Advanced Dependence Modeling of Loss Reserves: Integrating Recurrent Neural Networks and Seemingly Unrelated Regression Copula Mixed Models for Diversified Risk Capital

Advanced Dependence Modeling of Loss Reserves: Integrating Recurrent Neural Networks and Seemingly Unrelated Regression Copula Mixed Models for Diversified Risk Capital

BY
PENGFEI (FRANK) CAI, M.Sc.

A thesis
submitted to the Department of Mathematics & Statistics
and the school of graduate studies
of McMaster University
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Doctor of Philosophy (2025) McMaster University

(Statistics) Hamilton, Ontario, Canada

TITLE: Advanced Dependence Modeling of Loss Reserves: Integrating Recurrent Neural Networks and Seemingly Unrelated Regression Copula Mixed Models for Diversified Risk Capital

AUTHOR: Pengfei Cai
M.Sc. (Statistics)
McMaster University
Hamilton, Ontario, Canada

SUPERVISORS: Dr. Anas Abdallah and Dr. Pratheepa Jeganathan

NUMBER OF PAGES: xv, 122

Abstract

In the property and casualty (P&C) insurance industry, reserves comprise most of a company's liabilities. These reserves are the best estimates made by actuaries for future unpaid claims. The actuarial industry has developed both parametric and non-parametric methods for loss reserving. However, the use of machine learning tools to capture dependence between loss reserves from multiple LOBs and calculate the aggregated risk capital remains uncharted. This thesis introduces the use of the Deep Triangle (DT), a recurrent neural network, for multivariate loss reserving, incorporating an asymmetric loss function to combine incremental paid losses of multiple LOBs. Further, we extend generative adversarial networks (GANs) by transforming the two loss triangles into a tabular format and generating synthetic loss triangles to obtain the predictive distribution for reserves. We refer to the integration of DT for multivariate loss reserving and GAN for risk capital analysis as the extended Deep Triangle (EDT). As the second contribution of this thesis, we propose SUR copula mixed models to enhance SUR copula regression with multiple companies' data for loss prediction and risk capital analysis. Due to the heterogeneous history of losses between companies and different LOBs, we model this heterogeneity using random effects and select varying distributions for losses from each LOB. We model the development and accident year effects as fixed effects and apply shrinkage to make it more robust to the decreasing number of observations over accident year and development year. To illustrate EDT and SUR copula mixed models, we apply and calibrate these methods using data from

multiple companies from the National Association of Insurance Commissioners database. For validation, we compare the EDT and SUR copula mixed model to the SUR copula regression models and find that the EDT and SUR copula mixed model outperform the SUR copula regression models in predicting total loss reserve. Furthermore, with the obtained predictive distribution for reserves, we show that risk capital calculated from the EDT and SUR copula mixed model is smaller than that of the SUR copula regression models, suggesting a more considerable diversification benefit. We also confirmed these findings in simulation studies. Finally, we introduce a chapter on a hybrid semi-parametric approach, which bridges the interpretability of dependence structures with the flexibility to capture complex effects, including interactions; its deeper application and simulation studies are left for future work.

Acknowledgments

Foremost, I want to express my sincere gratitude to my exceptional supervisors, Dr. Anas Abdallah and Dr. Jeganathan Pratheepa. The past four years working with them have been invaluable, and their guidance and insightful suggestions have been crucial in conducting my research.

I'd also like to acknowledge Dr. N. Balakrishnan and Dr. John Maheu, members of my supervisory committee, for their valuable feedback and recommendations throughout my research. Additionally, Dr. Ben Bolker and Dr. Shui Feng provided helpful comments during statistics seminar presentations and discussions, for which I am very thankful. My thanks also go to friends and fellow graduate students, Lan Wang, Qingzhe Zeng, and Rajitha Senanayake for their assistance.

Last but not least, I am very grateful to my parents for their understanding and support. A special thanks also to my wife, Lei, and my daughter, Jennifer; their love and support provided me with the time needed to complete this work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Insurance firms are tasked with the crucial responsibility of establishing reserve funds to guarantee the future compensation of policyholders who have made claims. To meet their commitments, insurers maintain claim reserves, ensuring sufficient funds are available for all future payouts. These reserves are largely based on historical claim data, which helps in estimating future claims through various reserving methods. Loss reserving generally follows two approaches: a micro-level approach focusing on individual claims, or a macro-level approach dealing with claims in aggregate.

The macro-level approach aggregates individual claims, organizing them into loss triangles according to accident and development years. For loss reserving, the chain ladder method (Mack, 1993) has been widely used in practice with the assumption that claims will continue to develop similarly in the future. However, a notable limitation of this method is its exclusion of uncertainty in its calculations. Mack (1993) present a method to compute the distribution-free standard error of the reserve based on the chain ladder method to address this gap. For further insight into stochastic loss reserving methods, specifically for a single line

of business (LOB) and at the macro-level, the works of England and Verrall (2002) and Wüthrich and Merz (2008) provide comprehensive reviews.

Insurance companies typically assess risk measures across all their LOBs. These measures are calculated based on the predictive distribution of the total reserve, which includes reserves allocated for each LOB. Insurers must estimate this predictive distribution accurately, as it offers valuable insights for them, especially in risk management. When dealing with multiple LOBs, a common assumption is the independence of claims across different LOBs. In such scenarios, the portfolio's total reserve and risk measure is the sum of each LOB. This approach, known as the "silo" method (Ajne, 1994), does not account for any diversification benefits. However, insurance companies often operate across multiple LOBs, where claims can be related. For example, claims across different LOBs can be related due to a common factor like inflation, which impacts the cost of claims in different LOBs. When a claim involves different coverages from different lines of business, losses can also become correlated. Therefore, it becomes imperative for insurers to account for the dependencies between claims in different LOBs. Acknowledging these dependencies is essential for accurately estimating total reserves and effectively leveraging diversification benefits in calculating risk capital.

1.2 Reserve and Risk Capital

Let X_{ij} denote the incremental paid losses of all claims in accident year i ($1 \leq i \leq I$) and development year j ($1 \leq j \leq I$). The accident year refers to the year the insured event happened. The first accident year is denoted with 1, and the most recent accident year is denoted with I. The development year indicates the time the payment is made. The incremental paid loss refers to all payments in development year j for the claims in year i. For one company and one business line, the observed data X_{ij} for $i = 1, 2, ..., I$ and $j = 1, 2, ..., I - i + 1$ is shown in

the upper triangle of Table 1.1.

Table 1.1: The Loss Triangle

	Development year j				
Accident year i	1	2	...	I-1	I
1	X_{11}	X_{12}	...	$X_{1,I-1}$	$X_{1,I}$
2	X_{21}	X_{22}	...	$X_{2,I-1}$	
...	...	...	...		
I-1	$X_{I-1,1}$	$X_{I-1,2}$			
I	$X_{I,1}$				

Note: The upper triangle is the loss triangle. The rows are accident year, and the columns are development year. X_{ij} denotes the incremental paid loss in accident year i and development year j.

The incremental paid loss X_{ij} is adjusted for each LOB's exposure to ensure comparability across accident years. The exposure variable, such as premiums or the number of policies, provides a scaling factor. The standardized incremental paid loss is then defined as $Y_{ij} = X_{ij}/\omega_i$, where ω_i represents the exposure for the i^{th} accident year. In the case of multiple LOBs from one company, the standardized incremental paid loss for the ℓ^{th} LOB is denoted by $Y_{i,j}^{(l)}$, with its predicted value represented as $\hat{Y}_{ij}^{(\ell)}$.

To estimate the lower triangle values $X_{ij}^{(\ell)}$, we multiply $\hat{Y}_{ij}^{(\ell)}$ by the corresponding exposure $\omega_i^{(\ell)}$. This yields a point estimate of the outstanding claims for each LOB, given by $R^{(\ell)} = \sum_{i=2}^{I} \sum_{j=I-i+2}^{I} \omega_i^{(\ell)} \hat{Y}_{ij}^{(\ell)}$. Finally, the total reserve for the entire insurance portfolio is $R = \sum_{\ell=1}^{2} R^{(\ell)}$.

In actuarial practice, reserve estimation extends beyond point estimates to include measures of reserve variability. Given the predictive distribution of reserves, denoted by F_R, we compute commonly used actuarial risk measures, such as value at risk (VaR) and tail value at risk (TVaR). Risk measures evaluate potential loss and are crucial for determining the amount of capital to hold to cover severe losses.

The VaR_k is the $100 * k$ percentile of R, i.e., $\text{VaR}_k(R) = F_R^{-1}(k)$ while TVaR_k is the expected loss conditional on exceeding the VaR_k, i.e., $\text{TVaR}_k(R) = \mathbb{E}[R|R \geq$

$\text{VaR}_k(R)]$.

Tail Value-at-Risk (TVaR) is more informative than Value-at-Risk (VaR) in risk assessment. TVaR, a coherent risk measure captures the expected shortfall and adherence to the sub-additive property (Acerbi and Tasche, 2002). This means that for any two reserves R_1 and R_2, corresponding to two LOBs, the combined risk measure ρ of their sum is less than or equal to the sum of their individual risk measures. That is, $\rho(R_1+R_2) \leq \rho(R_1)+\rho(R_2)$. This ensures that the total risk measure does not exceed the sum of the individual risk measures, reflecting diversification benefits in risk assessment. Contrarily, VaR lacks this sub-additivity, especially in skewed distributions, making TVaR a more reliable indicator in risk management. From the insurance perspective, risk measures lacking the sub-additivity can be misleading because they can increase the company's liability, resulting in a larger tax deduction.

From the TVaR, we calculate the risk capital, which is the difference between the risk measure and the liability value. (see, e.g., Dhaene et al. (2006)). Risk capital is also set aside as a buffer against potential losses from extreme events. In practice, the risk measure is set at a high-risk tolerance k, and the liability value is set at a lower risk tolerance between 60% and 80%, according to the risk appetite. We set the risk tolerance at 60% for the reserve in our risk capital analysis.

We define risk capital associated with total reserve R as in (1.1).

$$\text{Risk capital (R)} = \text{TVaR}_k(R) - \text{TVaR}_{60\%}(R). \tag{1.1}$$

Moreover, exploring the diversification benefits between two LOBs is essential in risk management. The "silo" approach computes risk measures for each LOB independently and aggregates them, disregarding potential diversification benefits. In contrast, our study compares the risk capital estimates obtained using the proposed method with those from the "silo" approach, as defined in (1.2). This

comparison highlights the impact of recognizing interdependencies between LOBs in risk capital (RC) estimation. Further details on this approach can be found in Abdallah and Wang (2023).

$$\text{Gain} = \Big(\text{RC}_{\text{Silo}}(R) - \text{RC}_{\text{Copula}}(R)\Big)/\text{RC}_{\text{Silo}}(R). \tag{1.2}$$

1.3 Copula Model for Loss Reserving

1.3.1 Copulas

According to Sklar's theorem (Nelsen, 2006), any cumulative distribution function (cdf) $F(x_1, x_2)$ of a two-dimensional random vector (X_1, X_2) can be expressed as

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)), \tag{1.3}$$

where $F_1(\cdot)$ and $F_2(\cdot)$ are the marginal cdfs of X_1 and X_2, and C is a bivariate function, called a copula. If X_1 and X_2 are independent, then C is the product copula.

The most common measure of dependence between two random variables is Pearson's correlation coefficient, which only measures linear dependence. To measure nonlinear dependence, rank correlation coefficients such as Kendall's τ and Spearman's ρ are more suitable. They can be expressed in terms of the copula as

$$\tau(X_1, X_2) = 4 \iint_{[0,1]^2} C(u_1, u_2)\, dC(u_1, u_2) - 1 = 4\mathbb{E}[C(u_1, u_2)] - 1, \tag{1.4}$$

$$\rho(X_1, X_2) = 12 \iint_{[0,1]^2} C(u_1, u_2)\, dC(u_1, u_2) - 3 = 12\mathbb{E}[C(u_1, u_2)] - 3, \tag{1.5}$$

where (u_1, u_2) is a two-dimensional random vector on $[0, 1]^2$ and $C(u_1, u_2)$ is the corresponding cdf.

Next, we give examples of copulas and briefly summarize the main properties used in this study.

Gaussian copula

The Gaussian copula allows for positive and negative dependence. The Gaussian copula is defined as

$$C\left(u_1, u_2\right) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-r^2}} \exp\left[-\frac{x_1^2 + x_2^2 - 2rx_1x_2}{2\left(1-r^2\right)}\right] \mathrm{d}x_1 \, \mathrm{d}x_2, \tag{1.6}$$

where $-1 < r < 1$ is Pearson's correlation coefficient between x_1 and x_2 and Φ is the cdf of the standard normal random variable. Parameter r is related to Kendall's τ and Spearman's ρ coefficients by the relations $\tau = (2/\pi)\sin^{-1}(r)$ and $\rho = (6/\pi)\sin^{-1}(r/2)$.

Student's t copula

The Student's t copula allows for positive and negative dependence. Student's t copula takes the form

$$C\left(u_1, u_2\right) = \int_{-\infty}^{T_\nu^{-1}(u_1)} dx_1 \int_{-\infty}^{T_\nu^{-1}(u_2)} dx_2 \left[1 + \frac{x_1^2 - 2rx_1x_2 + x_2^2}{\nu\left(1-r^2\right)}\right]^{-\frac{\nu+2}{2}}, \tag{1.7}$$

where r is the correlation coefficient between x_1 and x_2 and T_ν is the cdf of a Student distribution with ν degrees of freedom. When ν goes to infinity, the T copula converges to the Gaussian copula.

Frank copula

The Frank copula allows for positive and negative dependence. The corresponding copula function is given by

$$C\left(u_1, u_2\right) = -\frac{1}{\theta}\ln\left(1 + \frac{\left(\exp\left(-\theta u_1\right) - 1\right)\left(\exp\left(-\theta u_2\right) - 1\right)}{\exp(-\theta) - 1}\right), \tag{1.8}$$

where $\theta \in (-\infty, +\infty)\backslash\{0\}$. Positive values of θ indicate positive dependence, whereas negative values indicate negative dependence. The independence copula is obtained when $\theta \to 0$. The relationship between rank and the Pearson correlation coefficient and θ is

$$\tau = 1 - \frac{4}{\theta} + 4\frac{D_1(\theta)}{\theta},$$

and

$$\rho = 1 - \frac{12}{\theta}\left[D_1(\theta) - D_2(\theta)\right],$$

where $D_k(\theta)$ is defined as

$$D_k(\theta) = \frac{k}{\theta^k}\int_0^\theta \frac{t^k}{\exp(t) - 1}dt, \quad k = 1, 2.$$

1.3.2 Copula Regression

Now we detail the background of the copula regression. Consider the cumulative distribution of $Y_{ij}^{(\ell)}$,

$$F_{ij}^{(\ell)} = \text{Prob}(Y_{ij}^{(\ell)} \leq y_{ij}^{(\ell)}) = F(y_{ij}^{(\ell)}; \eta_{ij}^{(\ell)}, \gamma^{(\ell)}), \tag{1.9}$$

where ℓ denote ℓ^{th} LOB, $\eta_{ij}^{(\ell)}$ denotes the systematic component, which determines the location and $\gamma^{(\ell)}$ determines the shape.

We assume $\gamma^{(\ell)}$ is the same for all the cells (i, j) for each loss triangle. Now we model the systematic component $\eta_{ij}^{(\ell)}$ using $\alpha_i^{(\ell)}(i \in 1, 2, ..., 10)$ and $\beta_j^{(\ell)}(j \in$

$1, 2, ..., 10$) as predictors that characterize the effect of the accident year and the development year corresponding to $Y_{ij}^{(\ell)}$ as in (1.10).

$$\eta_{ij}^{(\ell)} = \xi^{(\ell)} + \alpha_i^{(\ell)} + \beta_j^{(\ell)}, \tag{1.10}$$

where $\xi^{(\ell)}$ is the intercept and constraints are $\alpha_1^{(\ell)} = 0$ and $\beta_1^{(\ell)} = 0$ for parameter identification. We use the goodness-of-fit test to choose the distribution for $Y_{i,j}^{(\ell)}$.

In addition to the specified marginal densities, we assume that $Y_{ij}^{(\ell)}$ and $Y_{ij}^{(\ell')}$ from different LOBs with the same accident and development year are dependent. This is called pair-wise dependence. Moreover, we consider the copulas to model the dependence structure between the two lines of business (Shi and Frees, 2011).

Next, we write the joint distribution of $\left(Y_{ij}^{(\ell)}, Y_{ij}^{(\ell')}\right)$ using copulas based on Sklar's theorem (Nelsen, 2006) as follows

$$F_{ij}\left(y_{ij}^{(\ell)}, y_{ij}^{(\ell')}\right) = \text{Prob}\left(Y_{ij}^{(\ell)} \leq y_{ij}^{(\ell)}, Y_{ij}^{(\ell')} \leq y_{ij}^{(\ell')}\right) = C\left(F_{ij}^{(\ell)}(y_{ij}^{(\ell)}), F_{ij}^{(\ell')}(y_{ij}^{(\ell')}); \theta\right), \tag{1.11}$$

where $F_{ij}^{(\ell)}$ and $F_{ij}^{(\ell')}$ are the marginal distributions for $Y_{ij}^{(\ell)}$ and $Y_{ij}^{(\ell')}$, respectively, and $C(\cdot, \theta)$ is the copula function such that $C(\cdot, \theta) : [0, 1]^2 \mapsto [0, 1]$ with parameter θ.

By getting derivative of (1.11) with respect to $Y_{ij}^{(\ell)}$ and $Y_{ij}^{(\ell')}$, we get the join PDF for $\left(Y_{ij}^{(\ell)}, Y_{ij}^{(\ell')}\right)$ in (1.12).

$$f_{ij}(y_{ij}^{(\ell)}, y_{ij}^{(\ell')}) = c\left(F_{ij}^{(\ell)}, F_{ij}^{(\ell')}; \theta\right) \prod_{\ell=1}^{2} f_{ij}^{(\ell)}, \tag{1.12}$$

where $c(\cdot)$ denotes the PDF corresponding to copula $C(\cdot)$ and $f_{ij}^{(\ell)}$ denotes the PDF associated with the marginal distribution $F_{ij}^{(\ell)}$.

Next, we use the maximum likelihood method to estimate the parameters in the regression model in (1.10) using the copula density in (1.12). We denote estimators as $\hat{\mu}_{ij}^{(\ell)}$, $\hat{\sigma}^{(\ell)}$ and $\hat{\phi}^{(\ell)}$.

The log-likelihood for all joint $\left(Y_{ij}^{(\ell)}, Y_{ij}^{(\ell')}\right)$ is given by

$$\begin{aligned} L(\eta_{ij}^{(\ell)}, \gamma^{(\ell)}, \eta_{ij}^{(\ell')}, \gamma^{(\ell')}, \theta) = & \sum_{i=1}^{I} \sum_{j=1}^{I+1-i} \log\left(c\left(F_{ij}^{(\ell)}, F_{ij}^{(\ell')}; \theta\right)\right) \\ & + \sum_{i=1}^{I} \sum_{j=1}^{I+1-i} \sum_{\ell=1}^{2} \log\left(f_{ij}^{(\ell)}\right), \end{aligned} \tag{1.13}$$

where $\gamma^{(\ell)} = \sigma$, $\gamma^{(\ell')} = \phi$, $\eta_{ij}^{(\ell)} = \mu_{ij}^{(\ell)}$, $\eta_{ij}^{(\ell')} = \log\left(\mu_{ij}^{(\ell')}\phi\right)$, and $\eta_{ij}^{(\ell)}$ is a function of $\alpha_i^{(\ell)}$ and $\beta_j^{(\ell)}$ as in regression model (1.10).

1.4 Predictive Distribution of the Total Reserve

In practice, insurance companies are interested in understanding the uncertainty of reserves. The bootstrapping technique can provide this information and allows for the determination of the entire predictive distribution. The two most popular approaches to generating the predictive distribution of the reserve based on the copulas are simulation and parametric bootstrapping.

Simulation is based on the estimated copula regression model, in which we use the Monte Carlo simulation to generate the predictive distribution of the reserve. The simulation is summarized as the following procedure (Shi and Frees, 2011):

(1) Simulate $\left(u_{ij}^{(1)}, u_{ij}^{(2)}\right)$ $(i + j - 1 > I)$ from estimated copula function $C(\cdot; \hat{\theta})$.

(2) Transform $u_{ij}^{(\ell)}$ to predictions of the lower triangles by inverse function $y_{ij}^{*(\ell)} = F^{(\ell)(-1)}(u_{ij}^{(\ell)}; \hat{\eta}_{ij}^{(\ell)}, \hat{\gamma}^{(\ell)})$, where $\hat{\eta}_{ij}^{(\ell)} = \hat{\xi}^{(\ell)} + \hat{\alpha}_i^{(\ell)} + \hat{\beta}_j^{(\ell)}$.

(3) Obtain a prediction of the total reserve by

$$\sum_{\ell=1}^{2} \sum_{i=2}^{I} \sum_{j=I-i+2}^{I} \omega_i^{(\ell)} y_{ij}^{*(\ell)}.$$

Repeat Steps (1) - (3) many times to obtain the bootstrap replicates of R. However, the limitation of Monte Carlo simulation is the inability to incorporate estimated parameter uncertainty. To address this constraint, we consider the parametric bootstrapping.

In the parametric bootstrapping, we generate a new upper triangle for each simulation with estimated parameters and fit the corresponding copula regression model to this new upper triangle (Taylor and McGuire, 2007; Shi and Frees, 2011). The detailed algorithm is as follows:

(1) Simulate $\left(u_{ij}^{(1)}, u_{ij}^{(2)}\right)$ $(i+j-1 \leq I)$ from estimated copula cdf function $C(\cdot;\hat{\theta})$.

(2) Transform $u_{ij}^{(\ell)}$ to estimate the upper triangles by inverse transform $y_{ij}^{*(\ell)} = F^{(\ell)(-1)}(u_{ij}^{(\ell)};\hat{\eta}_{ij}^{(\ell)},\hat{\gamma}^{(\ell)})$, where $\hat{\eta}_{ij}^{(\ell)} = \hat{\xi}^{(\ell)} + \hat{\alpha}_i^{(\ell)} + \hat{\beta}_j^{(\ell)}$.

(3) Generate an estimate of the total reserve using $y_{ij}^{*(\ell)}$ from step (2).

- Estimate the parameters $\hat{\eta}_{ij}^{*(\ell)}$,$\hat{\gamma}^{*(\ell)}$ and $\hat{\theta}^{*}$ by performing MLE for the copula regression model for $y_{ij}^{*(\ell)}$.
- Use $\hat{\eta}_{ij}^{*(\ell)}$,$\hat{\gamma}^{*(\ell)}$ and $\hat{\theta}^{*}$ to simulate the lower triangle, $y_{ij}^{**(\ell)}$ using the simulation Steps (1) and (2).
- Obtain a prediction of the total reserve by

$$\sum_{\ell=1}^{2}\sum_{i=2}^{I}\sum_{j=I-i+2}^{I}\omega_i^{(\ell)}y_{ij}^{**(\ell)}.$$

Repeat Steps (1)-(3) many times to obtain bootstrap replicates of R.

1.5 Thesis Objectives

The motivation and objectives of the thesis are as follows:

- Develop machine learning tools for multivariate loss reserving that capture dependence between two lines of business (LOBs). Traditional copula regression is limited in flexibility for modeling the tail of the marginal distributions and does not account for time dependence in incremental paid losses. While machine learning techniques are increasingly used in loss reserving, few models capture dependence between LOBs using recurrent neural networks (RNNs). In this thesis, we develop a Deep Triangle (DT), a gated recurrent neural network framework for multivariate loss reserving.

- Develop machine learning methods to generate aggregated risk capital from the predictive distribution, capturing pairwise dependence between the two LOBs and leveraging diversification benefits. We use generative adversarial networks (GANs) to generate synthetic loss triangles and forecast the predictive distribution of reserves for the DT. The combination of DT and GAN, called extended Deep Triangle (EDT), provides a framework for multivariate loss reserving and risk capital analysis.

- Develop seemingly unrelated regression (SUR) copula mixed models to model dependence between LOBs and to address the heterogeneous history of losses across companies and LOBs. The SUR copula regression incorporates dependence between two LOBs through a copula using loss triangles from one company, but tends to produce a relatively large bias. We enhance this approach by developing SUR copula mixed models that incorporate multiple companies' data for improved loss prediction and risk capital analysis.

- Develop a sparse SUR copula mixed model to improve the robustness of the SUR copula mixed model. In the most recent accident and development years, the number of observed incremental paid losses decreases substantially. To address this, we combine the SUR copula mixed model with LASSO to shrink coefficients toward zero, thereby reducing variability.

- Develop a hybrid model combining the EDT and SUR copula mixed model to interpret the dependence between the two LOBs in the EDT. While the EDT is effective in prediction, it is limited in interpreting the sign and strength of dependence. In this thesis, we estimate dependence using the SUR copula mixed model applied to the residuals from the EDT. Due to heterogeneity in residuals across companies and LOBs, we incorporate random effects into the model.

1.6 Scope of the thesis

The work is organized as follows: In Chapter 2, we comprehensively describe the extended Deep Triangle (EDT) employed in this study for loss reserving and predictive distribution of reserves. We apply and calibrate the EDT and copula models using a dataset focused on personal and commercial automobile LOBs from 30 companies. Additionally, we conduct a comparative analysis of the computed risk capitals against other models, revealing that the EDT model yields smaller risk capital estimates. We also introduce a simulation study to illustrate that the EDT framework consistently generates smaller risk capital than the copula regression models. The copula regression incorporates the dependence between two LOBs through a copula and multiple company fixed effects and produces a relatively larger percentage of error compared to the EDT.

In Chapter 3, we propose Seemingly Unrelated Regression (SUR) copula mixed models to enhance SUR copula regression with multiple companies' data for loss prediction and risk capital analysis. Due to the heterogeneous history of losses between companies and different LOBs, we model this heterogeneity using random effects and select varying distributions for losses from each LOB. To overcome the computational complexity of the SUR copula mixed model, we develop a two-stage estimation approach to estimate the parameters for the proposed model .

This approach is illustrated with multiple pairs of loss triangles from the National Association of Insurance Commissioners database. We find that the SUR copula mixed model produces a smaller bias between predicted and actual reserves than the SUR copula regression model. Moreover, we generate the predictive distribution of the reserves using a modified bootstrap method and show that the SUR copula mixed models provide a larger risk capital gain than the SUR copula regression, indicating a greater diversification benefit.

Moving on to Chapter 4, we combine the SUR copula mixed model with the least absolute shrinkage and selection operator (LASSO) for loss reserving to reduce bias due to too many covariates. We first provide an overview of the LASSO for generalized linear models. Then we discuss the methodologies for loss reserving and predictive distribution estimation, with an emphasis on the sparse SUR copula mixed model approach. We apply and calibrate the sparse SUR copula mixed model using a dataset that includes personal and commercial automobile LOBs from multiple companies.

In Chapter 5, we combine the SUR copula mixed model with the EDT model to capture the dependence between the two LOBs. We first generate predicted losses from EDT for each LOB. Then we obtain the residuals of the predicted loss from EDT. We model the residual heterogeneity between companies and different LOBs using random effects. We estimate the dependence between the LOB through a copula.

Finally, Chapter 6 presents a summary of the thesis and discusses potential directions for future work.

Chapter 2

Recurrent Neural Networks for Multivariate Loss Reserving and Risk Capital Analysis

This chapter is adapted from a paper published by the North American Actuarial Journal (Cai et al., 2025). https://doi.org/10.1080/10920277.2025.2517149

2.1 Introduction

The non-parametric and the parametric approaches are the two primary approaches to modeling the dependence between two LOBs. In the non-parametric approach, the multivariate Mack model (Pröhl and Schmidt, 2005) extends the traditional Mack model to capture dependence across multiple LOBs. The multivariate additive model (Ludwig and Schmidt, 2010) uses flexible, data-driven methods to estimate dependence structures without assuming a specific functional form. In the parametric approach, Shi and Frees (2011) proposes a copula regression model for two LOBs, which links the claims with the same accident and development year with copulas. This model assumes that claims from different triangles with the

same accident year and development year are dependent, called pair-wise dependence. Moreover, studies have incorporated dependence between LOBs through Gaussian or Hierarchical Archimedean copulas and derived the predictive distribution of the reserve, the reserve ranges, and risk capital (Abdallah et al., 2015; Shi et al., 2012; De Jong, 2012). However, copula regression is limited in its flexibility in modeling the marginal distribution and does not account for time dependence in the incremental paid losses.

Various machine learning techniques have recently been developed in micro-level loss reserving for a single LOB. These methods are either tree-based learning methods or neural networks. The tree-based method is based on recursively splitting the claims into more homogeneous groups to predict the number of payments (Wüthrich, 2018a). It can include numerical and categorical attributes of the claimant, such as type of injury and payment history, as predictors. Moreover, Duval and Pigeon (2019) uses a gradient boosting algorithm with a regression tree as the base learner for loss reserving. Gabrielli et al. (2018) proposes separate over-dispersed Poisson models for claim counts and claim sizes embedded in neural network architecture. Neural networks are used as a boosting mechanism to learn the model structure.

In the context of neural networks, Mulquiney (2006) explored their use for predicting claim sizes, finding better performance compared to generalized linear models. Wüthrich (2018b) extended Mack's Chain-Ladder method using neural networks for individual claim reserving, modeling development year ratios with claim features but without prediction uncertainty. Taylor (2019) noted that neural networks can capture interactions between covariates with minimal feature selection, though at the cost of interpretability and prediction accuracy.

Machine learning techniques are increasingly used in loss reserving; however, few models have been developed to capture dependence across LOBs using recurrent neural networks (RNN), as noted by Cossette and Pigeon (2021). Kuo (2019)

introduced the Deep Triangle (DT) multitask learning framework for a single LOB, leveraging gated recurrent units (GRU) to model incremental paid losses and outstanding claims. DT is under-explored for reserve prediction for multiple LOBs. Moreover, the predictive distribution of the reserve and risk capital analysis for the DT is not straightforward and understudied in the current literature. In this chapter, we utilize the multi-task DT model for multivariate loss reserving and introduce an asymmetric loss function to reflect the volatility in the paid losses in different LOBs. Further, we propose to use GAN to generate the predictive distribution of reserves, which allows us to conduct risk capital analysis. Thus, the summary of our contributions is as follows

1. We propose an asymmetric loss function for DT, an unequal weighting scheme that uses the inverse of the standard deviation of the incremental paid losses in the sequence from each LOB to weight the prediction task, reflecting the volatility in the paid losses of that LOB.

2. We introduce a GAN-based technique to generate the predictive distribution for loss reserves. Specifically, we utilize conditional tabular GAN (CTGAN) and CopulaGAN to create synthetic loss triangles (Goodfellow et al., 2014; Patki et al., 2016; Xu et al., 2019; Cote et al., 2020). By integrating these approaches with DT, we propose two models: DT-CTGAN and DT-CopulaGAN, collectively referred to as the Extended Deep Triangle (EDT).

3. We investigate the optimal input sequence length for DT. Since accident years have varying lengths of development years, we examine the effect of different input sequence lengths and find that longer sequences generally yield improved performance.

4. We implement pre-trained model weight initialization to train DT on the GAN-generated samples (thousands of samples), thereby reducing the computational time in generating the predictive distribution of loss reserves.

Next, we describe the Deep Triangle model, which uses GRU to capture the complex dependence between two LOBs.

2.2 Methods

2.2.1 Deep Triangle Architecture for Multivariate Sequence Prediction

The DT framework of Kuo (2019) utilizes a multi-task framework to stabilize training. While it can be trivially extended to accept multivariate inputs, the differing volatilities of paid losses in the multiple-LOB setting necessitates a more nuanced approach. Hence, we propose moving away from the multi-task framework as a stabilizing mechanism, instead replacing it with a loss function which is asymmetric across the various LOBs.

Figure 2.1 illustrates the architecture of the DT model. We employ a vector sequence-to-sequence architecture to model the time series of incremental paid losses, effectively capturing both the pairwise dependence between two LOBs and the temporal dependence of incremental paid losses within each accident year (Sutskever et al., 2014; Srivastava et al., 2015). To our knowledge, this approach has not been previously explored in multivariate loss reserving analysis. As depicted in Figure 2.1, consider the i^{th} accident year and j^{th} development year. The input sequence is the pair of vectors: $\left(Y_{i,1}^{(1)}, Y_{i,2}^{(1)}, \ldots, Y_{i,j-1}^{(1)}\right)$ and $\left(Y_{i,1}^{(2)}, Y_{i,2}^{(2)}, \ldots, Y_{i,j-1}^{(2)}\right)$. The corresponding output sequence is the pair of vectors: $\left(Y_{i,j}^{(1)}, Y_{i,j+1}^{(1)}, \ldots, Y_{i,I}^{(1)}\right)$ and $\left(Y_{i,j}^{(2)}, Y_{i,j+1}^{(2)}, \ldots, Y_{i,I}^{(2)}\right)$. We predict $I-j+1$ time steps into the future for the j^{th} development year, resulting in an output sequence length of $I-1$. Note that we assume the standardized incremental paid loss is independent across accident years. Since there is only one value for the last accident year, we do not use that incremental paid loss for training.

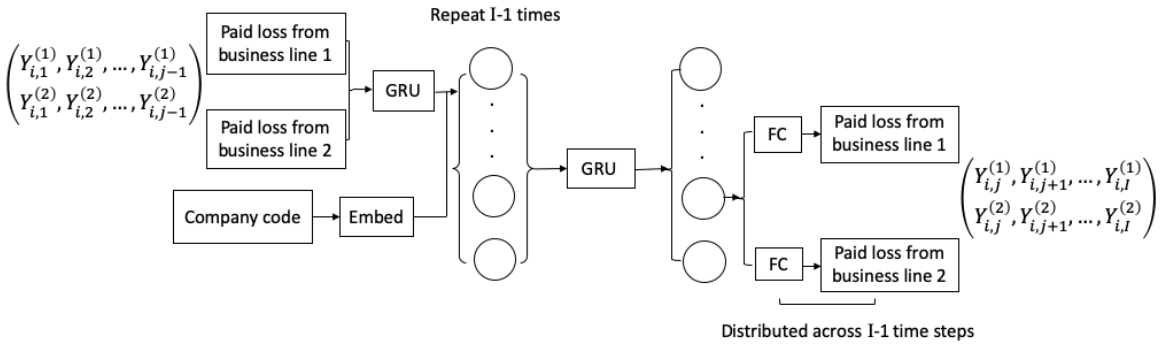

Figure 2.1: The DT architecture for multivariate sequence prediction.

DT uses GRU to handle the time series of incremental paid losses for each accident year i over development year j. GRUs are preferred over Long Short-Term Memory (LSTM) networks due to their fewer training parameters and faster execution (Goodfellow et al., 2016). The GRU processes each element in the input sequence vector and includes mechanisms to determine when a hidden state should be updated or reset at each time step. For each input sequence and for the current time step $\nu; \nu = 1, \ldots, j-1$, the input to the GRU is $q_\nu = (Y_{i\nu}^{(1)}, Y_{i\nu}^{(2)})$ along with the previous time step's hidden state $h_{\nu-1} = (h_{\nu-1}^{(1)}, h_{\nu-1}^{(2)})$. The GRU outputs reset and update gates, r_ν and z_ν respectively, which take values between 0 and 1.

The reset gate r_ν and update gate z_ν for time step ν are computed as follows:

$$r_\nu = \sigma(W_{re}[h_{\nu-1}, q_\nu] + b_{re}), \tag{2.1}$$

and

$$z_\nu = \sigma(W_z[h_{\nu-1}, q_\nu] + b_z), \tag{2.2}$$

where W_{re} and W_z are weight parameters, b_{re} and b_z are biases and h_ν is the hidden state value at ν. The weights and bias parameters are learned during training. In (2.1) and (2.2), the sigmoid function $\sigma(.)$ is used to transform input values to the interval $(0, 1)$. The candidate's hidden state at ν is of the form

$$\tilde{h}_\nu = \tanh(W_{hi}[r_\nu h_{\nu-1}, q_\nu] + b_{hi}). \tag{2.3}$$

The update gate z_ν determines the extent to which the new state h_ν is just the old state $h_{\nu-1}$ and how much of the new candidate state $\tilde{h}_\nu$ is used. The final update equation for the GRU is as follows:

$$h_\nu = z_\nu \tilde{h}_\nu + (1 - z_\nu) h_{\nu-1}. \tag{2.4}$$

When the update gate z_ν is close to 0, the information from q_ν is ignored, skipping time step ν in the dependency chain. However, when z_ν is close to 1, the new state h_ν approaches the candidate state $\tilde{h}_\nu$. These designs help better capture sequence dependencies for $\left(Y_{i,1}^{(1)}, Y_{i,2}^{(1)}, \ldots, Y_{i,j-1}^{(1)}\right)$ and $\left(Y_{i,1}^{(2)}, Y_{i,2}^{(2)}, \ldots, Y_{i,j-1}^{(2)}\right)$. The outputs of the decoder GRU are then passed to two sub-networks of fully connected layers, which correspond to LOB 1 and LOB 2. Each consists of a hidden layer of 64 units, followed by an output layer of 1 unit representing the incremental paid loss at a time step ν. The final output sequences are denoted by $(\hat{Y}_{i,j}^{(1)}, \hat{Y}_{i,j+1}^{(1)}, \ldots, \hat{Y}_{i,I}^{(1)})$ and $(\hat{Y}_{i,j}^{(2)}, \hat{Y}_{i,j+1}^{(2)}, \ldots, \hat{Y}_{i,I}^{(2)})$.

To enhance the robustness and generalizability of our model, we utilize data from multiple companies to train the DT model. We use c_{ij} to denote the company code associated with $Y_{ij}^{(\ell)}$, which is processed through an embedding layer. This layer converts each company code into a fixed-length vector, where the length is a predetermined hyperparameter. In our implementation, we set the length as $C-1$, the number of companies minus one. This embedding process is an integral component of the neural network and is trained with the network itself rather than as a separate pre-processing step. Consequently, companies with similar characteristics are mapped to vectors that exhibit proximity regarding Euclidean distance.

2.2.2 Learning/Validation

Training/Testing Setup

The input for the training sample associated with accident year i ($1 \leq i \leq I-1$) and development year j ($2 \leq j \leq I+1-i$) are the sequences (mask,..., mask, $Y_{i,1}^{(1)}$, $Y_{i,2}^{(1)}$, ..., $Y_{i,j-1}^{(1)}$) and (mask,..., mask, $Y_{i,1}^{(2)}$, $Y_{i,2}^{(2)}$, ..., $Y_{i,j-1}^{(2)}$). The assumption is that $Y_{i,j}^{(1)}$ and $Y_{i,j}^{(2)}$ are predicted using the past $I-1$ time steps. While RNNs can handle variable-length sequences, in practice, we use masks to fix sequence lengths for efficient batch processing. Note that there is no historical data before development year 1. Thus, we use a mask value where $j < 1$ and $j > I$. Masking selectively ignores certain parts of the sequences during training. If the value at a timestep is equal to the mask value, that timestep is skipped in subsequent calculations, including the computation of the loss for backpropagation.

The output for the training sample associated with accident year i ($1 \leq i \leq I-1$) and development year j ($2 \leq j \leq I+1-i$) are the sequences ($Y_{i,j}^{(1)}$, $Y_{i,j+1}^{(1)}$, ..., $Y_{i,I+1-i}^{(1)}$, mask, ..., mask) and ($Y_{i,j}^{(2)}$, $Y_{i,j+1}^{(2)}$, ..., $Y_{i,I+1-i}^{(2)}$, mask, ..., mask). Note that the output sequences also consist of $I-1$ time steps. We use a mask value because we do not have the lower part of the triangle.

The training data is randomly split into training and validation sets using an 80-20 split. When splitting, the training data corresponding to the same accident year and development year from different companies stay in the same training or validation sets. We train the DT model for a maximum of 1000 epochs, employing an early stopping scheme. If the loss on the validation set does not improve over a 100-epoch window, we stop training and keep the weights on the epoch with the lowest validation loss. In the DT, we initialize the neural networks with random weights using the He initialization technique (He et al., 2015), recommended for ReLU activation function (Murphy, 2022).

Next, we predict future incremental paid loss with the trained and validated

DT and obtain a point estimate of the reserve. The input for the testing sample associated with accident year i ($2 \leq i \leq I$) and development year j ($j = I + 2 - i$) are the sequences (mask, ..., mask,$Y_{i,1}^{(1)}$, $Y_{i,2}^{(1)}$, ..., $Y_{i,I+1-i}^{(1)}$) and (mask, ..., mask, $Y_{i,1}^{(2)}$, $Y_{i,2}^{(2)}$, ..., $Y_{i,I+1-i}^{(2)}$). There are $I-1$ testing samples whose accident year and development year satisfy $i+j = I+2$ ($2 \leq i \leq I$). For accident year 1, we have all the data from development year 1 to development year I. The input sequences for testing also consist of $I-1$ time steps. At each accident year and development year for which we have data, we predict future incremental paid loss ($\hat{Y}_{i,I+2-i}^{(1)}$, $\hat{Y}_{i,I+3-i}^{(1)}$, ..., $\hat{Y}_{i,I}^{(1)}$) and ($\hat{Y}_{i,I+2-i}^{(2)}$, $\hat{Y}_{i,I+3-i}^{(2)}$, ..., $\hat{Y}_{i,I}^{(2)}$). Next, we obtain a point estimate of the outstanding claims for each LOB by $R^{(\ell)} = \sum_{i=2}^{I} \sum_{j=I+2-i}^{I} \omega_i^{(\ell)} \hat{Y}_{ij}^{(\ell)}$.

Weighted Loss Function

For the DT model, the loss function is the average over the predicted time steps of the mean squared error of predictions. For each output sequence ($\hat{Y}_{i,j}^{(1)}$, $\hat{Y}_{i,j+1}^{(1)}$, ..., $\hat{Y}_{i,I+1-i}^{(1)}$, mask, ..., mask) and ($\hat{Y}_{i,j}^{(2)}$, $\hat{Y}_{i,j+1}^{(2)}$, ..., $\hat{Y}_{i,I+1-i}^{(2)}$, mask, ..., mask), the symmetric loss is defined as

$$\frac{1}{I-i+1-(j-1)} \sum_{\nu=j}^{I+1-i} \frac{(\hat{Y}_{i,\nu}^{(1)} - Y_{i,\nu}^{(1)})^2 + (\hat{Y}_{i,\nu}^{(2)} - Y_{i,\nu}^{(2)})^2}{2}. \tag{2.5}$$

We define an asymmetric loss function as

$$\frac{1}{I-i+1-(j-1)} \sum_{\nu=j}^{I+1-i} \frac{1}{2(\sigma_{i,j}^{(1)})^2} (\hat{Y}_{i,\nu}^{(1)} - Y_{i,\nu}^{(1)})^2 + \frac{1}{2(\sigma_{i,j}^{(2)})^2} (\hat{Y}_{i,\nu}^{(2)} - Y_{i,\nu}^{(2)})^2, \tag{2.6}$$

where $(\sigma_{i,j}^{(1)})^2$ and $(\sigma_{i,j}^{(2)})^2$ are variances for sequences ($Y_{i,j}^{(1)}$, $Y_{i,j+1}^{(1)}$, ..., $Y_{i,I+1-i}^{(1)}$, mask, ..., mask) and ($Y_{i,j}^{(2)}$, $Y_{i,j+1}^{(2)}$, ..., $Y_{i,I+1-i}^{(2)}$, mask, ..., mask), respectively. The volatilities in the paid losses are different between the two LOBs, and we use uncertainty-based weighting to balance the two prediction tasks. When calculating

the variances $\left(\sigma_{i,j}^{(1)}\right)^2$ and $\left(\sigma_{i,j}^{(2)}\right)^2$, we exclude the mask value from the sequences.

To optimize the parameters for the DT model, we employ the AMSGRAD method (Reddi et al., 2018), which is a variant of the Adaptive Moment Estimation (ADAM) algorithm. AMSGRAD is chosen specifically due to its ability to manage the high variability in gradients that arise from the small size of the training sample, a common issue in stochastic gradient descent (SGD) methods. AMSGRAD addresses this by incorporating the gradients' moment into the parameter update process, thus offering a more stable and effective optimization in scenarios with limited data.

Once total loss reserves are estimated using the DT model, our approach includes generating the predictive distribution of total loss reserves.

2.2.3 Predictive Distribution of the Total Reserve

We adapt Generative Adversarial Nets (GANs), as introduced by Goodfellow et al. (2014), to generate synthetic loss triangles and to generate the predictive distribution of the total reserve. This approach provides a novel way of applying advanced machine learning methods to the traditional actuarial problem of reserve estimation. GAN generates new data based on learned distributions from the original data. Bootstrap resamples the data with replacement to create multiple simulated data. Bootstrap may not sufficiently capture the underlying data distribution, especially in complex scenarios.

GAN simultaneously trains two models: a generative model, G, which captures the data distribution and generates new data, and a discriminative model, D, which outputs the probability of how likely the generated data belongs to the training data. Figure 2.2 shows the relationship between the generator G and the discriminator D. The generator G generates realistic samples while the discriminator D distinguishes between genuine and counterfeit samples. The generator G

takes some noise z as input and outputs a synthetic sample.

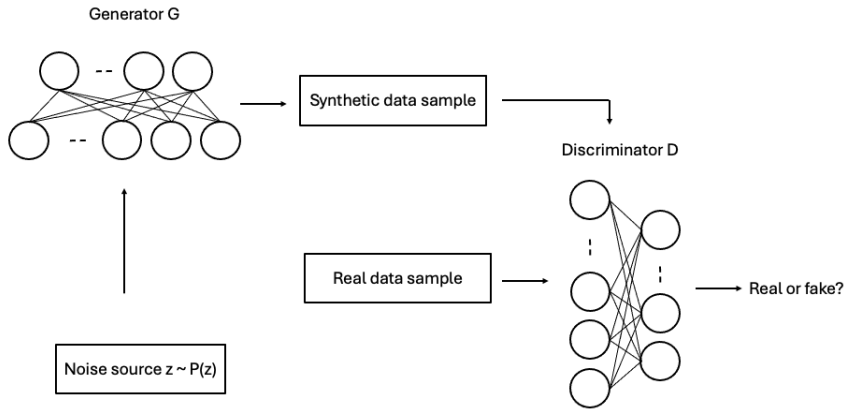

Figure 2.2: The architecture of GANs.

Generating New Samples

The traditional GAN model utilizes the latent variable z sampled from a standard multivariate normal distribution. However, for our purposes, we require a GAN approach that can capture the pairwise dependence between two LOBs and the sequential structure inherent in the standardized incremental paid losses. To address this, we utilize the GAN method for tabular data generation, considering the pairwise dependency and the sequential structure.

To generate synthetic data from loss triangles, we employ conditional tabular GAN (CTGAN), a GAN variant (Xu et al., 2019), which is adept at modeling dependencies in data. For CTGAN, numerical inputs $Y_{ij}^{(\ell)}$ are normalized to fit within the (-1, 1) range using mode-specific normalization. Each $Y_{ij}^{(\ell)}$ is represented as a one-hot vector $\beta_{ij}^{(\ell)}$, indicating the mode, and a scalar $\alpha_{ij}^{(\ell)}$, indicating the value within the mode. The company code categorical variable c_{ij} is represented as a one-hot vector d_{ij}.

CTGAN generates data conditioned on additional information by combining random noise z sampled from a standard normal vector with a condition (such as company code c_{ij}). This is done by concatenating the noise z and the con-

dition and then passing the combined input through the generator network G, which learns to generate synthetic data that satisfies the given condition. Additionally, to maintain the sequential integrity of the data, we generate synthetic data for each development year separately. This approach ensures that the sequential properties of standardized incremental paid losses are preserved, allowing for more accurate and realistic simulation of loss triangles across multiple LOBs. Generating a synthetic loss triangle involves a three-stage process:

(1) Combination of Data: For each development year j ($1 \leq j \leq I$), we combine $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ of all accident years i ($1 \leq i \leq I$) from all companies into one table. The first column of the table is $Y_{ij}^{(1)}$ of all accident years from the personal auto line. The predicted loss from the DT model is used if the $Y_{ij}^{(1)}$ is not available. Similarly, the data from the commercial auto line is used for the second column. The third column of the table is the company code.

(2) GAN Model Training: We train a GAN model for each development year j using the combined data from (1). The representation r_{ij} of a row in the combined table is the concatenation of the three columns: $r_{ij} = \alpha_{ij}^{(1)} \oplus \beta_{ij}^{(1)} \oplus \alpha_{ij}^{(2)} \oplus \beta_{ij}^{(2)} \oplus d_{ij}$. This training enables the GAN model to learn the underlying distribution of the combined data in (1).

(3) Sampling and Loss Triangle Formation: After training, we use the GAN to sample I new rows for each development year j for each company. These sampled data are then sequentially arranged according to their development years. We remove the lower triangle to form a new loss triangle, ensuring the structure aligns with the loss triangle format.

Note that the loss triangles from the two LOBs are on different scales, presenting a challenge for effective modeling. To address this, we employ a CopulaGAN (Patki et al., 2016), which leverages the scale-invariant property of copulas to define the covariance of z. In addition to the marginal distributions, CopulaGAN uses a

Gaussian copula. CopulaGAN is a variation of the CTGAN model that leverages the CDF-based transformation applied by the Gaussian Copula, making it easier for the underlying CTGAN model to fit the data. For computing, we use the SDV (Patki et al., 2016) library to build GAN models by constructing the input as follows.

(1) Let the marginal CDFs of columns $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ be F_1 and F_2, respectively.

(2) Go through the table row-by-row. Each row is denoted as $y = \left(Y_{ij}^{(1)}, Y_{ij}^{(2)}\right)$.

(3) Transform each row using the inverse probability transform:

$$z = \left[\Phi^{-1}\left(F_1\left(Y_{ij}^{(1)}\right)\right), \Phi^{-1}\left(F_2\left(Y_{ij}^{(2)}\right)\right)\right]$$

where $\Phi^{-1}(\cdot)$ is the inverse CDF of the Gaussian distribution.

(4) After all the rows are transformed, estimate the covariance matrix Σ of the transformed values.

The parameters for each column distribution and the covariance matrix Σ are used in the generative model for that table. The CDF transformed data $F_1(Y_{ij}^{(1)})$ and $F_2(Y_{ij}^{(2)})$ are then fed into the CTGAN architecture. After generating synthetic data in the transformed space, CopulaGAN uses the inverse CDF to bring the data back to the original space. For the new table generated using CopulaGAN, we consider only the upper loss triangle as a new sample.

Initializing Weights for the DT

Next, we introduce a new weight initialization mechanism to speed up the training of DT for the newly generated samples. In order to generate the predictive distribution of the reserve, we apply the DT model to the loss triangles generated as in Subsection 2.3.1. For each set of newly generated loss triangles, we obtain a

point estimate of the reserves. We repeat this procedure many times to construct a predictive distribution for the reserve. Given the high computational cost arising from the need to train DTs for each generated sample, we implement a more efficient approach. We propose to leverage a pre-trained DT model from the real data to fine-tune weights for new samples, instead of training DTs using random weight initialization for every generated sample.

Subsequently, the risk capital gain is calculated using the methodology outlined in (1.2). The proposed EDT method, which includes DT-CTGAN or DT-CopulaGAN, integrates the predictive capabilities of the DT model with the distributional insights provided by GAN, offering a comprehensive view of the reserve estimation and its associated risks.

2.3 Applications

2.3.1 Data Description

In this section, we demonstrate the proposed EDT approach on a real dataset. We use 30 companies' loss triangles from Schedule P of the National Association of Insurance Commissioners (NAIC) database (Meyers and Shi, 2011) to illustrate and compare the EDT. Each pair comprises two loss triangles from the personal and commercial auto LOBs and is associated with a company code. Each triangle contains incremental paid losses for accident years 1988-1997 and ten development years. Here, we demonstrate prediction and risk capital analysis for a major US property-casualty insurer. Table 2.1 and Table 2.2 show the incremental paid losses for this company's personal and commercial auto LOBs.

Table 2.1: Incremental paid losses ($X_{ij}^{(1)}$) for personal auto LOB.

year	premium	1	2	3	4	5	6	7	8	9	10
1988	4 711 333	1 376 384	1 211 168	535 883	313 790	168 142	79 972	39 235	15 030	10 865	4 086
1989	5 335 525	1 576 278	1 437 150	652 445	342 694	188 799	76 956	35 042	17 089	12 507	
1990	5 947 504	1 763 277	1 540 231	678 959	364 199	177 108	78 169	47 391	25 288		
1991	6 354 197	1 779 698	1 498 531	661 401	321 434	162 578	84 581	53 449			
1992	6 738 172	1 843 224	1 573 604	613 095	299 473	176 842	106 296				
1993	7 079 444	1 962 385	1 520 298	581 932	347 434	238 375					
1994	7 254 832	2 033 371	1 430 541	633 500	432 257						
1995	7 739 379	2 072 061	1 458 541	727 098							
1996	8 154 065	2 210 754	1 517 501								
1997	8 435 918	2 206 886									

Table 2.2: Incremental paid losses ($X_{ij}^{(2)}$) for commercial auto LOB.

year	premium	1	2	3	4	5	6	7	8	9	10
1988	267 666	33 810	45 318	46 549	35 206	23 360	12 502	6 602	3 373	2 373	778
1989	274 526	37 663	51 771	40 998	29 496	12 669	11 204	5 785	4 220	1 910	
1990	268 161	40 630	56 318	56 182	32 473	15 828	8 409	7 120	1 125		
1991	276 821	40 475	49 697	39 313	24 044	13 156	12 595	2 908			
1992	270 214	37 127	50 983	34 154	25 455	19 421	5 728				
1993	280 568	41 125	53 302	40 289	39 912	6 650					
1994	344 915	57 515	67 881	86 734	18 109						
1995	371 139	61 553	132 208	20 923							
1996	323 753	112 103	33 250								
1997	221 448	37 554									

2.3.2 Prediction of Total Reserve

First, we apply the DT model to the losses in the two LOBs from 30 companies. In the pairwise training sample, the first component is the incremental paid loss from the personal LOB, and the second component corresponds to the incremental paid loss from the commercial LOB.

To train the DT, we consider the incremental paid losses up to 1997, which is the current calendar year for this dataset. We use the lower part of the loss triangle to evaluate the EDT's predictive performance. In particular, we compare the percentage errors of actual and predicted loss reserves to evaluate the performance of different models. We predict the reserve using DT with both the symmetric loss function as in (2.5) and the asymmetric loss function as in (2.6). Table 2.4 displays the predicted reserves from DT alongside the actual reserves. In terms of the percentage error from the actual reserve, DT with the asymmetric loss function generates a more accurate estimation of the reserve, which is shown in Table 2.5.

As demonstrated in Table 2.5, introducing the asymmetric loss function leads to a substantial reduction in bias for the reserve of the commercial LOB, a sector characterized by more volatile incremental paid losses.

Next, we apply the copula regression model to the two loss triangles from the major US property-casualty insurer. For the marginal distribution, we use the log-normal and the gamma distributions for personal and commercial LOBs (Shi and Frees, 2011), respectively. We consider the systematic component $\eta_{ij} = \mu_{ij}$ for the log-normal distribution with location parameter μ_{ij} and shape parameter σ. For the gamma distribution with location parameter μ_{ij} and shape parameter ϕ, we use $\eta_{ij} = \log(\mu_{ij}\phi)$.

We use the Gaussian and Frank copulas to model the dependence between the two LOBs. These copula functions are specified using the R package `copula` (Hofert et al., 2020). The `gjrm` function from the R package `GJRM` estimates the copula regression model (Marra and Radice, 2023). The log-likelihood, Akaike information criterion (AIC), and Bayesian information criterion (BIC) for all copula models are provided in Table 2.3. According to Table 2.4, independence and copula models generate comparable point estimates for the total reserve, about 7 million dollars.

Table 2.3: Summary statistics for fitted copula regression models with product copula, Gaussian copula, and Frank copula.

	Copula		
	Product	Gaussian	Frank
Dependence Parameter (θ)	.	-0.3656	-2.7977
Log-Likelihood	346.6	350.4	350.3
AIC	-613.2	-618.9	-618.5
BIC	-505.2	-508.2	-507.8

In addition to DT, Table 3 shows the percentage error of prediction to the actual reserve for the copula regression model. Interestingly, the DT model provides a more accurate point estimation of the reserve for personal and commercial auto LOBs. This improved performance can be attributed to the neural network's

Table 2.4: Point estimates of the reserves from DT and copula regression models.

Model	Reserves		
	LoB 1, R_1	LoB 2, R_2	Total, R
DT (symmetric)	7 756 417	327 517	8 083 934
DT (asymmetric)	7 781 299	324 024	8 105 323
Product Copula	6 464 083	490 653	6 954 736
Gaussian Copula	6 423 246	495 925	6 919 171
Frank Copula	6 511 360	487 893	6 999 253
Actual Reserve	8 086 094	318 380	8 404 474

ability to learn complex non-linear relationships of incremental paid losses between LOBs and within accident years (Murphy, 2022).

Table 2.5: Performance comparison using percentage error of actual and estimated loss reserve.

LOB	DT (symmetric)	DT (asymmetric)	Product Copula	Gaussian Copula	Frank Copula
Personal Auto	-4.1%	**-3.8%**	-20.1%	-20.6%	-19.5%
Commercial Auto	2.9%	**1.8%**	54.1%	55.8%	53.2%
Total	-3.8%	**-3.6%**	-17.3%	-17.7%	-16.7%

Note: The best metric for each LOB is in bold.

For a fair comparison between copula regression and DT, we also utilize 30 companies' data for copula regression. The results in Appendix A.2 show similar results to those with one company. However, the company heterogeneity should be modeled as a random effect, and to our knowledge, there is no software implementation of such a model to be used in this analysis.

2.3.3 Predictive Distribution of Total Reserve

First, we obtain the marginal predictive distribution of reserves to evaluate diversification benefits. Assuming log-normal and gamma marginal distributions for personal and commercial LOBs, respectively, we use parametric bootstrapping to generate predictive reserve distributions separately for each LOB. Figure 2.3 illustrates that the personal LOB exhibits heavier tails compared to the commercial LOB. To capture the benefits of risk diversification for this data, we generate the predictive distributions of the total reserves.

To obtain the predictive distribution of the reserve for the DT model, we first

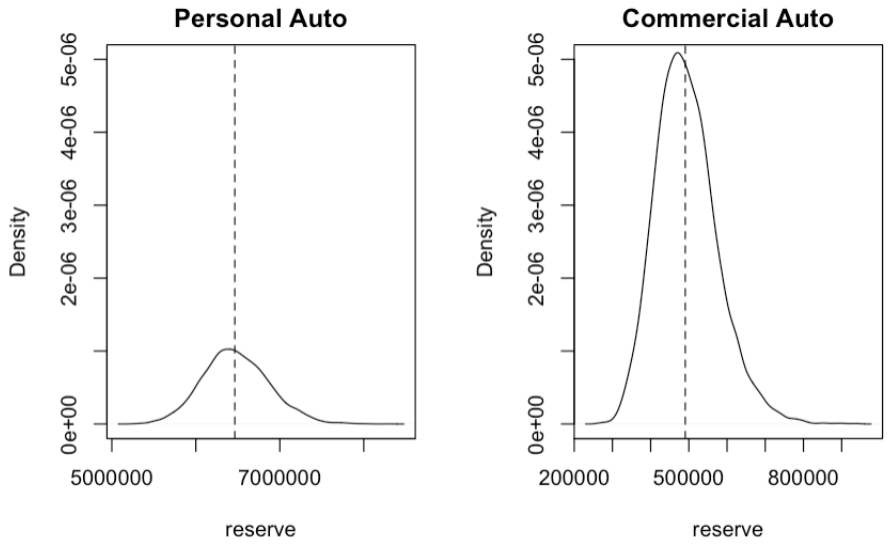

Figure 2.3: Marginal predictive distributions of the reserves from parametric bootstrapping.
Note: The vertical dotted lines indicate the maximum likelihood estimates of the reserve from the observed data.

utilize the CTGAN model. We train the CTGAN model with data from 30 companies for each development year. New incremental paid losses are generated for each of the ten development years. The newly generated data for each development year are then stacked in the order of development years to form new upper loss triangles. The DT model estimates reserves for these newly created loss triangles. This process is repeated multiple times to construct a predictive reserve distribution. In addition to CTGAN, CopulaGAN is employed to create new upper triangles to generate the predictive distribution of the reserve. Moreover, we use block bootstrapping to generate the predictive distribution of the reserve, with full details provided in Appendix A.3.

To reduce the computational expense associated with the EDT, stemming from training numerous DTs for GAN samples, we leverage the trained model on the observed data to fine-tune weights for new samples. For each newly generated sample, DT takes two minutes to run when trained from a random weight initialization, and about one minute when trained from a saved model that was previously fitted

on real data. It takes about 4 hours to obtain the predictive distribution using parallel computing with 32 CPUs.

For copula regression models, we used the parametric bootstrap to generate the predictive distributions of the reserves. Table 3.5 shows the estimated reserve, bias, and standard errors for the different models. The percentage bias is computed as the percentage error between the bootstrap mean reserves and estimated loss reserves. It has been observed that the standard deviations from the DT-GAN are smaller than those from the copula regression models, and the biases are slightly higher, likely due to heterogeneity across companies and between LOBs. This is also corroborated in Table 2.5, which shows positive and negative bias for the commercial and personal LOB, respectively. CTGAN uses Gaussian mixtures to model distributions when generating synthetic loss triangles (Xu et al., 2019). These modeling assumptions in GANs introduce a certain bias compared to the bootstrapping method. However, the DT-GAN models, by capturing the inter-LOB dependence, generate a smaller risk capital than the copula regression models, as discussed in the next subsection. The coefficient of variation (CV) can be used to measure the risks when the insurance company has more than one LOB. Based on the CV in Table 3.5, all the copula regression models and the EDT have CVs smaller than one, which complies with the insurance standards. However, the EDT stands out because its CVs are consistently smaller than the copula regression models.

Table 2.6: Bias, Standard deviation, Coefficient of variation (CV)

	Reserve	Bootstrap mean reserve	Bias	Std. dev.	CV
DT-CTGAN	8 105 323	8 261 718	1.93%	197 465	0.024
DT-CopulaGAN	8 105 323	8 255 638	1.85%	196 791	0.023
Product Copula	6 954 736	6 972 792	0.26%	399 758	0.057
Gaussian Copula	6 919 171	6 941 806	0.33%	368 555	0.053
Frank Copula	6 999 253	7 043 309	0.63%	388 357	0.056

We compare the predictive distributions of the reserves for the EDT and the

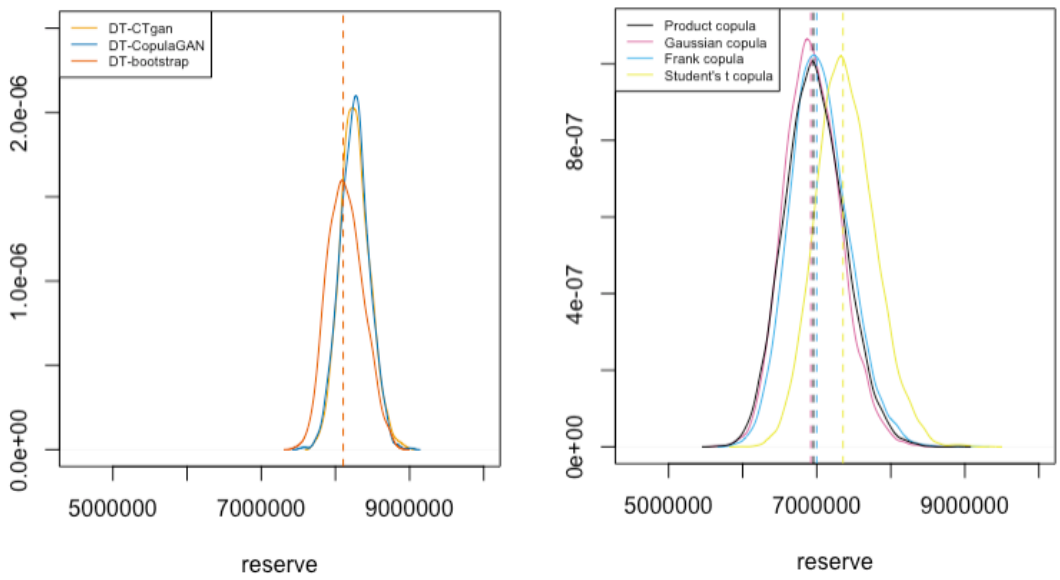

Figure 2.4: Predictive distributions of total reserves from the EDT and copula regression.
Note: The vertical dotted lines indicate the estimated reserves for each model.

copula regression in Figure 2.4, which shows that the bootstrap mean reserves of these models are pretty similar. In summary, for this insurer, the dependence between triangles only results in small changes in the point estimates of reserves. Though the point estimates are similar, the dependencies between LOBs affect the reserve's predictive distribution, which helps diversify risk within the portfolio.

Using the predictive distributions we generated, Table 2.7 and Figure 2.5 showcase the 95% confidence interval of the total reserve, where the lower bound is the 2.5th percentile of the predictive distribution and the upper bound is the 97.5th percentile of the predictive distribution. We observe that the EDT models generate the narrowest confidence intervals.

Table 2.7: 95% confidence intervals for the total reserve using the predictive distribution.

	Lower bound	Upper bound
DT-CTGAN	7 900 272	8 683 653
DT-CopulaGAN	7 875 828	8 653 414
Product Copula	6 241 016	7 756 950
Gaussian Copula	6 280 339	7 715 924
Frank Copula	6 315 438	7 807 835

Figure 2.5: 95% confidence interval for the total reserves for different models.

2.3.4 Risk Capital Calculation

In actuarial practice, a common method to calculate the risk measure for the entire portfolio, which here includes both personal auto and commercial auto LOBs, is the "silo" method, as in the Methods section. We denote "Silo-GLM" and "Silo-DT" as the aggregate of risk measures derived from copula regression models and EDT model, respectively.

We compared these silo methods with the risk measures calculated from both the EDT and copula regression models to evaluate the benefits of diversification. Table 3.6 indicates that the risk measures from both the EDT and copula regression models are lower than their respective silo totals. This might be due to the fact that there is a negative association between the incremental paid losses of the two LOBs. Table 2.3 and Appendix A.1 show negative dependence through the copula parameter estimates of the copula regression and Kendall's tau for the dependence, respectively. The dependence estimated by the Student's t copula regression is not significant because its confidence interval includes zero. All these results suggest that incorporating the interdependencies between LOBs into the risk measurement process can yield lower risk estimates, highlighting the potential value of these more integrated approaches in risk management.

Next, we calculate risk capital as defined in (1.1) from the predictive distribu-

Table 2.8: Risk capital estimation for different methods.

Risk measure	TVaR (60%)	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
Silo-DT	8 105 434	8 217 523	8 258 835	8 317 629	8 399 375	8 558 981
DT-CTGAN	8 452 870	8 549 295	8 582 802	8 627 247	8 699 935	8 846 865
DT-CopulaGAN	8 443 684	8 532 939	8 564 483	8 606 020	8 666 096	8 789 649
Silo-GLM	7 442 692	7 671 633	7 756 992	7 872 138	8 060 489	8 460 435
Product copula	7 367 695	7 553 768	7 621 203	7 710 435	7 847 773	8 126 433
Gaussian copula	7 313 951	7 490 387	7 556 029	7 644 886	7 782 646	8 054 737
Frank copula	7 424 807	7 616 405	7 685 514	7 776 754	7 921 574	8 202 695
Risk capital						
Silo-DT		112 089	153 401	212 195	293 941	453 547
DT-CTGAN		96 425	129 932	174 377	247 065	393 995
DT-CopulaGAN		89 255	120 799	162 336	222 412	345 965
Silo-GLM		228 941	314 300	429 446	617 797	1 017 743
Product copula		186 073	253 508	342 740	480 078	758 738
Gaussian copula		176 436	242 078	330 935	468 695	740 786
Frank copula		191 598	260 707	351 947	496 767	777 888

tion of the reserves to set up a buffer from extreme losses. Table 3.6 shows that the risk capital required under the Silo-DT method is less than that of Silo-GLM, suggesting that the EDT is instrumental in reducing the insurer's risk capital. It is noteworthy that while the silo method tends to yield a more conservative estimate of risk capital, both the EDT and copula models lean towards a more aggressive estimation.

Furthermore, when comparing the risk capital from different models, those derived from the EDT are consistently smaller than those from the copula regression models. This is attributable to the EDT's ability to capture pairwise dependencies between the two LOBs and the time dependencies of incremental paid losses. Notably, among all models evaluated, DT-CopulaGAN produces the smallest risk capital. This could be due to its use of flexible marginals, such as truncated Gaussians with varying parameters and a Gaussian copula for capturing dependencies between these marginals. Thus, insurers can leverage the EDT as an effective tool for risk management, particularly in reducing the required risk capital.

Next, we compute the risk capital gain defined in (1.2). Note that the risk capital gains for both the EDT and copula regression models are computed using silo-GLM, which is the industry standard, as the base. Table 3.7 shows that

risk capital gain is large when we capture the association between the two LOBs. Further, the larger risk capital gains are obtained for the EDT models compared to the copula regression models. We can associate these gains with the diversification effect in the insurance portfolio. For example, to better take advantage of the diversification effect, the insurer can increase the volume of the commercial LOB, which is smaller than the personal LOB.

Table 2.9: Risk capital gain for different methods.

Risk capital gain	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
DT-CTGAN vs Silo-GLM	58.36%	58.89%	59.64%	61.06%	61.67%
DT-CopulaGAN vs Silo-GLM	**61.45%**	**61.78%**	**62.42%**	**64.94%**	**66.34%**
Product copula vs Silo-GLM	18.72%	19.34%	20.19%	22.29%	25.45%
Gaussian copula vs Silo-GLM	22.93%	22.98%	22.97%	24.13%	27.21%
Frank copula vs Silo-GLM	16.31%	17.05%	18.05%	19.59%	23.57%

Note: The largest risk capital gain for each risk level is in bold.

2.4 Simulation Case Study

In this section, we further validate our conclusions that the EDT produces reduced risk capital through simulation studies. We simulate pairs of loss triangles of personal and commercial auto LOBs with ten accident and development years for each simulation run. The details of the simulation setup are provided below.

We begin with the estimated copula regression model for the real data in Section 3.4. In this model, we use log-normal and gamma densities for the marginal distributions of standardized incremental paid losses from the personal and commercial LOBs, respectively. To simulate these losses in the loss triangles $\left(Y_{ij}^{(1)}, Y_{ij}^{(2)}\right)$, we first calculate the systematic component $\eta_{ij}^{(\ell)}(\ell = 1, 2)$ from the accident year effect $\alpha_i^{(\ell)}$ (Table A.3) and development year effect $\beta_i^{(\ell)}$ (Table A.4). Next, we simulate $u_{ij}^{(\ell)}(\ell = 1, 2)$ $(i + j - 1 \leq I)$ from Gaussian copula model $c(\cdot; \theta)$ with dependence parameter $\theta = -0.36$. Then, we transform $u_{ij}^{(\ell)}$ to the upper triangles by inverse function $y_{ij}^{(\ell)} = F^{(\ell)(-1)}(u_{ij}^{(\ell)}; \eta_{ij}^{(\ell)}, \gamma^{(\ell)})$, where $\eta_{ij}^{(\ell)} = \xi^{(\ell)} + \alpha_i^{(\ell)} + \beta_j^{(\ell)}(\ell = 1, 2)$. Here, we set the shape parameter $\gamma^{(1)} = 0.089$, as estimated in the copula regres-

sion model, and larger $\gamma^{(2)} = 2$ to account for higher volatility in the commercial LOB. Moreover, $\eta_{ij}^{(\ell)}$ are derived from the marginal distribution parameters as follows. $\eta_{ij} = \mu_{ij}$ for a log-normal distribution with location parameter μ_{ij} and shape parameter $\gamma^{(1)} = \sigma$. For a gamma distribution with location parameter μ_{ij} and shape parameter $\gamma^{(2)} = \phi$, we use the form $\eta_{ij} = \log(\mu_{ij}\phi)$. Finally, the incremental paid losses, $\left(X_{ij}^{(1)}, X_{ij}^{(2)}\right)$ are obtained by multiplying the simulated $y_{ij}^{(\ell)}$ by the premium for the i-th accident year. Note that in Table A.4, most of the development year effects are negative, which indicates that the incremental paid losses are decreasing with development years.

Using the above procedure, we simulate the upper and lower parts of each loss triangle. The sum of the lower triangle represents the actual reserve for each loss triangle. We retain only the upper part of all simulated loss triangles to apply the proposed EDT and compare its results with copula regression models. To reflect the multiple companies of real data, we simulate 50 pairs of loss triangles per simulation run.

For each simulation run, we train the DT model using 50 pairs of loss triangles. We use the symmetric loss function for DT in the simulation study because the simulated data was generated using a Gaussian copula, which assumes symmetric dependencies between loss triangles. Since the data does not exhibit inherent asymmetry, applying an asymmetric loss function in this setting would not provide meaningful improvements and could introduce unnecessary distortions. However, in the real-data analysis of the previous section, we implemented the asymmetric loss function to capture better potential skewness and tail risks observed in empirical loss triangles. The trained model is then used to predict the reserve for each pair. Additionally, we generate predictive distributions of the total reserve using CTGAN and CopulaGAN for each pair of simulated loss triangles.

Through the simulation study, we examined the impact of input and output sequence lengths on the DT model performance. We generated input and output

sequences of varying lengths for each accident year. As shown in Figure 2.6, the cross-validation error was evaluated across different sequence lengths, with a length of nine identified as optimal for the DT model.

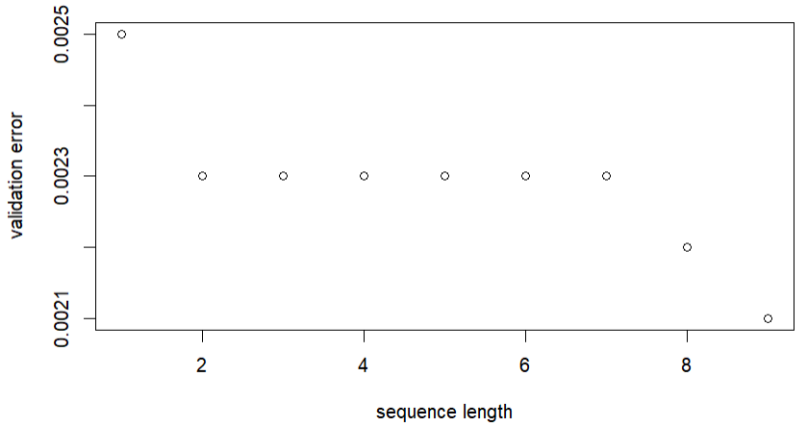

Figure 2.6: Cross-validation errors for different input sequence lengths in the DT.

Next, we apply the copula regression model separately to each pair of 50 simulated loss triangles. We assume log-normal and gamma distributions for the marginals and evaluate different copula structures, including the product copula, Gaussian copula, and Frank copula.

In Table 2.10, we evaluate the performance of all copula regression and DT models in estimating the total reserve. To compute the true reserve $R^{(1)}$ and $R^{(2)}$, we sum up the expected values of the lower triangle $\exp\left(\mu_{ij}^{(1)} + \frac{1}{2}(\sigma)^2\right)$ and $\mu_{ij}^{(2)}\phi$, respectively.

Here, we compare the predicted reserves with the actual reserves for each pair across all models, including copula regression and DT. In each simulation run, we evaluate 50 pairs of predicted and actual reserves. To quantify the prediction error for each LOB ℓ we compute the mean absolute percentage error (MAPE) as defined in (2.7).

$$\text{MAPE}_\ell = \frac{1}{50}\sum_{b=1}^{50}\left|\frac{\hat{R}_b^{(\ell)} - R^{(\ell)}}{R^{(\ell)}}\right|, \tag{2.7}$$

where $\hat{R}_b^{(\ell)}$ is the predicted reserves for b^{th} loss triangle from ℓ^{th} LOB and $R^{(\ell)}$ is

the true reserve for the ℓ^{th} LOB. Table 2.10 shows that the DT model outperforms the copula regression models for both LOBs and is particularly effective for the more volatile commercial auto LOB. The MAPE is almost negligible in personal LOB because the DT has increased flexibility than the copula regression models and captures the time dependence in the sequence input. The MAPE is relatively large for the commercial LOB, which is set to be more volatile than the personal LOB in the simulation setting.

The performance of the copula regression model can be attributed to its limited flexibility in capturing both the sequence dependence and the pairwise dependence. In particular, the model underestimates the shape parameter of the gamma marginal distribution for the commercial LOB, which plays a crucial role in controlling the dispersion and tail behavior of the distribution. This underestimation leads to significant errors in the predicted reserve, especially for the more volatile commercial LOB.

Table 2.10: Performance comparison using the mean absolute percentage error.

LOB	DT	Product Copula	Gaussian Copula	Frank Copula
Personal Auto	**0.63%**	5.28%	5.11%	5.07%
Commercial Auto	**18.67%**	27.28%	31.85%	31.59%

Note: The best metric for each LOB is in bold. The actual reserves for the personal and commercial LOBs are 6 423 246 and 495 925, respectively.

For each simulated loss triangle pair, we use the DT model's predicted full triangle as input to the GAN models. We then apply CTGAN and CopulaGAN to generate 1,000 synthetic loss triangles per pair and use the DT model to predict reserves for each synthetic triangle.

For each of the 50 simulated loss triangle pairs, we construct the predictive distribution of reserves using 1,000 predicted reserves from the corresponding synthetic loss triangles. Based on the EDT models, the corresponding 95% confidence intervals for the total reserve are presented in Figure 2.7, where the horizontal line

represents the true reserve of the simulated loss triangles. Notably, we observe that the true reserve falls within all 95% confidence intervals for all models. Thus, the coverage exceeds the nominal value of 95%. This over-coverage may be due to the EDT relying on the predicted lower triangle from the DT as input for the GAN models, which could lead to conservative uncertainty estimates, or an insufficient number of synthetic loss triangles, limiting the variability captured in the predictive distribution. Among all the EDT models, DT-CopulaGAN produces the narrowest confidence interval. We also expect the coverage of the interval of the GAN to become close to the nominal value when the GAN is modified to accept missing values in the lower triangle.

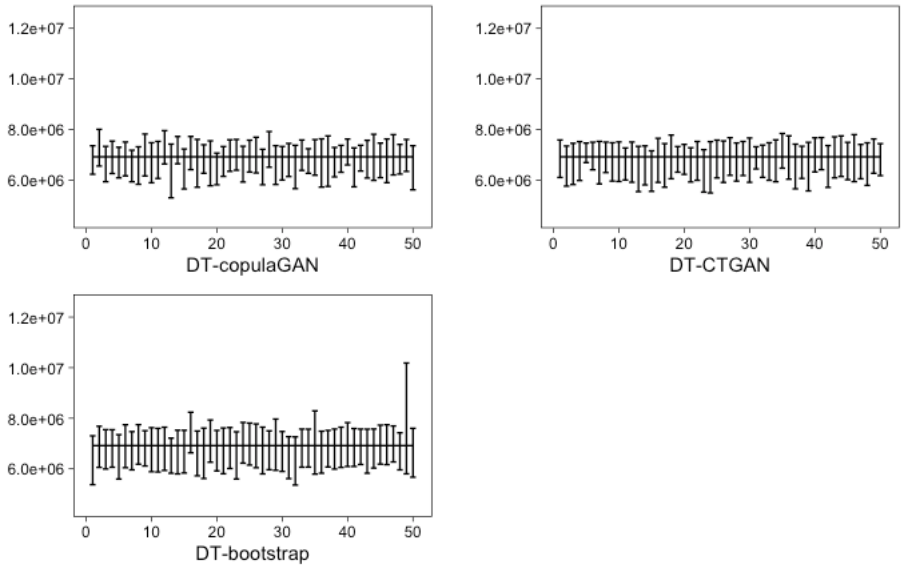

Figure 2.7: 95% confidence interval for total reserves for different EDT models. Note: The horizontal line indicates the true reserve. The true reserve is within all the 95% confidence intervals. The average length of confidence intervals are 1 413 034 and 1 498 851, respectively.

For copula regressions, we generate the predictive distribution of the total reserve for each of the 50 simulated loss triangles. For each copula regression model, we conduct 1000 bootstrap simulations to generate the predictive distribution of the total reserve. We present in Figure 2.8 the 95% confidence interval for the total reserve. We observe that, for all models, the true reserve falls within most of

the 95% confidence intervals. Among them, the product copula regression model provides the highest coverage, approaching the nominal 95%, but at the cost of the widest confidence intervals, indicating greater uncertainty. In contrast, the Gaussian copula regression model yields the narrowest confidence intervals, but with a lower coverage rate of 90%, suggesting it may underestimate reserve variability.

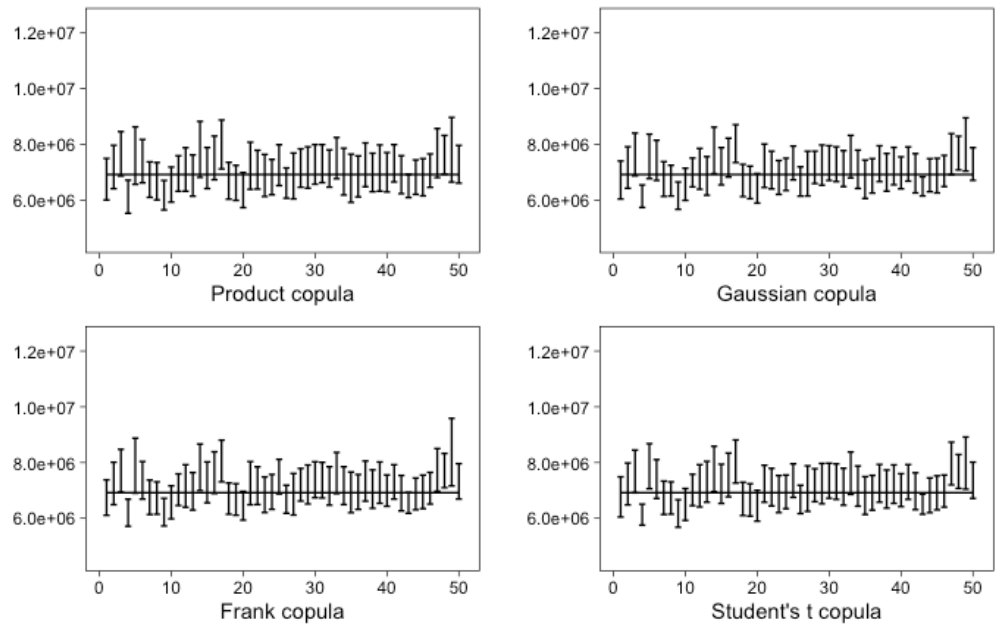

Figure 2.8: 95% confidence interval for the total reserves for different copula models.
Note: The horizontal line indicates the true reserve. The coverage for product copula, Gaussian copula, and Frank copula are 94%, 90%, and 88%, respectively. The average length of confidence intervals are 1 457 826, 1 291 627, and 1 329 655.

Next, we compute the risk measure for all 50 simulated loss triangle pairs based on their predictive reserve distributions. Figure 2.9 illustrates the box plot of the TVaR measure at the 99% confidence level for each pair. The boxplots provide a comprehensive comparison across different models: Silo-GLM, Product copula, Gaussian copula, Frank copula, Silo-DT, DT-CTGAN, and DT-CopulaGAN.

We observe that the median risk measures from all copula regression models are smaller than those from Silo-GLM, while the median risk measures from all the EDT models are smaller than those from Silo-DT. This trend can be attributed to the negative association between the two LOBs. Specifically, the copula-based

models (Product copula, Gaussian copula, and Frank copula) demonstrate moderate medians and variability, whereas the DT-based models (DT-CTGAN and DT-CopulaGAN) show the smallest spread, indicating more consistent and lower risk measures.

Furthermore, the spread of the data and the presence of outliers vary across models. Silo-GLM and Silo-DT exhibit larger variability in their risk measures. In contrast, the DT-CTGAN and DT-CopulaGAN models have fewer and lower extreme values, contributing to their overall consistency in risk assessment.

Overall, the choice of model significantly influences the assessment of risk measures, with the copula and the EDT models providing more conservative and consistent estimates compared to the silo approaches. Between the EDT and copula models, the EDT models generally exhibit lower variability and fewer extreme values, suggesting that the EDT models might offer a more robust and reliable assessment of risk in this context.

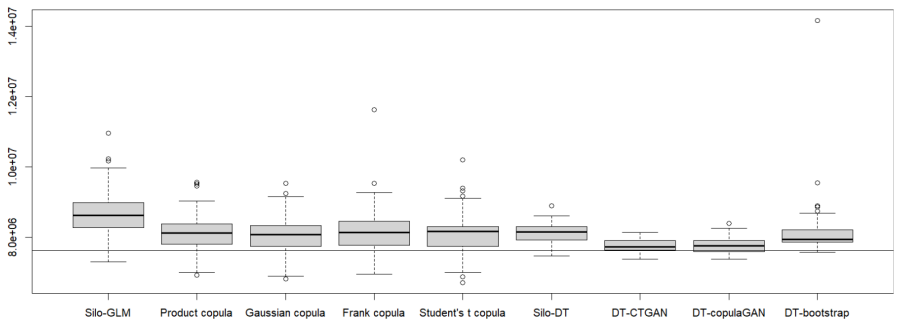

Figure 2.9: The risk measures at 99% for different models. The horizontal line indicates the true risk measure.

Additionally, we present the risk capital using the average of the risk measures derived from the 50 simulated loss triangle pairs in Table 2.11. Similar to the real data application, we calculate the risk measures for Silo-GLM and Silo-DT, respectively. Once again, we observe that the risk capital for Silo-DT is smaller than that for Silo-GLM. Furthermore, DT-CopulaGAN consistently generates the smallest risk capital among all the EDT models, aligning with the real data ap-

Table 2.11: Average risk capital from 50 simulated loss triangles.

Risk capital	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
Silo-DT	198 839	272 500	372 127	531 067	863 805
DT-CTGAN	141 922	192 690	258 984	363 501	576 167
DT-CopulaGAN	140 300	190 851	257 575	361 720	569 043
Silo-GLM	256 262	354 572	486 496	702 533	1 180 878
Product copula	179 938	246 664	335 529	477 132	758 015
Gaussian copula	158 158	217 078	295 256	422 683	700 221
Frank copula	165 128	227 321	311 099	449 677	751 712
True risk capital	106 371	146 047	199 655	282 214	467 289

plication findings. It's important to note that we do not calculate the risk capital gain in this context because the risk capitals for the silo method vary across the 50 simulated loss triangles. We show the risk capital percentage errors for different models in Table 2.12. We find that DT-CTGAN and DT-CopulaGAN generate risk capitals that are closest to the true risk capital, particularly in the tail, when compared to all other models.

Table 2.12: Risk capital percentage error for different methods.

Risk capital	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
Silo-DT	86.93%	86.58%	86.38%	88.18%	84.85%
DT-CTGAN	33.42%	31.94%	29.72%	28.80%	23.30%
DT-CopulaGAN	**31.90%**	**30.68%**	**29.01%**	**28.17%**	**21.78%**
Silo-GLM	140.91%	142.78%	143.67%	148.94%	152.71%
Product copula	69.16%	68.89%	68.05%	69.07%	62.22%
Gaussian copula	48.68%	48.64%	47.88%	49.77%	49.85%
Frank copula	55.24%	55.65%	55.82%	59.34%	60.87%

Note: The smallest risk capital percentage error for each risk level is in bold.

2.5 Summary and Discussion

When applied to multiple LOBs, the DT model leverages the dependence between loss triangles from different LOBs. Specifically, we use the incremental paid losses from different LOBs as training inputs, with the model designed to minimize asymmetric prediction errors.

The DT model estimates reserves for different LOBs, while DT-CTGAN and

DT-CopulaGAN generate predictive distributions for the total reserve. We demonstrate that DT requires sequence-based input, whereas GAN-based models operate on tabular data, ensuring both approaches effectively capture dependencies and improve reserve estimation.

A crucial aspect of this chapter is assessing the diversification benefits of the EDT in multivariate loss reserving and risk capital analysis. This is achieved by comparing risk measures and risk capital derived from the "silo" method against those obtained from the EDT or copula regression approaches. This comparison highlights the potential advantages of employing the EDT, a more interconnected and sophisticated modeling technique, in managing insurance portfolio risks.

To evaluate the practical effectiveness of the EDT, we apply it to loss triangles from 30 companies in the NAIC database. For comparison, we also include copula regression in our study. The EDT outperforms alternative models by yielding the smallest bias between predicted and true reserves for both LOBs. Moreover, both DT-CTGAN and DT-CopulaGAN consistently produce lower risk capital estimates than industry standards. These results highlight the potential benefits of integrating advanced modeling techniques for more accurate reserve estimation and enhanced risk management in the insurance industry.

Through the real data applications and simulations study, we discerned certain limitations of copula regression. One reason for the relatively large bias in copula regression is using a single pair of loss triangles. Modeling loss triangles from multiple companies using fixed effects in copula regression still generates a larger bias. Future extensions may involve seemingly unrelated regressions and mixed models to address the heterogeneity in data from multiple companies (Zellner, 1962). Another notable constraint is that complex dependencies may not be captured in copula regression attributed to the Fréchet-Hoeffding bounds on the dependence parameter (Schweizer and Sklar, 2011). Additionally, the potential over-parameterization of the copula regression model, particularly given the lim-

ited pairs of observations available in the LOBs, poses a challenge. To mitigate this, regularization techniques such as the least absolute shrinkage and selection operator (LASSO) can be applied. By imposing an L_1 norm constraint on the parameters within the loss function, LASSO facilitates parameter shrinkage, effectively reducing overfitting (Tibshirani, 1996). These regularization methods, akin to those used in generalized linear models (Taylor, 2019), play a crucial role in improving the robustness of copula regression models, especially when working with limited data.

In summary, the EDT framework shows strength and potential in multivariate loss reserving and risk capital analysis by providing a smaller bias in reserve prediction and larger diversification benefits. This flexible framework allows its application in various settings in actuarial science, such as rate-making and reinsurance. This adaptability showcases the EDT versatility across diverse insurance scenarios by complementing the DT model.

Chapter 3

Seemingly Unrelated Regression (SUR) Copula Mixed Models

3.1 Introduction

Dependence modeling of loss ratios between multiple loss triangles is critical in predicting loss reserves and risk capital analysis, which depends on the predictive distribution of the reserve. Incorporating dependency into reserve leverages the risk diversification benefit between incremental paid losses of different lines of business (LOBs) for the insurer. For example, Cai et al. (2025) show that capturing the pairwise and sequence dependence in multiple loss triangles reduces risk capital. On the other hand, Abdallah et al. (2015) demonstrate that a parametric approach that models some of these relationships can accurately determine reserve ranges and the amount of risk capital needed.

Neural networks-based method utilizes GAN techniques to generate the predictive distribution, and parametric models utilize model-based and rank-based bootstrapping to generate the predictive distribution. In the presence of dependence, copula GAN leads to the largest risk capital gain (Cai et al., 2025). Among parametric approaches, the rank-based bootstrapping yields the largest risk capital

gain and highest risk diversification benefit (Abdallah and Wang, 2023).

Both neural network-based and parametric approaches have notable limitations. Neural methods, while flexible and capable of quantifying predictive uncertainty as demonstrated in the Extended Deep Triangle (EDT) model, often lack interpretability, particularly in how they represent dependence structures (Cai et al., 2025). This hinders their usefulness for decision-making in actuarial contexts where understanding pairwise dependence is essential. Parametric models, on the other hand, are more interpretable but may suffer from model misspecification and fail to fully leverage multiple company data, leading to biased reserve and capital estimates (Shi and Frees, 2011; Abdallah et al., 2015).

Kuo (2019) introduced the Deep Triangle (DT), a recurrent neural network framework that leverages loss triangles from multiple companies to enhance the predictive accuracy of traditional stochastic reserving methods. Building on this, the Extended Deep Triangle (EDT) incorporates dependence between two LOBs by modeling pairwise and sequential relationships in loss ratios. To address heterogeneity across companies, the EDT encodes company identifiers such that companies with similar incremental paid loss patterns are represented with similar codes. Additionally, dropout is applied within the RNN to induce sparsity and improve prediction. Trained on data from multiple companies, the EDT generates reserve estimates that closely align with actual reserves. Moreover, by learning development year patterns and latent dependence structures directly from data, EDT yields lower risk capital than copula regression models, which rely on fixed effects to represent development year behavior and company heterogeneity.

While SUR-based models provide a flexible regression framework to capture dependencies, they have typically been limited to single-company settings or often model company and development year effects as fixed. For example, Zhang (2010) extend the classical chain ladder to a multivariate SUR framework, while Shi and Frees (2011) and Abdallah et al. (2015) use SUR copula regression to account

for dependence between two LOBs. However, these models do not fully leverage multiple company data and may suffer from biased estimates due to heterogeneity in company data.

We propose a SUR copula mixed model that extends SUR copula regression to accommodate hierarchical data from multiple companies to address this gap. The model includes fixed effects to capture accident and development year patterns, and random effects to represent heterogeneity across companies. Moreover, we propose to estimate the model parameters using a two-stage iterative maximum likelihood approach.

The chapter is organized as follows: Section 2 provides an overview of the preliminaries on SUR and mixed models. Section 3 discusses the methodologies for loss reserving and predictive distribution estimation, with an emphasis on the SUR copula mixed model approach. Section 4 applies and calibrates the SUR copula mixed model using a dataset that includes personal and commercial automobile LOBs from multiple companies. Section 5 presents a simulation study that further demonstrates the superior performance of the SUR copula mixed model. Finally, Section 6 summarizes our findings.

3.2 Methods: Preliminaries

3.2.1 Seemingly Unrelated Regression

Suppose we have a set of M regression equations

$$y_{mk} = \boldsymbol{x}_{mk}^T \boldsymbol{\beta}_m + \varepsilon_{mk}, \tag{3.1}$$

where $m = 1, \ldots, M$ is the euqation number and $k = 1, \ldots, N$ is the individual observation. The vector $\boldsymbol{x}_{mk}$ denotes the regressor, $\boldsymbol{\beta}_m$ represents the coefficients vector, and ε_{mk} is the error term.

If we stack all the observations for the m^{th} equation into vectors and matrices, then the model can be written as

$$\boldsymbol{y}_m = \mathbf{X}_m \boldsymbol{\beta}_m + \boldsymbol{\varepsilon}_m. \tag{3.2}$$

Next, we stack the M vector equations on top of each other and obtain the following form

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & 0 & \dots & 0 \\ 0 & \mathbf{X}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{X}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{3.3}$$

The model assumes that ε_{mk} are independent within equations but may have cross-equation correlations. Thus, we have $E[\varepsilon_{mk}\varepsilon_{mk'}|\mathbf{X}] = 0 (k \neq k')$ whereas $E[\varepsilon_{mk}\varepsilon_{m'k}|\mathbf{X}] = \sigma_{mm'}$. For the k^{th} observation, the covariance matrix of the error terms $(\varepsilon_{1k}, \varepsilon_{2k}, \dots, \varepsilon_{Mk})$ is denoted as $\boldsymbol{\Sigma} = [\sigma_{ij}]$. The covariance matrix of the stacked error terms $\boldsymbol{\varepsilon}$ equals to

$$\boldsymbol{\Omega} \equiv \mathrm{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T|\mathbf{X}\right] = \boldsymbol{\Sigma} \otimes \mathbf{I}_N, \tag{3.4}$$

where $\mathbf{I}_N$ is the identity matrix of dimension N and $\otimes$ denotes the matrix Kronecker product.

For the SUR model, the generalized least squares (GLS) estimator takes the form

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = \{\mathbf{X}^T(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_N)\mathbf{X}\}^{-1}\mathbf{X}^T(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_N)\boldsymbol{y}. \tag{3.5}$$

In most situations, the covariance $\boldsymbol{\Sigma}$ needed in GLS is unknown. Feasible generalized least squares (FGLS) estimate the elements of $\boldsymbol{\Sigma}$ by $\hat{\sigma}_{jk} = \hat{\boldsymbol{\varepsilon}}_j^T \hat{\boldsymbol{\varepsilon}}_k / N$, where $\hat{\boldsymbol{\varepsilon}}_j$ is the residual vector of the j th equation obtained from ordinary least

squares and then replace $\boldsymbol{\Sigma}$ in GLS by the resulting estimator $\hat{\boldsymbol{\Sigma}}$.

Alternatively, maximum likelihood estimators (MLE) can be considered, assuming the errors are multivariate normal (Srivastava and Giles, 1987; Peremans and Van Aelst, 2018). Under the assumption that the errors are normally distributed, the log-likelihood of the SUR model is given by

$$\ell(\boldsymbol{\beta},\, \boldsymbol{\Sigma}|\mathbf{X}, \boldsymbol{y}) = -\frac{MN}{2}\ln(2\pi) - \frac{N}{2}\ln(|\boldsymbol{\Sigma}|) - \frac{1}{2}(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta})^T(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_N)(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta}). \quad (3.6)$$

Maximizing this log-likelihood with respect to $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ yields the estimators $(\hat{\boldsymbol{\beta}}_{\text{MLE}},\, \hat{\boldsymbol{\Sigma}}_{\text{MLE}})$, which are the solutions of the equations

$$\hat{\boldsymbol{\beta}}_{\text{MLE}} = \{\mathbf{X}^T(\hat{\boldsymbol{\Sigma}}_{\text{MLE}}^{-1} \otimes \mathbf{I}_N)\mathbf{X}\}^{-1}\mathbf{X}^T(\hat{\boldsymbol{\Sigma}}_{\text{MLE}}^{-1} \otimes \mathbf{I}_N)\boldsymbol{y}, \quad (3.7)$$

$$\hat{\boldsymbol{\Sigma}}_{\text{MLE}} = (\mathbf{Y} - \tilde{\mathbf{X}}\hat{\boldsymbol{\mathcal{B}}}_{\text{MLE}})^T(\mathbf{Y} - \tilde{\mathbf{X}}\hat{\boldsymbol{\mathcal{B}}}_{\text{MLE}})/N, \quad (3.8)$$

where $\mathbf{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_M)$, $\tilde{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M)$ and $\hat{\boldsymbol{\mathcal{B}}}_{\text{MLE}}$ is the block diagonal form of $\hat{\boldsymbol{\beta}}_{\text{MLE}}$.

3.2.2 Linear Mixed Model

In general, a linear mixed-effects model satisfies

$$\begin{cases} \boldsymbol{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \\ \boldsymbol{b}_i \sim N(\mathbf{0}, \mathbf{D}), \\ \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i), \\ \boldsymbol{b}_1, \ldots, \boldsymbol{b}_N, \boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_N \text{ independent}, \end{cases} \quad (3.9)$$

where $\boldsymbol{y}_i$ is the n_i -dimensional response vector for subject i ($1 \leq i \leq N$) and N is the number of subjects. $\boldsymbol{\beta}$ is a p -dimensional fixed effects vector and $\boldsymbol{b}_i$ is the q -dimensional random effects vector. $\mathbf{X}_i$ and $\mathbf{Z}_i$ are $(n_i \times p)$ and $(n_i \times q)$ dimensional design matrices for the fixed effects and random effects, respectively. $\boldsymbol{\varepsilon}_i$ is an n_i

-dimensional vector of errors.

$\mathbf{D}$ is the covariance matrix for the q-dimensional random effects vector. $\mathbf{\Sigma}_i$ is the covariance matrix for the error vector. The set of unknown parameters in $\mathbf{\Sigma}_i$ will not depend upon i.

Conditional on the random effect $\boldsymbol{b}_i$, $\boldsymbol{y}_i$ is normally distributed with mean vector $\mathbf{X}_i\beta + \mathbf{Z}_i\boldsymbol{b}_i$ and covariance matrix $\mathbf{\Sigma}_i$. The marginal density of $\boldsymbol{y}_i$ can be shown to be the density of an n_i-dimensional normal distribution with mean vector $\mathbf{X}_i\boldsymbol{\beta}$ and covariance matrix $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \mathbf{\Sigma}_i$.

Let $\boldsymbol{\alpha}$ denote the vector of all variance and covariance parameters found in $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \mathbf{\Sigma}_i$. We can estimate the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by maximizing the marginal likelihood function

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}|\mathbf{X}_i, \boldsymbol{y}_i) = \prod_{i=1}^{N} \left\{ (2\pi)^{-n_i/2} \left|\mathbf{V}_i(\boldsymbol{\alpha})\right|^{-\frac{1}{2}} \times \exp\left(-\frac{1}{2} \left(\boldsymbol{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)^T \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \left(\boldsymbol{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right) \right) \right\}. \tag{3.10}$$

We first assume $\boldsymbol{\alpha}$ is known. The maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$, obtained from maximizing (3.10), conditional on $\boldsymbol{\alpha}$ is then given by (Verbeke and Molenberghs, 2009)

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \mathbf{X}_i^T \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^{N} X_i^T \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \boldsymbol{y}_i \right). \tag{3.11}$$

The maximum likelihood estimation (MLE) of $\boldsymbol{\alpha}$ is obtained by maximizing (3.10) with respect to $\boldsymbol{\alpha}$, after $\boldsymbol{\beta}$ is replaced by (3.11). This approach arises naturally when we consider maximizing the joint likelihood (3.10) to obtain the estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ simultaneously. The MLE involves a precision matrix $\mathbf{V}_i^{-1}(\boldsymbol{\alpha})$, so we will work with a precision matrix instead of a covariance matrix in the optimization.

3.3 SUR Mixed Model for Loss Reserving

Let Y_{ijc} denote the standardized incremental claims for accident year i $(1 \leq i \leq I)$, development year j $(1 \leq j \leq I)$, and company c $(1 \leq c \leq C)$. In the case of two LOBs, we use $Y_{ijc}^{(\ell)}$ and $Y_{ijc}^{(\ell')}$ to denote the standardized incremental claim from ℓ^{th} and ℓ'^{th} LOB. Now we model $\log(Y_{ijc}^{(\ell)})$ using $\boldsymbol{\alpha}^{(\ell)} = (\alpha_1^{(\ell)}, \alpha_2^{(\ell)}, ..., \alpha_I^{(\ell)})$ and $\boldsymbol{\lambda}^{(\ell)} = (\lambda_1^{(\ell)}, \lambda_2^{(\ell)}, ..., \lambda_I^{(\ell)})$ as predictors that characterize the accident and the development year effects, and the company effect $\boldsymbol{b}^{(\ell)} = (b_1^{(\ell)}, b_2^{(\ell)}, ..., b_C^{(\ell)})$ as in (3.12) and (3.13). Working on the logarithm scale of y_{ijc}, we ensure the predicted incremental claims are positive.

$$\log(y_{ijc}^{(\ell)}) = \xi^{(\ell)} + \boldsymbol{x}_i^{(\ell)}\boldsymbol{\alpha}^{(\ell)} + \boldsymbol{x}_j^{(\ell)}\boldsymbol{\lambda}^{(\ell)} + \boldsymbol{z}_c^{(\ell)}\boldsymbol{b}^{(\ell)} + \varepsilon_{ijc}^{(\ell)}, \tag{3.12}$$

$$\log(y_{ijc}^{(\ell')}) = \xi^{(\ell')} + \boldsymbol{x}_i^{(\ell')}\boldsymbol{\alpha}^{(\ell')} + \boldsymbol{x}_j^{(\ell')}\boldsymbol{\lambda}^{(\ell')} + \boldsymbol{z}_c^{(\ell')}\boldsymbol{b}^{(\ell')} + \varepsilon_{ijc}^{(\ell')}. \tag{3.13}$$

- $\boldsymbol{x}_i^{(\ell)}$ and $\boldsymbol{x}_i^{(\ell')}$ represent the observed accident year for the two LOBs. $\boldsymbol{x}_j^{(\ell)}$ and $\boldsymbol{x}_j^{(\ell')}$ represent the observed development year for the two LOBs.
- $\xi^{(\ell)}$ and $\xi^{(\ell')}$ are the intercepts for the two LOBs. We set $\alpha_1 = 0$ and $\lambda_1 = 0$ for parameter identification.
- We assume the following for the company effects. $b_c^{(\ell)} \sim N(0, \tau_1)$ and $b_c^{(\ell')} \sim N(0, \tau_2)$. The company effects are random, and they are independent and normally distributed with means of zero and standard deviations τ_1 and τ_2. The company effects are identical within each LOB.

In Table 3.1, we give an example for the regressor $\boldsymbol{x}_i^{(1)}$ and $\boldsymbol{x}_j^{(1)}$. Similarly, an example for the regressor z_c is in table 3.2.

If we stack observations corresponding to the 1st and 2nd equations into vectors and matrices, we write the model in vector form as

	$x_{i2}^{(1)}$	$x_{i3}^{(1)}$	$x_{i4}^{(1)}$	...	$x_{j2}^{(1)}$	$x_{j3}^{(1)}$	$x_{j4}^{(1)}$	...
$y_{111}^{(1)}$	0	0	0	...	0	0	0	...
$y_{121}^{(1)}$	0	0	0	...	1	0	0	...
$y_{131}^{(1)}$	0	0	0	...	0	1	0	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$y_{211}^{(1)}$	1	0	0	...	0	0	0	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$y_{331}^{(1)}$	0	1	0	...	0	1	0	...
$y_{341}^{(1)}$	0	1	0	...	0	0	1	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 3.1: An example for the regressor $\boldsymbol{x}_i^{(1)}$ and $\boldsymbol{x}_j^{(1)}$

	$z_c^{(1)}$	$z_c^{(2)}$
$y_{111}^{(1)}$	1	0
$y_{121}^{(1)}$	1	0
$y_{131}^{(1)}$	1	0
$\vdots$	$\vdots$	$\vdots$
$y_{214}^{(1)}$	1	0
$\vdots$	$\vdots$	$\vdots$
$y_{338}^{(2)}$	0	1
$y_{349}^{(2)}$	0	1
$\vdots$	$\vdots$	$\vdots$

Table 3.2: An example for the regressor z_c

$$\log(\boldsymbol{y}_c^{(\ell)}) = \mathbf{X}_c^{(\ell)}\boldsymbol{\beta}^{(\ell)} + \mathbf{Z}_c^{(\ell)}\boldsymbol{b}^{(\ell)} + \boldsymbol{\varepsilon}_c^{(\ell)}, \tag{3.14}$$

$$\log(\boldsymbol{y}_c^{(\ell')}) = \mathbf{X}_c^{(\ell')}\boldsymbol{\beta}^{(\ell')} + \mathbf{Z}_c^{(\ell')}\boldsymbol{b}^{(\ell')} + \boldsymbol{\varepsilon}_c^{(\ell')}. \tag{3.15}$$

Next, we stack the two vector equations on top of each other and transform (3.14) and (3.15) into the following form

$$\begin{bmatrix} \log(\boldsymbol{y}_c^{(\ell)}) \\ \log(\boldsymbol{y}_c^{(\ell')}) \end{bmatrix} = \begin{bmatrix} \mathbf{X}_c^{(\ell)} & 0 \\ 0 & \mathbf{X}_c^{(\ell')} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^{(\ell)} \\ \boldsymbol{\beta}^{(\ell')} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_c^{(\ell)} & 0 \\ 0 & \mathbf{Z}_c^{(\ell')} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}^{(\ell)} \\ \boldsymbol{b}^{(\ell')} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_c^{(\ell)} \\ \boldsymbol{\varepsilon}_c^{(\ell')} \end{bmatrix}. \tag{3.16}$$

Suppose $\boldsymbol{y}_c$ denote the $I(I+1)\times 1$ vector of incremental paid losses for the c^{th} company. The linear mixed model can be written as

$$\log(\boldsymbol{y}_c) = \mathbf{X}_c\boldsymbol{\beta} + \mathbf{Z}_c\boldsymbol{b} + \boldsymbol{\varepsilon}_c, \quad c = 1, \ldots, C. \tag{3.17}$$

where $\boldsymbol{\beta}$ represents all fixed-effects parameters and $\boldsymbol{b}$ represents the random effect for all companies. The company random effects $\boldsymbol{b}$ and errors $\boldsymbol{\varepsilon}_c$ are independent.

In the next section, we will detail parametric models for the error terms. We consider the multivariate normal distribution to model the errors.

3.3.1 SUR Multivariate Normal Mixed model

We assume the errors are independent within each equation but correlated between equations. One model for the pairs of errors, $\begin{pmatrix} \varepsilon_{ijc}^{(1)} \\ \varepsilon_{ijc}^{(2)} \end{pmatrix}$, is bivariate normal. That is,

$$\begin{pmatrix} \varepsilon_{ijc}^{(1)} \\ \varepsilon_{ijc}^{(2)} \end{pmatrix} \sim N(0, \boldsymbol{\Sigma}). \tag{3.18}$$

The covariance matrix $\boldsymbol{\Sigma}$ is defined by $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$. The covariance of the pair of errors are the same regardless of the accident and development years.

The stacked error terms $\boldsymbol{\varepsilon}_c$ follow a multivariate normal with covariance $\boldsymbol{\Sigma}_c = \boldsymbol{\Sigma} \otimes \mathbf{I}_{I(I+1)/2}$.

$$\boldsymbol{\varepsilon}_c \sim N(0, \boldsymbol{\Sigma} \otimes \mathbf{I}_{I(I+1)/2}), \tag{3.19}$$

where $\mathbf{I}_{I(I+1)/2}$ is the identity matrix of dimension $I(I+1)/2$ (total number of observations in the upper triangle) and $\otimes$ denotes the Kronecker product.

Next, we derive the log-likelihood function with the above multivariate normal assumption on errors. Consider y_c conditional on b_c.

$$\log(\boldsymbol{y}_c) \mid \boldsymbol{b}_c \sim N\left(\mathbf{X}_c\boldsymbol{\beta} + \mathbf{Z}_c\boldsymbol{b}_c, \boldsymbol{\Sigma}_c\right),$$

$$\boldsymbol{b}_c \sim N\left(0, \mathbf{D}\right).$$

The covariance matrix $\mathbf{D}$ is defined by $\mathbf{D} = \begin{bmatrix} \tau_1^2 & \\ & \tau_2^2 \end{bmatrix}$, and the design matrix $\mathbf{Z}_c$ is given by

$\mathbf{Z}_c = \begin{bmatrix} \mathbf{1}_{I(I+1)/2} & \\ & \mathbf{1}_{I(I+1)/2} \end{bmatrix}$, where $\mathbf{1}_{I(I+1)/2}$ is a vector of all ones.

Then, $\log(\boldsymbol{y}_c)$ has a multivariate normal distribution with mean $\mathbf{X}_c\boldsymbol{\beta}$ and covariance matrix $\mathbf{V}$.

$$\log(\boldsymbol{y}_c) \sim N\left(\mathbf{X}_c\boldsymbol{\beta}, \mathbf{V}\right),$$

where the covariance matrix $\mathbf{V}$ is given by

$$\mathbf{V} = \mathbf{V}_1 + \mathbf{V}_2,$$

$\mathbf{V}_1$ is defined by

$$\begin{aligned} \mathbf{V}_1 &= \boldsymbol{\Sigma} \otimes \mathbf{I}_{I(I+1)/2} \\ &= \begin{bmatrix} \sigma_1^2\mathbf{I}_{I(I+1)/2} & \rho\sigma_1\sigma_2\mathbf{I}_{I(I+1)/2} \\ \rho\sigma_1\sigma_2\mathbf{I}_{I(I+1)/2} & \sigma_2^2\mathbf{I}_{I(I+1)/2} \end{bmatrix}, \end{aligned}$$

and $\mathbf{V}_2$ is defined by

$$\mathbf{V}_2 = \mathbf{Z}_c\mathbf{D}\mathbf{Z}_c' = \begin{bmatrix} \tau_1^2\mathbf{1}_{I(I+1)/2}\mathbf{1}'_{I(I+1)/2} & \\ & \tau_2^2\mathbf{1}_{I(I+1)/2}\mathbf{1}'_{I(I+1)/2} \end{bmatrix}.$$

Combining $\mathbf{V}_1$ and $\mathbf{V}_1$, we obtain $\mathbf{V}$ as

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2\mathbf{I}_{I(I+1)/2} + \tau_1^2\mathbf{1}_{I(I+1)/2}\mathbf{1}'_{I(I+1)/2} & \rho\sigma_1\sigma_2\mathbf{I}_{I(I+1)/2} \\ \rho\sigma_1\sigma_2\mathbf{I}_{I(I+1)/2} & \sigma_2^2\mathbf{I}_{I(I+1)/2} + \tau_2^2\mathbf{1}_{I(I+1)/2}\mathbf{1}'_{I(I+1)/2} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2+\tau_1^2 & \tau_1^2 & \dots & \tau_1^2 & \rho\sigma_1\sigma_2 & & & \\ \tau_1^2 & \sigma_1^2+\tau_1^2 & \dots & \vdots & & \rho\sigma_1\sigma_2 & & \\ \dots & \dots & \dots & \tau_1^2 & & & \dots & \\ \tau_1^2 & \dots & \tau_1^2 & \sigma_1^2+\tau_1^2 & & & & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & & & & \sigma_2^2+\tau_2^2 & \tau_2^2 & \dots & \tau_2^2 \\ & \rho\sigma_1\sigma_2 & & & \tau_2^2 & \sigma_2^2+\tau_2^2 & \dots & \vdots \\ & & \dots & & \dots & \dots & \dots & \tau_2^2 \\ & & & \rho\sigma_1\sigma_2 & \tau_2^2 & \dots & \tau_2^2 & \sigma_2^2+\tau_2^2 \end{bmatrix}.$$

The corresponding multivariate normal probability density function for company c is: $f(\log(\boldsymbol{y}_c); \boldsymbol{\beta}, \tau_1, \tau_2, \sigma_1, \sigma_2, \rho)$ is:

$$f(\log(\boldsymbol{y}_c); \boldsymbol{\beta}, \tau_1, \tau_2, \sigma_1, \sigma_2, \rho) = (2\pi)^{-I(I+1)/2}|\mathbf{V}(\tau_1, \tau_2, \sigma_1, \sigma_2, \rho)|^{-1/2}\cdot$$
$$\exp(-0.5 \times (\log(\boldsymbol{y}_c) - \mathbf{X}_c\boldsymbol{\beta})'\mathbf{V}^{-1}(\tau_1, \tau_2, \sigma_1, \sigma_2, \rho)(\log(\boldsymbol{y}_c) - \mathbf{X}_c\boldsymbol{\beta})). \quad (3.20)$$

From (3.20), the likelihood function for company c is:

$$L_c(\boldsymbol{\beta}, \tau_1, \tau_2, \sigma_1, \sigma_2, \rho \mid \log(\boldsymbol{y}_c)) = (2\pi)^{-I(I+1)/2}|\mathbf{V}(\tau_1, \tau_2, \sigma_1, \sigma_2, \rho)|^{-1/2}\cdot$$
$$\exp(-0.5 \times (\log(\boldsymbol{y}_c) - \mathbf{X}_c\boldsymbol{\beta})'\mathbf{V}^{-1}(\tau_1, \tau_2, \sigma_1, \sigma_2, \rho)(\log(\boldsymbol{y}_c) - \mathbf{X}_c\boldsymbol{\beta})). \quad (3.21)$$

We write the likelihood function, $L(\boldsymbol{\beta}, \tau_1, \tau_2, \sigma_1, \sigma_2, \rho \mid \log(\boldsymbol{y}_1), \log(\boldsymbol{y}_2), \dots, \log(\boldsymbol{y}_C))$, for all the companies as the product of the C independent contributions from the companies (c=1,..., C) :

$$
\begin{aligned}
&L(\boldsymbol{\beta}, \tau_1, \tau_2, \sigma_1, \sigma_2, \rho \mid \log(\boldsymbol{y}_1), \log(\boldsymbol{y}_2), \ldots, \log(\boldsymbol{y}_C)) \\
&= \prod_c L_c(\boldsymbol{\beta}, \tau_1, \tau_2, \sigma_1, \sigma_2, \rho \mid \log(\boldsymbol{y}_c)) \\
&= \prod_c (2\pi)^{\frac{-I(I+1)}{2}} |\mathbf{V}(\tau_1, \tau_2, \sigma_1, \sigma_2, \rho)|^{\frac{1}{2}} \\
&\times \exp(-\frac{1}{2}(\log(\boldsymbol{y}_c) - \mathbf{X}_c\boldsymbol{\beta})'\mathbf{V}^{-1}(\tau_1, \tau_2, \sigma_1, \sigma_2, \rho)(\log(\boldsymbol{y}_c) - \mathbf{X}_c\boldsymbol{\beta})). \qquad (3.22)
\end{aligned}
$$

The corresponding log-likelihood function, $\ell(\boldsymbol{\beta}, \tau_1, \tau_2, \sigma_1, \sigma_2, \rho \mid \log(\boldsymbol{y}_1), \ldots, \log(\boldsymbol{y}_C))$ is defined as

$$
\ell(\boldsymbol{\beta}, \tau_1, \tau_2, \sigma_1, \sigma_2, \rho \mid \log(\boldsymbol{y}_1), \log(\boldsymbol{y}_2), \ldots, \log(\boldsymbol{y}_C)) = -\frac{1}{2}[I(I+1) \cdot C \cdot \log(2\pi) - C \log |\mathbf{V}(\tau_1, \tau_2, \sigma_1, \sigma_2, \rho)| + \sum_{c=1}^{C} (\log(\boldsymbol{y}_c) - \mathbf{X}_c\boldsymbol{\beta})'\mathbf{V}^{-1}(\tau_1, \tau_2, \sigma_1, \sigma_2, \rho)(\log(\boldsymbol{y}_c) - \mathbf{X}_c\boldsymbol{\beta})]. \qquad (3.23)
$$

Next, we summarize the steps to estimate the parameters in (3.23) as follows:

(1) Give initial values $\tau_1^0, \tau_2^0, \sigma_1^0, \sigma_2^0, \rho^0$, and set k=0.

(2) Maximize equation (3.23) with respect to $\boldsymbol{\beta}$ and obtain $\boldsymbol{\beta}^k = \left(\sum_{c=1}^{C} \mathbf{X}_c'\mathbf{V}(\tau_1^k, \tau_2^k, \sigma_1^k, \sigma_2^k, \rho^k)^{-1}\mathbf{X}_c\right)^{-1} \left(\sum_{c=1}^{C} \mathbf{X}_c'\mathbf{V}(\tau_1^k, \tau_2^k, \sigma_1^k, \sigma_2^k, \rho^k)^{-1} \log(\boldsymbol{y}_c)\right)$.

(3) Given $\boldsymbol{\beta}^k$, Maximize (3.23) to obtain $\tau_1^{k+1}, \tau_2^{k+1}, \sigma_1^{k+1}, \sigma_2^{k+1}\, \rho^{k+1}$. update k=k+1.

(4) Repeat steps (2) and (3) until it meets the stopping criterion, where the absolute difference of the fixed effects parameters in two consecutive iterations is negligible.

After we obtain the estimated fixed effects $\hat{\boldsymbol{\beta}}$, we estimate the random effects by $\hat{\boldsymbol{b}}_c = \mathbf{D}\mathbf{Z}_c'\mathbf{V}^{-1}(\log(\boldsymbol{y}_c) - \mathbf{X}_c\hat{\boldsymbol{\beta}})$, an average of the estimated effect.

Next, we detail an alternative way to model pairs of errors using copulas.

3.3.2 SUR Copula Mixed Model

In this section, we use copulas to model the dependence between the different LOBs. For each LOB, ℓ, we assume that the loss ratios, $Y_{ijc}^{(\ell)}$, are independent and follow a distribution belonging to the exponential family. Let $\mu_{ijc}^{(\ell)}$ be the expected value of $Y_{ijc}^{(\ell)}$, and it can be expressed as $\mu_{ijc}^{(\ell)} = g^{-1}(\eta_{ijc}^{(\ell)})$, where $\eta_{ijc}^{(\ell)}$ is the linear predictor and $g(\cdot)$ is the link function.

Following Liang and Zeger (1986), we model $\eta_{ijc}^{(\ell)}$ using $\boldsymbol{\alpha}^{(\ell)}$, $\boldsymbol{\lambda}^{(\ell)}$, and $\boldsymbol{b}^{(\ell)}$ as in (3.24).

$$\eta_{ijc}^{(\ell)} = \xi^{(\ell)} + \boldsymbol{x}_i^{(\ell)}\boldsymbol{\alpha}^{(\ell)} + \boldsymbol{x}_j^{(\ell)}\boldsymbol{\lambda}^{(\ell)} + \boldsymbol{z}_c^{(\ell)}\boldsymbol{b}^{(\ell)}. \tag{3.24}$$

Let $\boldsymbol{\beta}^{(\ell)}$ denote the intercept and all fixed-effects parameters and $\boldsymbol{b}^{(\ell)}$ be the random effect for companies. Then,

$$\eta_{ijc}^{(\ell)} = \boldsymbol{x}_{ij}^{(\ell)}\boldsymbol{\beta}^{(\ell)} + \boldsymbol{z}_c^{(\ell)}\boldsymbol{b}^{(\ell)}. \tag{3.25}$$

Let $\boldsymbol{y}_c^{(\ell)}$ be the $I(I+1) \times 1$ vector of incremental paid losses for the c^{th} company from ℓ^{th} LOB. Then, the joint probability density function (PDF) of $\boldsymbol{y}_1^{(\ell)}, \boldsymbol{y}_2^{(\ell)}, \ldots, \boldsymbol{y}_C^{(\ell)}$ is

$$f(\boldsymbol{y}_1^{(\ell)}, \boldsymbol{y}_2^{(\ell)}, \ldots, \boldsymbol{y}_C^{(\ell)}; \boldsymbol{\beta}^{(\ell)}, \sigma_\ell, \tau_\ell) = \int_{[-\infty,\infty]^C} f(\boldsymbol{y}_1^{(\ell)}, \boldsymbol{y}_2^{(\ell)}, \ldots, \boldsymbol{y}_C^{(\ell)} \mid b_1^{(\ell)}, b_2^{(\ell)}, \ldots, b_C^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \sigma_\ell) \cdot f(b_1^{(\ell)}, b_2^{(\ell)}, \ldots, b_C^{(\ell)}; \tau_\ell) db_1^{(\ell)} db_2^{(\ell)} \ldots db_C^{(\ell)}, \tag{3.26}$$

where C is the number of companies, σ_ℓ is the shape parameter of the marginal distribution, and τ_ℓ is the standard deviation of the company random effect $b_c^{(\ell)}$.

Assuming $\boldsymbol{y}_c^{(\ell)}$ from each company is independent, we can rewrite (3.26) as

$$\begin{aligned} f(\boldsymbol{y}_1^{(\ell)}, \boldsymbol{y}_2^{(\ell)}, \ldots, \boldsymbol{y}_C^{(\ell)}; \boldsymbol{\beta}^{(\ell)}, \sigma_\ell, \tau_\ell) &= \prod_{c=1}^{C} \int_{-\infty}^{\infty} f(\boldsymbol{y}_c^{(\ell)} \mid b_c^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \sigma_\ell) f(b_c^{(\ell)}; \tau_\ell) db_c^{(\ell)} \\ &= \prod_{c=1}^{C} \int_{-\infty}^{\infty} \prod_{i=1}^{I} \prod_{j=1}^{I+1-i} f(y_{ijc}^{(\ell)} | b_c^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \sigma_\ell) f(b_c^{(\ell)}; \tau_\ell) db_c^{(\ell)}, \end{aligned} \tag{3.27}$$

where $f(y_{ijc}^{(\ell)} | b_c^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \sigma_\ell)$ denotes the conditional density of $Y_{ijc}^{(\ell)}$ given $b_c^{(\ell)}$ and $f(b_c^{(\ell)}; \tau_\ell)$ denotes the marginal density of the company effect $b_c^{(\ell)}$.

From (3.27), a general formula for the log-likelihood for each LOB, ℓ, from all companies C

$$L^{(\ell)}(\boldsymbol{\beta}^{(\ell)}, \sigma_\ell, \tau_\ell; \boldsymbol{y}_1^{(\ell)}, \boldsymbol{y}_2^{(\ell)}, \ldots, \boldsymbol{y}_C^{(\ell)}) = \sum_{c=1}^{C} \log \int_{-\infty}^{\infty} \prod_{i=1}^{I} \prod_{j=1}^{I+1-i} f(y_{ijc}^{(\ell)} | b_c^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \sigma_\ell) f(b_c^{(\ell)}; \tau_\ell) db_c^{(\ell)}. \tag{3.28}$$

Suppose $Y_{ijc}^{(\ell)}$ follows log-normal distribution with location $\mu_{ijc}^{(\ell)}$ and the shape σ_ℓ, (3.28) becomes

$$\begin{aligned} L^{(\ell)}(\boldsymbol{\beta}^{(\ell)}, \sigma_\ell, \tau_\ell \mid \boldsymbol{y}_1^{(\ell)}, \boldsymbol{y}_2^{(\ell)}, \ldots, \boldsymbol{y}_C^{(\ell)}) = \sum_{c=1}^{C} \log \int_{-\infty}^{\infty} \prod_{i=1}^{I} \prod_{j=1}^{I+1-i} \frac{1}{y_{ijc}^{(\ell)} \sqrt{2\pi} \sigma_\ell} e^{-\frac{1}{2}\left(\frac{\log\left(y_{ijc}^{(\ell)}\right) - \mu_{ijc}^{(\ell)}}{\sigma_\ell}\right)^2} \\ \frac{1}{\sqrt{2\pi}\tau_\ell} e^{-\frac{1}{2}\left(\frac{b_c^{(\ell)}}{\tau_\ell}\right)^2} db_c^{(\ell)}. \end{aligned} \tag{3.29}$$

Suppose $Y_{ijc}^{(\ell)}$ follows gamma distribution with the scale $\gamma_{ijc}^{(\ell)}$ and the shape σ_ℓ,

(3.28) becomes

$$L^{(\ell)}(\boldsymbol{\beta}^{(\ell)}, \sigma_\ell, \tau_\ell \mid \boldsymbol{y}_1^{(\ell)}, \boldsymbol{y}_2^{(\ell)}, \ldots, \boldsymbol{y}_C^{(\ell)}) = \sum_{c=1}^{C} \log \int_{-\infty}^{\infty} \prod_{i=1}^{I} \prod_{j=1}^{I+1-i} \left(\frac{y_{ijc}^{(\ell)}}{\gamma_{ijc}^{(\ell)}} \right)^{\sigma_\ell} \frac{e^{-\frac{y_{ijc}^{(\ell)}}{\gamma_{ijc}^{(\ell)}}}}{\Gamma(\sigma_\ell)\, y_{ijc}^{(\ell)}} \frac{1}{\sqrt{2\pi}\tau_\ell} e^{-\frac{1}{2}\left(\frac{b_c^{(\ell)}}{\tau_\ell}\right)^2} db_c^{(\ell)}. \tag{3.30}$$

Based on Sklar's theorem (Nelsen, 2006), the joint density function of a pair $(Y_{ijc}^{(\ell)}, Y_{ijc}^{(\ell')})$ is

$$f(y_{ijc}^{(\ell)}, y_{ijc}^{(\ell')}) = f(y_{ijc}^{(\ell)}) f(y_{ijc}^{(\ell')}) c(F(y_{ijc}^{(\ell)}), F(y_{ijc}^{(\ell')}); \theta_c), \tag{3.31}$$

where $c(\cdot)$ denotes the copula conditional density corresponding to conditional copula $C(\cdot)$.

Note that the joint PDF of all pairs of $(Y_{ijc}^{(\ell)}, Y_{ijc}^{(\ell')})$ from all companies is then given by

$$\begin{aligned} &f((\boldsymbol{y}_1^{(\ell)}, \boldsymbol{y}_1^{(\ell')}), \ldots, (\boldsymbol{y}_C^{(\ell)}, \boldsymbol{y}_C^{(\ell')}); \boldsymbol{\beta}^{(\ell)}, \sigma_\ell, \tau_\ell, \boldsymbol{\beta}^{(\ell')}, \sigma_{\ell'}, \tau_{\ell'}) \\ &= f((y_{111}^{(\ell)}, y_{111}^{(\ell')}), (y_{121}^{(\ell)}, y_{121}^{(\ell')}), \ldots, (y_{1IC}^{(\ell)}, y_{1IC}^{(\ell')}); \boldsymbol{\beta}^{(\ell)}, \sigma_\ell, \tau_\ell, \boldsymbol{\beta}^{(\ell')}, \sigma_{\ell'}, \tau_{\ell'}). \end{aligned} \tag{3.32}$$

In the case of two LOBs, let $\ell = 1$ and $\ell' = 2$, we derive the log-likelihood function from (3.31) and (3.32) as

$$
\begin{aligned}
&L(\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \sigma_1, \sigma_2, \tau_1, \tau_2, \theta_c \mid (\boldsymbol{y}_1^{(1)}, \boldsymbol{y}_1^{(2)}), \ldots\ldots, (\boldsymbol{y}_C^{(1)}, \boldsymbol{y}_C^{(2)})) \\
&= \sum_{\ell=1}^{2} \sum_{c=1}^{C} \log \int_{-\infty}^{\infty} \prod_{i=1}^{I} \prod_{j=1}^{I+1-i} f(y_{ijc}^{(\ell)} \mid b_c^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \sigma_\ell) f(b_c^{(\ell)}; \tau_\ell) db_c^{(\ell)} + \sum_{c=1}^{C} \sum_{i=1}^{I} \sum_{j=1}^{I+1-i} \log c(F_{ijc}^{(1)}, F_{ijc}^{(2)}; \theta_c) \\
&= \sum_{\ell=1}^{2} L^{(\ell)}(\boldsymbol{\beta}^{(\ell)}, \sigma_\ell, \tau_\ell \mid \boldsymbol{y}_1^{(\ell)}, \ldots, \boldsymbol{y}_C^{(\ell)}) + \sum_{c=1}^{C} \sum_{i=1}^{I} \sum_{j=1}^{I+1-i} \log c(F_{ijc}^{(1)}, F_{ijc}^{(2)}; \theta_c), \qquad (3.33)
\end{aligned}
$$

where $L^{(\ell)}(\boldsymbol{\beta}^{(\ell)}, \sigma_\ell, \tau_\ell \mid \boldsymbol{y}_1^{(\ell)}, \ldots, \boldsymbol{y}_C^{(\ell)})$ is defined in (3.28).

We adopt an iterative two-stage approach, in which marginal models, including systematic effects and company random effects, are estimated in one step, and copula parameters are estimated from rank-transformed residuals in the next. Then, the copula parameters are plugged into the complete likelihood to optimize it with respect to the marginal distribution parameters. This process is repeated until the convergence of marginal distribution parameter estimates.

We first estimate $\boldsymbol{\beta}^{(\ell)}$, σ_ℓ, and τ_ℓ for each LOB by maximizing (3.33). As a starting value, we compute these estimates for each LOB using the restricted maximum likelihood (REML) implemented in the R package `lme4` (Bates et al., 2015) for a generalized linear mixed model.

In the second step, we maximize the second term of (3.33) with respect to the copula parameter θ given the marginal parameter estimates from the first step. The copula likelihood term is given by

$$
L_2(\theta) = \sum_{c=1}^{C} \sum_{i=1}^{I} \sum_{j=1}^{I+1-i} \log c(F_{ijc}^{(1)}, F_{ijc}^{(2)}; \theta_c).
$$

We replace the marginal cumulative distribution functions with pseudo-observations derived from the ranks of the residuals. We use the AIC statistic to choose the marginal distribution for $Y_{ijc}^{(\ell)}$. For example, we define rank-based pseudo-observations for log-normal and gamma marginals as follows:

Suppose $Y_{ijc}^{(\ell)}$ follows log-normal, we compute the pseudo-residuals given the parameter estimates in step 1 (Davison and Hinkley, 1997, pp. 331-340).

$$\epsilon_{ijc}^{(\ell)} = \frac{\log y_{ijc}^{(\ell)} - \hat{\mu}_{ijc}^{(\ell)}}{\hat{\sigma}_1}, \tag{3.34}$$

where $\hat{\mu}_{ijc}^{(\ell)}$ and $\hat{\sigma}_1$ are estimates.

Similarly if $Y_{ijc}^{(\ell)}$ follows gamma distribution, we compute the pseudo-residuals as

$$\epsilon_{ijc}^{(\ell)} = \frac{y_{ijc}^{(\ell)} - \hat{\mu}_{ijc}^{(\ell)}}{(\hat{\mu}_{ijc}^{(\ell)} * \hat{\gamma}_{ijc}^{(\ell)})^{1/2}}, \tag{3.35}$$

where $\hat{\mu}_{ijc}^{(\ell)}$ and $\hat{\gamma}_{ijc}^{(\ell)}$ are estimates from step 1.

Next, we use the empirical cumulative distribution function (CDF) to get the ranks of pseudo-residuals. The rank $R_{ijc}^{(\ell)}$ of the residual $\epsilon_{ijc}^{(\ell)}$ is given by

$$R_{ijc}^{(\ell)} = \frac{1}{I(I+1)/2+1} \sum_{i^*=1}^{I} \sum_{j^*=1}^{I+1-i^*} \mathbf{1}\left(\varepsilon_{i^*j^*c}^{(\ell)} \leq \varepsilon_{ijc}^{(\ell)}\right),$$

where $\mathbf{1}$ is the indicator function.

Now, we approximate the copula term $c(F(y_{ijc}^{(\ell)}), F(y_{ijc}^{(\ell')}); \theta_c)$ by $c(R_{ijc}^{(\ell)}, R_{ijc}^{(\ell')}; \theta_c)$ and then maximize the second term in (3.33) to obtain copula parameter θ.

We iterate step 1 and step 2 until the convergence on $\boldsymbol{\beta}^{(\ell)}$ is reached. Note that we choose the marginal distribution for each LOB using the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).

We summarize the steps to estimate the parameters in (3.33) as follows:

(1) Given initial values $\boldsymbol{\beta}^{(\ell)(0)}$, $\sigma_\ell^{(0)}$, and $\tau_\ell^{(0)}$, which are from generalized linear mixed model. log-likelihood function (3.33) is a function of copula parameter θ: $L_2(\theta) = \sum_{c=1}^{C} \sum_{i=1}^{I} \sum_{j=1}^{I+1-i} \log c(F(y_{ijc}^{(\ell)}), F(y_{ijc}^{(\ell')}); \theta_c)$. Set iteration $k = 0$.

(2) Maximize pseudo log-likelihood $L_3(\theta)$ with respect to θ and obtain θ^k. Here $L_3(\theta) = \sum_{c=1}^{C} \sum_{i=1}^{I} \sum_{j=1}^{I+1-i} \log c(R_{ijc}^{(\ell)}, R_{ijc}^{(\ell')}; \theta_c)$.

(3) Given θ^k, maximize (3.33) with respect to $\boldsymbol{\beta}^{(\ell)}$, σ_ℓ, and τ_ℓ. update k=k+1.

(4) Repeat steps (2) and (3) until it meets the stopping criterion: $||\boldsymbol{w}^{k+1} - \boldsymbol{w}^k||_2 \leq \epsilon$, where we group all parameters $\boldsymbol{\beta}^{(\ell)}, \sigma_\ell, \tau_\ell, \theta$ under vector $\boldsymbol{w}$.

3.3.3 Predictive Distribution of the Total Reserve

In practice, actuaries are interested in understanding the uncertainty of reserves. The bootstrapping technique provides this information and allows for the determination of the entire predictive distribution. In the bootstrap procedure, we use the pseudo residuals as defined in (3.35). Following the resampling approach outlined in Davison and Hinkley (1997), we summarize the steps of our bootstrap as follows:

(1) Fit the SUR copula mixed model to the observed incremental paid losses $y_{ijc}^{(\ell)}$, generating the residuals for the incremental paid losses using:

$$\epsilon_{ijc}^{(\ell)} = \frac{y_{ijc}^{(\ell)} - \hat{\mu}_{ijc}^{(\ell)}}{(\hat{\mu}_{ijc}^{(\ell)} * \hat{\gamma}_{ijc}^{(\ell)})^{1/2}}, \tag{3.36}$$

where $\hat{\gamma}_{ijc}^{(\ell)} = \exp(\boldsymbol{x}_{ij}^{(\ell)}\hat{\boldsymbol{\beta}}^{(\ell)} + \boldsymbol{z}_c^{(\ell)}\hat{\boldsymbol{b}}^{(\ell)})/\sigma_\ell$ is the estimated scale parameter.

(2) Begin the iterative loop, to be repeated N times (e.g., N = 1000):

- Simulate $\left(u_{ijc}^{(1)}, u_{ijc}^{(2)}\right)$ $(i + j - 1 \leq I)$ from estimated copula function $C(\cdot; \hat{\theta})$.
- Transform $u_{ijc}^{(\ell)}$ to the residuals by transform $\epsilon_{ijc}^{*(\ell)} = Q^{(\ell)}(u_{ijc}^{(\ell)})$, where $Q^{(\ell)}$ is the empirical quantile function of the residuals.
- Generate a set of pseudo incremental paid losses $y_{ijc}^{*(\ell)}$, which is given by $y_{ijc}^{*(\ell)} = \epsilon_{ijc}^{*(\ell)} * (\hat{\mu}_{ijc}^{(\ell)} * \hat{\gamma}_{ijc}^{(\ell)})^{1/2} + \hat{\mu}_{ijc}^{(\ell)}$.
- Estimate the parameters $\hat{\boldsymbol{\beta}}^{*(\ell)}$, $\hat{\sigma_\ell}^*$, and $\hat{\tau_\ell}^*$ and $\hat{\theta}^*$ for $y_{ijc}^{*(\ell)}$ using the SUR copula mixed model.

– Obtain a prediction of the total reserve for company c by

$$\sum_{\ell=1}^{2}\sum_{i=2}^{I}\sum_{j=I-i+2}^{I}\omega_{ic}^{(\ell)}\cdot\exp(\hat{\eta}_{ijc}^{*(\ell)}),$$

where $\hat{\eta}_{ijc}^{*(\ell)}=\boldsymbol{x}_{ij}^{(\ell)}\hat{\boldsymbol{\beta}}^{*(\ell)}+\boldsymbol{z}_{c}^{(\ell)}\hat{\boldsymbol{b}}^{*(\ell)}$ and $\omega_{ic}^{(\ell)}$ is the premium for accident year i and company c in LOB ℓ.

3.4 Applicaton

To illustrate the SUR mixed model in loss reserving, we analyze a dataset of 30 pairs of loss triangles from Schedule P of the National Association of Insurance Commissioners (NAIC) database (Meyers and Shi, 2011). Each pair of loss triangles is from the same company and consists of personal auto LOB and commercial auto LOB. Each triangle includes incremental claims data for accident years 1988 to 1997 and spans ten development years. We consider the reserve prediction and risk capital analysis for a major US property-casualty insurer. We use the upper part of the loss triangles to train the SUR mixed model and evaluate the predictive performance of the model on the lower part of the loss triangles.

We model the accident year and development year effects as fixed and we assume a specific effect of each level of these categorical variables (accident year and development year). With fixed effects, the estimated parameters for each accident year and development year have a direct interpretation. For the company level heterogeneity, we model it with random effects and assume to follow a normal distribution. Random effects are more parsimonious than fixed effects, leading to a more robust model.

For the SUR normal mixed model, we work with the logarithm of the loss ratios for both LOBs. Since the dataset is historical, the actual reserve is known and can be compared with predictions obtained from the SUR normal mixed model.

We first apply the SUR normal mixed model to the 30 pairs of loss triangles. By maximizing the log-likelihood function in 3.23, we obtain the estimated standard deviations: $\sigma_1 = 0.85$ and $\sigma_2 = 1.12$. The estimated standard deviations for the company random effects are $\tau_1 = 0.28$ and $\tau_2 = 0.41$. The variation induced by the company is larger in the commercial LOB than in the personal LOB. For the SUR normal mixed model, the estimated correlation coefficient is 0.23, which indicates a positive association between the two LOBs. The predicted reserves for the personal LOB and commercial LOB are 10 891 437 and 562 111, respectively.

Table 3.3: Point estimates of the reserves.

Model	Reserves		
	LoB 1, R_1	LoB 2, R_2	Total, R
SUR Gaussian copula	6 823 325	378 386	7 364 511
SUR normal mixed	10 891 437	562 111	11 453 548
SUR copula mixed	7 166 266	378 217	7 544 483

Table 3.4: Performance comparison using percentage error of actual and estimated loss reserve.

LOB	SUR copula	SUR normal mixed	SUR copula mixed
Personal Auto	-15.6%	34.7%	-11.4%
Commercial Auto	19.0%	76.5%	18.8%
Total	-12.4%	36.3%	-10.2%

For the SUR copula mixed model, we use a gamma distribution as the marginal for both LOBs based on the AIC statistic from 30 pairs of loss triangles. We use a Gaussian copula to capture the dependence between the two LOBs. The estimated shape parameters for the two LOBs are $\sigma_1 = 2.01$ and $\sigma_2 = 1.10$, respectively. The estimated standard deviations for the company random effects in the two LOBs are $\tau_1 = 0.33$ and $\tau_2 = 0.40$, respectively. The estimated dependence parameter between the two LOBs for the major US property-casaulty company is around -0.20. As shown in Table 3.3, the estimated reserve for the personal LOB is 7 246 135 while the reserve for the commercial LOB is 377 324.

Table 3.4 presents a comparison of various models based on the percentage error between the actual and the estimated reserves. Note that we use a gamma distribution as the marginal for both LOBs and model the company effects using fixed effects in the SUR copula model. The SUR copula mixed model produces a smaller bias between the predicted reserve and the true reserve than the SUR copula for both LOBs. This shows the effectiveness of SUR copula mixed models in handling heterogeneity in data from multiple companies. The subtle improvement in the predicted reserve in the commercial LOB may be due to the limited flexibility of the marginal distribution for the commercial LOB.

Table 3.5: Bias, Standard deviation, Coefficient of variation (CV) from the predictive distribution using parametric bootstrapping.

	Reserve	Bootstrap reserve	Bias	Std. dev.	CV
SUR copula	7 364 511	7 118 647	3.32%	1 865 284	0.262
SUR copula mixed	7 544 483	7 464 656	1.13%	599 454	0.080

Insurance companies are concerned with the expected unpaid loss or reserve, its standard deviation, and other risk measures. These measures are defined based on the reserve's predictive distribution, such as the Tail Value-at-Risk (TVaR). This measure is more informative than the value at risk (VaR), and the subadditivity of VaR is not generally guaranteed.

We employ the bootstrap method to obtain the predictive distribution of the reserve. Table 3.5 shows the estimated reserve, bias, and standard errors for different models. It has been observed that the standard deviation from the SUR copula mixed model is smaller than that from the SUR copula regression models, showing the effectiveness of the SUR copula mixed model in handling heterogeneity across companies and between LOBs.

The predictive distribution is particularly useful for assessing the risk capital of an insurance portfolio. The risk capital is the difference between the risk measure and the expected unpaid losses of the portfolio. We show in Table 3.6 the calcu-

lated risk capitals for SUR copula mixed models with dependence captured by the Gaussian copula, the SUR copula, and the silo method, which is widely used in industry. Using the 30 pairs of loss triangles, we model the company effects using fixed effects in the SUR copula method. For the silo method, we use the simple sum of the risk measures from each subportfolio (i.e., the personal auto line and the commercial auto line) as the risk measure for the entire portfolio. The silo method does not account for any diversification effect in the portfolio.

Table 3.6: Risk capital estimation for different methods.

Risk measure	TVaR (60%)	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
Silo-GLM	9 632 534	10 799 117	11 248 347	11 863 905	12 950 137	15 168 393
SUR copula	8 992 101	9 975 813	10 325 124	10 793 817	11 526 223	13 179 997
SUR copula mixed	8 110 263	8 436 316	8 555 644	8 723 725	9 009 793	9 597 029
Risk capital						
Silo-GLM		1 166 583	1 615 813	2 231 371	3 317 603	5 535 859
SUR copula		983 712	1 333 023	1 801 716	2 534 122	4 187 896
SUR copula mixed		326 053	445 381	613 462	899 530	1 486 766

Table 3.7: Risk capital gain for different methods.

Risk capital gain	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
SUR copula vs Silo-GLM	15.68%	17.50%	19.26%	23.62%	24.35%
SUR copula mixed vs Silo-GLM	70.73%	72.00%	72.32%	72.86%	73.24%

We show in Table 3.7 the gain in terms of risk capital for SUR copula mixed models compared to the silo method. The SUR copula mixed model captures the dependence between the two LOBs with one or multiple dependence parameters, as well as the heterogeneity across companies and between LOBs, resulting in the greatest gain in risk capital.

3.5 Simulation Study

We begin with the estimated SUR Gaussian copula mixed model for the real data in Section 3.4. In the model, the conditional distributions of the marginals given the random effects are gamma. Figure 3.1 and Figure 3.2 illustrate the box plot of the loss ratios for each LOB from 30 companies. Based on the box plots, the

commercial LOB exhibits a larger number of outliers than the personal LOB, suggesting higher variability across companies.

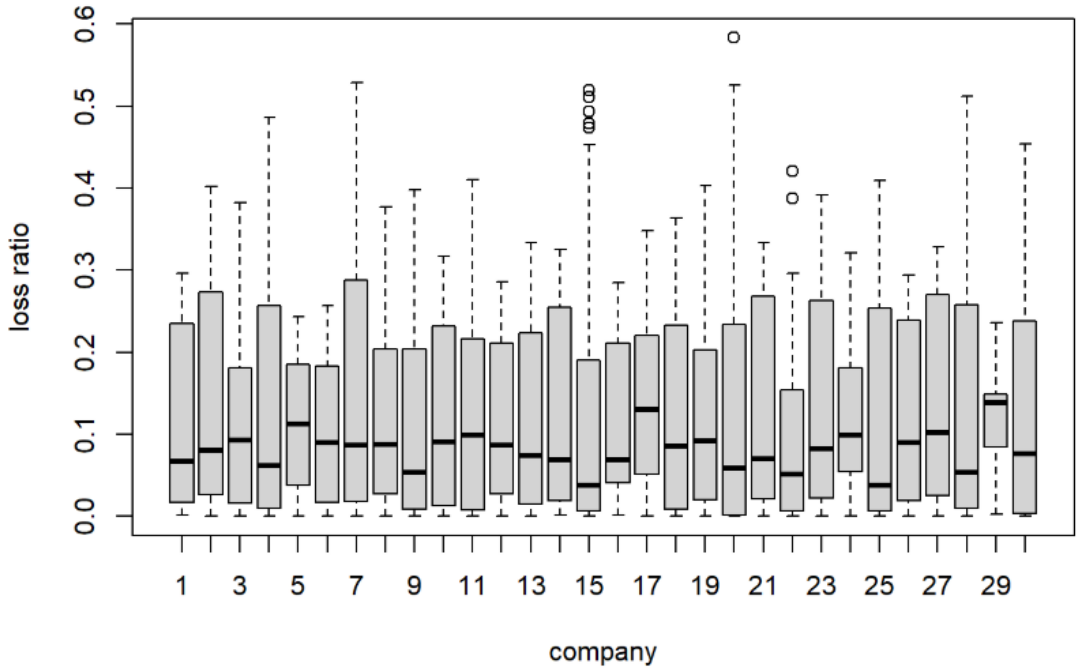

Figure 3.1: The loss ratios for different companies in personal LOB.

We also perform a heterogeneity analysis to capture variability in loss ratios across companies, using the heterogeneity measure I^2 to quantify variation both across companies and between LOBs. Table 3.8 shows the calculated I^2 for the loss ratios in each development year. We observe large I^2 values, indicating the presence of heterogeneity across companies in both LOBs for all development years.

Table 3.8: The heterogeneous measure I^2 for the loss ratios across companies.

	Dev 1	Dev 2	Dev 3	Dev 4	Dev 5	Dev 6	Dev 7	Dev 8	Dev 9
Personal auto	0.9815	0.9308	0.9052	0.9674	0.9675	0.9599	0.9679	0.9580	0.9986
Commercial auto	0.9769	0.7891	0.8265	0.8674	0.9514	0.9579	0.9856	0.9967	0.9998

For each company, we also calculate I^2 between LOBs for the loss ratios in Table 3.9. We find that some companies exhibit low heterogeneity between LOBs, whereas others show higher levels of heterogeneity. For insurers, low heterogeneity may indicate stronger correlations between LOBs, potentially limiting diversification benefits, while high heterogeneity could offer greater opportunities to reduce

overall risk capital through diversification.

Table 3.9: The heterogeneous measure I^2 for the loss ratios between LOBs.

Company code	1	353	388	620	1066	1090	1538	1767	3240	4839
I^2	0.000000	0.005894	0.341242	0.656587	0.000000	0.676405	0.000000	0.588694	0.760000	0.214274
Company code	5185	6947	7080	8427	10022	13420	13439	13641	13889	14044
I^2	0.000000	0.795116	0.899615	0.000000	0.904714	0.200141	0.000000	0.159789	0.949554	0.000000
Company code	14176	14257	15199	18163	25275	27022	27065	29440	31550	34606
I^2	0.844223	0.000000	0.000000	0.000000	0.876128	0.410227	0.000000	0.000000	0.478415	0.904328

To account for the above results, we consider two different scenarios for the company random effects: (1) lower heterogeneity and (2) higher heterogeneity, subsequently referred to as Simulation 1 and Simulation 2. The simulation settings, such as the accident year and development year effects, are included in Appendix B.2.

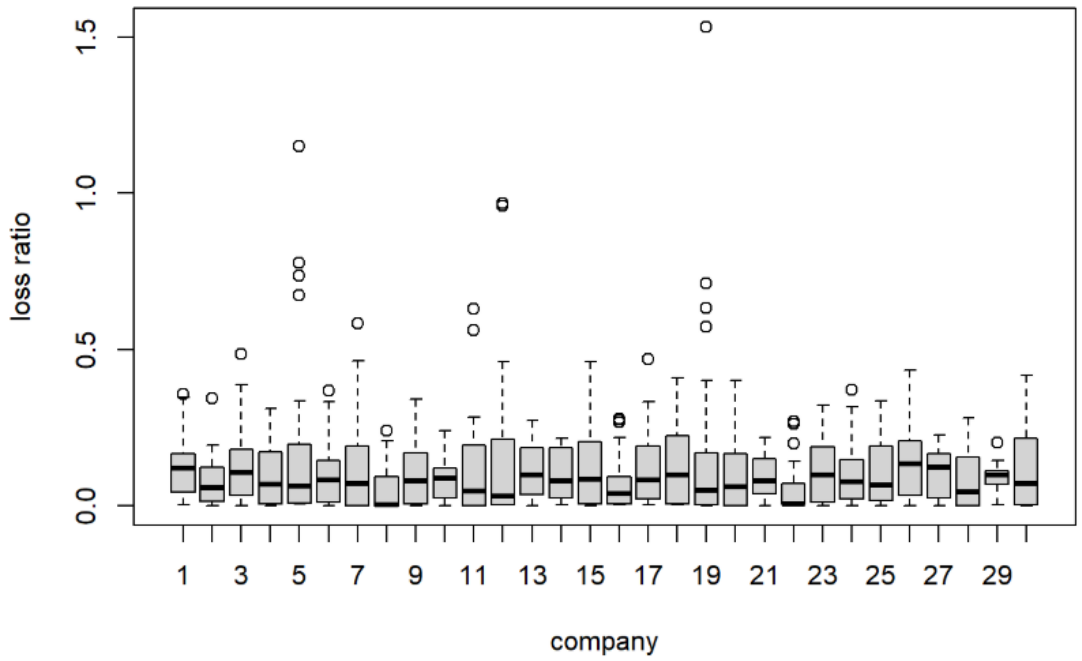

Figure 3.2: The loss ratios for different companies in commercial LOB.

In each scenario, we simulate 30 pairs of loss triangles to reflect multiple companies' data. For each company c, we simulate the company effects from $b_c^{(1)} \sim N(0, \tau_1)$ and $b_c^{(2)} \sim N(0, \tau_2)$ with $\tau_1 = 0.2$ and $\tau_2 = 0.3$. To simulate the losses in the loss triangles $\left(Y_{ijc}^{(1)}, Y_{ijc}^{(2)}\right)$, we first calculate the systematic component $\eta_{ijc}^{(\ell)}(\ell = 1, 2)$ from the accident year and development year effect $\boldsymbol{\beta}^{(\ell)}$ and

company random effect $\boldsymbol{b}^{(\ell)}$.

Next, we simulate $u_{ij}^{(\ell)}(\ell = 1, 2)$ $(i + j - 1 \leq I)$ from Gaussian copula model $c(\cdot; \theta)$ with dependence parameter $\theta = -0.3$. Then, we transform $u_{ij}^{(\ell)}$ to the upper triangles by inverse function $y_{ijc}^{(\ell)} = F^{(\ell)(-1)}(u_{ij}^{(\ell)}; \eta_{ijc}^{(\ell)}, \gamma^{(\ell)})$, where $\eta_{ijc}^{(\ell)} = \boldsymbol{x}_{ij}^{(\ell)}\boldsymbol{\beta}^{(\ell)} + \boldsymbol{z}_c^{(\ell)}\boldsymbol{b}^{(\ell)}$.

Table 3.10: Point estimates of the reserves for Simulation 1 (lower heterogeneity).

Model	Reserves		
	LoB 1, R_1	LoB 2, R_2	Total, R
SUR Gaussian copula	7 575 901	1 058 039	8 633 940
SUR copula mixed	7 522 203	1 033 518	8 555 721

Table 3.11: Point estimates of the reserves for Simulation 2 (higher heterogeneity).

Model	Reserves		
	LoB 1, R_1	LoB 2, R_2	Total, R
SUR Gaussian copula	64 495	139 129	203 624
SUR copula mixed	65 881	132 708	198 589

We apply the SUR copula mixed model to the simulated loss triangles and compare the estimated reserves with those from the SUR Gaussian copula model in Table 3.10 and Table 3.11. We also evaluated the percentage error of the estimated reserves and actual reserves in Table 3.12 and Table 3.13. The SUR copula mixed model outperforms the SUR Gaussian copula models for both LOBs.

Table 3.12: Performance comparison using percentage error of actual and estimated loss reserve for Simulation 1 (lower heterogeneity).

LOB	SUR Gaussian Copula	SUR copula Mixed
Personal Auto	6.5%	5.7%
Commercial Auto	5.1%	2.7%
Total	6.3%	5.3%

For both the SUR copula mixed model and the SUR Gaussian copula model, we conduct bootstrap simulations to generate the predictive distribution of the total reserve. We then computed the risk measures and risk capitals in Table 3.16 and Table 3.17.

Table 3.13: Performance comparison using percentage error of actual and estimated loss reserve for Simulation 2 (higher heterogeneity).

LOB	SUR Gaussian Copula	SUR copula Mixed
Personal Auto	-3.5%	-1.5%
Commercial Auto	11.9%	6.8%
Total	6.5%	3.9%

Table 3.14: Bias, Standard deviation, Coefficient of variation (CV) from the predictive distribution using parametric bootstrapping for Simulation 1 (lower heterogeneity).

	Reserve	Bootstrap reserve	Bias	Std. dev.	CV
SUR Gaussian copula	8 633 940	8 679 885	0.53%	896 505	0.10
SUR copula mixed	8 555 721	8 579 244	0.27%	392 197	0.05

Table 3.15: Bias, Standard deviation, Coefficient of variation (CV) from the predictive distribution using parametric bootstrapping for Simulation 2 (higher heterogeneity).

	Reserve	Bootstrap reserve	Bias	Std. dev.	CV
SUR Gaussian copula	203 624	204 317	0.34%	31 888	0.16
SUR copula mixed	198 589	199 008	0.21%	13 894	0.07

Table 3.16: Risk capital estimation for different methods for Simulation 1 (lower heterogeneity).

Risk measure	TVaR (60%)	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
Silo-GLM	9 759 583	10 269 193	10 470 030	10 756 851	11 209 133	12 182 469
SUR Gaussian copula	9 552 057	9 978 909	10 135 782	10 356 717	10 715 211	11 332 526
SUR copula mixed	8 957 174	9 138 074	9 202 294	9 287 100	9 422 704	9 697 197
Risk capital						
Silo-GLM		509 610	710 447	997 268	1 449 550	2 422 886
SUR Gaussian copula		426 852	583 725	804 660	1 163 154	1 780 469
SUR copula mixed		180 900	245 120	329 926	465 530	740 023

Table 3.17: Risk capital estimation for different methods for Simulation 2 (higher heterogeneity).

Risk measure	TVaR (60%)	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
Silo-GLM	249 075	272 437	281 655	294 005	314 755	355 305
SUR Gaussian copula	235 253	252 084	258 278	266 523	277 778	297 145
SUR copula mixed	212 687	219 275	221 677	224 699	228 961	235 865
Risk capital						
Silo-GLM		23 362	32 580	44 930	65 680	106 230
SUR Gaussian copula		16 831	23 025	31 270	42 525	61 892
SUR copula mixed		6 588	8 990	12 012	16 274	23 178

Compared to the SUR Gaussian copula model, the SUR copula mixed model achieves a smaller bias and a notably lower standard deviation for both cases, re-

Table 3.18: Risk capital gain for different methods for Simulation 1 (lower heterogeneity).

Risk capital gain	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
SUR Gaussian copula vs Silo-GLM	15.68%	17.50%	19.26%	23.62%	24.35%
SUR copula mixed vs Silo-GLM	72.05%	72.44%	72.51%	72.89%	73.14%

Table 3.19: Risk capital gain for different methods for Simulation 2 (higher heterogeneity).

Risk capital gain	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
SUR Gaussian copula vs Silo-GLM	27.96%	29.33%	30.40%	35.25%	41.74%
SUR copula mixed vs Silo-GLM	71.80%	72.41%	73.27%	75.22%	78.18%

sulting in a reduced CV. These results indicate that the SUR copula mixed model produces more accurate reserve predictions, effectively capturing both heterogeneity across companies and LOBs. Consistently, the SUR copula mixed model produces a larger risk capital gain than the SUR Gaussian copula, as shown in Table 3.18 and Table 3.19. The SUR copula mixed model leverages the heterogeneity across companies and obtains a larger risk diversification benefit.

3.6 Summary and Discussion

We have proposed SUR copula mixed models to extend SUR copula regression by incorporating data from multiple companies for improved loss prediction and risk capital analysis, and developed a two-stage approach to estimate the parameters for SUR copula mixed models. In the SUR copula mixed model framework, the model contains the fixed effects and random effect, which characterize the variation induced in the response by different companies. In addition to the point estimate of the reserves, we generate the predictive distribution of the reserves by residual bootstrapping.

To determine whether to include the company random effect, we can use a likelihood ratio test. We fit two models: a null model without the random effect (with only fixed effects) and an alternative model with the random effects (a mixed-effects model). We calculate a test statistic based on the difference in the

log-likelihoods between the two models. The test statistic approximately follows a chi-squared distribution. This test allows to calculate a p-value to determine if the random effect is a significant addition to the model.

Our analysis of real data and simulation studies revealed some limitations of SUR copula mixed models. For the most recent accident and development years, we have progressively fewer observed incremental paid losses. To address this, regularization techniques such as the least absolute shrinkage and selection operator (LASSO) in Tibshirani (1996) can be applied to handle the shrinkage of model parameters. Next, we combine the SUR copula mixed model with LASSO to mitigate the impact of reduced observed paid losses in later accident and development years.

Chapter 4

Sparse SUR (SSUR) Copula Mixed Models

4.1 Introduction

Regularized regression has recently gained traction in the loss reserving literature to handle high-dimensional predictors and improve model interpretability. Williams et al. (2015) applied the elastic net penalty, a convex combination of L1 and L2 penalties, to a dataset with over 350 initial covariates, enhancing predictive performance in insurance claims modeling. McGuire et al. (2018) introduced a loss reserving LASSO framework (Tibshirani, 1996), modeling individual claim data with complex features and applying L1 regularization to stabilize the estimation of loss development factors. Regularization is particularly beneficial in loss reserving, where collinearity and sparse observations can compromise the robustness of the model. Notably, it is straightforward to incorporate LASSO regularization into the SUR framework, enabling effective variable selection while preserving model interpretability.

We propose a sparse SUR copula mixed model to increase the robustness of the SUR copula mixed model. Sparsity is introduced through regularization on

the fixed effects, which promotes variable selection and improves interpretability.

Parametric bootstrapping has been a widely used approach for resampling incremental paid losses, followed by generating the predictive distribution of the loss reserve. Davison and Hinkley (1997) provides a comprehensive framework for bootstrap methodologies in the context of generalized linear models using the pseudo residuals. Kirschner et al. (2008) employs a synchronized parametric bootstrap to model dependencies between correlated lines of business, capturing correlations that existed in the loss triangles when producing the pseudo loss triangles. Taylor and McGuire (2007) further develops this approach in the context of generalized linear models. Most recently, Abdallah et al. (2015) applies the parametric bootstrapping to generate the predictive distribution of the reserve while modeling dependencies between loss triangles using copulas.

While parametric bootstrap remains a natural choice for simulating future reserves, it requires adaptation for sparse models. In the sparse SUR copula mixed model, standard bootstrap methods may fail when some of the components of fixed effects coefficients are zero, leading to instability in the resampling process. To address this, we adopt a modified residual bootstrap based on the method of Chatterjee and Lahiri (2011) to construct a more robust resampling procedure that accommodates sparsity. This approach preserves that similar coefficients are set to zero in all resampling, enabling the stable estimation of the predictive distribution of reserves in the presence of regularized effects.

4.2 Background

4.2.1 The Lasso for Linear Models

First, we summarize the LASSO for linear models. We denote the predictor used in linear regression by $\mathbf{X}$. The parameter vector $\boldsymbol{\beta}$ of the linear regression can be

obtained using the least squares method. Let $\mathbf{y}$ denote the M-dimensional response vector; then the lasso can be written as follows using the Lagrangian form

$$\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \right\}, \tag{4.1}$$

where $\lambda > 0$. Now, let $L(\boldsymbol{\beta}) = \frac{1}{2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1$. That is,

$$\begin{aligned} L(\boldsymbol{\beta}) &= f(\boldsymbol{\beta}) + \lambda \|\boldsymbol{\beta}\|_1 \\ &= \frac{1}{2} \sum_{i=1}^{M} \left[y^{(i)} - \sum_{j=1}^{n} \beta_j x_j^{(i)} \right]^2 + \lambda \sum_{j=1}^{n} |\beta_j| . \end{aligned} \tag{4.2}$$

4.2.2 The Coordinate Descent Method

As in Hastie et al. (2015), we apply the coordinate descent procedure to perform the numerical computation of the solution for the lasso problem. Now we take the derivative of $f(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, and we have

$$\begin{aligned} \frac{\partial}{\partial \beta_j} f(\boldsymbol{\beta}) &= -\sum_{i=1}^{M} x_j^{(i)} \left[y^{(i)} - \sum_{j=1}^{n} \beta_j x_j^{(i)} \right] \\ &= -\sum_{i=1}^{M} x_j^{(i)} \left[y^{(i)} - \sum_{k \neq j}^{n} \beta_k x_k^{(i)} - \beta_j x_j^{(i)} \right] \\ &= -\sum_{i=1}^{M} x_j^{(i)} \left[y^{(i)} - \sum_{k \neq j}^{n} \beta_k x_k^{(i)} \right] + \beta_j \sum_{i=1}^{M} \left(x_j^{(i)} \right)^2 \\ &\triangleq -\rho_j + \beta_j z_j. \end{aligned} \tag{4.3}$$

To perform coordinate descent, we must also isolate β_j for the L_1 term.

$$\lambda \sum_{j=0}^{n} |\beta_j| = \lambda |\beta_j| + \lambda \sum_{k \neq j}^{n} |\beta_k| . \tag{4.4}$$

Optimizing this equation as a function of β_j reduces it to a univariate optimization problem.

$$\partial_{\beta_j}\lambda\sum_{j=0}^{n}|\beta_j| = \partial_{\beta_j}\lambda\,|\beta_j| = \begin{cases}\{-\lambda\} & \text{if } \beta_j < 0\\ [-\lambda,\lambda] & \text{if } \beta_j = 0\\ \{\lambda\} & \text{if } \beta_j > 0\end{cases}. \tag{4.5}$$

Next, we compute the derivative of the Lasso cost function and equate it to zero to find the minimum:

$$\begin{aligned}\partial_{\beta_j} f(\boldsymbol{\beta}) &= \partial_{\beta_j} f(\boldsymbol{\beta}) + \partial_{\beta_j}\lambda\|\boldsymbol{\beta}\|_1\\ 0 &= -\rho_j + \beta_j z_j + \partial_{\beta_j}\lambda\,\|\beta_j\|\\ 0 &= \begin{cases}-\rho_j + \beta_j z_j - \lambda & \text{if } \beta_j < 0\\ [-\rho_j - \lambda, -\rho_j + \lambda] & \text{if } \beta_j = 0\\ -\rho_j + \beta_j z_j + \lambda & \text{if } \beta_j > 0.\end{cases}\end{aligned} \tag{4.6}$$

For the second case, we must ensure the closed interval contains zero

$$\begin{aligned}0 &\in [-\rho_j - \lambda, -\rho_j + \lambda]\\ -\rho_j - \lambda &\leq 0\\ -\rho_j + \lambda &\geq 0\\ -\lambda \leq \rho_j &\leq \lambda.\end{aligned}$$

Solving for β_j for all three cases, we have

$$\begin{cases}\beta_j = \frac{\rho_j+\lambda}{z_j} & \text{for } \rho_j < -\lambda\\ \beta_j = 0 & \text{for } -\lambda \leq \rho_j \leq \lambda\\ \beta_j = \frac{\rho_j-\lambda}{z_j} & \text{for } \rho_j > \lambda\end{cases}. \tag{4.7}$$

Lasso is an effective tool to eliminate noisy covariates from a large set of candidates since its loss function can force many components β_k to zero. The term

$\lambda||\beta||_1$ serves as a penalty for the parameters included in the model. The penalty increases with increasing λ. When λ is close to zero, there is no elimination of covariates. When λ goes to infinity, we eliminate all covariates.

For lasso regression, we have a sequence of models corresponding to different choices of λ. It is important to choose λ since it controls the bias-variance trade-off. A suitable λ can improve the prediction accuracy and interpretability. If the regularization is too strong, many important covariates may be omitted, which decrease prediction accuracy. The optimal λ is the one that minimizes the mean squared error in the validation set.

We can use cross-validation to evaluate the prediction accuracy of the model produced by lasso. It consists of the following steps:

1 Randomly select one n-th of the data set as a validation sample;

2 Train the model on the remainder of the data set;

3 Use the trained model to generate fitted values on the test data;

4 Compute the mean squared error between the fitted values and actual values for the validation set;

5 Repeat steps [1] to [4] many times. The cross-validation error is the average of the mean squared errors.

4.2.3 The Lasso for Generalized Linear Models

We can fit GLM by maximizing the likelihood, or equivalently, minimizing the negative log-likelihood along with an ℓ_1 penalty

$$\min_{\boldsymbol{\beta}} \left\{ -\frac{1}{M} L\left(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X}\right) + \lambda \|\boldsymbol{\beta}\|_1 \right\}, \tag{4.8}$$

where $\mathbf{y}$ is the M-vector of outcomes and $\mathbf{X}$ is the $M \times p$ matrix of predictors. The specific form of the log-likelihood L depends on the GLM used. Some examples are the following.

If the responses follow a Gaussian distribution, we have $L(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) = \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2$. Then the optimization problem in (4.8) corresponds to the ordinary linear least-squares lasso.

If the response is binary, we estimate the probability $P(y = 1)$. Then, the negative log likelihood with ℓ_1 penalty takes the form

$$\begin{aligned} & -\frac{1}{M}\sum_{i=1}^{M}\left\{y_i \log \mathrm{P}\left(Y=1 \mid x_i\right)+\left(1-y_i\right) \log \mathrm{P}\left(Y=0 \mid x_i\right)\right\}+\lambda\|\boldsymbol{\beta}\|_1 \\ & =-\frac{1}{M}\sum_{i=1}^{M}\left\{y_i\left(\beta_0+x_i^T \boldsymbol{\beta}\right)-\log\left(1+e^{\beta_0+x_i^T \beta}\right)\right\}+\lambda\|\boldsymbol{\beta}\|_1 \\ & =-\ell(\beta_0, \boldsymbol{\beta})+\lambda\|\boldsymbol{\beta}\|_1. \end{aligned} \tag{4.9}$$

We can use iteratively reweighted least squares (Holland and Welsch, 1977) to maximize the log-likelihood $\ell(\beta_0, \boldsymbol{\beta})$. Given the current estimates of the parameters $(\tilde{\beta}_0, \tilde{\boldsymbol{\beta}})$, we form a quadratic approximation (Hastie et al., 2015) to the log-likelihood

$$\ell_Q\left(\beta_0, \boldsymbol{\beta}\right)=-\frac{1}{2M}\sum_{i=1}^{N} w_i\left(z_i-\beta_0-x_i^T \boldsymbol{\beta}\right)^2+C\left(\tilde{\beta}_0, \tilde{\boldsymbol{\beta}}\right)^2, \tag{4.10}$$

where

$$z_i=\tilde{\beta}_0+x_i^T \tilde{\boldsymbol{\beta}}+\frac{y_i-\tilde{p}\left(x_i\right)}{\tilde{p}\left(x_i\right)\left(1-\tilde{p}\left(x_i\right)\right)}, \tag{4.11}$$

$$w_i=\tilde{p}\left(x_i\right)\left(1-\tilde{p}\left(x_i\right)\right), \tag{4.12}$$

where $\tilde{p}\,(x_i)$ is evaluated at the current parameters $(\tilde{\beta}_0, \tilde{\boldsymbol{\beta}})$ and the term $C\left(\tilde{\beta}_0, \tilde{\boldsymbol{\beta}}\right)^2$ is constant.

For each value of λ, we compute the quadratic approximation ℓ_Q at the current parameters $(\beta_0, \boldsymbol{\beta})$. We use coordinate descent to solve the penalized weighted least-squares problem

$$-\min_{\boldsymbol{\beta}} \left\{\ell_Q(\beta_0, \boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1\right\}. \tag{4.13}$$

This requires a sequence of three nested loops:

- outer loop: Decrement λ.
- middle loop: Form the quadratic function $\ell_Q(\beta_0, \boldsymbol{\beta})$ at the current parameters $(\tilde{\beta}_0, \tilde{\boldsymbol{\beta}})$
- inner loop: Solve the penalized weighted least squares problem using the coordinate descent algorithm.

As an example, if a response variable is non-negative integer, we often use the Poisson distribution. To ensure the positivity of the mean, we usually choose the log link. Thus, the GLM is

$$\log \mu = \beta_0 + x^T\boldsymbol{\beta}. \tag{4.14}$$

Then the negative log likelihood with ℓ_1 penalty takes the form

$$-\frac{1}{M}\sum_{i=1}^{M}\left\{y_i\left(\beta_0 + \boldsymbol{\beta}^T x_i\right) - e^{\beta_0+\boldsymbol{\beta}^T x_i}\right\} + \lambda\|\boldsymbol{\beta}\|_1. \tag{4.15}$$

4.3 Proposed Sparse SUR Model

For the most recent accident and development years, the number of observed incremental paid losses progressively decreases. To improve the robustness of the

SUR copula mixed model under this sparse history of incremental paid losses, we propose applying shrinkage techniques to the fixed effects parameters. Specifically, we incorporate the least absolute shrinkage and selection operator (LASSO) into the SUR copula mixed model to shrink the fixed effects coefficients toward zero and reduce their variability.

We construct the loss function as the negative log-likelihood of equation (3.33) combined with an $\mathcal{L}_1$ penalty on $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$. The Lagrangian form of the penalized loss function includes tuning parameters λ_1 and λ_2 corresponding to the penalties on $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$, respectively.

$$
\begin{aligned}
&J_{\lambda_1,\lambda_2}(\boldsymbol{\beta}^{(1)},\boldsymbol{\beta}^{(2)},\sigma_1,\sigma_2,\tau_1,\tau_2,\theta \mid \boldsymbol{y}_1^{(1)},\boldsymbol{y}_2^{(1)},\ldots,\boldsymbol{y}_C^{(1)},\boldsymbol{y}_1^{(2)},\boldsymbol{y}_2^{(2)},\ldots,\boldsymbol{y}_C^{(2)}) \\
&= -\sum_{\ell=1}^{2} L^{(\ell)}(\boldsymbol{\beta}^{(\ell)},\sigma_\ell,\tau_\ell \mid \boldsymbol{y}_1^{(\ell)},\ldots,\boldsymbol{y}_C^{(\ell)}) - \sum_{c=1}^{C}\sum_{i=1}^{I}\sum_{j=1}^{I+1-i} \log c(R_{ijc}^{(1)},R_{ijc}^{(2)};\theta_c) \\
&+ \lambda_1||\boldsymbol{\beta}^{(1)}||_1 + \lambda_2||\boldsymbol{\beta}^{(2)}||_1 \\
&= -\sum_{\ell=1}^{2}\sum_{c=1}^{C} \log \int_{-\infty}^{\infty} \prod_{i=1}^{I}\prod_{j=1}^{I+1-i} f(y_{ijc}^{(\ell)}|b_c^{(\ell)},\boldsymbol{\beta}^{(\ell)},\sigma_\ell) f(b_c^{(\ell)};\tau_\ell) db_c^{(\ell)} + \lambda_1||\boldsymbol{\beta}^{(1)}||_1 + \lambda_2||\boldsymbol{\beta}^{(2)}||_1 \\
&- \sum_{c=1}^{C}\sum_{i=1}^{I}\sum_{j=1}^{I+1-i} \log c(R_{ijc}^{(1)},R_{ijc}^{(2)};\theta_c). \qquad (4.16)
\end{aligned}
$$

We aim to estimate the parameters $\boldsymbol{\beta}^{(\ell)}$, σ_ℓ, τ_ℓ and θ by minimizing the penalized loss function in (4.16). The estimation procedure follows the two-stage iterative procedure as detailed in section 3.3.2, but with a modification to step 1. As an initialization, we obtain estimates of $\boldsymbol{\beta}^{(\ell)}$, σ_ℓ, and τ_ℓ by fitting a generalized linear mixed model to the chosen distribution of $Y_{ijc}^{(\ell)}$.

In step 1, the penalized loss function in (4.16) depends on the regularization parameters λ_1 and λ_2, which control the degree of shrinkage for $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$, respectively. We select these parameters using the Akaike Information Criterion

(AIC), defined as

$$\text{AIC} = -2\log L(\hat{\boldsymbol{\beta}}^{(1)}, \hat{\boldsymbol{\beta}}^{(2)}, \hat{\sigma}_1, \hat{\sigma}_2, \hat{\tau}_1, \hat{\tau}_2, \hat{\theta}) + 2 \cdot \hat{df}_{\lambda_1,\lambda_2}, \tag{4.17}$$

where, $\hat{df}_{\lambda_1,\lambda_2} = |\{1 \leq k \leq I : \hat{\beta}_k^{(\ell)} \neq 0\}| + 5$, accounting for the number of nonzero fixed-effect coefficients, the two variance parameters, the two dispersion parameters, and the dependence parameter.

For each pair of λ_1 and λ_2, we maximize (4.16) with respect to $\boldsymbol{\beta}^{(\ell)}$ and select the pair that minimizes AIC in (4.17). Given the optimal $\boldsymbol{\beta}^{(\ell)}$, we then itertavely update remianing parameters σ_ℓ, and τ_ℓ in step 1 and θ in step 2.

We summarize the steps to estimate the parameters in (4.16) as follows:

(1) Given initial values $\boldsymbol{\beta}^{(\ell)(0)}$, $\sigma_\ell^{(0)}$, and $\tau_\ell^{(0)}$, which are from generalized linear mixed model, log-likelihood function (4.16) is a function of copula parameter θ: $L_2(\theta) = \sum_{c=1}^{C}\sum_{i=1}^{I}\sum_{j=1}^{I+1-i} \log c(F(y_{ijc}^{(\ell)}), F(y_{ijc}^{(\ell')}); \theta_c)$. Set iteration $k = 0$.

(2) Maximize pseudo log-likelihood $L_3(\theta)$ with respect to θ and obtain $\theta^{(k)}$. Here $L_3(\theta) = \sum_{c=1}^{C}\sum_{i=1}^{I}\sum_{j=1}^{I+1-i} \log c(R_{ijc}^{(\ell)}, R_{ijc}^{(\ell')}; \theta_c)$.

(3) Given $\theta^{(k)}$, for different λ_1 and λ_2, maximize (4.16) with respect to $\boldsymbol{\beta}^{(\ell)}$. λ_1 and λ_2 are selected using AIC/BIC.

(4) Given $\theta^{(k)}$ and $\boldsymbol{\beta}^{(\ell)(k)}$, maximize (4.16) with respect to σ_ℓ and τ_ℓ. update k=k+1.

(5) Repeat steps (2) and (4) until it meets the stopping criterion: $||\boldsymbol{w}^{k+1} - \boldsymbol{w}^k||_2 \leq \epsilon$, where we group parameters $\theta, \lambda_1, \lambda_2$ under vector $\boldsymbol{w}$.

The AIC criterion selects the model that minimizes

$$AIC(\lambda) = -2L_\lambda + 2 \cdot df_\lambda, \tag{4.18}$$

where L_λ is the maximum log-likelihood for the λth model and df_λ is the sum of the number of nonzero fixed-effect coefficients and the number of covariance parameters.

The BIC has a similar form as the AIC, with the exception that the log-likelihood is penalized by $\log n$ instead of 2, where n is the number of samples. The BIC criterion selects the model that minimizes:

$$BIC(\lambda) = -2L_\lambda + \log(n) \cdot df_\lambda. \tag{4.19}$$

Based on a modified version of the bootstrap lasso estimator by Chatterjee and Lahiri (2011), we implement the following bootstrapping steps to generate the reserve's predictive distribution.

(1) Fit the sparse SUR copula mixed model to the observed incremental paid losses $y_{ijc}^{(\ell)}$ and choose the optimal penalization parameter $\hat{\lambda}_1$ and $\hat{\lambda}_2$.

(2) Calculate the thresholded coefficients $\tilde{\boldsymbol{\beta}}$, where we force components of $\hat{\boldsymbol{\beta}}$ to be exactly zero whenever they are close to zero.

(3) Generate the residuals for the incremental paid losses using:

$$\epsilon_{ijc}^{(\ell)} = \frac{y_{ijc}^{(\ell)} - \tilde{\mu}_{ijc}^{(\ell)}}{(\tilde{\mu}_{ijc}^{(\ell)} * \hat{\gamma}_{ijc}^{(\ell)})^{1/2}}, \tag{4.20}$$

where $\tilde{\mu}_{ijc}^{(\ell)} = \exp(\boldsymbol{x}_{ij}^{(\ell)}\tilde{\boldsymbol{\beta}}^{(\ell)} + \boldsymbol{z}_c^{(\ell)}\hat{\boldsymbol{b}}^{(\ell)})$.

(4) Begin the iterative loop, to be repeated N times (e.g., N = 1000):

- Simulate $\left(u_{ijc}^{(1)}, u_{ijc}^{(2)}\right)$ $(i + j - 1 \leq I)$ from estimated copula function $C(\cdot; \hat{\theta})$.
- Transform $u_{ijc}^{(\ell)}$ to the residuals by transform $\epsilon_{ijc}^{*(\ell)} = Q^{(\ell)}(u_{ijc}^{(\ell)})$, where $Q^{(\ell)}$ is the empirical quantile function of the residuals.

- Generate a set of pseudo incremental paid losses $y_{ijc}^{*(\ell)}$, which is given by $y_{ijc}^{*(\ell)} = \epsilon_{ijc}^{*(\ell)} * (\hat{\mu}_{ijc}^{(\ell)} * \hat{\gamma}_{ijc}^{(\ell)})^{1/2} + \hat{\mu}_{ijc}^{(\ell)}$.
- Estimate the parameters $\hat{\boldsymbol{\beta}}^{*(\ell)}$, $\hat{\sigma_\ell}^*$, and $\hat{\tau_\ell}^*$ and $\hat{\theta}^*$ for $y_{ijc}^{*(\ell)}$ using the sparse SUR copula mixed model with the optimal penalization parameter $\hat{\lambda}_1$ and $\hat{\lambda}_2$.
- Obtain a prediction of the total reserve for company c by

$$\sum_{\ell=1}^{2} \sum_{i=2}^{I} \sum_{j=I-i+2}^{I} \omega_{ic}^{(\ell)} \cdot \exp(\hat{\eta}_{ijc}^{*(\ell)}),$$

where $\hat{\eta}_{ijc}^{*(\ell)} = \boldsymbol{x}_{ij}^{(\ell)} \hat{\boldsymbol{\beta}}^{*(\ell)} + \boldsymbol{z}_c^{(\ell)} \hat{\boldsymbol{b}}^{*(\ell)}$ and $\omega_{ic}^{(\ell)}$ is the premium for accident year i and company c in LOB ℓ.

4.4 Application

We apply the sparse SUR copula mixed model to a dataset of 30 pairs of loss triangles from Schedule P of the National Association of Insurance Commissioners (NAIC) database (Meyers and Shi, 2011). We demonstrate the reserve prediction and risk capital analysis for a major US property-casualty insurer. We model the loss ratios in both LOBs with a gamma distribution. By minimizing the loss function in (4.16), we obtain the estimated shape parameters for the two LOBs: $\sigma_1 = 2.01$ and $\sigma_2 = 1.10$. The estimated standard deviations for the company random effects are $\tau_1 = 0.30$ and $\tau_2 = 0.36$, respectively. These indicate that the volatility in the commercial LOB is higher, and the variation induced by the company is larger in the commercial LOB than in the personal LOB. We use a Gaussian copula to model the dependence between the two LOBs. For the sparse SUR copula mixed model, the estimated dependence parameter is -0.27, which indicates a negative association between the two LOBs. As shown in Table 4.1, the estimated reserve for the personal LOB is 7 289 760 while the reserve for the

commercial LOB is 372 992.

Table 4.1: Point estimates of the reserves.

Model	Reserves		
	LoB 1, R_1	LoB 2, R_2	Total, R
Sparse SUR copula mixed	7 295 694	372 761	7 668 455
SUR Gaussian copula	6 823 325	378 386	7 364 511
Actual reserve	8 086 094	318 380	8 404 474

Table 4.2: Performance comparison using percentage error of actual and estimated loss reserve.

	Personal Auto	Commercial Auto	Total
Sparse SUR copula mixed	-9.8%	17.1%	-8.8%
SUR copula mixed	-10.3%	18.5%	-9.3%
SUR Gaussian copula	-15.6%	19.0%	-12.4%

Table 4.2 compares the models by showing the percentage error between their estimated reserves and the actual reserves. The sparse SUR copula mixed model consistently demonstrates a smaller error than the SUR copula mixed model when predicting reserves for both LOBs. This indicates that the sparse SUR copula mixed model effectively handles heterogeneity in data from multiple companies and the shrinkage of model parameters.

Table 4.3: Bias, Standard deviation, Coefficient of variation (CV) from the predictive distribution using parametric bootstrapping.

	Reserve	Bootstrap reserve	Bias	Std. dev.	CV
SUR copula mixed	7 623 460	7 530 255	1.22%	612 947	0.082
Sparse SUR copula mixed	7 668 319	7 576 596	1.19%	514 371	0.068

Table 4.3 compares the bias, standard deviation, and coefficient of variation (CV) of the predictive distribution for different models. The sparse SUR copula mixed model achieves similar bias but lowers the standard deviation, indicating greater stability in predictions. With the predictive distribution we generated, Table 4.4 show the 95% confidence interval of the total reserve, where the lower bound and upper bound are the 2.5th and 97.5th percentiles of the predictive

distribution, respectively. We observe that the actual reserve (8 404 474) is within the 95% confidence interval of the reserve for both models.

Table 4.4: 95% confidence intervals for the total reserve using the predictive distribution.

	Lower bound	Upper bound
SUR copula mixed	6 548 624	8 892 605
Sparse SUR copula mixed	6 748 291	8 608 556

We compared the predictive distribution of the reserve from sparse SUR copula mixed model with that from the SUR copula mixed model. Though Figure 4.1 doesn't clearly depict the differences in the predictive distribution, Figure 4.2 shows that the sparse SUR copula mixed model generates a shorter tail than SUR copula mixed model.

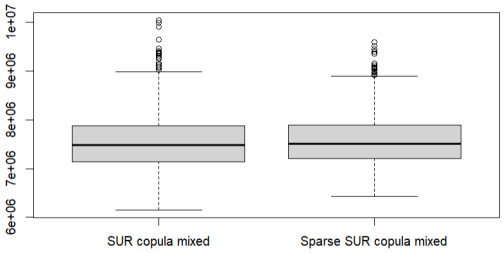

Figure 4.1: Boxplot of the predictive distribution of reserves for different models.

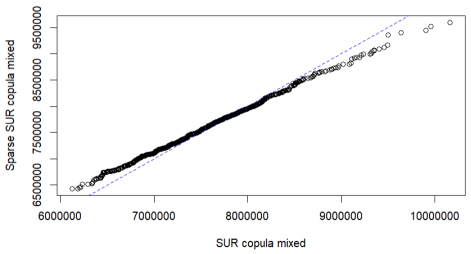

Figure 4.2: QQ plot of the predictive distribution of reserves for different models.

Table 4.5: Risk capital estimation for different methods.

Risk measure	TVaR (60%)	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
Silo-GLM	9 632 534	10 799 117	11 248 347	11 863 905	12 950 137	15 168 393
SUR Gaussian copula	8 992 101	9 975 813	10 325 124	10 793 817	11 526 223	13 179 997
Sparse SUR copula mixed	8 035 364	8 300 898	8 400 387	8 525 030	8 742 768	9 224 643
Risk capital						
Silo-GLM		1 166 583	1 615 813	2 231 371	3 317 603	5 535 859
SUR Gaussian copula		983 712	1 333 023	1 801 716	2 534 122	4 187 896
Sparse SUR copula mixed		265 534	365 023	489 666	707 404	1 189 279

Table 4.6: Risk capital gain for different methods.

Risk capital gain	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
SUR Gaussian copula vs Silo-GLM	15.68%	17.50%	19.26%	23.62%	24.35%
Sparse SUR copula mixed vs Silo-GLM	76.16%	77.05%	77.91%	78.03%	78.59%

We show in Table 4.6 the gain in terms of risk capital for sparse SUR copula mixed models compared to the silo method. The sparse SUR copula mixed models captures the dependence between the two LOBs with one or multiple dependence parameters and generates larger risk capital gain compared to the SUR copula model. The sparse SUR copula mixed model effectively reduces the effect of model parameter shrinkage, resulting in the largest gain in risk capital. We found that the sparse SUR copula mixed model provides the greatest diversification benefits and the most efficient use of capital.

4.5 Simulation Study

In this simulation study, we simulate 30 pairs of loss triangles to represent data from multiple companies. Each pair consists of one loss triangle for the personal LOB and one for the commercial LOB. For each company c, we simulate the company random effects using $b_c^{(1)} \sim N(0, \tau_1)$ and $b_c^{(2)} \sim N(0, \tau_2)$ with $\tau_1 = 0.2$ and $\tau_2 = 0.3$. To simulate the losses in the loss triangles $\left(Y_{ijc}^{(1)}, Y_{ijc}^{(2)}\right)$, we first calculate the systematic component $\eta_{ijc}^{(\ell)}(\ell = 1, 2)$ from the accident year and development year effect $\boldsymbol{\beta}^{(\ell)}$ and company random effect $\boldsymbol{b}^{(\ell)}$. One of the accident year effect parameters is set to 0 to reflect the sparsity in the simulated data; this scenario is referred to as Simulation Setting 1. In another scenario, we set one of the development year parameters to 0, which we refer to as Simulation Setting 2. In Simulation Setting 3, we set an accident year effect parameter to 0 and a development year parameter to 0.

We use Gaussian copula model $c(\cdot; \theta)$ with dependence parameter $\theta = -0.3$ to simulate $u_{ij}^{(\ell)}(\ell = 1, 2)$ $(i + j - 1 \leq I)$. Then we obtain the upper triangles by inverse function $y_{ijc}^{(\ell)} = F^{(\ell)(-1)}(u_{ij}^{(\ell)}; \eta_{ijc}^{(\ell)}, \gamma^{(\ell)})$, where $\eta_{ijc}^{(\ell)} = \boldsymbol{x}_{ij}^{(\ell)}\boldsymbol{\beta}^{(\ell)} + \boldsymbol{z}_c^{(\ell)}\boldsymbol{b}^{(\ell)}$. Finally, the incremental paid losses, $\left(X_{ijc}^{(1)}, X_{ijc}^{(2)}\right)$ are obtained by multiplying the simulated $y_{ijc}^{(\ell)}$ by the premium for the i-th accident year.

Table 4.7: Point estimates of the reserves for Simulation Setting 1 (sparsity in accident years).

	Reserves		
Model	LoB 1, R_1	LoB 2, R_2	Total, R
Sparse SUR copula mixed	7 387 615	994 914	8 382 530
SUR copula mixed	7 465 672	1 017 500	8 483 172
SUR Gaussian copula	8 597 787	762 678	9 360 465

Table 4.8: Point estimates of the reserves for Simulation Setting 2 (sparsity in development years).

	Reserves		
Model	LoB 1, R_1	LoB 2, R_2	Total, R
Sparse SUR copula mixed	7 230 840	1 100 206	8 331 046
SUR copula mixed	7 253 686	1 110 404	8 364 090
SUR Gaussian copula	5 884 783	1 220 056	7 104 839

Table 4.9: Point estimates of the reserves for Simulation Setting 3 (sparsity in accident and development years).

	Reserves		
Model	LoB 1, R_1	LoB 2, R_2	Total, R
Sparse SUR copula mixed	6 703 487	1 013 406	7 716 893
SUR copula mixed	6 688 865	1 017 392	7 716 257
SUR Gaussian copula	6 019 797	1 267 551	7 287 349

We apply the sparse SUR copula mixed model to the simulated loss triangles and compare the estimated reserves with those from the SUR copula model in Table 4.7, Table 4.8, and Table 4.9. The estimated dependence parameter from the sparse SUR copula mixed model is -0.29, which is close to the true dependence parameter of -0.3.

Table 4.10: Performance comparison using percentage error of actual and estimated loss reserve for Simulation Setting 1 (sparsity in accident years).

LOB	Sparse SUR copula Mixed	SUR copula Mixed	SUR Gaussian Copula
Personal Auto	4.2%	4.9%	20.8%
Commercial Auto	1.8%	3.3%	-22.6%
Total	3.9%	4.7%	15.6%

Next, we compute the percentage error between the estimated reserves and the actual reserves in Table 4.10, Table 4.11, and Table 4.12. The sparse SUR

Table 4.11: Performance comparison using percentage error of actual and estimated loss reserve for Simulation Setting 2 (sparsity in development years).

LOB	Sparse SUR copula Mixed	SUR copula Mixed	SUR Gaussian Copula
Personal Auto	1.6%	1.9%	17.3%
Commercial Auto	10.7%	11.8%	22.9%
Total	2.7%	3.2%	12.3%

Table 4.12: Performance comparison using percentage error of actual and estimated loss reserve for Simulation Setting 3 (sparsity in accident and development years).

LOB	Sparse SUR copula Mixed	SUR copula Mixed	SUR Gaussian Copula
Personal Auto	-5.7%	-6.0%	-15.4%
Commercial Auto	4.2%	4.7%	30.4%
Total	-4.5%	-4.6%	-9.9%

copula mixed model produces smaller percentage errors in both LOBs than the SUR copula model. Comparing the percentage errors in Table 4.10 and Table 4.11, we find that the sparse SUR copula mixed model effectively accounts for the reduced number of incremental paid losses history in recent development and accident years.

Table 4.13: Bias, Standard deviation, Coefficient of variation (CV) from the predictive distribution using parametric bootstrapping for Simulation Setting 1 (sparsity in accident years).

	Reserve	Bootstrap reserve	Bias	Std. dev.	CV
SUR Gaussian copula	9 360 465	9 473 081	1.21%	1 430 188	0.15
Sparse SUR copula mixed	8 418 298	8 362 206	0.67%	834 910	0.10

Table 4.14: Bias, Standard deviation, Coefficient of variation (CV) from the predictive distribution using parametric bootstrapping for Simulation Setting 2 (sparsity in development years).

	Reserve	Bootstrap reserve	Bias	Std. dev.	CV
SUR Gaussian copula	7 104 839	7 159 126	0.76%	1 133 194	0.16
Sparse SUR copula mixed	8 331 046	8 442 766	1.34%	849 607	0.11

To generate the predictive distribution of the reserves, we perform the proposed modified bootstrap as outlined in section 4.3 for both the SUR copula model

Table 4.15: Bias, Standard deviation, Coefficient of variation (CV) from the predictive distribution using parametric bootstrapping for Simulation Setting 3 (sparsity in accident and development years).

	Reserve	Bootstrap reserve	Bias	Std. dev.	CV
SUR Gaussian copula	7 287 349	7 351 670	0.88%	1 047 167	0.14
Sparse SUR copula mixed	7 716 893	7 751 918	0.45%	743 528	0.09

Table 4.16: Risk capital estimation for different methods for Simulation Setting 1 (sparsity in accident years).

Risk measure	TVaR (60%)	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
Silo-GLM	10 885 418	11 723 992	12 057 173	12 527 821	13 269 161	15 100 527
SUR Gaussian copula	10 858 436	11 608 310	11 899 370	12 259 013	12 884 829	14 109 723
Sparse SUR copula mixed	9 155 447	9 564 475	9 696 652	9 879 661	10 166 494	10 828 871
Risk capital						
Silo-GLM		838 574	1 171 755	1 642 403	2 383 743	4 215 109
SUR Gaussian copula		749 874	1 040 934	1 400 577	2 026 393	3 251 287
Sparse SUR copula mixed		409 028	541 205	724 214	1 011 047	1 673 424
True risk capital		363 090	491 570	649 916	883 450	1 339 463

Table 4.17: Risk capital estimation for different methods for Simulation Setting 2 (sparsity in development years).

Risk measure	TVaR (60%)	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
Silo-GLM	8 430 836	9 113 243	9 381 361	9 749 298	10 351 488	11 742 042
SUR Gaussian copula	8 267 915	8 876 446	9 108 225	9 398 868	9 895 646	11 031 386
Sparse SUR copula mixed	9 266 126	9 682 246	9 840 694	10 053 377	10 392 631	11 170 969
Risk capital						
Silo-GLM		682 407	950 525	1 318 462	1 920 652	3 311 206
SUR Gaussian copula		608 531	840 310	1 130 953	1 627 731	2 763 471
Sparse SUR copula mixed		416 120	574 568	787 251	1 126 505	1 904 843
True risk capital		327 789	450 061	613 978	884 873	1 408 852

Table 4.18: Risk capital estimation for different methods for Simulation Setting 3 (sparsity in accident and development years).

Risk measure	TVaR (60%)	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
Silo-GLM	8 753 481	9 520 564	9 820 696	10 232 329	10 900 412	12 165 368
SUR Gaussian copula	8 369 106	8 935 070	9 143 298	9 426 176	9 895 287	10 743 208
Sparse SUR copula mixed	8 466 662	8 841 829	8 985 442	9 175 143	9 454 174	10 008 244
Risk capital						
Silo-GLM		767 083	1 067 215	1 478 848	2 146 931	3 411 887
SUR Gaussian copula		565 964	774 192	1 057 070	1 526 181	2 374 102
Sparse SUR copula mixed		375 167	518 780	708 481	987 512	1 541 582
True risk capital		313 950	429 705	582 822	827 869	1 360 640

Table 4.19: Risk capital gain for different methods for Simulation Setting 1 (sparsity in accident years).

Risk capital gain	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
SUR Gaussian copula vs Silo-GLM	10.58%	11.16%	14.72%	14.99%	22.87%
Sparse SUR copula mixed vs Silo-GLM	51.22%	53.81%	55.91%	57.59%	60.30%

Table 4.20: Risk capital gain for different methods for Simulation Setting 2 (sparsity in development years).

Risk capital gain	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
SUR Gaussian copula vs Silo-GLM	10.83%	11.60%	14.22%	15.25%	16.54%
Sparse SUR copula mixed vs Silo-GLM	39.02%	39.55%	40.29%	41.35%	42.47%

Table 4.21: Risk capital gain for different methods for Simulation Setting 3 (sparsity in accident and development years).

Risk capital gain	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
SUR Gaussian copula vs Silo-GLM	26.22%	27.46%	28.52%	28.91%	30.42%
Sparse SUR copula mixed vs Silo-GLM	51.09%	51.39%	52.09%	54.00%	54.82%

Table 4.22: Risk capital percentage error for different methods for Simulation Setting 1 (sparsity in accident years).

Risk capital	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
Silo-GLM	130.95%	138.37%	152.71%	169.82%	214.69%
SUR Gaussian copula	106.53%	111.76%	115.50%	129.37%	142.73%
Sparse SUR copula mixed	12.65%	10.10%	11.43%	14.44%	24.93%

Table 4.23: Risk capital percentage error for different methods for Simulation Setting 2 (sparsity in development years).

Risk capital	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
Silo-GLM	108.18%	111.20%	114.74%	117.05%	135.03%
SUR Gaussian copula	85.65%	86.71%	84.20%	83.95%	96.15%
Sparse SUR copula mixed	26.95%	27.66%	28.22%	27.31%	35.21%

Table 4.24: Risk capital percentage error for different methods for Simulation Setting 3 (sparsity in accident and development years).

Risk capital	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
Silo-GLM	144.33%	148.36%	153.74%	159.33%	150.76%
SUR Gaussian copula	80.27%	80.17%	81.37%	84.35%	74.48%
Sparse SUR copula mixed	19.50%	20.73%	21.56%	19.28%	13.30%

and the sparse SUR copula mixed model. Table 4.13, Table 4.14, and Table 4.15 show the bias and standard deviations of the total loss reserve from the predictive distribution. The sparse SUR copula mixed model effectively handles both variations in reduced number of histories in loss ratios across different LOBs., leading to smaller standard deviations compared to the SUR copula model.

For both the sparse SUR copula mixed model and the SUR Gaussian copula model, we compute the risk measures and risk capitals for different risk levels in Table 4.16, Table 4.17, and Table 4.18. Similar to real data application, the

sparse SUR copula mixed model produces a larger risk capital gain than the SUR Gaussian copula, as shown in Table 4.19, Table 4.20, and Table 4.21.

We also compared the calculated risk capital with the true risk capital for each simulation setting. As shown in Table 4.22, Table 4.23, and Table 4.24, the sparse SUR copula mixed model generates risk capital closest to the true risk capital.

4.6 Summary and Discussion

We have proposed the sparse SUR copula mixed model to incorporate data from multiple companies and handle the shrinkage of model parameters, thereby improving predictions of reserves and risk capital. The model consists of three components: fixed accident year and development year effects, company random effects, and a copula to model dependence between LOBs. We estimate the parameters for the model using a two-stage iterative approach that alternates between estimating fixed and random effects and estimating the dependence parameter. We apply coefficient thresholding in the bootstrapping to generate the predictive distribution of the reserves.

We demonstrate the method using both real data from NAIC database and simulation studies. Empirical and simulation results show that the sparse SUR copula mixed model generates smaller prediction errors, reduced variability, and larger risk capital gains than the SUR copula mixed model without sparsity and the SUR Gaussian copula model. This is due to its ability to capture dependence between LOBs, and account for variability across companies, and remain robust under sparsity.

Both Gaussian and Frank copulas can model a wide range of dependence, from positive to negative. Both copulas primarily captures the dependence structure in the center of the distribution. However, they are limited in modeling extreme events: they exhibit no tail dependence. Another limitation of the current formu-

lation is that we assume the errors within equations are independent. Errors may exhibit autocorrelation within the equation due to the development year effect over time. A natural next step is to develop a hybrid model of the recurrent neural networks (EDT) (Cai et al., 2025) and the SUR copula mixed model, where we can interpret the dependence using the copula and estimate other components of the model using EDT.

Chapter 5

Hybrid Modeling of RNN and SUR Copula Mixed Models

5.1 Introduction

This chapter presents preliminary results from our ongoing work. We will revise our methods so that fixed effects are captured by DT, while the dependence structure is captured using the SUR copula mixed model.

Traditionally, property and casualty (P&C) insurance companies have used generalized linear models (GLMs) for loss reserving. Insurers operate across multiple lines of business (LOBs) where claims can be related. Copula regression accounts for the dependence between incremental paid losses in different LOBs, leading to large risk capital reduction. However, copula regression has certain limitations, such as its limited flexibility in modeling marginal distributions. The incremental paid losses are assumed to be independent and follow a distribution belonging to the exponential family. Schelldorfer and Wuthrich (2019) discusses the strategy of using the capabilities of neural networks to improve the GLM, which is referred to as a hybrid model.

Wüthrich and Merz (2019) demonstrates the importance of embedding classical

actuarial models like GLM into a neural net, known as the Combined Actuarial Neural Net (CANN) approach. According to Schelldorfer and Wuthrich (2019), GLMs can be seen as a starting point of neural network models for both regression and classification tasks. The benefit of this is that we receive better run times in model calibration, and we can explicitly identify deficiencies in GLMs. Wilson et al. (2024) shows that combined models work more effectively than single models and suggests that combining GLM and neural network performs better as it aids in maximizing the advantages of both techniques. Saad et al. (2024) combine a deep neural network architecture with hierarchical Bayesian modeling for complex spatiotemporal fields, reducing the prediction error across several benchmarks.

We propose a hybrid model of Recurrent Neural Networks (RNN) and SUR copula mixed model to improve the interpretability of the dependence between LOBs while modeling the complex fixed effects including interactions. We train the DT model for each LOB and obtain the corresponding residuals, and then model these residuals using the SUR copula mixed model. We let the SUR copula mixed model compute the dependence between the two LOBs, which is not available in the EDT predicted results. The SUR copula mixed model takes into account the heterogeneity across companies in the residuals.

5.2 Method

Let $Y_{ijc}^{(\ell)}$ denote the standardized incremental paid loss for accident year i ($1 \leq i \leq I$) and development year j ($1 \leq j \leq I$) in company c. We train the DT model using $Y_{ijc}^{(\ell)}$ from all companies for the ℓth LOB. DT captures the fixed effects through neurons, company effects through embeddings, and pair-wise dependencies through paired sequence input.

For all the $Y_{ijc}^{(\ell)}$, we calculate the residual as $\epsilon_{ijc}^{(\ell)} = Y_{ijc}^{(\ell)} - \hat{Y}_{ijc}^{(\ell)}$, where $\hat{Y}_{ijc}^{(\ell)}$ is the predicted incremental paid loss from the DT.

We model the residuals $\epsilon_{ijc}^{(\ell)}$ with the SUR copula mixed model. Let $\mu_{ijc}^{(\ell)}$ be the expected value of $\epsilon_{ijc}^{(\ell)}$. We model $\mu_{ijc}^{(\ell)}$ using the company effect $\boldsymbol{b}^{(\ell)} = (b_1^{(\ell)}, b_2^{(\ell)}, ..., b_C^{(\ell)})$ as in (5.1).

$$\mu_{ijc}^{(\ell)} = \boldsymbol{z}_c^{(\ell)} \boldsymbol{b}^{(\ell)}. \tag{5.1}$$

The probability density for all the data is

$$f(\boldsymbol{\epsilon}_1^{(\ell)}, \boldsymbol{\epsilon}_2^{(\ell)}, \ldots, \boldsymbol{\epsilon}_C^{(\ell)}; \tau_\ell, \sigma_\ell) = \int_{[-\infty,\infty]^C} f(\boldsymbol{\epsilon}_1^{(\ell)}, \boldsymbol{\epsilon}_2^{(\ell)}, \ldots, \boldsymbol{\epsilon}_C^{(\ell)} \mid b_1^{(\ell)}, b_2^{(\ell)}, \ldots, b_C^{(\ell)}; \sigma_\ell) \cdot f(b_1^{(\ell)}, b_2^{(\ell)}, \ldots, b_C^{(\ell)}; \tau_\ell) db_1^{(\ell)} db_2^{(\ell)} \ldots db_C^{(\ell)}, \tag{5.2}$$

where $\boldsymbol{\epsilon}_c^{(\ell)}$ is the $I(I+1) \times 1$ vector of residuals for the c^{th} company from ℓ^{th} LOB.

Assuming the residuals from each company are independent, we can write the probability density as

$$\begin{aligned} f(\boldsymbol{\epsilon}_1^{(\ell)}, \boldsymbol{\epsilon}_2^{(\ell)}, \ldots, \boldsymbol{\epsilon}_C^{(\ell)}; \sigma_\ell, \tau_\ell) &= \prod_{c=1}^{C} \int_{-\infty}^{\infty} f(\boldsymbol{\epsilon}_c^{(\ell)} \mid b_c^{(\ell)}; \sigma_\ell) f(b_c^{(\ell)}; \tau_\ell) db_c^{(\ell)} \\ &= \prod_{c=1}^{C} \int_{-\infty}^{\infty} \prod_{i=1}^{I} \prod_{j=1}^{I+1-i} f(\epsilon_{ijc}^{(\ell)} | b_c^{(\ell)}) f(b_c^{(\ell)}; \tau_\ell) db_c^{(\ell)}, \end{aligned} \tag{5.3}$$

where $f(\epsilon_{ijc}^{(\ell)} | b_c^{(\ell)}; \sigma_\ell)$ denotes the conditional density of $\epsilon_{ijc}^{(\ell)}$ given $b_c^{(\ell)}$ and $f(b_c^{(\ell)}; \tau_\ell)$ denotes the density of the company effect $b_c^{(\ell)}$. τ_ℓ is the standard deviation of the company effect $b_c^{(\ell)}$.

For each LOB, we have the following log-likelihood function

$$L^{(\ell)}(\tau_\ell, \sigma_\ell \mid \boldsymbol{\epsilon}_1^{(\ell)}, \boldsymbol{\epsilon}_2^{(\ell)}, \ldots, \boldsymbol{\epsilon}_C^{(\ell)}) = \sum_{c=1}^{C} \log \int_{-\infty}^{\infty} \prod_{i=1}^{I} \prod_{j=1}^{I+1-i} f(\epsilon_{ijc}^{(\ell)} | b_c^{(\ell)}; \sigma_\ell) g(b_c^{(\ell)}; \tau_\ell) db_c^{(\ell)}. \tag{5.4}$$

The joint PDF for all $(\epsilon_{ijc}^{(\ell)}, \epsilon_{ijc}^{(\ell')})$ from all companies is then given by

$$\begin{aligned}
&f(\epsilon_{111}^{(\ell)}, \epsilon_{111}^{(\ell')}, \epsilon_{121}^{(\ell)}, \epsilon_{121}^{(\ell')}, \ldots, \epsilon_{1IC}^{(\ell)}, \epsilon_{1IC}^{(\ell')}; \tau_\ell, \sigma_\ell, \tau_{\ell'}, \sigma_{\ell'}) \\
=&\prod_{c=1}^{C} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{i=1}^{I} \prod_{j=1}^{I+1-i} f(\epsilon_{ijc}^{(\ell)} \mid b_c^{(\ell)}; \sigma_\ell) f(\epsilon_{ijc}^{(\ell')} \mid b_c^{(\ell')}; \sigma_{\ell'}) \\
&c(F(\epsilon_{ijc}^{(\ell)} \mid b_c^{(\ell)}), F(\epsilon_{ijc}^{(\ell')} \mid b_c^{(\ell')}); \theta) f(b_c^{(\ell)}; \tau_\ell) f(b_c^{(\ell')}; \tau_{\ell'}) db_c^{(\ell)} db_c^{(\ell')} \qquad (5.5)
\end{aligned}$$

Here $c(\cdot)$ denotes the PDF corresponding to copula $C(\cdot)$.

We then write the copula part in terms of the ranks of pseudo-residuals for $\epsilon_{ijc}^{(\ell)}$ and $\epsilon_{ijc}^{(\ell')}$, conditional on the random effect $b_c^{(\ell)}$ and $b_c^{(\ell')}$, respectively.

Suppose $\epsilon_{ijc}^{(\ell)} \mid b_c^{(\ell)}$ follows normal distribution. We define the pseudo-residuals as

$$s_{ijc}^{(\ell)} = \frac{\epsilon_{ijc}^{(\ell)} - \hat{\mu}_{ijc}^{(\ell)}}{\hat{\sigma}_\ell}. \qquad (5.6)$$

Next, we use the empirical cumulative distribution function (CDF) to get ranks of pseudo-residuals. The rank $R_{ijc}^{(\ell)}$ of the residual $s_{ijc}^{(\ell)}$ is given by

$$R_{ijc}^{(\ell)} = \frac{1}{I(I+1)/2+1} \sum_{i^*=1}^{I} \sum_{j^*=1}^{I+1-i^*} \mathbf{1}\left(s_{i^*j^*c}^{(\ell)} \leq s_{ijc}^{(\ell)}\right),$$

where $\mathbf{1}$ is the indicator function.

We approximate $F(\epsilon_{ijc}^{(\ell)} \mid b_c^{(\ell)})$ and $F(\epsilon_{ijc}^{(\ell')} \mid b_c^{(\ell')})$ in (5.5) with $R_{ijc}^{(\ell)}$ and $R_{ijc}^{(\ell')}$, respectively. The copula term $c(F(\epsilon_{ijc}^{(\ell)} \mid b_c^{(\ell)}), F(\epsilon_{ijc}^{(\ell')} \mid b_c^{(\ell')}); \theta)$ are replaced by $c(R_{ijc}^{(\ell)}, R_{ijc}^{(\ell')}; \theta)$.

In the case of two LOBs, let $\ell = 1$ and $\ell' = 2$, we obtain the following log-likelihood function

$$
\begin{aligned}
&L(\sigma_1, \sigma_2, \tau_1, \tau_2, \theta \mid \boldsymbol{\epsilon}_1^{(1)}, \boldsymbol{\epsilon}_2^{(1)}, \ldots, \boldsymbol{\epsilon}_C^{(1)}, \boldsymbol{\epsilon}_1^{(2)}, \boldsymbol{\epsilon}_2^{(2)}, \ldots, \boldsymbol{\epsilon}_C^{(2)}) \\
&= \sum_{\ell=1}^{2} L^{(\ell)}(\sigma_\ell, \tau_\ell \mid \boldsymbol{\epsilon}_1^{(\ell)}, \boldsymbol{\epsilon}_2^{(\ell)}, \ldots, \boldsymbol{\epsilon}_C^{(\ell)}) + \sum_{c=1}^{C} \sum_{i=1}^{I} \sum_{j=1}^{I+1-i} \log c(R_{ijc}^{(\ell)}, R_{ijc}^{(\ell')}; \theta) \\
&= \sum_{\ell=1}^{2} \sum_{c=1}^{C} \log \int_{-\infty}^{\infty} \prod_{i=1}^{I} \prod_{j=1}^{I+1-i} f(\epsilon_{ijc}^{(\ell)} | b_c^{(\ell)}) f(b_c^{(\ell)}; \tau_\ell) db_c^{(\ell)} \\
&+ \sum_{c=1}^{C} \sum_{i=1}^{I} \sum_{j=1}^{I+1-i} \log c(R_{ijc}^{(\ell)}, R_{ijc}^{(\ell')}; \theta). \qquad (5.7)
\end{aligned}
$$

We apply the iterative two-stage estimation approach developed in Chapter 3 to estimate the parameters by maximizing (5.7). After fitting the SUR copula mixed model to the residuals $\epsilon_{ijc}^{(\ell)}$, we obtain the fitted residuals, which are denoted as $\hat{\epsilon}_{ijc}^{(\ell)}$.

In the hybrid model, we define the loss for each sample in the DT as

$$
\frac{(\hat{Y}_{ijc}^{(1)} + \hat{\epsilon}_{ijc}^{(1)} - Y_{ijc}^{(1)})^2 + (\hat{Y}_{ijc}^{(2)} + \hat{\epsilon}_{ijc}^{(2)} - Y_{ijc}^{(2)})^2}{2}, \qquad (5.8)
$$

where $\hat{Y}_{ijc}^{(1)}$ and $\hat{Y}_{ijc}^{(2)}$ are the predicted incremental paid losses from the DT. The $\hat{\epsilon}_{ijc}^{(\ell)}$ represent the structure captured by the SUR copula mixed model. Finally, we utilize the AMSGRAD method (Reddi et al., 2018) to optimize the parameters in the DT and then obtain the estimated loss reserve.

5.3 Application

To illustrate the hybrid model, we consider the same data as used in Cai et al. (2025), which are from the Schedule P of the NAIC database. We use multiple pairs of loss triangles of paid losses in Schedule P for the year 1997. Each pair consists of personal auto and commercial auto lines of business.

We train the DT model using the incremental paid losses $y_{ijc}^{(\ell)}$ ($1 \leq i \leq 10$,

$1 \leq j \leq 10$, $1 \leq c \leq 30$) from 30 companies for the ℓ^{th} ($\ell = 1, 2$) LOB. During training, we obtain the predicted losses for the upper triangles. Then we calculate the residuals $\epsilon_{ijc}^{(\ell)}$ by $\epsilon_{ijc}^{(\ell)} = y_{ijc}^{(\ell)} - \hat{y}_{ijc}^{(\ell)}$, where $\hat{y}_{ijc}^{(\ell)}$ is the predicted loss for the upper triangle in the ℓ^{th} LOB from the DT model.

As shown in (5.1), we model the residuals $\epsilon_{ijc}^{(\ell)}$ with the SUR copula mixed model and capture the dependence between the two LOBs through a Gaussian copula. The estimated standard deviations for the company random effects in the two LOBs are $\tau_1 = 0.010$ and $\tau_2 = 0.013$, respectively. We show the estimated reserves and dependence parameter for the major US property and casualty insurer. The estimated reserves from a single run of the hybrid model for the two LOBs are 7 747 946 and 333 877, respectively. We compare the estimated reserve with those from other models in Table 5.1. The estimated reserves from the hybrid model are close to that from the Deep Triangle, which shows that the hybrid model can also be used to improve the reserve prediction while also improve the interpretability of the dependence structure between the two LOBs.

Table 5.1: Point estimates of the reserves.

Model	Reserves		
	LoB 1, R_1	LoB 2, R_2	Total, R
Hybrid model	7 747 946	333 877	8 081 823
Deep Triangle	7 781 299	324 024	8 105 323
Sparse SUR copula mixed	7 295 694	372 761	7 668 455
SUR copula mixed	7 246 135	377 324	7 623 460
SUR copula	6 823 325	378 386	7 364 511
Actual reserve	8 086 094	318 380	8 404 474

Table 5.2: Performance comparison using percentage error of actual and estimated loss reserve.

	Personal Auto	Commercial Auto	Total
Hybrid model	-4.2%	4.8%	-3.9%
Deep Triangle	-4.1%	2.9%	-3.8%
Sparse SUR copula mixed	-9.7%	16.8%	-8.7%
SUR copula mixed	-10.3%	18.5%	-9.3%
SUR copula	-15.6%	19.0%	-12.4%

Table 5.3: Dependence comparison between models. SUR copula mixed model and sparse SUR copula mixed model are abbreviated to SURCMM and sSURCMM, respectively.

	Hybrid	EDT	sSURCMM	SURCMM	SUR copula
dependence parameter	-0.24	na	-0.19	-0.20	-0.36

Next, we compute the percentage errors between the estimated reserve and the true reserve in Table 5.2. We find that the hybrid model and Deep Triangle generate the smallest percentage errors among all the models. The SUR copula mixed models generates smaller percentage errors than the SUR copula model.

The estimated dependence parameter between the two LOBs is around -0.24, which indicates a negative association between the two LOBs. As shown in Table 5.3, the negative association is also consistent with the result from SUR copula and SUR copula mixed models. This dependence structure information is valuable for the insurer to make strategic business decisions.

5.4 Simulation Study

To further validate our result on the hybrid model for computing the dependence structure between two LOBs. We simulate 30 pairs of loss triangles using the fixed effects $\boldsymbol{\beta}^{(\ell)}$ estimated from one of the 30 pairs of loss triangles in Chapter 4. We assume the company's random effects follow $b_c^{(1)} \sim N(0, \tau_1)$ and $b_c^{(2)} \sim N(0, \tau_2)$ with $\tau_1 = 0.2$ and $\tau_2 = 0.3$.

We then simulate $u_{ij}^{(\ell)} (\ell = 1, 2)$ $(i + j - 1 \leq I)$ from a Gaussian copula model $c(\cdot; \theta)$ with dependence parameter $\theta = -0.3$. The upper triangles are obtained by inverse function $y_{ijc}^{(\ell)} = F^{(\ell)(-1)}(u_{ij}^{(\ell)}; \eta_{ijc}^{(\ell)}, \gamma^{(\ell)})$, where $\eta_{ijc}^{(\ell)} = \boldsymbol{x}_{ij}^{(\ell)} \boldsymbol{\beta}^{(\ell)} + \boldsymbol{z}_c^{(\ell)} \boldsymbol{b}^{(\ell)}$.

Following the procedure in the application section, we train the DT on the simulated loss triangles for each LOB. During training, we generated the predicted loss for the upper triangles, enabling us to obtain the residuals for each LOB. We apply the SUR copula mixed model to the residuals for the simulated loss triangles.

The estimated dependence parameter, -0.25, has a consistent sign with the actual dependence parameter, -0.3.

5.5 Summary and Discussion

We integrate the SUR copula mixed model with the extended Deep Triangle to interpret the dependence between LOBs. Specifically, we model the heterogeneous residuals from the DT using a SUR copula mixed model. The heterogeneity across companies and between LOBs is handled by the random effect, and the dependence is captured by a Gaussian copula.

To evaluate the proposed integration method, we apply it to multiple loss triangles from the NAIC database. The hybrid modeling of the DT and SUR copula mixed model reveals a negative association between the personal and commercial LOBs, consistent with the findings from the SUR copula mixed model. A simulation study further highlights the benefits of integrating the DT and SUR copula mixed model in interpreting the dependency between LOBs. We will generate synthetic loss triangles by resampling the errors from the SUR copula mixed model and adding the fixed effects from the DT. This approach will allow us to generate the predictive distribution of the reserve and perform risk capital analysis.

Chapter 6

Conclusions and Future Work

6.1 Summary and Discussion

Estimating unpaid claims is crucial for an insurer's operations in property and casualty (P&C) insurance. Insurance companies often engage in multiple interrelated lines of business (LOBs), and accounting for dependence between LOBs is essential in accurately determining an insurance company's reserve ranges and the amount of risk capital needed. Incorporating dependency into reserve calculations helps the insurer determine the appropriate amount of risk capital and leverage diversification benefits. The actuarial industry has developed parametric and non-parametric methods for loss reserving. However, few methods effectively capture the dependency between loss reserves while balancing interpretability and predictive accuracy. In particular, there is a lack of hybrid approaches that integrate neural networks with copula-based models to leverage the strengths of both methods. The SUR copula regression incorporates the dependence between two LOBs through a copula using loss triangles from one company, producing a relatively large bias, due to modeling single-company effects as fixed effects, restrictive marginal assumptions, and the omission of sequential dependence in development year effects. In this thesis, we introduce the use of the Deep Triangle (DT), a

recurrent neural network, for multivariate loss reserving. We also propose SUR copula mixed models that extend SUR copula regression to incorporate multiple companies' data, improving both loss prediction and risk capital analysis. Furthermore, we introduce a hybrid approach that combines neural networks with copula-based models to balance interpretability and predictive accuracy; however, a comprehensive simulation study and in-depth application of this method are left for future work.

In Chapter 2, we introduce the Extended Deep Triangle (EDT) framework, which tailors the Deep Triangle (DT), a gated recurrent neural network, for multivariate loss reserving with bivariate loss triangles of incremental paid losses. We also introduce an asymmetric loss function to account for the varying volatility across different lines of business (LOBs). By investigating the impact of input sequence length, we find that longer sequences generally improve predictive performance. Furthermore, we propose GAN-based techniques to generate predictive distributions of reserves, yielding larger risk capital gains. To generate these predictive distributions, we integrate DT with a copula-based generative adversarial network (copula GAN) that produces synthetic pairs of loss triangles. In addition, we reduce the computational cost of generating predictive distributions by initializing training with pre-trained model weights on GAN simulated samples. We validate EDT through simulation studies and an empirical application using real data from the National Association of Insurance Commissioners (NAIC) database. Results demonstrate that EDT consistently outperforms copula regression in predicting loss reserves and produces larger risk capital gains.

While neural network based approaches such as EDT achieve strong predictive performance, their interpretability is limited. Since the ultimate goal of this thesis is to develop models that balance predictive accuracy with interpretability, in Chapter 3, we turn to parametric methods. To integrate the SUR copula mixed model within a hierarchical structure, we focus on random effects to cap-

ture heterogeneity across companies and lines of business (LOBs). Importantly, the dependence structure between LOBs is interpretable through the sign of the estimated dependence parameters, providing insurers with insight into how different lines are related. We develop a two-stage iterative approach to estimate the parameters of the SUR copula mixed model and illustrate the method using multiple pairs of loss triangles from the NAIC database. Our results show that the SUR copula mixed model produces smaller bias between predicted and actual reserves compared to the SUR copula regression model. In addition, by generating the predictive distribution of reserves, we demonstrate that the SUR copula mixed model provides larger risk capital gains than SUR copula regression, reflecting a greater diversification benefit. Finally, we validate these findings through a simulation study.

Continuing from the model in Chapter 3, in Chapter 4, we investigate the shrinkage of model parameters in the SUR copula mixed model and develop the sparse SUR copula mixed method. In this work, we incorporate the least absolute shrinkage and selection operator (LASSO) regularization for the fixed effects to mitigate the impact of limited data in the tail of the loss triangles. We also adapt the bootstrap approach to account for sparsity by applying coefficient thresholding during the resampling step, ensuring that the predictive distribution of reserves reflects the penalized estimates. We demonstrate the estimation method and bootstrapping procedures using both a real data application and a simulation study. Compared to the SUR copula mixed model, the sparse SUR copula mixed model produces reserve estimates closer to the true values and generates larger risk capital gains. One limitation of using the Gaussian copula is its inability to capture tail dependence. We may consider Student's t copula for modeling extreme dependence.

In Chapter 5, we explore a hybrid modeling framework that integrates the Deep Triangle (DT), a gated recurrent neural network, with the SUR copula mixed

model for loss reserving. We first compute residuals from the DT, and then feed these residuals into the SUR copula mixed model, using only its random effect and copula components to capture dependence across lines of business (LOBs). The estimated loss ratio is obtained by summing the DT output with the SUR copula mixed model output, and the loss function is calculated by subtracting the predicted values from the observed values. We demonstrate the proposed hybrid approach using real data, which reveals a negative association between the two LOBs. This framework bridges interpretability and flexibility, allowing us to capture complex accident year and development year effects with DT while simultaneously modeling interpretable dependence structures through the SUR copula mixed model. This chapter presents preliminary results and remains incomplete, with a comprehensive simulation study and further evaluation left for future work.

6.2 Future work

We assume the errors within equations are independent in the current formulation of the SUR copula mixed model. In practice, errors may exhibit autocorrelation within the equation due to the development year effect over time. Although AR(1) or higher order dependence structures could be incorporated into the SUR copula mixed model, doing so would substantially increase the complexity of the estimation procedure.

The copula component is currently capturing the cell-level residual dependence between LOBs, while we are not capturing the company-level dependence in the reserves due to the independence assumption of the random effects. Another extension would be to relax the independence assumption of the company random effects and introduce a bivariate random effect, allowing for correlation between the random effects of different LOBs.

Future extensions may further develop the hybrid model, retaining the copula

component for dependence interpretation while estimating the other elements of the model using flexible architectures such as our recent work with recurrent neural networks (EDT). The hybrid model could provide a valuable framework for analyzing the dependence structures between different LOBs and simultaneously modeling fixed effects for accident year and development year, company random effects, and their interactions. Moreover, within the hybrid framework, resampling the residuals to generate synthetic loss triangles could be used to construct the predictive distribution of reserves. Finally, the hybrid model can be enhanced to capture cross-LOB dependence primarily through the SUR copula mixed model by introducing a penalty term in the loss function that discourages the neural network from absorbing this dependence, ensuring that the interpretability of the dependence structure is preserved.

To enhance the prediction for the reserve, we can use weighted averaging from all the models' predictions. One approach would be to set weights proportional to the model's performance on a validation set.The better the model's performance, the higher its weight. For example, we can use the inverse of the mean squared error (MSE) on the validation set to assign weight to each model's prediction.

In addition to the accident year, development year, and company effects, we could also consider other macroeconomic conditions, such as inflation and interest rates, in the models. For the EDT model, the macroeconomic conditions could also be formatted as another input sequence to the GRU module. As for the SUR copula mixed model, we may add the macroeconomic conditions to the systematic component of the marginal distribution. For example, the mean of the marginal distribution can be expressed as a function of the macroeconomic factors. By incorporating more information, these models can capture a broader range of factors that influence incremental paid losses, potentially leading to more accurate reserve predictions.

Bibliography

Abdallah, A., Boucher, J.-P., and Cossette, H. (2015). Modeling dependence between loss triangles with Hierarchical Archimedean copulas. *ASTIN Bulletin: The Journal of the IAA*, 45(3):577–599.

Abdallah, A. and Wang, L. (2023). Rank-based multivariate Sarmanov for modeling dependence between loss reserves. *Risks*, 11(11):187.

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503.

Ajne, B. (1994). Additivity of chain-ladder projections. *ASTIN Bulletin: The Journal of the IAA*, 24(2):311–318.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Cai, P., Abdallah, A., and Jeganathan, P. (2025). Recurrent neural networks for multivariate loss reserving and risk capital analysis. *North American Actuarial Journal*, pages 1–25.

Chatterjee, A. and Lahiri, S. N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association*, 106(494):608–625.

Cossette, H. and Pigeon, M. (2021). *Using Open Files for Individual Loss Reserving in Property and Casualty Insurance*. Université du Québec à Montréal.

Cote, M.-P., Hartman, B., Mercier, O., Meyers, J., Cummings, J., and Harmon, E. (2020). Synthesizing property & casualty ratemaking datasets using generative adversarial networks. *arXiv preprint arXiv:2008.06110*.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*. Number 1. Cambridge university press.

De Jong, P. (2012). Modeling dependence between loss triangles. *North American Actuarial Journal*, 16(1):74–86.

Dhaene, J., Vanduffel, S., Goovaerts, M. J., Kaas, R., Tang, Q., and Vyncke, D. (2006). Risk measures and comonotonicity: a review. *Stochastic models*, 22(4):573–606.

Duval, F. and Pigeon, M. (2019). Individual loss reserving using a gradient boosting-based approach. *Risks*, 7(3).

England, P. D. and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3):443–518. Publisher: Cambridge University Press.

Gabrielli, A., Richman, R., and Wüthrich, M. V. (2018). Neural network embedding of the over-dispersed Poisson reserving model. *Available at SSRN 3288454*.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT Press. `http://www.deeplearningbook.org`.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). Statistical learning with sparsity. *Monographs on statistics and applied probability*, 143(143):8.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *IEEE International Conference on Computer Vision (ICCV 2015)*, 1502.

Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2020). *copula: Multivariate Dependence with Copulas*. R package version 1.0-1.

Holland, P. W. and Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Communications in Statistics-theory and Methods*, 6(9):813–827.

Kirschner, G., Kerley, C., and Isaacs, B. (2008). Two approaches to calculating correlated reserve indications across multiple lines of business. *Variance*, 2(1):15–38.

Kuo, K. (2019). Deeptriangle: A deep learning approach to loss reserving. *Risks*, 7(3):97.

Lahiri, S. and Lahiri, S. (2003). *Resampling methods for dependent data*. Springer Science & Business Media.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Ludwig, A. and Schmidt, K. D. (2010). *Calendar year reserves in the multivariate additive model*. Professoren des Inst. für Math. Stochastik.

Mack, T. (1993). Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates. *ASTIN Bulletin: The Journal of the IAA*, 23(2):213–225. Publisher: Cambridge University Press.

Marra, G. and Radice, R. (2023). GJRM: Generalised joint regression modelling. *URL: https://cran.r-project.org/web/packages/GJRM*. R package version 0.2-6.4.

McGuire, G., Taylor, G., and Miller, H. (2018). Self-assembling insurance claim models using regularized regression and machine learning. *Available at SSRN 3241906*.

Meyers, G. G. and Shi, P. (2011). Loss reserving data pulled from NAIC Schedule P. *URL: https://www. casact. org/publications-research/research/research-resources/loss-reserving-data-pulled-naic-schedule-p*, 5.

Mulquiney, P. (2006). Artificial neural networks in insurance loss reserving. In *JCIS*. Citeseer.

Murphy, K. P. (2022). *Probabilistic machine learning: an introduction*. MIT press.

Nelsen, R. B. (2006). *An Introduction to Copulas (Springer Series in Statistics)*. Springer-Verlag, Berlin, Heidelberg.

Patki, N., Wedge, R., and Veeramachaneni, K. (2016). The synthetic data vault. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 399–410. IEEE.

Peremans, K. and Van Aelst, S. (2018). Robust inference for seemingly unrelated regression models. *Journal of Multivariate Analysis*, 167:212–224.

Pröhl, C. and Schmidt, K. D. (2005). *Multivariate chain ladder*. Professoren des Inst. für Math. Stochastik.

Reddi, S., Kale, S., and Kumar, S. (2018). On the convergence of Adam and beyond. In *International Conference on Learning Representations*.

Saad, F., Burnim, J., Carroll, C., Patton, B., Köster, U., A. Saurous, R., and Hoffman, M. (2024). Scalable spatiotemporal prediction with bayesian neural fields. *Nature Communications*, 15(1):7942.

Schelldorfer, J. and Wuthrich, M. V. (2019). Nesting classical actuarial models into neural networks. *Available at SSRN 3320525*.

Schweizer, B. and Sklar, A. (2011). *Probabilistic metric spaces*. Courier Corporation.

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

Shi, P. and Frees, E. W. (2011). Dependent loss reserving using copulas. *ASTIN Bulletin: The Journal of the IAA*.

Srivastava, N., Mansimov, E., and Salakhudinov, R. (2015). Unsupervised learning of video representations using LSTMs. In *International Conference on Machine Learning*, pages 843–852. PMLR.

Srivastava, V. K. and Giles, D. E. (1987). *Seemingly unrelated regression equations models: Estimation and inference*, volume 80. CRC press.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Taylor, G. (2019). Loss reserving models: Granular and machine learning forms. *Risks*, 7(3):82.

Taylor, G. and McGuire, G. (2007). A synchronous bootstrap to account for dependencies between lines of business in the estimation of loss reserve prediction error. *North American Actuarial Journal*, 11(3):70–88.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Verbeke, G. and Molenberghs, G. (2009). *Linear Mixed Models for Longitudinal Data*. Springer Science & Business Media.

Wilson, A. A., Nehme, A., Dhyani, A., and Mahbub, K. (2024). A comparison of generalised linear modelling with machine learning approaches for predicting loss cost in motor insurance. *Risks*, 12(4):62.

Wüthrich, M. V. (2018a). Machine learning in individual claims reserving. *Scandinavian Actuarial Journal*, 2018(6):465–480.

Wüthrich, M. V. (2018b). Neural networks applied to chain–ladder reserving. *European Actuarial Journal*, 8(2):407–436.

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

Wüthrich, M. V. and Merz, M. (2019). Yes, we cann! *ASTIN Bulletin: The Journal of the IAA*, 49(1):1–3.

Xu, L., Skoularidou, M., Cuesta-Infante, A., and Veeramachaneni, K. (2019). Modeling tabular data using conditional GAN. *Advances in Neural Information Processing Systems*, 32.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298):348–368.

Zhang, Y. (2010). A general multivariate chain ladder model. *Insurance: Mathematics and Economics*, 46(3):588–599.

Appendix A

A.1 Dependence Analysis

We compute Kendall's tau on the residuals of the marginal fits, where the marginals are the log-normal and gamma regression models. Note that the analysis is performed on the residuals because we want to remove the accident year and development year effects. For the log-normal, the residual is $\hat{\epsilon}_{ij}^{(1)} = (\ln y_{ij}^{(1)} - \hat{\mu}_{ij}^{(1)})/\hat{\sigma}$, and for gamma $\hat{\epsilon}_{ij}^{(2)} = y_{ij}^{(2)}/\hat{\mu}_{ij}^{(2)}$. The computed Kendall's tau is -0.1562, suggesting a negative association between personal and commercial LOBs.

A.2 Copula Regression Using Loss Triangles from 30 Companies

Here we consider modeling the systematic component η_{ijc} using accident year effect $\alpha_i (i \in 1, 2, ..., 10)$, development year effect $\beta_j (j \in 1, 2, ..., 10)$, and company effect $b_c (c \in 1, 2, .., 30)$ as in (A.1).

$$\eta_{ijc} = \xi + \alpha_i + \beta_j + b_c, \tag{A.1}$$

where b_c is an additional predictor that characterizes the company effect.

We identify that $Y_{i,j}^{(1)}$ and $Y_{i,j}^{(2)}$ follow log-normal and gamma distributions,

respectively. Let's consider the probability density function (PDF) of the log-normal distribution for $Y_{ij}^{(1)}$

$$f_{ij}^{(1)}\left(y_{ij}^{(1)}\right) = \frac{1}{y_{ij}^{(1)}\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log\left(y_{ij}^{(1)}\right)-\mu_{ij}^{(1)}}{\sigma}\right)^2}, \quad y_{ij}^{(1)} > 0, \tag{A.2}$$

where $\mu_{ij}^{(1)}$ is the location and $\sigma > 0$ is the shape. Thus, the systematic component is $\eta_{ij}^{(1)} = \mu_{ij}^{(1)}$.

Next, the gamma PDF for $Y_{ij}^{(2)}$ is given by

$$f_{ij}^{(2)}\left(y_{ij}^{(2)}\right) = \left(\frac{y_{ij}^{(2)}}{\mu_{ij}^{(2)}}\right)^{\phi} \frac{e^{-\frac{y_{ij}^{(2)}}{\mu_{ij}^{(2)}}}}{\Gamma(\phi)\, y_{ij}^{(2)}}, \quad y_{ij}^{(2)} > 0, \tag{A.3}$$

where $\phi > 0$ is the shape and $\mu_{ij}^{(2)} > 0$ is the location. Thus, the systematic component is $\eta_{ij}^{(2)} = \log\left(\mu_{ij}^{(2)}\phi\right)$ (Abdallah et al., 2015), ensuring $\mu_{ij}^{(2)}$ is positive.

For the log-normal distribution, the $Y_{ij}^{(1)}$ is estimated by $\hat{Y}_{ij}^{(1)} = \exp\left(\hat{\mu}_{ij}^{(1)} + \frac{1}{2}(\hat{\sigma})^2\right)$ and for the gamma distribution, $\hat{Y}_{ij}^{(2)} = \hat{\mu}_{ij}^{(2)}\hat{\phi}$.

We use Gaussian copula to capture the dependence between the two LOBs, and the estimated reserves are 6 823 325 and 370 386, respectively. The percentage errors of actual and estimated reserves for the two LOBs are -15.62% and 16.33%, respectively.

A.3 Block Bootstrapping for Predictive Distribution of the Reserve

We consider block bootstrap as another way to generate samples to compute the predictive distributions of the reserve based on DT. The block bootstrap resamples consecutive blocks of observations, treating these blocks as exchangeable. As a result, the original dependence structure of the data is preserved within each block (Lahiri and Lahiri, 2003). Nevertheless, the data generated by block bootstrapping

input sequence vectors might not capture the same sampling uncertainty as seen in methods like the copula regression parametric bootstrap or GAN-based schemes unless the appropriate block size is obtained based on the bias-variance tradeoff in approximating the predictive distribution.

To select a suitable block size, we evaluated the validation error of DT across different sequence lengths and found that the longest sequence (I) minimized the validation error (Figure 2.6). We adopt this length, assuming it best captures the within-block temporal dependence, thereby justifying the approximate exchangeability of blocks. Specifically, we resample the training data of sequencing length I using the bootstrapping of blocks, leveraging this exchangeability to construct the predictive reserve distribution, referred to as DT-bootstrap.

In particular, first, we randomly split the training data into training and validation sets using an 80-20 split described in Section 2.2.2. Suppose $I = 10$. For each company, we have 36 training sequences and 9 validation sequences. Let $\boldsymbol{X}_n$ $(1 \leq n \leq 36)$ denote the training input sequences (mask,..., mask, $Y_{i,1}^{(1)}$, $Y_{i,2}^{(1)}$, ..., $Y_{i,j-1}^{(1)}$) and (mask ,..., mask, $Y_{i,1}^{(2)}$, $Y_{i,2}^{(2)}$, ..., $Y_{i,j-1}^{(2)}$) from one company. We also let $\boldsymbol{Y}_n$ $(1 \leq n \leq 36)$ denote the training output sequences ($Y_{i,j}^{(1)}$, $Y_{i,j+1}^{(1)}$, ..., $Y_{i,11-i}^{(1)}$, mask, ... , mask) and ($Y_{i,j}^{(2)}$, $Y_{i,j+1}^{(2)}$, ..., $Y_{i,11-i}^{(2)}$, mask, ... , mask). Our original training data are ($\boldsymbol{X}_1$,$\boldsymbol{Y}_1$) , ... , ($\boldsymbol{X}_{36}$,$\boldsymbol{Y}_{36}$). We draw bootstrap training data ($\boldsymbol{X}_1^*$,$\boldsymbol{Y}_1^*$) , ... , ($\boldsymbol{X}_{36}^*$,$\boldsymbol{Y}_{36}^*$) randomly with replacement from the original training set. The training data with the same accident year and development year of different companies stay together during bootstrapping. We apply the same procedure to the validation data.

Table A.1 shows that the standard deviation from the DT-bootstrap is smaller than that from the copula regression models. DT-bootstrap also has a CV that is smaller than one, which also complies with the insurance standards. In addition, Table A.2 depicts that the DT-bootstrap generates a smaller risk capital by capturing the inter-LOB dependence than the copula regression models.

Table A.1: Bias, Standard deviation, Coefficient of variation (CV) of the loss reserve when we use DT-bootstrap and copula regression models.

	Reserve	Bootstrap mean reserve	Bias	Std. dev.	CV
DT-bootstrap	8 105 323	8 137 107	0.39%	235 304	0.029
Product Copula	6 954 736	6 972 792	0.26%	399 758	0.057
Gaussian Copula	6 919 171	6 941 806	0.33%	368 555	0.053
Frank Copula	6 999 253	7 043 309	0.63%	388 357	0.056

Table A.2: Risk capital estimation comparisons for DT-bootstrap and copula regression models.

Risk measure	TVaR (60%)	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
DT-Bootstrap	8 370 792	8 480 870	8 519 400	8 565 251	8 632 670	8 743 527
Silo-GLM	7 442 692	7 671 633	7 756 992	7 872 138	8 060 489	8 460 435
Product copula	7 367 695	7 553 768	7 621 203	7 710 435	7 847 773	8 126 433
Gaussian copula	7 313 951	7 490 387	7 556 029	7 644 886	7 782 646	8 054 737
Frank copula	7 424 807	7 616 405	7 685 514	7 776 754	7 921 574	8 202 695
Risk capital						
DT-Bootstrap		110 078	148 608	194 459	261 878	372 735
Silo-GLM		228 941	314 300	429 446	617 797	1 017 743
Product copula		186 073	253 508	342 740	480 078	758 738
Gaussian copula		176 436	242 078	330 935	468 695	740 786
Frank copula		191 598	260 707	351 947	496 767	777 888

We further validate our conclusion that DT-bootstrap reduces risk capital through simulation studies, as with the setup detailed in Section 4. Figure A.1 indicates that using the largest block size yields interval properties similar to those of DT-GAN, with the exception of coverage. We anticipate that reducing the block size may improve coverage, bringing it closer to the nominal level.

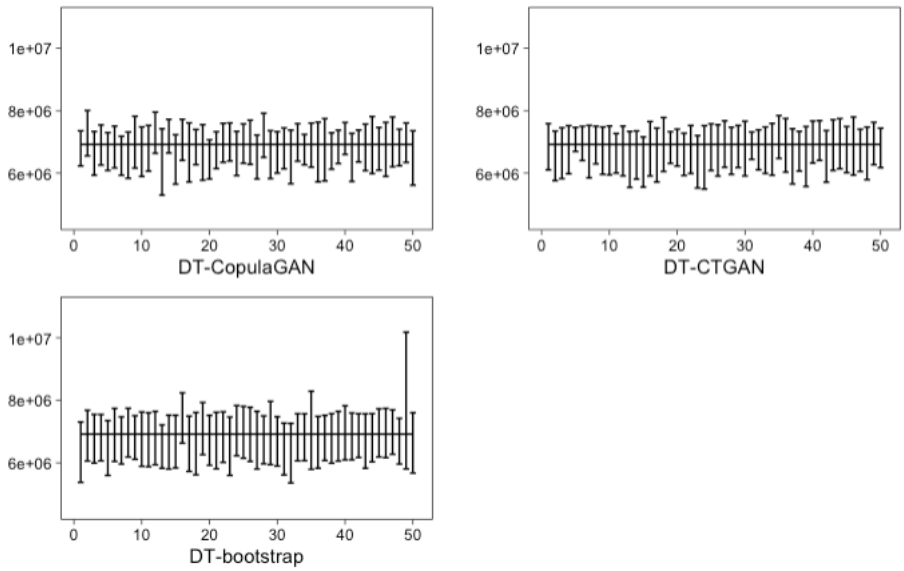

Figure A.1: 95% confidence interval for total reserves for EDT.
Note: The horizontal line indicates the true reserve. The true reserve is within all the 95% confidence intervals.

A.4 Simulation Setting

We present the true values of the parameters used in the simulation study in Chapter 2.

Table A.3: Accident year effect α_i

	personal auto	commercial auto
year 2	-0.03	-0.14
year 3	-0.03	-0.15
year 4	-0.13	-0.30
year 5	-0.17	-0.29
year 6	-0.18	-0.27
year 7	-0.18	-0.14
year 8	-0.24	-0.10
year 9	-0.27	0.17
year 10	-0.21	-0.12

Table A.4: Development year effect β_j

	personal auto	commercial auto
dev 2	-0.23	0.20
dev 3	-1.05	-0.02
dev 4	-1.65	-0.41
dev 5	-2.26	-1.06
dev 6	-3.02	-1.47
dev 7	-3.68	-2.10
dev 8	-4.50	-2.81
dev 9	-4.91	-3.12
dev 10	-5.92	-4.18

Table A.5: Premium ω_i

	personal auto	commercial auto
year 1	4 711 333	267 666
year 2	5 335 525	274 526
year 3	5 947 504	268 161
year 4	6 354 197	276 821
year 5	6 738 172	270 214
year 6	7 079 444	280 568
year 7	7 254 832	344 915
year 8	7 739 379	371 139
year 9	8 154 065	323 753
year 10	8 435 918	221 448

A.5 Fréchet-Hoeffding bounds

According to Fréchet-Hoeffding theorem (Schweizer and Sklar, 2011), for any bivariate copula $C : [0,1]^2 \to [0,1]$, the following bounds hold:

$$W(u_1, u_2) \leq C(u_1, u_2) \leq M(u_1, u_2)$$

The function W is called the lower Fréchet-Hoeffding bound and is defined as

$$W(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}.$$

The function M is called the upper Fréchet-Hoeffding bound and is defined as

$$M(u_1, u_2) = \min\{u_1, u_2\}.$$

The upper bound is reached for comonotone random variables, which are perfectly positive dependent. The lower bound corresponds to countermonotonic random variables, which are perfectly negative dependent.

Appendix B

B.1 Estimated Reserves from SUR Copula Mixed Model

This section provides supplementary results on the estimated reserves and risk capital analysis from the SUR copula mixed model and sparse SUR copula mixed model using a single dependence parameter.

Table B.1: Point estimates of the reserves.

Model	Reserves		
	LoB 1, R_1	LoB 2, R_2	Total, R
SUR copula mixed	7 246 135	377 324	7 623 460
Sparse SUR copula mixed	7 296 308	371 920	7 668 227

Table B.2: Bias, Standard deviation, Coefficient of variation (CV) from the predictive distribution using parametric bootstrapping.

	Reserve	Bootstrap reserve	Bias	Std. dev.	CV
SUR copula mixed	7 623 460	7 530 255	1.22%	612 947	0.082
Sparse SUR copula mixed	7 662 748	7 571 496	1.19%	514 371	0.068

Table B.3: Risk capital estimation for different methods.

Risk measure	TVaR (60%)	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
SUR copula mixed	8 123 673	8 442 581	8 562 167	8 736 343	9 030 721	9 621 658
Sparse SUR copula mixed	8 078 775	8 344 914	8 451 227	8 593 281	8 829 193	9 285 882
Risk capital						
SUR copula mixed		318 908	438 494	612 670	907 048	1 497 985
Sparse SUR copula mixed		266 139	372 452	514 506	750 418	1 207 107

Table B.4: Risk capital gain for different methods.

Risk capital gain	TVaR (80%)	TVaR (85%)	TVaR (90%)	TVaR (95%)	TVaR(99%)
SUR copula mixed vs Silo-GLM	71.37%	72.43%	72.36%	72.83%	73.03%
Sparse SUR copula mixed vs Silo-GLM	76.11%	76.59%	76.69%	76.79%	78.27%

B.2 Parameters for Simulation Settings

We present the true values of the parameters used in Simulation Setting 1.

Table B.5: Accident year effect α_i

	personal auto	commercial auto
year 2	-0.03	-0.14
year 3	-0.03	-0.15
year 4	-0.13	-0.30
year 5	-0.17	-0.29
year 6	-0.18	-0.27
year 7	-0.18	-0.14
year 8	-0.24	-0.10
year 9	-0.27	0.17
year 10	-0.21	-0.12

Table B.6: Development year effect β_j

	personal auto	commercial auto
dev 2	-0.23	0.20
dev 3	-1.05	-0.02
dev 4	-1.65	-0.41
dev 5	-2.26	-1.06
dev 6	-3.02	-1.47
dev 7	-3.68	-2.10
dev 8	-4.50	-2.81
dev 9	-4.91	-3.12
dev 10	-5.92	-4.18

We present the true values of the parameters used in Simulation Setting 2.

B.3 Accident Year and Development Year Effects

Table B.7: Premium ω_i

	personal auto	commercial auto
year 1	4 711 333	267 666
year 2	5 335 525	274 526
year 3	5 947 504	268 161
year 4	6 354 197	276 821
year 5	6 738 172	270 214
year 6	7 079 444	280 568
year 7	7 254 832	344 915
year 8	7 739 379	371 139
year 9	8 154 065	323 753
year 10	8 435 918	221 448

Table B.8: Accident year effect α_i

	personal auto	commercial auto
Year 2	-0.31	-0.18
Year 3	-0.21	-0.79
Year 4	-0.25	-1.28
Year 5	-0.40	-2.28
Year 6	-0.33	-2.84
Year 7	-0.32	-4.19
Year 8	-0.30	-4.46
Year 9	-0.26	-5.68
Year 10	-0.29	-6.46

Table B.9: Development year effect β_j

	personal auto	commercial auto
dev 2	-0.19	-0.01
dev 3	-0.46	-0.19
dev 4	-0.24	-0.38
dev 5	-0.30	-1.26
dev 6	-0.40	-2.19
dev 7	-0.25	-2.81
dev 8	-0.10	-4.38
dev 9	-0.17	-5.61
dev 10	-0.07	-8.81

Table B.10: Premium ω_i

	personal auto	commercial auto
year 1	48 731	30 224
year 2	49 951	35 778
year 3	52 434	42 257
year 4	58 191	47 171
year 5	61 873	53 546
year 6	63 614	58 004
year 7	63 807	64 119
year 8	61 157	68 613
year 9	62 146	74 552
year 10	68 003	78 855

Table B.11: Estimates for SUR Gaussian copula (model 1) and SUR copula mixed (model 2).

	LOB 1		LOB 2	
	model 1	model 2	model 1	model 2
(Intercept)	-1.12353	-0.98298	-1.35199	-1.56252
year2	-0.01881	0.01879	0.12451	0.18451
year3	-0.09658	-0.09476	0.15081	0.14636
year4	-0.14320	-0.16042	-0.01204	0.00775
year5	-0.15018	-0.14025	0.04106	0.06353
year6	-0.14554	-0.14294	-0.01037	0.02795
year7	-0.15722	-0.14295	0.05926	0.08909
year8	-0.17019	-0.15376	0.02306	0.05249
year9	-0.15105	-0.12807	-0.00835	0.04677
year10	-0.13720	-0.11038	-0.01279	0.07024
dev2	-0.31421	-0.32753	-0.24138	-0.24398
dev3	-1.02508	-1.04632	-0.65307	-0.64914
dev4	-1.62031	-1.67526	-1.05445	-1.09029
dev5	-2.17539	-2.24411	-1.69377	-1.69201
dev6	-3.01263	-3.09300	-2.17012	-2.25405
dev7	-3.91055	-3.96892	-2.88883	-2.97663
dev8	-4.42991	-4.48746	-3.82081	-3.98401
dev9	-5.74510	-5.83263	-3.70529	-3.87516
dev10	-5.93063	-5.94626	-4.35880	-4.46641