84	75,489.60
	CTE 90%	209,761.38	233,321.29	223,497.02	174,000.54
	CTE 95%	274,996.31	295,939.47	278,098.80	186,720.87

Table 7: Comparison of predictive performances among three Bayesian hierarchical GLMs—Model (A), (B), and (C)—and the GLM-based SIMEX, based on the LGPIF data with a covariate error rate of $R_{\epsilon_x} = 0.40$ and a scaling factor of $\zeta = 0.5$.

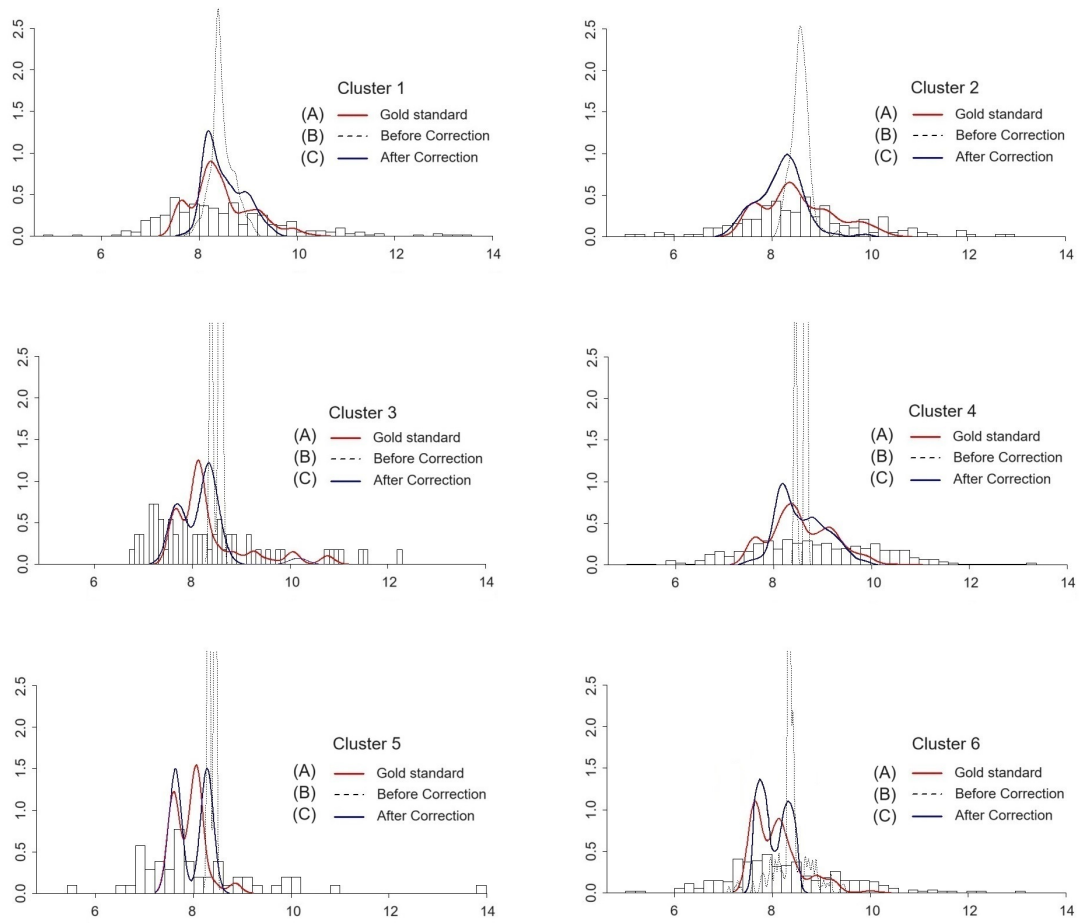


Figure 8: Curve alignment under the condition of ‘error rate=0.40’ with $\zeta = 0.5$: Cluster-wise histograms (for $j = 1, \dots, 6$) of the observed claim amount Y_h on a log scale and the out-of-sample predictive densities obtained from Model(A), (B), and (C)

5 Discussion

This paper proposes a novel hierarchical GLM framework to estimate insurance risk premium with the incorporation of the heterogeneity and NDB covariate, outlined in research questions RQ1 and RQ2. Our proposed model presents overall good empirical performances in retrieving the original error-free parameter values, which suggests that it is worth considering the Gustafson correction boosted with the hierarchical Bayesian framework to avoid the impact of the NDB errors. Ultimately, we seek to develop practical guidelines for selecting the optimal scaling factor with the Gustafson correction in the absence of gold standard data.

Regarding **RQ1**, we initially analyzed the partial pooling effect of a hierarchical GLM as a solution for addressing heterogeneity. Given the predetermined risk clusters $j = 1, \dots, 6$, we observed the highest LPPD value from the gold standard model, indicating an improvement in predictive power compared to other conventional GLM models. This suggests that the effect of heterogeneity is effectively mitigated. For **RQ2**, we expanded the model by incorporating Gustafson correction into the hierarchical GLM framework to mitigate the model risk linked to the NDB covariate. Throughout our experiments, we demonstrated the impact of the model risk when using a naïve model built on the NDB covariate. This model misspecification led to significant distortion in the spread and created extreme modality in the predictive distribution. In this context, we showed how applying the Gustafson correction within the same framework effectively ‘restores’ the predictive distribution’s original spread and characteristics. The NDB error mitigation was further confirmed by additional performance metrics such as LPPD, SSPE, SAPE, D_{KL} , etc.

A key aspect of our hybrid modeling framework to handle NDB covariates is the choice of scaling factor ζ , which our findings demonstrate to be crucial in error correction. Specifically, ζ moderates the contribution of $\hat{\lambda}^2 : V(\mathbf{x}^*|\mathbf{z})$ to explaining $\mathfrak{I}^2 : V(\mathbf{x}^*|\mathbf{x})$, guiding adjustments of the erroneous parameters. Notably, we found that both the scaling factor ζ and the error rates R_{ϵ_x} in the NDB covariate are crucial, requiring careful calibration for optimal results. To determine the ideal ζ , we have devised a rule of thumb relying on the maximum LPPD value, providing a path forward without access to gold-standard data. For the error rates R_{ϵ_x} in the NDB covariate, our experimental results indicate that the hierarchical GLM equipped with the Gustafson correction performs consistently well across different error rate scenarios, particularly within the range of scaling factor such as $0.5 \leq \zeta \leq 0.7$. However, these numbers suggest a relatively strong degree of relation between $\hat{\lambda}^2 : V(\mathbf{x}^*|\mathbf{z})$ and $\mathfrak{I}^2 : V(\mathbf{x}^*|\mathbf{x})$, as reflected by the prior knowledge: $\mathfrak{I}_j^2 = (1 - \zeta)\hat{\lambda}_j^2$ in Equation (34). Furthermore, the optimal range of ζ appears to be dataset-dependent, as it is likely unique to each one. Varying relationships among covariates can lead to distinct ranges for ζ . Indeed, we have yet to explore other datasets where this degree of relation is characterized by a different range of ζ . It can be important to investigate how the performance of the correction varies across different error rate scenarios when the optimal ζ is ranged differently. This leaves room for further research to fully understand the implications of ζ on model performance.

5.1 Future Work

There are several concerns regarding our hierarchical GLM risk premium modeling equipped with Gustafson correction.

- (a) **Dimensionality:** In our current analysis, we have limited our focus to two covariates - one binary and one continuous - for the sake of simplicity. However, it is imperative to explore more intricate datasets that encompass a greater variety of covariates. In the Bayesian framework, as the number of covariates increases, the complexity of the likelihood components - essentially the models representing these covariates - also expands. This escalation may introduce additional noise or unobserved structures that could compromise the stability of the resulting predictive distributions (Gelman and Hill, 2007). Therefore, conducting further investigations into the challenges posed by high-dimensional covariates within the hierarchical GLM framework presents an opportunity for enhancing model robustness and accuracy.
- (b) **Reliability of risk clusters:** As a parametric Bayesian approach, fixed risk clusters in the hierarchical GLM framework have several downsides that merit further exploration. Above all, if the predetermined risk clusters are not entirely accurate, this can lead to additional bias and possibly reduce model validity (Gelman and Carlin, 2013). However, when the fixed clusters are indeed correct, the partial pooling mechanism within the hierarchical GLM framework demonstrates its superiority, effectively leveraging the structure of the data while improving parameter estimates. Thus, future research should investigate more flexible Bayesian clustering methods that allow for dynamic cluster adjustments.
- (c) **Granularity of error rates:** In the development of simulation data, we sought to determine the error rate at which the Gustafson correction becomes ineffective. Our findings indicate that for error rates R_{ϵ_x} below 0.01, the application of the correction is generally unwarranted, as model risk remains minimal under such conditions. In the range of error rates from 0.10 to 0.50, the correction demonstrated its utility. Specifically, it was effective when the scaling factor $0.5 \leq \zeta \leq 0.7$, as the hidden relation between $\hat{\lambda}^2 : V(\mathbf{x}^*|\mathbf{z})$ and $\mathfrak{T}^2 : V(\mathbf{x}^*|\mathbf{x})$ in the LGPIF dataset is explained by the range of ζ . However, we hypothesize that at error rates exceeding 0.50, the correction may encounter considerable limitations in its effectiveness, despite our identification of the optimal range for ζ . It is noteworthy that this specific scenario has not yet been explored in our study, indicating a potential avenue for future research utilizing different datasets.
- (d) **Scalability of posterior simulation:** Lastly, we propose an examination of the scalability of posterior simulations utilizing our Gibbs sampler. Our empirical analysis, particularly with the LGPIF dataset, demonstrates that the hierarchical GLM framework maintains stable performance with sample sizes up to $H \leq 2,000$. However, the impact of larger sample sizes, specifically those exceeding 10,000, has not been addressed within the scope of this paper. As the volume of data increases, it raises critical questions regarding the computational efficiency of our hierarchical GLM framework, particularly given the heightened demand for computational resources and the potential for performance degradation, as noted in studies of Ni et al. (2020). This aspect becomes increasingly significant in contexts where insurance loss data is anticipated to accumulate

over time, necessitating a thorough exploration of the framework's capacity to scale effectively while maintaining performance.

Variable Definitions

The following variables and functions are used in this manuscript:

$i = 1, \dots, N_h$	observation index i in policy h .
$h = 1, \dots, H$	policy index h with sample (policy) size H .
$j = 1, \dots, J$	cluster index for J clusters.
n_j	number of observations in cluster j .
Y_{hi}, N_h	i th individual loss amount and loss count in a policy h .
$Y_{j(hi)}, N_{j(h)}$	i th individual loss amount and loss count in a policy h in a cluster j .
S_h	outcome variable as $\sum_i Y_{hi}$ in a policy h .
$\mathbf{X} = \{\mathbf{X}^F, \mathbf{X}^S\}$	list of covariate matrices (including $\mathbf{X}^S, \mathbf{X}^S$) for both frequency and severity.
$\mathbf{X}^F : \{\mathbf{x}^F, \mathbf{z}^F\}$	matrix of covariates (including $\mathbf{x}^F, \mathbf{z}^F$) for claim count outcome (Frequency).
$\mathbf{X}^S : \{\mathbf{x}^S, \mathbf{z}^S\}$	matrix of covariates (including $\mathbf{x}^S, \mathbf{z}^S$) for claim amount outcome (Severity). Focusing solely on severity, we omit the superscript 's' for simplicity.
$\mathbf{X} : \{\mathbf{x}, \mathbf{z}\}$	matrix of covariates (including $\mathbf{z}^S, \mathbf{x}^S$) for claim amount outcome (Severity). Focusing solely on severity, we omit the superscript 's' for simplicity.
$\mathbf{X}_h^F : \{x_h^F, z_h^F\}$	vector of covariates in policy h (Frequency).
\mathbf{x}^F	vector of continuous covariate (Frequency).
\mathbf{z}^F	vector of binary covariate (Frequency).
x_h^F	individual value of continuous covariate in policy h (Frequency).
z_h^F	individual value of binary covariate in policy h (Frequency).
$\mathbf{X}_h^S : \{x_h^S, z_h^S\}$	vector of covariates in policy h (Severity).
\mathbf{x}^S	vector of continuous covariate, and \mathbf{x}^* indicates the mismeasured (Severity).
\mathbf{z}^S	vector of binary covariate (Severity).
x_h^S	individual value of continuous covariate, and x_h^* indicates the mismeasured in policy h (Severity).
z_h^S	individual value of binary covariate in policy h (Severity).
$\mathbf{X}_h : \{x_h, z_h\}$	vector of covariates in policy h .
\mathbf{x}	vector of continuous covariate, and \mathbf{x}^* indicates the mismeasured.
\mathbf{z}	vector of binary covariate.
x_h	individual value of continuous covariate, and x_h^* indicates the mismeasured in policy h .
z_h	individual value of binary covariate in policy h .
$p_0(\cdot)$	parameter model (for prior).
$p(\cdot)$	parameter model (for posterior).
$f_0(\cdot)$	data model (for continuous cluster).
$f(\cdot)$	data model (for discrete cluster).
$E[\cdot], V[\cdot]$	Expectation and variance as point estimates.

$\phi(\cdot)$	probability density function of Standard Gaussian density.
$\Phi(\cdot)$	Cumulative density function of Standard Gaussian density.
θ_j	set of parameters - β, σ^2 - associated with the outcome model $f(Y \mathbf{X})$ for j cluster (posterior sample: $\theta_j^{(*)}$).
w_j	set of parameters - π, μ, λ^2 - associated with the covariate models $f(\mathbf{X})$ for j cluster (posterior sample: $w_j^{(*)}$).
ω_j	cluster weights (mixing coefficient) for j cluster (finalized sample: $\omega_j^{(*)}$).
$\beta_j : \{\beta_{j0}, \beta_{j1}, \beta_{j2}\}$	regression coefficient vector for a mean outcome estimation.
$\beta_0, \Sigma_{\beta_0}$	vector of initial regression coefficients and variance-covariance matrix, i.e. $\hat{\sigma}^2(\mathbf{X}^T \mathbf{X})^{-1} = \mathbf{X}^T \mathbf{X} (\Sigma Y - \Sigma \hat{Y})^T (\Sigma Y - \Sigma \hat{Y}) / (n - p)$ obtained from the baseline multivariate Gamma regression of $\Sigma \hat{Y} > 0$.
σ_j^2	cluster-wise variance or scale parameter for the outcome.
π_j	proportion parameter for Bernoulli covariate.
μ_j	location parameter for Gaussian covariate \mathbf{x} .
λ_j^2	dispersion parameter for Gaussian covariate \mathbf{x} .
$\kappa_j : \{\kappa_{j0}, \kappa_{j1}\}$	regression coefficient vectors to explain the mean of the unobserved Gaussian covariate $\mathbf{x} \mathbf{z}$.
\mathfrak{I}_j^2	variance parameter for $\mathbf{x}^* \mathbf{x}$ to indicate the contamination level in the measurement model.
m_0, δ	hyperparameters of Multivariate Normal for β_0^S .
q_0, Λ	hyperparameters of Inverse Wishart density for $\Sigma_{\beta_0}^S$.
u_0, v_0	hyperparameters of Inverse Gamma density for σ_j^2 .
ρ_{u1}, ρ_{u2}	hyperparameters of Fink's function for u_0 .
ρ_{v1}, ρ_{v2}	hyperparameters of Gamma density for v_0 .
μ_0, λ_j^2	hyperparameters of Gaussian density of μ_j .
c_0, d_0	hyperparameters of Inverse Gamma density for λ_j^2 .
g_0, h_0	hyperparameters of Beta density for π_j .
$\mathbb{K}_1 : \begin{pmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_{n_j} \end{pmatrix}$	$n_j \times 2$ matrix to compute $\sum_{h=1}^{n_j} (x_h^* - \hat{\kappa}_{j0} - \hat{\kappa}_{j1} z_h)^2$.
$\mathbb{K}_2 : \begin{pmatrix} \sum_{h=1}^{n_j} x_h^* \\ \sum_{h=1}^{n_j} x_h^* z_h \end{pmatrix}$	2×1 matrix to compute $\sum_{h=1}^{n_j} (x_h^* - \hat{\kappa}_{j0} - \hat{\kappa}_{j1} z_h)^2$.
$\beta_j^F : \{\beta_{j0}^F, \beta_{j1}^F, \beta_{j2}^F\}$	regression coefficient vector for a mean claim count (Frequency) estimation.
$\beta_j^S : \{\beta_{j0}^S, \beta_{j1}^S, \beta_{j2}^S\}$	regression coefficient vector for a mean claim amount (Severity) estimation.
ξ_h, ψ	parameters - ξ_h (number of failure) and ψ (number of success) - for Negative Binomial.
$\beta_0^F, \Sigma_{\beta_0}^F$	vector of initial regression coefficients and variance-covariance matrix obtained from the baseline multivariate Poisson regression of $\hat{N} > 0$.
$\beta_0^S, \Sigma_{\beta_0}^S$	vector of initial regression coefficients and variance-covariance matrix obtained from the baseline multivariate Gamma regression of $\hat{Y} > 0$.

u_0^F, v_0^F	hyperparameters of Inverse Gamma density for ψ_j .
u_0^S, v_0^S	hyperparameters of Inverse Gamma density for σ_j^2 .
$\underline{m}_0, \underline{\delta}$	hyperparameters of Multivariate Normal for β_0^F .
$\underline{q}_0, \underline{\Delta}$	hyperparameters of Inverse Wishart density for $\Sigma_{\beta_0}^F$.
$\underline{\rho}_{-u1}, \underline{\rho}_{-u2}$	hyperparameters of Fink's function for u_0^F .
$\underline{\rho}_{v1}, \underline{\rho}_{v2}$	hyperparameters of Gamma density for v_0^F .
$\beta_0^{S+}, \Sigma_{\beta_0}^{S+}, u_0^{S+}, v_0^{S+}$	communal hyperparameters for partial pooling in a hierarchical GLM (claim amount).
$\beta_0^{F+}, \Sigma_{\beta_0}^{F+}, u_0^{F+}, v_0^{F+}$	communal hyperparameters for partial pooling in a hierarchical GLM (claim count).
$\hat{\beta}_j : \{\hat{\beta}_{j0}, \hat{\beta}_{j1}, \hat{\beta}_{j2}\}$	regression coefficient vector for a mean outcome estimation based on NDB covariate \mathbf{x}^* (before correction).
$\hat{\sigma}_j^2$	cluster-wise variance or scale parameter for the outcome based on NDB covariate \mathbf{x}^* (before correction).
$\hat{\xi}_j$	skewness parameter for log skew-normal outcome based on NDB covariate \mathbf{x}^* (before correction).
$\hat{\lambda}_j^2$	dispersion parameter for Gaussian covariate $\mathbf{x} \mathbf{z}$ based on NDB covariate \mathbf{x}^* (before correction).
$\hat{\kappa}_j : \{\hat{\kappa}_{j0}, \hat{\kappa}_{j1}\}$	regression coefficient vectors to explain the mean of the unobserved Gaussian covariate $\mathbf{x} \mathbf{z}$ based on NDB covariate \mathbf{x}^* (before correction).
$\tilde{\kappa}, \tilde{\Sigma}_{\kappa}$	hyperparameters of Multivariate Normal for κ .
R_{ϵ_x}	error rate, representing the proportion of the total noise relative to the total true values within an NDB covariate.
$\sigma_{j\epsilon}^2$	variance of an NDB error.
$\rho_{j(x, x^*)}$	correlation between the true covariate and the NDB covariate in cluster j .
ϵ	NDB measurement error
$1/\delta$	Variance inflation factor as a ratio of the virtual sample size to the observation sample size, representing the impact of the prior. The default choice is $1/\delta = 100$. See Sharples 1990 Sharples (1990).
$L(\cdot)$	likelihood function of the current model.
$P(\cdot)$	likelihood function of the target model.

Acronym List

NDB	Non-Differential Berkson Error
LPPD	Log Pointwise Predictive Density
D_{KL}	Kullback-Leibler Divergence
SSPE	Sum of Square Prediction Error
SAPE	Sum of Absolute Prediction Error
CTE	Conditional Tail Expectation

NB	Negative Binomial Distribution
N	Normal (Gaussian) Distribution
LogN	Lognormal Distribution
MVN	Multivariate Normal Distribution
Ga	Gamma Distribution
Beta	Beta Distribution
InvGa	Inverse Gamma Distribution
IW	Inverse Wishart Distribution
LGPIF	Local Government Property Insurance Fund
GLM	Generalized Linear Model
GAM	Generalized Additive Model
MARS	Multivariate Adaptive Regression Spline
RC	Regression Calibration
SIMEX	Simulation Extrapolation
WLS	Weighted Least Squares

References

- Ankur Aggarwal, Michael B Beck, Matthew Cann, Tim Ford, Dan Georgescu, Nirav Morjaria, A Smith, Yvonne Taylor, Andreas Tsanakas, Louise Witts, et al. Model risk–daring to open up the black box. *British Actuarial Journal*, 21(2):229–296, 2016.
- D Anderson and K Burnham. Model selection and multi-model inference. *Second*. NY: Springer-Verlag, 63(2020):10, 2004.
- Etti Baranoff et al. Risk management for enterprises and individuals. *Journal of organizational Management*, 3(2):23–35, 2006.
- Philip J Boland. Statistical methods in general insurance. In *International Conference on Teaching Statistics, ICOTS-7*, pages 1–6, 2006.
- Nicolas Bousquet. Diagnostics of prior-data agreement in applied Bayesian analysis. *Journal of Applied Statistics*, 35(9):1011–1029, 2008.
- Vytaras Brazauskas, Bruce L Jones, Madan L Puri, and Ričardas Zitikis. Estimating conditional tail expectation with actuarial applications in view. *Journal of Statistical Planning and Inference*, 138(11):3590–3604, 2008.
- Stephen P Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998.

- Andrew JG Cairns. A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics*, 27(3):313–330, 2000.
- Raymond J Carroll, David Ruppert, Leonard A Stefanski, and Ciprian M Crainiceanu. *Measurement error in nonlinear models: a modern perspective*. Chapman and Hall/CRC, 2006.
- John R Cook and Leonard A Stefanski. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical association*, 89(428):1314–1328, 1994.
- Denis Cousineau and Teresa Allan. Likelihood and its use in parameter estimation and model comparison. *Mesure et évaluation en éducation*, 37(3):63–98, 2015.
- Michel Denuit and Stefan Lang. Non-life rate-making with bayesian gams. *Insurance: Mathematics and Economics*, 35(3):627–647, 2004.
- Zoe Fewell. *Causal modelling in epidemiology and health services research*. PhD thesis, University of Bristol, 2007.
- Daniel Fink. A compendium of conjugate priors. *Environmental Statistics Group*, 46, 1997.
- Louise Francis. Martian chronicles: is mars better than neural networks? In *Casualty Actuarial Society Forum*, pages 75–102, 2003.
- Laurence S Freedman, Vitaly Fainberg, Victor Kipnis, Douglas Midthune, and Raymond J Carroll. A new method for dealing with measurement error in explanatory variables of regression models. *Biometrics*, 60(1):172–181, 2004.
- Andrew Gelman and John Carlin. *Bayesian data analysis*. CRC press, 2013.
- Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, 2007.
- Y Yi Grace, Aurore Delaigle, and Paul Gustafson. *Handbook of measurement error models*. CRC Press, 2021.
- Paul Gustafson. *Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments*. CRC Press, 2008.
- Sabine Hoffmann, Estelle Rage, Dominique Laurier, Pierre Laroche, Chantal Guihenneuc, and Sophie Ancelet. Accounting for berkson and classical measurement error in radon exposure using a bayesian structural approach in the analysis of lung cancer mortality in the french cohort of uranium miners. *Radiation Research*, 187(2):196–209, 2017.
- Liang Hong and Ryan Martin. Dirichlet process mixture models for insurance loss data. *Scandinavian Actuarial Journal*, 2018(6):545–554, 2018.
- Rob Kaas, Marc Goovaerts, Jan Dhaene, and Michel Denuit. *Modern actuarial risk theory: using R*, volume 128. Springer Science & Business Media, 2008.
- Marc C Kennedy and Anthony O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

- Simon Klau, Sabine Hoffmann, Chirag J Patel, John PA Ioannidis, and Anne-Laure Boulesteix. Examining the robustness of observational associations to model, measurement and sampling uncertainty with the vibration of effects framework. *International Journal of Epidemiology*, 50(1):266–278, 2021.
- Richard McElreath. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC, 2018.
- John M Neuhaus and Charles E McCulloch. Separating between-and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(5):859–872, 2006.
- Yang Ni, Yuan Ji, and Peter Müller. Consensus monte carlo for random subsets using shared anchors. *Journal of Computational and Graphical Statistics*, 29(4):703–714, 2020.
- Eric J Oh, Bryan E Shepherd, Thomas Lumley, and Pamela A Shaw. Considerations for analysis of time-to-event outcomes measured with error: Bias and correction with SIMEX. *Statistics in medicine*, 37(8):1276–1289, 2018.
- Esbjörn Ohlsson and Björn Johansson. *Non-life insurance pricing with generalized linear models*, volume 174. Springer, 2010.
- Pietro Parodi. *Pricing in general insurance*. Chapman and Hall/CRC, 2023.
- Karl Pearson. Method of moments and method of maximum likelihood. *Biometrika*, 28(1/2):34–59, 1936.
- Alexandra Romann. *Evaluating the performance of simulation extrapolation and Bayesian adjustments for measurement error*. PhD thesis, University of British Columbia, 2008.
- David PM Scollnik. Actuarial modeling with mcmc and bugs. *North American Actuarial Journal*, 5(2):96–124, 2001.
- Linda Sharples. Identification and accommodation of outliers in general hierarchical models. *Biometrika*, 77(3):445–453, 1990.
- Anders Skrondal and Jouni Kuha. Improved regression calibration. *Psychometrika*, 77(4):649–669, 2012.
- Geoff Werner and Claudine Modlin. Basic ratemaking. In *Casualty Actuarial Society*, volume 4, pages 1–320, 2010.
- Rainer Winkelmann. *Econometric analysis of count data*. Springer Science & Business Media, 2008.
- Mario V Wuthrich. Non-life insurance: mathematics & statistics. *Available at SSRN 2319328*, 2020.