

A HYBRID APPROACH IN CONSTRUCTING AN INTERNAL SOLVENCY
MODEL

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ETKİN HASGÜL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
FINANCIAL MATHEMATICS

SEPTEMBER 2023

Approval of the thesis:

**A HYBRID APPROACH IN CONSTRUCTING AN INTERNAL SOLVENCY
MODEL**

submitted by **ETKİN HASGÜL** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Financial Mathematics Department, Middle East Technical University** by,

Prof. Dr. A. Sevtap Selçuk-Kestel
Dean, Graduate School of **Applied Mathematics**

Prof. Dr. A. Sevtap Selçuk-Kestel
Head of Department, **Financial Mathematics**

Prof. Dr. A. Sevtap Selçuk-Kestel
Supervisor, **Actuarial Sciences, METU**

Examining Committee Members:

Prof. Dr. Ceylan Yozgatlıgil
Statistics, METU

Prof. Dr. A. Sevtap Selçuk-Kestel
Actuarial Sciences, METU

Assoc. Prof. Dr. Uğur Karabey
Actuarial Sciences, Hacettepe University

Assist. Prof. Dr. Başak Bulut Karageyik
Actuarial Sciences, Hacettepe University

Assist. Prof. Dr. Büşra Zeynep Temoçin
Actuarial Sciences, METU

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ETKİN HASGÜL

Signature :

ABSTRACT

A HYBRID APPROACH IN CONSTRUCTING AN INTERNAL SOLVENCY MODEL

HASGÜL, ETKİN

Ph.D., Department of Financial Mathematics

Supervisor : Prof. Dr. A. Sevtap Selçuk-Kestel

September 2023, 96 pages

This thesis presents a innovative hybrid approach combining Time Series, Artificial Neural Network (ANN), and Copula models for calculating the Solvency Capital Requirement (SCR) concerning Non-Life Premium Risk across multiple lines of businesses (LoB). The loss ratio is formulated as $Z_i = \hat{X}_i + \hat{Y}_i + \varepsilon_i$, where \hat{X}_i represents the Time Series component, and \hat{Y}_i denotes the ANN component. Initially, the loss ratios are subjected to modeling through suitable time series models to capture the linear component of the model. Subsequently, the appropriate autoregressive neural network (NNAR) model is applied to the residuals resulting from the time series modeling for each LoB, representing the non-linear component. Lastly, the residuals of the combined model are modeled using an R-vine structure. By utilizing these models and the copula structure, simulated loss ratios are generated for the selected LoBs, enabling the analysis of Non-Life Premium Risk. The comparison with the Standard SCR model and the proposed model incorporating VaR and TVaR is performed to assess the efficacy of the proposed approach.

Keywords: SARIMA, ANN, NNAR, COPULA, R-vine, VaR, TVaR, SCR, Premium Risk

ÖZ

İÇSEL YETERLİLİK MODELİNİN OLUŞTURULMASINDA HİBRİT BİR YAKLAŞIM

HASGÜL, ETKİN

Doktora, Finansal Matematik Bölümü

Tez Yöneticisi : Prof. Dr. A. Sevtap Selçuk-Kestel

Eylül 2023, 96 sayfa

Bu tez, çeşitli hayatdışı sigortalar branşlarında ele alınan Prim Riski için Solvency Sermaye Gerekliliğini (SCR) hesaplamak için Zaman Serisi, Yapay Sinir Ağı (YSA) ve Kopula modellerini birleştiren yeni bir hibrit yaklaşım sunmaktadır. Hasar prim oranı $Z_i = \hat{X}_i + \hat{Y}_i + \varepsilon_i$ formülü ile tanımlanmıştır. Bu formülde \hat{X}_i zaman serisi bileşenini, \hat{Y}_i ise YSA bileşenini temsil eder. İlk olarak, hasar prim oranları uygun zaman serisi modelleri kullanılarak modellenir ve modelin lineer bileşeni oluşturulur. Sonrasında, uygun otoregresif sinir ağı (NNAR) modeli, her bir branş için uyumlandırılan zaman serisi modellemesinin sonucunda oluşan artıklara uygulanarak, modelin doğrusal olmayan bileşenini tespit etmek için kullanılır. Son olarak, birleşik modelin artıkları R-vine yapısı kullanılarak modellenir. Zaman serisi, otoregresif sinir ağı ve kopula yapısı kullanılarak, seçilen branşlar için hasar prim oranı simülasyonları elde edilir ve Hayat Dışı Prim Riski analizi yapılabilir. Önerilen yaklaşımın etkinliğini değerlendirmek için Standart SCR modeli ve önerilen hibrit model sonucunda ortaya çıkan VaR ve TVaR değerleriyle karşılaştırma yapılır.

Anahtar Kelimeler: SARIMA, YSA, NNAR, COPULA, R-vine, VaR, TVaR, SCR, Prim Riski

To My Family

ACKNOWLEDGMENTS

I would like to articulate my profound gratitude to my thesis supervisor Prof. A. Sevtap Selçuk-Kestel for her patient mentorship, wholehearted encouragement, constructive feedback and valuable advices during the development and preparation of this thesis. Her willingness to dedicate her time and share her insights has enhanced my journey.

I wish to express my gratitude to Prof. Ceylan Yozgatlıgil and Assoc. Prof. Uğur Karabey for their valuable comments through the thesis monitoring meetings.

I would like to thank my colleagues and friends, Dr. Ozan Evkaya, Dr. Özenç Murat Mert, Anıl Gülveren, Oğuz Koç and Buse Uysal for imparting their expertise and providing support to me.

Lastly, I wish to convey my gratitude to my parents Rıza Hasgül and Mühibe Hasgül, my brother Eren Hasgül, my wife Eylül Fitoz Hasgül and our beloved friend Ruby for their endless support and patience during the thesis phase.

TABLE OF CONTENTS

ABSTRACT	vii
ÖZ	ix
ACKNOWLEDGMENTS	xi
TABLE OF CONTENTS	xiii
LIST OF TABLES	xvii
LIST OF FIGURES	xix
LIST OF ALGORITHMS	xxi
LIST OF ABBREVIATIONS	xxii
CHAPTERS	
1 INTRODUCTION	1
1.1 Organization of the thesis	4
2 SOLVENCY II AND TURKISH INSURANCE MARKET	5
2.1 Solvency II requirements in Türkiye	9
3 PRELIMINARIES IN METHODS	13
3.1 Loss Ratio	13
3.2 Time Series Modeling	15

3.2.1	Components of Time Series Process	17
3.2.2	Trend	17
3.2.3	Seasonality	19
3.2.4	Random/Noise Term	19
3.2.5	Time Series Modeling Algorithm	19
3.3	Time Series Residuals	21
3.4	Artificial Neural Network (ANN)	21
3.4.1	Neural Network Autoregressive Model (NNAR)	21
3.5	Performance Metrics	23
3.6	Comperative Measures	24
3.7	Copula Models	25
3.7.1	Bivariate Copulas	26
3.7.1.1	Elliptical Copulas	26
3.7.1.2	Archimedean Copulas	27
3.7.1.3	Non-parametric Copulas	29
3.7.2	Vine Structure	29
3.7.3	Value-at-Risk & Tail Value-at-Risk	33
4	INTERNAL SOLVENCY MODEL	35
4.1	Copula Modeling Algorithm	36
4.2	Simulation	36
4.2.1	SCR: Non-Life Premium Risk	38

4.2.2	Proposed Model Algorithm	38
4.3	Limitations	42
5	CASE STUDY: NON-LIFE PREMIUM RISK	43
5.1	Data	43
5.2	Time Series Modeling	46
5.3	Artificial Neural Network - NNAR	52
5.4	Copula Modeling	55
5.5	Application Utilizing the Box-Cox Transformation	59
5.6	Simulation	63
5.7	Ruin Probability	68
6	CONCLUDING COMMENTS	71
APPENDICES		
A	TIME SERIES ANALYSES	75
A.1	Accident	75
A.2	Fire	78
A.3	MTPL	81
A.4	Land Vehicle	84
A.5	Miscellaneous	87
REFERENCES	91
CURRICULUM VITAE	95

LIST OF TABLES

Table 2.1 Standard Deviations for Premium Risk (SII)	8
Table 2.2 Non-Life Correlation Table	8
Table 2.3 Non-Life Correlation Table (Adjusted)	8
Table 5.1 Descriptive Statistics: Loss Ratios	44
Table 5.2 Comparison of Time Series Models: Health	48
Table 5.3 Stationary Tests for LoBs' original loss ratios and differences	50
Table 5.4 Best Fitted Time Series Models of LoBs	51
Table 5.5 Diagnostic Tests Results: P-values	51
Table 5.6 Comparison of Time Series and TS+ANN Models	53
Table 5.7 Normality Tests for TS and TS+ANN	54
Table 5.8 Best Performed Hybrid Models: TS+ANN	54
Table 5.9 R-vine structure: Bivariate distributions of edges	58
Table 5.10 Comparison of Time Series and Hybrid Models (training set)	60
Table 5.11 Comparison of Time Series and Hybrid Models (testing set)	61
Table 5.12 Best Performed Hybrid Models for Transformed Data	62
Table 5.13 R-vine structure: Bivariate distributions of edges (data with Box-Cox transformation)	62
Table 5.14 Standard Deviations used in the Standard Formula	63
Table 5.15 Calculations with the estimated volumes for 2021 and equally weighted volumes (each 100)	64
Table 5.16 SCR ratios with the estimated volumes for 2021 and equally weighted volumes	65

Table 5.17 Sensitivity analysis for 5% shock in each loss ratio	68
Table 5.18 Ruin probability comparison	69
Table A.1 Comparison of Time Series Models: Accident	75
Table A.2 Comparison of Time Series Models: Fire	78
Table A.3 Comparison of Time Series Models: MTPL	81
Table A.4 Comparison of Time Series Models: Land Vehicle	84
Table A.5 Comparison of Time Series Models: Miscellaneous	87

LIST OF FIGURES

Figure 2.1 Gross Written Premiums in Turkish Insurance Industry 2015-2022, BLN TRY. Premiums are taken from Insurance Association of Turkey, Industry Reports [2]	10
Figure 2.2 Gross Written Premiums in Turkish Insurance Industry 2015-2022, BLN USD. Premiums are taken from Insurance Association of Turkey, Industry Reports [2]. Annual Averages of USD/TRY Parity is used for conversion, the rates are retrieved from TCMB EVDS [1].	10
Figure 3.1 Autoregressive Neural Network	22
Figure 4.1 Internal Model Diagram	41
Figure 5.1 Loss Ratios of the chosen LoBs	45
Figure 5.2 Health Insurance loss ratios and its components, 2009-2018	47
Figure 5.3 Coefficient Test for Health Time Series Model	48
Figure 5.4 Health Residuals	49
Figure 5.5 Residuals of the ultimate models	55
Figure 5.6 Copula pairs of the residuals of six LoBs	56
Figure 5.7 R-vine fitting for six dimensional structure	57
Figure 5.8 R-vine Contour	59
Figure 5.9 Graphical representation of Solvency Capital Requirements	66
Figure 5.10 Graphical representation of Solvency Capital Requirements after the Box-Cox transformation is applied	67
Figure A.1 Accident Insurance loss ratios and its components, 2009-2018	76
Figure A.2 Diagnostic checks of Accident branch	77

Figure A.3	Fire Insurance loss ratios and its components, 2009-2018	79
Figure A.4	Diagnostic checks of Fire branch	80
Figure A.5	MTPL Insurance loss ratios and its components, 2009-2018	82
Figure A.6	Diagnostic checks of MTPL branch	83
Figure A.7	Land Vehicle Insurance loss ratios and its components, 2009-2018	85
Figure A.8	Diagnostic checks of Land Vehicle branch	86
Figure A.9	Miscellaneous Insurance loss ratios and its components, 2009-2018	88
Figure A.10	Diagnostic checks of Miscellaneous branch	89

LIST OF ALGORITHMS

ALGORITHMS

Algorithm 1	Time Series Modeling	20
Algorithm 2	Sequential method	32
Algorithm 3	Copula Modeling	36
Algorithm 4	Proposed Model Algorithm	39
Algorithm 5	Proposed Model Algorithm, continued	40

LIST OF ABBREVIATIONS

Acc	Accident
ACF	Autocorrelation Function
ADF	Augmented Dickey-Fuller
AIC	Akaike information criterion
ANN	Artificial Neural Network
AR	Autoregressive
ARCH	Autoregressive Conditional Heteroskedasticity
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BIC	Bayesian information criterion
BP	Breusch-Pagan
CVaR	Conditional Value at Risk
Fir	Fire
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GDP	Gross Domestic Product
GOF	Goodness of fit
Hea	Health
IM	Internal Model
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
Lan	Land Vehicle
LoB	Line of business
NNAR	Autoregressive Neural Network
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MCR	Minimum Capital Requirement
Mis	Miscellaneous
MSE	Mean Squared Error

Mtp	Motor Third Party Liability
MTPL	Motor Third Party Liability
PACF	Partial Autocorrelation Function
RMSE	Root Mean Squared Error
R-vine	Regular Vine
RWWD	Random Walk With Drift
SARIMA	Sasonal Autoregresive Integrated Moving Average
SCR	Solvency Capital Requirement
SW	Shapiro-Wilk
TCMB	Central Bank of the Republic of Türkiye
TS	Time Series
TVaR	Tail Value at Risk
VaR	Value at Risk
$\rho_{i,j}$	Correlation between LoBs i and j
$P_i^{t,written}$	The estimate of net written premium for LoB i and period t
$P_i^{t,earned}$	The estimate of net earned premium for LoB i and period t
NL_p	The capital requirement for the premium risk
$V_{prem,lob}$	Volume calculated as an exposure for the premium risk
$Z_{i,t}$	Loss ratio for LoB i and period t
$L_i^{t,incurred}$	The sum of claim payments for LoB i and period t
$E_i^{t,attributable}$	The expenses attributable to the losses
$\hat{X}_{i,t}$	TS model estimation for LoB i and period t
$\hat{Y}_{i,t}$	NNAR model estimation of TS residual for LoB i and period t
$w_{i,t}$	Residual of time series model for LoB i and period t
$\varepsilon_{i,t}$	Residual of hybrid model for LoB i and period t
$C(u, v)$	Copula function of u and v
$R_{i,T+1}$	The required reserve amount for LoB i and period T+1
$S_{i,T+1}$	The required reserve amount allocated to SCR amount for LoB i and period T+1

CHAPTER 1

INTRODUCTION

The concept denoted as solvency pertains to the capacity of an (re)insurance company to sustain its operations over an extended duration, fulfilling all commitments, and possessing sufficient equity to ensure the continuation of its activities. Studies on efficient reserving have been conducted in a regulative basis as well. Although Solvency I has brought strong bases for a realistic minimum capital requirement determination, there were areas that needed to be improved such as diversification in accumulation of multiple lines of businesses and valuation methods of assets and liabilities. In 2016, Solvency II came into force as a regulatory framework for EU countries [12]. The Solvency II Directive aims to ensure the enduring financial solvency of (re)insurance company with regards to their capacity to fulfill obligations towards insured parties. Apart from regulatory and auditing aspects, the Solvency II framework primarily focuses on managing underwriting, market, credibility, operational, and liquidity risks.

Assessing the right amount of reserve has always been one of the challenges in (re)insurance business. Solvency II provides a solid framework for risk accumulation and a standard formula for capital requirements as a holistic approach. Despite having accumulated multiple lines of businesses and risks accurately in symmetric distributions, standard formula for Solvency Capital Requirement (SCR) has shortcomings in skewed distributions. Known that most of the loss distributions are skewed, developing internal models is of great importance. Although calibration methodologies are published by the European Insurance and Occupational Pension Authority (EIOPA), we believe that employing Copulas, known as link functions between variables, is

superior to all other methods.

Moreover, although the utilization of copulas effectively explains the dependence structure within a multivariate scenario, relying solely on copulas for modeling remains insufficient due to the time dimension inherent in the data used. A comprehensive assessment should also be done for both linear and non-linear time components to mitigate the influence of time-dependent effects on the dependence structure. It provokes curiosity to assess the feasibility of constructing a model capable of estimating a capital requirement based on the solvency criteria, wherein due consideration is given to the historical loss ratios.

Copulas, "bond" or "link" functions are introduced by Sklar in 1959 [33, 34]. Following the publication of the theorem, copulas are used in many areas, such as risk aggregation in financial studies. In 1998, Wang proposes a set of tools for aggregation of correlated risks using Monte Carlo Simulations or fast Fourier transform [38]. Frees and Valdez (1998) employ various Copulas to insurance loss data and provide fitting processes [15]. Having intensified works with Solvency II, Sandström (2007) set forth a calibration study regarding SCR calculation and shows the effect of non-calibrated studies on skewed data [29]. In 2011, Savelli and Clemente publish their work on aggregation of premium risk by using Copulas in hierarchical structures [30]. Nguyen and Molinari (2011) draw the attention to the importance of using Copulas in SCR calculations as the required amount is decreased [24]. On the other hand, Haug et al. (2011) focus on testing extreme value copulas and tail copulas on the extreme losses [16]. Ismail (2018) uses mixture of copulas in aggregation of insurance risks to achieve a better fit so that overestimation or underestimation is avoided while calculating SCR [20]. In 2019, Pellecchia and Perciaccante propose an alternative methodology to existing standard formula given in Solvency II Directive, which is based on employing Copulas in order to avoid underestimation of diversification effect [27]. Eling and Jung (2020) compare standard models used in Solvency II with internal models and conclude that standard models results in 50% higher solvency capital requirement [13].

Taking into account the time-dependency of data, incorporating time parameters in Copula models becomes a significant requirement in financial modeling. Patton (2001)

presents some results on Copula models indicating time-varying conditional expectation [25] and reviews his studies in 2012 [26]. Oleg and Dick (2011) proposed a Copula-based time series model to forecast time series [35]. Zhao et al. (2021) publish a Copula-based multivariate time series model [40].

Having calculated premium and reserve risks in SCR, in some of the studies, time-dependent loss ratios are needed to be engaged to detect volatility. Araichia et al. (2017) present a General Autoregressive Conditional Sinistrality model, used in reserve modeling and risk aggregation, coping with temporal dependence [3].

Several researches in Turkey have been conducted pertaining to Solvency II and Copula modeling. Höbek (2016) centers on the computation of non-life premium and reserve risks by employing both the standard formula and the copula based internal model across three different size insurance companies [17]. The outcome of this study includes a comparison on SCR. Erdener (2016) proposed a method to estimate the present value of the mean pure premium per policy through simulations utilizing the Clayton copula probability distribution function, which simultaneously specifies the dependence structure [36].

While constructing internal models, it becomes essential to incorporate the company's risk attributes encompassing its operational environment, portfolio composition, and other factors. As a quantitative measure, the loss ratio stands as a prominent representation of a company's risk profile, especially for the premium risk. Hence, the central focus of this thesis is the modeling of the loss ratios. Considered both seasonality and trend of loss ratios of a specific line of business (LoB), time series models or parameters help to detect solvency requirements more accurately. Aligned with this concept, a hybrid internal model is suggested for the purpose of estimating SCR using historical loss ratios. This process involves employing appropriate models to explain time components and clarifying the characterization of dependence among LoBs in a more sophisticated way. Thus, the premium risk of the chosen lines of businesses are modeled using a hybrid approach incorporating suitable time series models and autoregressive neural network models. Additionally, R-vine copula structure is used through the accumulation processes.

1.1 Organization of the thesis

In Chapter 2, the basic definition of Solvency II and some information regarding the current perspective of the Turkish Insurance System are given. The calculation methodology for the standard solvency capital requirement is outlined. In addition, the current status of the Turkish Insurance industry is provided.

Chapter 3 gives the descriptions of the models and algorithms utilized in the proposed model. Firstly, time series models including regular and seasonal components are expressed. Secondly, definition of autoregressive neural network model is made. Next, copula models including R-vine structure is illustrated. Lastly, VaR and TVaR measurements are defined. Furthermore, the algorithms for the selection criteria of the models and performance metrics are also covered in this chapter.

In Chapter 4, a comprehensive description of the proposed model is provided. The internal solvency model regarding premium risk is addressed, and the approach followed during the modeling process is explained. The general algorithm of the proposed approach is also given in this chapter.

Chapter 5 covers the implementation part of the thesis. The approach detailed in Chapter 4 is followed through employing the methods explained in Chapter 3. The outcomes of the models and the subsequent comparisons are presented within this chapter, alongside the conducted simulations.

CHAPTER 2

SOLVENCY II AND TURKISH INSURANCE MARKET

The Solvency II framework encompasses primary considerations involving underwriting, market, credibility, operational, and liquidity risks, alongside regulatory and auditing concerns. The Solvency II Directive is structured around three pillars, wherein all the aforementioned risk evaluations are encompassed within Pillar I. This pillar involves quantitative assessments, the application of quantitative models, their validation, and the computation of capital requisites, including the Solvency Capital Requirement (SCR) and Minimum Capital Requirement (MCR). Meanwhile, Pillar II emphasizes the necessity of a well-structured internal audit, as well as the insurers' adeptness in proficiently managing their own risks and fostering sound corporate governance. Pillar III pertains to the firm's obligation to furnish information to external auditors and to ensure transparency for the public.

The Solvency II directive posits that the economic capital, as ascertained through the SCR, is mandated to maintain a probability of ruin constrained to 0.5%, with the assessment grounded in the application of volatility metrics such as Value at Risk (VaR) or Conditional Value at Risk (CVaR) [24]. The calculation of a common SCR for a composition of various branches involves more than just the aggregation of these individual branches; it also covers the assessment of potential dependencies. Hence, it is essential to have an understanding of the dependence structure among risk factors associated with the respective branches. The expression of the common SCR formula, functioning as an aggregation formula, is presented in the following equation:

$$\text{SCR}_{\text{common}} = \sqrt{\sum_{i,j} \rho_{i,j} \text{SCR}_i \text{SCR}_j} \quad (2.1)$$

where SCR_i and SCR_j are the solvency capital requirements for i^{th} and j^{th} branches the insurance company operates in and $\rho_{i,j}$ expresses the correlation coefficient between LoBs i and j .

In the scope of this thesis, the non-life premium risk segment is studied. $\text{SCR}_{\text{common}}$ represents the SCR for aggregate premium risk of all branches. Premium risk is a form of risk that arises from variations in the timing, frequency, and severity of events that are insured. It applies to both new insurance policies and unexpired risks on the existing contracts. A significant aspect of the premium risk is that the provisions allocated by an insurance company to cater for claims may be inadequate, or require an increase (QIS5 SCR.9.9. [12]). This occurrence is likely when the insured events turn out to be more frequent or severe than initially expected. Premium risk encompasses not only the risk that arises due to variations in the timing, frequency, and severity but also encompasses the risk arising from the fluctuation of expense disbursements. This form of risk is particularly material for certain lines of business and must be accurately represented within the module calculations. Expense risk is considered to be an implicit component of premium risk (QIS5 SCR.9.10. [12]). On the other hand, Reserve risk is a type of risk that arises from the variability in the timing and the magnitude of claim settlements (QIS5 SCR.9.11. [12]).

Standard deviation (σ) and volume (V) are two main variables taken into account to be able to calculate the amount for each branch's capital requirement.

The estimate of net written premium ($P_i^{t,\text{written}}$) for each LoB i during the for the next year and the actual net written premium ($P_i^{t-1,\text{written}}$) of the previous year should be known. In addition, the estimate of net earned premium ($P_i^{t,\text{earned}}$) should also be known for the calculation of premium risk (QIS5 SCR.9.12. [12]).

The capital requirement, NL , for the premium risk is shown by:

$$NL_p = \rho(\sigma) \cdot V \quad (2.2)$$

where V is the volume measure, σ is the combined standard deviation and $\rho(\cdot)$ is a function which converts σ to a corresponding value to be used for the Value at Risk (VaR) calculation (QIS5 SCR.9.16. [12]) and shown as:

$$\rho(\sigma) = \frac{\exp(N_{0.995} \cdot \sqrt{\log(\sigma^2 + 1)})}{\sqrt{(\sigma^2 + 1)}} - 1 \quad (2.3)$$

where $N_{0.995}$ is the 99.5th percentile of the standard normal distribution (QIS5 SCR.9.17. [12]).

Resulted from the calibration and the lognormal distribution assumption on QIS5 Technical Specifications publication, $\rho(\sigma)$ converges to $3 \cdot \sigma$ for $\text{VaR}_{99.5}$ (QIS5 SCR.9.18. [12]). The proposed simplification is very useful for the calculations.

The function of volume measure for the premium (P) risk for a LoB i is as follows (QIS5 SCR.8.70. [12]):

$$V_{(i,prem)} = \max((P_i^{t,written}); (P_i^{t-1,written}); (P_i^{t,earned})) \quad (2.4)$$

On the other hand, the standard deviations for premium risk of LoBs are also listed in the regulation and given in Table 2.1. It should be noted that Health (Health&Sickness) and Accident branches are being classified as Non-Life segments in Türkiye. Thus, those two branches are analyzed as non-life underwriting risks.

LoBs characterized by higher risk exhibit correspondingly elevated standard deviations. The data in Table 2.1 indicates that Miscellaneous branch represents the most risky LoB, while Health stands out as comparatively less risky within the the studied sample. These rates contain the historical insights of the European insurance industry, as established by the European Insurance and Occupational Pensions Authority (EIOPA) through Solvency II regulation.

Table 2.1: Standard Deviations for Premium Risk (SII)

LoB	Standard Deviation
Motor vehicle liability-MTPL	10%
Other motor - Land Vehicle	8%
Fire	8%
Miscellaneous	13%
Accident	4%
Health	4%

While calculating aggregate capital requirements covering multiple LoBs, the consideration of the dependence structure becomes necessary to facilitate diversification. It is where the correlation matrix is needed to be activated, which assumes significant role in this process. The correlation values given in Table 2.2 are provided by EIOPA through Solvency II regulation [12].

Table 2.2: Non-Life Correlation Table

Corr Matrix	1	2	3	4
1: Motor vehicle liability	1	.	.	.
2: Other motor	0.5	1	.	.
3: Fire	0.25	0.25	1	.
4: Miscellaneous	0.5	0.5	0.5	1

A correlation coefficient of 0.5 is attributed to the relationship between the Health and Accident LoBs, as the Health segment demonstrates a correlation of 0.5 between Sickness and Accident. Furthermore, in alignment with the directive's proposition, the correlation coefficient between the Health segment and the Non-Life is determined as 0. Under our revised arrangements in relation to utilized LoBs, the modified correlation table is presented as follows:

Table 2.3: Non-Life Correlation Table (Adjusted)

Corr Matrix	1	2	3	4	5	6
1: Motor vehicle liability	1
2: Other motor	0.5	1
3: Fire	0.25	0.25	1	.	.	.
4: Miscellaneous	0.5	0.5	0.5	1	.	.
5: Health	0	0	0	0	1	.
6: Accident	0	0	0	0	0.5	1

The risk of financial losses associated with unearned premiums arises during the period in which the insurer remains accountable for the underwritten risks. Uncertainty relating to claims, engaging factors such as frequency, severity, and occurrence time, manifests as premium risk under the context of Solvency. This premium risk incorporates both the risks underwritten during the period and the unexpired contracts carried over from the previous period.

Since only premium risk is studied in this thesis, there is an assumption that $NL_{pr} = NL_p = NL$. Henceforth, our focus is directed towards SCR arising from premium risk. The formula used to calculate NL is given as follows:

$$NL_{\text{common}} = \sqrt{\sum_{i,j} \rho_{i,j} NL_i NL_j} \quad (2.5)$$

where NL_i is the capital requirement of premium risk for i^{th} branch the insurance company operates in and $\rho_{i,j}$ represents the correlation coefficient between LoBs i and j .

2.1 Solvency II requirements in Türkiye

The Turkish insurance industry has a well-established presence, having operated for many decades. Nevertheless, its magnitude remains relatively modest in relation to the Turkish Gross Domestic Product (GDP). With the written premiums to GDP ratio standing at approximately 2%, this value appears small in contrast to European countries. Consequently, the Turkish insurance sector demonstrates significant potential, given its current standing.

Figure 2.1 illustrates the gross written premiums spanning from 2015 to 2022. An inference can be drawn from the data, indicating that the Non-Life sector in Turkey is significantly larger in size compared to the Life business. It also seems that the industry has a continuous growth in terms of TRY

Additionally, there appears to be a consistent growth within the industry in terms of Turkish Lira (TRY). Despite encountering inflation shocks in recent years (2021-

2022), Turkish insurance industry exhibited a corresponding revenue increase and manages to maintain its share of the GDP.

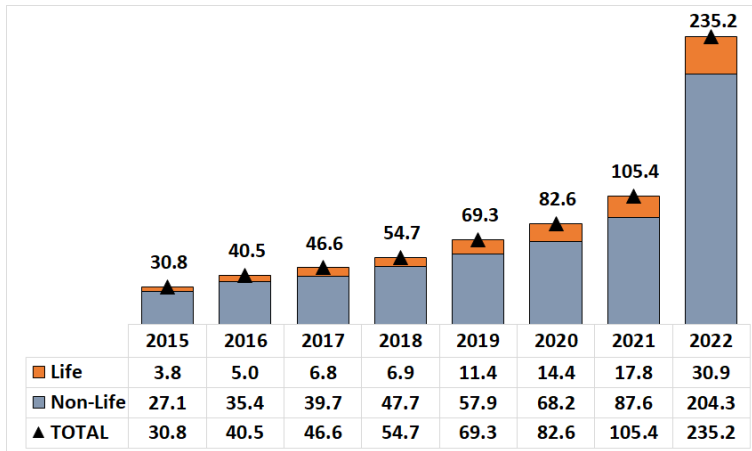


Figure 2.1: Gross Written Premiums in Turkish Insurance Industry 2015-2022, BLN TRY. Premiums are taken from Insurance Association of Turkey, Industry Reports [2]

Concerning Gross Written Premiums in USD, Figure 2.2 displays a relatively consistent flat line pattern, maintaining a range of USD 11.3 to 14.2 billion.

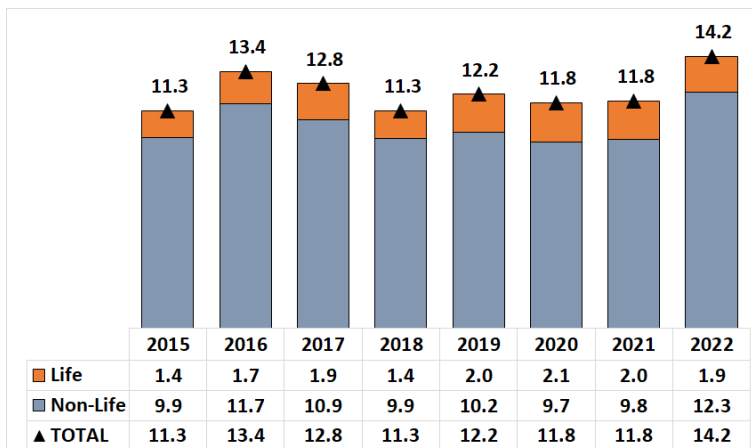


Figure 2.2: Gross Written Premiums in Turkish Insurance Industry 2015-2022, BLN USD. Premiums are taken from Insurance Association of Turkey, Industry Reports [2]. Annual Averages of USD/TRY Parity is used for conversion, the rates are retrieved from TCMB EVDS [1].

As Türkiye is not a European Union (EU) member, the implementation of Solvency II is not obligatory for the companies in the country. Nonetheless, in the framework of the European Union Accession Process and in pursuit of fostering a more secure environment, the integration of practices compatible with Solvency II appears to be a pragmatic approach for Turkish regulatory authorities.

In the study by İbiş and Çoban (2017), the necessary regulatory modifications for conformity with Solvency II are highlighted. In the context of this thesis, the subsequent modifications are examined within the current system to ensure alignment with Solvency II: describing the risks in details for inclusion in calculations, incorporating the dependence structure among risks, and accounting for the influence of risk diversification on calculations, along with granting flexibility to choose between the standard formula and full/partial internal models [19].

CHAPTER 3

PRELIMINARIES IN METHODS

The calculations for SCR, arising from premium risk, based on the standard formula rely on fixed standard deviation values for each LoB and a predetermined correlation matrix that governs the dependencies among LoBs. This method adopts a purely linear approach to model the interrelationships among LoBs. The determination of premium risk can be obtained by adopting the methodology proposed in QIS3 (2007), which involves the use of historical loss ratios [8]. In this approach, the standard deviation of historical loss and reserve ratios specific to a given LoB is regarded as the risk margin, expressed as a percentage of net earned premium. The proposed model incorporates suitable copula structures and functions to effectively model the relationships among LoBs, subsequent to the fitting of the optimal time series model for the linear part and neural network for the non-linear part of loss ratios. In this section, a detailed explanation of the methodology is provided, outlining the key components and steps involved.

3.1 Loss Ratio

The loss ratio assumes a pivotal role in actuarial practice, functioning as a fundamental metric to evaluate the profitability and underwriting efficiency of insurance companies. It covers the relationship between incurred losses, denoting the aggregate claims, and earned premiums over a defined time period. From an academic standpoint, actuarial professionals precisely analyze the loss ratio to gather the valuable insights into risk management strategies, pricing methodologies, claims management practices, and underwriting procedures. Such detailed examination supports the in-

formed decision-making within the insurance field. In addition to various other dimensions, the loss ratio has a critical role within the Solvency II framework, too. It serves as an indispensable tool for assessing premium and reserve risk. The loss ratio, expressed as a percentage, is derived through a defined formula as follows:

$$\begin{aligned}
 Z_{i,t} &= \frac{\text{Incurred Losses}}{\text{Earned Premium}} \\
 &= \frac{L_i^{t,incurred}}{P_i^{t,earned}}, \tag{3.1}
 \end{aligned}$$

where $L_i^{t,incurred}$ refer to the aggregate sum of claim payments disbursed by an insurance company for LoB i during a specified time frame, t . Conversely, $P_i^{t,earned}$ denote the total amount of premiums earned by the insurance company from LoB i over the same defined time period, t .

In the context of equation 3.1, an associated adjustment is conventionally made to enhance the numerator by incorporating the expenses attributable to losses. This modification of the ratio acknowledges the inclusion of relevant expenditures linked to the incurred losses, thereby offering a more comprehensive and accurate representation of the financial impact arising from claims activities. The revised form of equation 3.1 is expressed as follows:

$$\begin{aligned}
 Z_{i,t} &= \frac{\text{Incurred Losses} + \text{Attributable Expenses}}{\text{Earned Premium}} \\
 &= \frac{L_i^{t,incurred} + E_i^{t,attributable}}{P_i^{t,earned}}, \tag{3.2}
 \end{aligned}$$

with the assumption that the expenses attributable to losses for the Turkish market constitute 70% of the total expenses.

The volatility exhibited by the loss ratio pertaining to a specific line of business provides significant insights regarding the underwriting risk associated with that particular segment. Consequently, the utilization of the standard deviation of the loss ratio assumes remarkable importance as a fundamental parameter. It is asserted that the standard deviation for each LoB is calculated by

$$\sigma_i^2 = \sum_{t=1}^T \frac{(Z_{i,t} - \bar{Z}_i)^2}{T-1}, \quad (3.3)$$

where $\bar{Z}_i = \frac{\sum_{t=1}^T Z_{i,t}}{T}$, t represents the period and T is the total number of periods.

3.2 Time Series Modeling

Given the quarterly nature of loss ratio datasets, we postulate that the genuine relationships among LoBs can be discerned by studying time series components. To accomplish this, a meticulous time series analysis needs to be conducted, aiming to mitigate the influence of temporal factors and uncover the underlying dependencies. By modeling time-related effects, such as trend and seasonality, the analysis facilitates a clearer assessment of the true relationships among the LoBs.

Consistent with the aforementioned concept, an examination of $Z_{i,t}$'s is aimed to be conducted utilizing suitable time series models. Assume that X_t represents the linear component of $Z_{i,t}$ in the context of time series model definitions.

Definition 3.2.1. The autoregressive model AR(p) is expressed as

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \cdots + \phi_p(x_{t-p} - \mu) + \omega_t, \quad (3.4)$$

where x_t is stationary, and ϕ_j has constant values for $j = 1, 2, \dots, p$, knowing that $\phi_p \neq 0$. Here, we assume that noise term, ω_t , has a Gaussian distribution with mean μ and standard deviation σ_w . In case that $\mu = 0$, ω_t turns into a Gaussian white noise series [32].

AR(p) can also be represented as

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t, \quad (3.5)$$

where $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$.

If AR(p) model is represented with autoregressive operator, the representation becomes

$$\phi(B)x_t = w_t \quad (3.6)$$

where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (3.7)$$

and B is the backshift operator which has a form of

$$B^j x_t = x_{t-j} \quad \text{for} \quad j = 1, 2, \dots, p. \quad (3.8)$$

Definition 3.2.2. The **moving average model** MA(q) is

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q} \quad (3.9)$$

where there are q lags in the moving average (order q) with parameters $\theta_1, \theta_2, \dots, \theta_q$ knowing that $\theta_q \neq 0$. It is assumed that w_t follows a normal distribution with mean zero and variance σ_w .

Equation 3.9 can also be represented by using backshift operator B as

$$x_t = \theta(B)w_t, \quad (3.10)$$

where

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q, \quad (3.11)$$

with the condition that the moving average process maintains stationarity across all parameter values θ_j for $j = 1, 2, \dots, q$.

Definition 3.2.3. The **autoregressive moving average model** ARMA(p,q) is expressed as

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q}, \quad (3.12)$$

where $\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$ and $\phi_p \neq 0, \theta_q \neq 0$ and $\sigma_w^2 > 0$. It is assumed that w_t follows a normal distribution with mean zero and variance σ_w . If x_t has a zero mean, equation 3.12 turns into

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q}, \quad (3.13)$$

The backshift operator B expression of ARMA(p,q) is

$$\phi(B)x_t = \theta(B)w_t, \quad (3.14)$$

where

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q, \quad (3.15)$$

Definition 3.2.4. x_t process is called as **autoregressive integrated moving average model**, ARIMA(p,d,q), if the following is satisfied:

$$\nabla^d x_t = (1 - B)^d x_t \quad (3.16)$$

which turns into ARMA(p,q) process. In general, the ARIMA can be expressed as follows:

$$\phi(B)(1 - B)^d x_t = \delta + \theta(B) \quad (3.17)$$

knowing that $E(\nabla^d x_t) = \mu$ and $\delta = \mu(1 - \phi_1 - \dots - \phi_p)$.

Definition 3.2.5. The multiplicative seasonal **autoregressive integrated moving average model** SARIMA(p,d,q)(P,D,Q)[s] model is given by

$$\Phi_P(B^s)\phi(B)\nabla_s^D \nabla^d x_t = \delta + \Theta_Q(B^s)\theta(B)w_t \quad (3.18)$$

where w_t is the conventional Gaussian white noise process. Here the orders P and Q are shown by $\Phi_P(B)$ and $\Theta_Q(B)$, respectively. Seasonal difference is shown by $\nabla_s^D = (1 - B^s)^D$.

Seasonality can be defined as multiplicative or additive according to dynamics of the effect.

3.2.1 Components of Time Series Process

The primary objective of this section is to eliminate the influence of factors such as trends and seasonality on the loss ratio, which serves as our time series data. By doing so, we aim to isolate the random component and facilitate the analysis of the solvency margin derived from the non-life segment through premium risk.

3.2.2 Trend

In order to achieve stationarity in the process, it is necessary to address the presence of a trend in the series. By detrending the process, an important step is taken towards

obtaining a random volatility component. Trends observed in time series processes can be classified into two distinct types: Deterministic trend and stochastic trend.

Deterministic trends are eliminated via model fitting such as linear, logarithmic, exponential, and etc. On the other hand, stochastic trends are removed via differencing method which is discussed in the integrated models. Differencing is implemented in ARIMA and SARIMA models with the specified parameters "d" and "D" (d for regular trend, D for seasonal trend which will be given in the following subsection).

Trends can be categorized, in another dimension, as global or local. Global trends exist in whole time while local trends only affects a specific period in timeline of the process. Hence, it is crucial to examine the nature of the trend in both dimensions. By incorporating a model to capture the underlying trend in the time series, the process becomes trend-stationary.

A time series typically consists of a nonstationary trend component along with a stationary component characterized by a mean of zero. In the context of this model, we consider the following assumption,

$$x_t = \mu_t + y_t \quad (3.19)$$

where $\mu_t = \beta_0 + \beta_1 t$ while y_t is the stationary component. In order to induce stationarity in this process, it is necessary to undertake specific transformations by employing difference function ∇ . Consequently, a stationary process is obtained such as:

$$\begin{aligned} \nabla x_t &= x_t - x_{t-1} \\ &= \beta_1 + y_t - y_{t-1} \\ &= \beta_1 + \nabla y_t. \end{aligned} \quad (3.20)$$

In the scenario where the nonstationary component is also stochastic in nature, as indicated by $\mu_t = \mu_{t-1} + v_t$ knowing that v_t is stationary, the difference function becomes

$$\nabla x_t = v_t + \nabla y_t, \quad (3.21)$$

which holds stationary condition. In case that μ_t has a k-th order polynomial form, $\nabla^k y_t$ becomes stationary. Nonetheless, stochastic trend models may need higher order differencing to provide stationarity [32].

3.2.3 Seasonality

In time series analyses, various periodic patterns may manifest, including annual, quarterly, monthly, weekly, daily, or intra-daily cycles, among others. Once the presence of seasonality is detected, it becomes imperative to incorporate appropriate type of model to ensure accurate estimations and predictions. Parametric methods or difference method can be both applied for de-seasoning (or season adjustment). By eliminating the cyclical patterns inherent in the data, process becomes season-stationary.

3.2.4 Random/Noise Term

Fitted the appropriate model, the series obtained by taking the difference between the observed values and the fitted values is regarded as the residual or noise term, representing the random fluctuations in the data. Noise term should be normally distributed with mean zero and a constant variance for the model to be valid. This component provides an accurate representation of the underlying volatility of the time series data, which necessitates modeling through the use of copulas.

3.2.5 Time Series Modeling Algorithm

There are various types of time series models, and the selection of an appropriate model is pivotal to achieving a desired fit with the data. In the model selection process undertaken in this thesis, specific guidelines and tests are employed. Candidate models are chosen based on insights derived from ACF and PACF plots, in addition to outcomes from stationary tests, aimed at identifying potential requirements for regular and seasonal differencing. Furthermore, the "auto.arima" function from the "forecast" package of R contributes an additional candidate to the list. Subsequent to modeling, assessments encompass stationarity, serial autocorrelation, constant variance, and normality assumptions, conducted through corresponding tests. Additionally, AIC and BIC criteria are employed to identify the best-fitted models that fulfill the requisite assumptions. The algorithm to be utilized for the modeling is delineated as follows:

Algorithm 1: Time Series Modeling

1. Visualize the series and conduct an exploratory examination for potential outliers.
2. If necessary, employ data transformation techniques such as Box-Cox to stabilize the variance.
3. Examine the stationarity of the transformed series through rigorous analysis.
 - 3.1. Utilize the following tests to determine the stationarity status of the time series: Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS).
 - 3.2. Utilize Osborn-Chui-Smith-Birchenhall (OCSB) Test to determine the seasonal unit roots.
 - 3.3. Utilize Breusch-Pagan (BP) test to define heterokedasticity
4. If the series is not stationary, employ the necessary transformations to assure stationarity
 - 4.1. For non-seasonal time series, apply regular differencing
 - 4.2. For seasonal time series, apply seasonal differencing and if necessary regular differencing
5. Identify the seasonal model by analyzing the seasonal coefficients of the ACF and PACF
6. Identify the regular component by exploring the ACF and PACF of the residuals of seasonal model
7. Evaluate the statistical significance of the coefficients.
8. Residual checks:
 - 8.1. Test for serial correlation (Ljung-Box test)
 - 8.2. Plot the ACF and PACF functions
 - 8.3. Utilize the Shapiro-Wilk normality test as a means of assessing the adherence of the residuals to the assumption of normality.
9. Perform a comparison among the alternative candidate models using AIC and/or BIC

3.3 Time Series Residuals

Definition 3.3.1. Let $\hat{X}_{i,t}$ be the fitted value for $Z_{i,t}$ by employing the best fitted time series model. Herewith, the following equation holds for each LoB:

$$Z_{i,t} = \hat{X}_{i,t} + w_{i,t}, \quad (3.22)$$

where $\hat{X}_{i,t}$ is estimated by AR, MA, ARMA, ARIMA or SARIMA model and it is assumed that the residuals of time series model, $w_i \sim \mathbb{N}(0, \sigma_{w_i})$.

3.4 Artificial Neural Network (ANN)

In light of the potential complexity within time series data, which may encapsulate both linear and nonlinear components, traditional time series models may not fully capture all the information contained within the data. Consequently, the integration of an Artificial Neural Network (ANN) subsequent to a time series model could prove instrumental in acquiring the nonlinear information as well. The hybrid model concept, proposed by Zhang (2003), suggests that such a combined approach would lead to more robust and stable results compared to exclusive use of single time series modeling techniques [39].

$$w_{i,t} = \hat{Y}_{i,t} + \varepsilon_{i,t}, \quad (3.23)$$

where $\hat{Y}_{i,t}$ represents the fitted values of w_i by using ANN model and ε_i follows a normal distribution with zero mean.

3.4.1 Neural Network Autoregressive Model (NNAR)

Given that the data structure under consideration is a time series, the utilization of the Neural Network Autoregressive (NNAR) model is considered suitable. NNAR is employed in order to model the residuals coming from time series modeling.

Definition 3.4.1. Neural Network Autoregressive Model (NNAR) is a specific type model of Artificial Neural Networks. NNAR models incorporate neural networks to model temporal dependencies and grasp the nonlinear information accommodated

within the time series data. It is a feedforward type of neural network which has one hidden layer. Multi-step forecasts are generated based on a recursive approach. Among the several activation functions considered, the Sigmoid (or Logistic) function emerges as one of the effective choices within the model, and its mathematical representation is as follows:

$$g(x) = [1 - \exp(-x)]^{-1}. \quad (3.24)$$

The residuals obtained from the time series model is presented as follows:

$$w_t = f(w_{t-1}, w_{t-2}, \dots, w_{t-p}) + \varepsilon_t \quad (3.25)$$

where f denotes the neural network with the lagged values w_{t-k} where $k = 1, 2, \dots, p$ and p is the number of lags. The diagram illustrating ANN model is presented as follows:

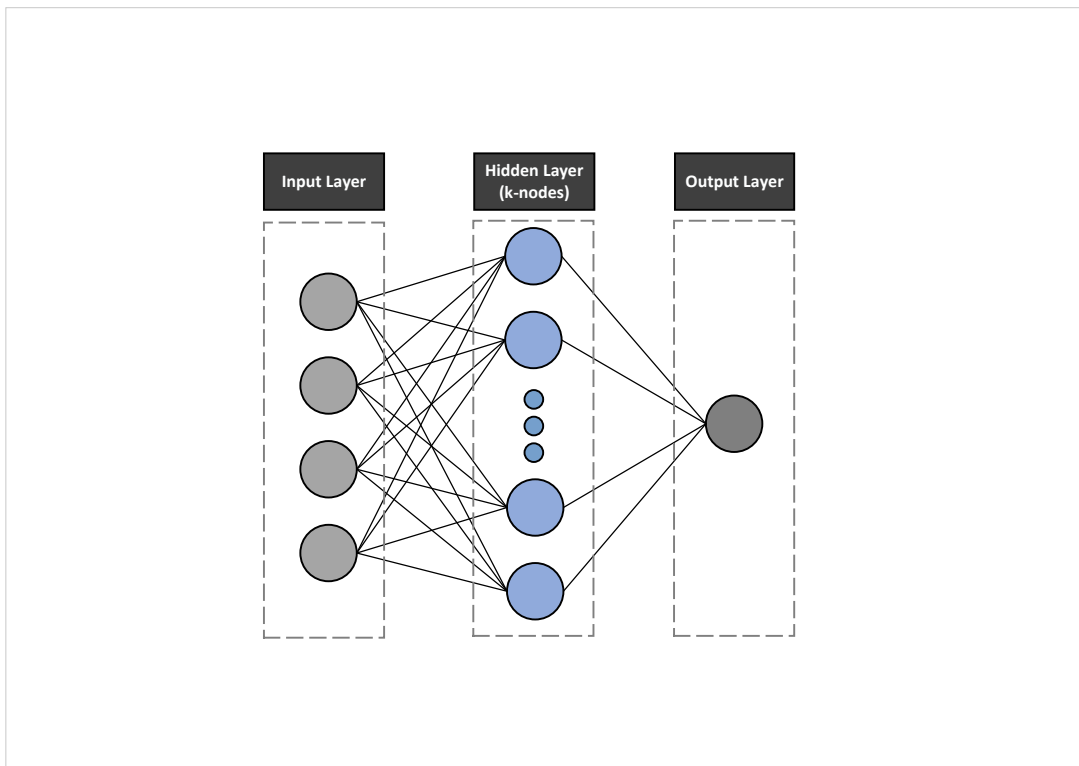


Figure 3.1: Autoregressive Neural Network

The implementation of NNAR model involves the utilization of the 'nnetar' function from the 'forecast' package in the R programming environment. The function for the model is represented as $\text{NNAR}(p, P, k)[m]$ which encompasses key parameters, where

'p' denotes the number of regular lags to be considered, 'P' signifies the number of seasonal lags, 'm' serves as the seasonal frequency parameter, and 'k' represents the number of nodes in the hidden layer [18]. During the model selection process, regular lags, seasonal lags, and seasonal frequency of corresponding time series models are utilized to account for the parameters 'p,' 'P,' and 'm', respectively. Additionally, the number of nodes (k) is determined by evaluating the model's performance by utilizing appropriate metrics, as well as meeting the normality assumption requirement. Furthermore, an investigation is conducted by exploring diverse numerical configurations for the number of iteration within neural networks in order to ensure the selected model exhibits robust stability.

3.5 Performance Metrics

If the residuals of the time series model meet the requisite assumptions, as determined by the designated criteria, the decision to employ an ANN model to the residuals of Time Series model is made by performance comparison between time series and Time Series-ANN model. The performance analysis is conducted using the following metrics:

Mean Absolute Error (MAE) provides a measure via taking average of the absolute errors on the forecasts. While this method is straightforward to employ and interpret, it is important to note that the scale of the metric varies for each dataset. MAE is described as follows:

$$MAE = \frac{1}{n} \sum_{t=1}^n (100 \cdot |Z_t - \hat{Z}_t|), \quad (3.26)$$

where n is the number of observation projected, Z_t is the actual value while \hat{Z}_t is the forecast at time t .

Mean Absolute Percentage Error (MAPE) provides a percentage-based measure via taking average of the relative errors on the forecasts. Thus, this method gives us the opportunity to make comparisons regardless of the scales. On the other hand, this method tends to yield unstable results when the actual values approach zero. MAPE

is expressed as:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left(100 \cdot \frac{|Z_t - \hat{Z}_t|}{Z_t} \right), \quad (3.27)$$

where n is the number of observation projected, Z_t is the actual value while \hat{Z}_t is the forecast at time t .

Mean Squared Error (MSE) quantifies the average of the squared errors, representing the average squared difference between the estimated values and the actual ones. It gives more weight to significant deviations and provides a comprehensive measure of the overall magnitude of the discrepancies between the estimated and actual values and it is shown as:

$$MSE = \frac{1}{n} \sum_{t=1}^n (Z_t - \hat{Z}_t)^2. \quad (3.28)$$

Here, n is the number of observation projected, Z_t is the actual value while \hat{Z}_t is the forecast at time t .

Root Mean Squared Error (RMSE) is a performance metric calculated as the square root of the average of squared errors (MSE). It considers the magnitude of errors by assigning proportionally larger weights to larger errors. RMSE is sensitive to outliers, as their larger squared errors contribute significantly to the overall metric. This characteristic makes RMSE suitable for assessing the overall accuracy of a forecast model but sensitive to extreme values in the data and is shown as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Z_t - \hat{Z}_t)^2}, \quad (3.29)$$

where n is the number of observation projected, Z_t is the actual value while \hat{Z}_t is the forecast at time t .

3.6 Comparative Measures

During the selection of suitable copula models, comparison metrics such as AIC and BIC are utilized. The copula model with the lowest scores indicates the optimal fit for the bivariate case, considering the following AIC and BIC calculations.

$$\text{AIC} := -2 \sum_{i=1}^N \ln[c(u_{i,1}, u_{i,2}|\theta)] + 2k, \quad (3.30)$$

$$\text{BIC} := -2 \sum_{i=1}^N \ln[c(u_{i,1}, u_{i,2}|\theta)] + \ln(N)k. \quad (3.31)$$

It is important to acknowledge that $k=1$ corresponds to copulas with a single parameter, while $k=2$ pertains to those with two parameters [7].

3.7 Copula Models

Within the realm of statistics and probability theory, copulas serve as multivariate distribution functions that establish a connection and clarify the joint dependence structure exhibited by a collection of random variables. Copulas play a fundamental role in modeling the dependence between variables. Moreover, copulas enable a more flexible and comprehensive approach to modeling complex dependencies, offering insights into the joint behavior of random variables.

Theorem 3.7.1. Fundamental Theorem for bivariate Copula: Sklar Theorem Let F be a two-dimensional distribution function with F_1 and F_2 margins. Subsequently, a two-dimensional copula denoted as C exists, satisfying the condition for all $x \in \bar{\mathbb{R}}^2$, which is shown as:

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)). \quad (3.32)$$

Provided that F_1 and F_2 are continuous distributions then C is unique, alternatively C is exclusively determined on $\text{Ran}F_1 \times \text{Ran}F_2$. Contrarily, if C represents a copula and F_1 and F_2 denote distribution functions, then F constitutes a two-dimensional distribution with F_1 and F_2 as its marginal components[9].

Definition 3.7.1. Let X and Y denote random variables with joint cumulative distribution:

$$F(x, y) = P[X \leq x, Y \leq y], \quad (3.33)$$

for all $x, y \in \mathbb{R}$ with the marginals: $F_1(x) = P[X \leq x]$ and $F_2(y) = P[Y \leq y]$.

The copula of a two-dimensional random vector is a bivariate function that defines its dependence structure and is unaffected by the marginal distributions. The copula pair

and the marginal distributions provide an alternative representation for the random vector, expressed as follows:

$$F(x, y) = C[F_1(x), F_2(y)]. \quad (3.34)$$

Given that $F_1(x)$ and $F_2(y)$ are continuous then the copula becomes unique and can be calculated via the distribution function and the following margins [9]:

$$C[F_1(x), F_2(y)] = F[F_1^{-1}(x), F_2^{-1}(y)]. \quad (3.35)$$

Copula of x and y is equal to copula of $f(x)$ and $g(y)$ if f and g are two strictly increasing functions [9].

Theorem 3.7.2. Fundamental Theorem for multivariate Copula: Sklar Theorem

Let F be a d -dimensional distribution function with F_1, \dots, F_d margins. Subsequently, a d -dimensional copula denoted as C exists, satisfying the condition for all $x \in \bar{\mathbb{R}}^d$, which is shown as:

$$F(x_1, x_2, \dots, x_d) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)). \quad (3.36)$$

Provided that F_1, \dots, F_d are continuous distributions then C is unique, alternatively C is exclusively determined on $\text{Ran}F_1 \times \dots \times \text{Ran}F_d$. Contrarily, if C represents a copula and F_1, \dots, F_d denote distribution functions, then F constitutes a d -dimensional distribution with F_1, \dots, F_d as its marginal components [9].

3.7.1 Bivariate Copulas

In this thesis, our attention is directed towards two distinct classes of copula families: Elliptical and Archimedean. Each of these family types possesses its own set of advantages and limitations. Definitions along with several instances for both families are illustrated in the following subsections.

3.7.1.1 Elliptical Copulas

The family of elliptical copula offers diverse resources for multivariate contexts, featuring several manageable characteristics of multivariate normal distributions. This

facilitates feasible modeling for instances involving multivariate extreme cases [14]. Nonetheless, this methodology necessitates the assumption of normality, naturally symmetry. Additionally, a limitation of elliptical copulas is the absence of a closed-form representation. Multivariate cases demand needs numerical methodologies for computation. Given the utilization of bivariate models within the R-vine structure, the analysis necessitates the incorporation of bivariate copula forms. Gaussian and t-copulas are the most commonly known elliptical forms.

The bivariate expression of **Gaussian copula** is given as follows:

$$\begin{aligned} C(u, v) &= \Phi_{\rho}(\Phi^{-1}(u), \Phi^{-1}(v)) \\ &= \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right] ds dt, \end{aligned} \quad (3.37)$$

where u and v are uniform marginal distributions to be modeled, ρ is the ordinary linear correlation coefficient and Φ represents The joint distribution function which corresponds to the bivariate standard normal distribution with a given correlation coefficient ρ [14].

The bivariate representation of **t-Copula** is shown as follows:

$$\begin{aligned} C(u, v) &= t_{d,\rho}(t_d^{-1}(u), t_d^{-1}(v)) \\ &= \int_{-\infty}^{t_d^{-1}(u)} \int_{-\infty}^{t_d^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left[1 + \frac{s^2 - 2\rho st + t^2}{d(1-\rho^2)}\right]^{-\frac{d+2}{2}} ds dt, \end{aligned} \quad (3.38)$$

where u and v are uniform marginal distributions to be modeled, ρ is the ordinary linear correlation coefficient and t_d represents the t-distribution with d degrees of freedom [14].

3.7.1.2 Archimedean Copulas

Archimedean copulas offer a wider scope for constructing dependencies, as they do not require normality or symmetry conditions. Unlike elliptical copulas, archimedean copulas do not originate from multivariate distribution functions through Sklar's theorem. Thus, multivariate extensions of archimedean copulas have disadvantages such as suffering from a lack of free parameter choice [14]. The decision on choosing

which copula type to use in each step of multivariate structure is made via goodness of fit test (GOF, BIC & AIC). The most appropriate bivariate copula types for each combining unit are selected in order to produce copulas which belong to a higher level of current unit in the multivariate structures. Hence, each case study consists of n different combination of bivariate copulas considering the results of tests in order to catch best fit while n is the number of combining unit. The general multivariate expressions for archimedean copulas are derived through an inductive process. All the multivariate structures provide the same result in terms of general formulas in case that the same copula family is used for each combination unit.

The general formulas are given for Frank, Clayton, Gumbel and Joe copulas for bivariate cases.

One of the most commonly used archimedean copulas is **Frank**, which has the bivariate distribution expressed as follows [14]:

$$C(u, v) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)} \right) \quad (3.39)$$

where $u, v \in [0, 1]$ and $\theta \in (-\infty, \infty) \setminus 0$.

Another common archimedean copula type is **Clayton**. It is associated with capturing lower tail dependence in bivariate distributions. Bivariate form is described as [31]:

$$C(u, v) = (\max\{u^{-\theta} + v^{-\theta} - 1; 0\})^{-1/\theta} \quad (3.40)$$

where $u, v \in [0, 1]$ and $\theta \in (0, \infty)$.

Gumbel is an upper tail copula with the subsequent bivariate distribution [31].

$$C(u, v) = \exp[-((-\ln u)^{-\theta} + (-\ln v)^{-\theta})^{1/\theta}] \quad (3.41)$$

where $u, v \in [0, 1]$ and $\theta \in [1, \infty)$.

The next copula to be described is the **Joe** copula, which likewise exhibits dependence in the upper tail. It has a bivariate form illustrated as [21]:

$$C(u, v) = 1 - [(1 - u)^\theta + (1 - v)^\theta - (1 - u)^\theta(1 - v)^\theta]^{1/\theta} \quad (3.42)$$

where $u, v \in [0, 1]$ and $\theta \in [1, \infty)$.

3.7.1.3 Non-parametric Copulas

If two random variables are **independent** of each other, the appropriate copula function to employ is given as follows [31]:

$$C(u, v) = u \cdot v, \quad (3.43)$$

where $u, v \in [0, 1]$.

3.7.2 Vine Structure

Despite the existence of numerous bivariate copulas, those with dimensions equal to or greater than 3 are notably limited in scope. Given that Archimedean copulas serve as models for multivariate scenarios, it becomes evident that a combination of diverse copula types is essential to enhance the representation of dependence structures more effectively. The impetus behind the development of vine copulas was to devise a methodology for constructing multivariate copulas exclusively employing bivariate copulas as fundamental building blocks. The conditioning methodology is employed to facilitate the generation of the required construction [10].

Bedford and Cooke (2001) proposed density decomposition by vine constructions for conditionally dependent random variables [4]. Within the vine construction, they defined tree form for the multiple bivariate couple of variables.

Definition 3.7.2. Vine: \mathcal{V} constitutes a vine structure on a set of d elements, provided that the ensuing three criteria are met [5]:

- 1 $\mathcal{V}=(T_1, \dots, T_j)$.
- 2 T_1 is a tree consisting of nodes $N_1=\{1, \dots, d\}$ and E_1 , the set of edges.
- 3 T_i is a tree consisting of nodes $N_i \subset N_1 \cup E_1 \cup E_2 \cup \dots \cup E_{i-1}$ and E_i , the set of edges for $i = 2, \dots, j$.

Definition 3.7.3. Regular Vine: \mathcal{V} is a regular vine on d elements if the following three conditions are satisfied, additional to the conditions in Definition 3.7.2 [5].

- 1 $j=d$
- 2 $N_i=E_{i-1}$ while $N_i=d - (i - 1)$
- 3 T_i is a tree consisting of nodes $N_i \subset N_1 \cup E_1 \cup E_2 \cup \dots \cup E_{i-1}$ and E_i , the set of edges for $i = 1, \dots, d$.

A regular vine represents a systematic approach for delineating a collection of conditional bivariate constraints. The conditional bivariate constraint corresponding to each edge is established in the following manner: the set of variables accessible from a particular edge through the membership relation is termed the constraint set for that edge. As two edges are connected via an edge from the subsequent tree, the intersection of their respective constraint sets comprises the conditioning variables, while the symmetric differences of the constraint sets constitute the conditioned variables. To elaborate further, the constraint, conditioning, and conditioned sets of an edge can be rigorously described as follows [23]:

Definition 3.7.4. Conditioned Set

$$\{C_{e,r}, C_{e,s}\} = \{U_r^* \setminus D_e, U_s^* \setminus D_e\} \quad (3.44)$$

is a conditioned set with e where, $D_e = U_r^* \cap U_s^*$ is the conditioning set with e , if $e = \{r, s\}$ for $i = 1, \dots, d - 1$ and $e \in E_i$. It is also noted that the constraint set linked with e is the complete union U_e^* of e , thus, the subset of $\{1, \dots, d\}$ can also be attainable from e by using the membership relation [23].

Following the establishment of the structure, a triplet of $(\mathcal{F}, \mathcal{V}, \mathcal{B})$ is defined to represent the R-vine distribution created for an $d - dimensional$ random variables, $\mathcal{X}=(X_1, \dots, X_d)$ [10].

Definition 3.7.5. Vine Distribution Triplet: $(\mathcal{F}, \mathcal{V}, \mathcal{B})$ [10]

- 1 $\mathcal{F} = (F_1, \dots, F_d)$ is a set of marginal distributions representing the random variables X_1, \dots, X_d .
- 2 \mathcal{V} represents an R-vine tree sequence on d elements.

3 $\mathcal{B} = \{C_e | e \in E_i\}$ for all $i = 1, \dots, d - 1$, where C_e is a bivariate copula and E_i is the edge set of T_i in \mathcal{V} .

4 C_e is the copula pertaining to the conditional distribution of $X_{C_{e,r}}$ and $X_{C_{e,s}}$ given X_{D_e} for each $e \in E_i$ where $i = 1, \dots, d - 1$ and $e = \{r, s\}$.

Theorem 3.7.3. *R-vine distribution existence [5, 10]. Should $(\mathcal{F}, \mathcal{V}, \mathcal{B})$ fulfill the first three properties outlined in Definition 3.7.5, the existence of a d -dimensional F distribution is assured. The expression for the density of F is as follows:*

$$f_{1,\dots,d}(x_1, \dots, x_d) = \left(\prod_{h=1}^d f_h(x_h) \right) \times \prod_{k=1}^{d-1} \prod_{e \in E_i} C_{C_{e,r}, C_{e,s} | D_e} \left(F_{C_{e,r} | D_e}(x_{C_{e,r}} | x_{D_e}), F_{C_{e,s} | D_e}(x_{C_{e,s}} | x_{D_e}) \right) \quad (3.45)$$

For each $e \in E_i$ and $i = 1, \dots, d - 1$ where $e = \{r, s\}$, the distribution function pertaining to $X_{C_{e,r}}$ and $X_{C_{e,s}}$ given X_{D_e} is expressed as

$$F_{C_{e,r} C_{e,s} | D_e}(x_{C_{e,r}}, x_{C_{e,s}} | x_{D_e}) = C_e \left(F_{C_{e,r} | D_e}(x_{C_{e,r}} | x_{D_e}), F_{C_{e,s} | D_e}(x_{C_{e,s}} | x_{D_e}) \right) \quad (3.46)$$

Moreover, $F_i(x_i)$ where $i = 1, \dots, d$ represents the one-dimensional margins of F .

The resulting distribution becomes an R-vine copula if all margins are standard uniform [10].

In order to ascertain the structure and identify the most suitable bivariate copula models for the edges between nodes, the "rvinecopulib" package authored by Nagler and Vatter (2020) is employed [37]. The selection criteria of the best structure rely on Dissman's Algorithm. They present a comprehensive selection methodology that sequentially determines the tree representation and copula type for each copula term from an extensive range of bivariate copula families, while simultaneously estimating the associated parameters. This process involves employing any graph-theoretic algorithm capable of identifying a maximum spanning tree. This approach is used to choose the tree structure that optimizes the summation of absolute empirical Kendall's taus [11].

Algorithm 2: Sequential method

1. Compute the empirical Kendall's tau, $\hat{\tau}_{r,s}$, for all possible combination of variables $\{r, s\}$, where $1 \leq r < s \leq d$.
2. Choose the spanning tree that maximizes the summation of absolute empirical Kendall's taus, denoted as

$$\max_{e=\{r,s\} \text{ in spanning tree}} \sum |\hat{\tau}_{r,s}| \quad (3.47)$$

3. Designate an appropriate copula and estimate the associated parameter(s) by minimizing the AIC for each edge $\{r, s\}$ within the chosen spanning tree.
4. Apply Independence test for two variables. If two variables demonstrate independence according to the independence test, the link function employed is determined as the independence copula. Otherwise, modify $\hat{F}_{r|s}(x_{jr}|x_{js})$ and $\hat{F}_{s|r}(x_{js}|x_{jr})$ for each realization $j = 1, \dots, I$ by utilizing the fitted copula.
5. Apply the items 6-7 for $i = 2, \dots, d - 1$ where d is the number of random variables.
6. Compute the empirical Kendall's tau for all conditional variable pairs $\{r, s|D\}$ that are eligible to be included in tree T_i .
7. Considering the above mentioned edges, choose the spanning tree maximizing the summation of absolute empirical Kendall's taus, described as follows:

$$\max_{e=\{r,s|D\} \text{ in spanning tree}} \sum |\hat{\tau}_{r,s|D}|. \quad (3.48)$$

8. Choose a conditional copula by employing the same approach as described in Item 3-4 and estimate the associated parameter(s) for each edge $\{r, s|D\}$ incorporated within the determined spanning tree. Thereafter modify, $\hat{F}_{r|s \cup D}(x_{jr}|x_{js}, x_{jD})$ and $\hat{F}_{s|r \cup D}(x_{js}|x_{jr}, x_{jD})$ for each realization $j = 1, \dots, I$ by utilizing the fitted copula.
-

3.7.3 Value-at-Risk & Tail Value-at-Risk

As per the Solvency II regulation, the 99.5th percentile holds significant importance in terms of the safety level, which is attained from the perspective of Value-at-Risk (VaR). Risk measures constitute a crucial component of risk management across various fields, particularly within the realm of insurance. Consequently, the assessment and application of VaR and Tail Value-at-Risk (TVaR) measures are conducted on the summation of exceeding values for $\{t + 1\}$ th period, $S_{\{total,t+1\}}$

Definition 3.7.6. VaR formula is defined as follows:

$$\text{VaR}[S; q] = F^{-1}(S) = \inf\{s : F_S(s) \geq q\} \quad (3.49)$$

where S is the underlying variable for the risk and q is the level of confidence [22].

A risk measure is categorized as coherent if it fulfills the following axioms: Subadditivity, Monotonicity, Translation Invariance and Positive Homogeneity. VaR is not a coherent risk measure since it does not satisfy Subadditivity axiom. However, given that the standard approach in Solvency II is aligned with the VaR, it is, therefore, also analyzed in this study for the purpose of comparison.

Definition 3.7.7. Tail Value-at-Risk (TVaR) formula is defined as follows:

$$\text{TVaR}[S; q] = \frac{1}{1 - q} \int_q^1 \text{VaR}[S; h] dh \quad (3.50)$$

$$= \frac{1}{n} \sum_{j=1}^n \inf\{S_j \geq \text{VaR}[S; q]\} \quad (3.51)$$

where n is the number of observations [22].

CHAPTER 4

INTERNAL SOLVENCY MODEL

Let Z_t denote the loss ratio defined in Equation 3.2. In this thesis, we propose a composition of methods which captures both the time impact and dependence structure of LoB in Solvency II framework. To do so, we implement the proposed methodology on Turkish Insurance dataset to illustrate the development of capturing the contributing variables.

Subsequently in the thesis, the type of LoB is denoted by the variable i , where i takes values from 1 to d , and the time period is represented by the variable t , ranging from 1 to T .

Proposition 4.0.1. *The correlation coefficients calculated using Loss Ratios includes trend and seasonality effects, too. In order to remove these effects and determine dependence, we work with residuals after time series and ANN fitting, $\varepsilon_{i,t}$ for $i = 1, 2, \dots, d$.*

Herewith, the Loss Ratio equation is defined as:

$$Z_{i,t} = \hat{X}_{i,t} + \hat{Y}_{i,t} + \varepsilon_{i,t}, \quad (4.1)$$

where $\hat{X}_{i,t}$ component represents the fitted values of time series model and $\hat{Y}_{i,t}$ is the estimation of applied ANN model to the residuals of employed time series model, denoting the nonlinear time component. Thus, the remaining noise part, the residual of hybrid Time series-ANN model ($\varepsilon_{i,t}$) is expected to be distributed normally with a mean zero.

4.1 Copula Modeling Algorithm

The following algorithm is utilized to fit copula models to the residuals of Hybrid Time series-ANN Model. Algorithm 3 involves a series of procedures consisting of the transformation of initial variables, determination of trees, and selection of copulas, as detailed in the subsequent algorithm.

Algorithm 3: Copula Modeling

Let $\varepsilon_{i,t}$ be the residuals of hybrid model for LoBs for $i = 1, 2, \dots, d$ and $t = 1, 2, \dots, T$.

for $i \leftarrow 1$ **to** d **by** 1 **do**

 Transform $\varepsilon_{i,t}$ to $u_{i,t}$ from $[-\infty, +\infty]$ to $[0,1]$ via using corresponding quantiles where $t = 1, 2, \dots, T$;

end

Save the $u_{i,t}$ for $i = 1, 2, \dots, d$ and $t = 1, 2, \dots, T$.

Utilize the Algorithm 2 on $u_{i,t}$: Sequential Estimation Approach to select the suitable tree structure and copula functions for the edges connecting nodes across all LoBs.

It is noteworthy that the "pobs" function within the R:VineCop package can serve for transformation purposes, while the "vinecop" function in the R:rvinecopulib package is employed to establish tree structures and copula functions [37].

4.2 Simulation

Upon completion of copula modeling, we proceed to perform simulations under the determined copula structure. This simulation process can be achieved through several methods, including the Inverse Rosenblatt Transform (1952) [28] or Recursion over conditional distributions, as proposed by Czado (2022) [10] by referencing Joe (1996) [21].

The Inverse Rosenblatt Transform method involves taking random uniform marginal sets and converting them into dependent variables through the use of a predetermined copula. The following definition for Inverse Rosenblatt Transform was provided in

Czado's paper in 2022 [10].

Definition 4.2.1. Inverse Rosenblatt Transform: Let $L(U) = (K_1, \dots, K_d) = K$ be a random vector consisting of independent uniform variables. The following inversion $U = L^{-1}(K)$ transform the independent variables K into a set of U characterized by a copula C . The conditional function for U is described as follows:

$$U_i = C_{i|i-1, \dots, 1}^{-1}(K_i | K_{i-1}, \dots, K_1), \quad (4.2)$$

where $i = 1, \dots, d$. The recursion formula defined by Czado and Nagler (2022) is given as follows[10]:

$$F_{A|B}(\cdot|b) = \frac{\partial C_{A, B_i; B_{-i}}(F_{A|B_i}(a|b_{-i}), F_{B_i|B_{-i}}(b|b_{-i}))}{\partial F_{B_i|B_{-i}}(b_i|b_{-i})} \quad (4.3)$$

where $C_{A, B_i; B_{-i}}(\cdot, \cdot)$ is the copula associated to (A, B_i) given B_{-i} .

Ultimately, one of the proposed approaches is utilized to simulate a uniform distribution set that represents the quantiles for the residuals of the SARIMA-ANN model.

Proposition 4.2.1. *The derived quantiles are transformed into normally distributed residuals, characterized by a mean of zero and a standard deviation computed from the original residuals. Furthermore, Inverse Transform Method is employed for the transformation process.*

Subsequently, let $\varepsilon_{i, T+1}$ be the representative of residual of LoB_i for the next time period where T is the last period analyzed. The expression of the loss ratio of LoB_i for the same time period is defined as follows:

$$Z_{i, T+1} = \hat{X}_{i, T+1} + \hat{Y}_{i, T+1} + \varepsilon_{i, T+1}, \quad (4.4)$$

where $\hat{X}_{i, T+1}$ is the Time Series component and $\hat{Y}_{i, T+1}$ is the ANN component.

Upon simulating $Z_{i, T+1}$ for n iterations, a collection of n loss ratios is obtained for each LoB. Considering the volumes of LoBs, the required reserve amount to cover the losses in the upcoming time slot is expressed as follows:

$$\begin{aligned} R_{i, T+1} &= Z_{i, T+1} \cdot \hat{V}_{i, T+1} \\ &= (\hat{X}_{i, T+1} + \hat{Y}_{i, T+1} + \varepsilon_{i, T+1}) \cdot \hat{V}_{i, T+1} \end{aligned} \quad (4.5)$$

Since $\hat{V}_{i,T+1}$, $\hat{X}_{i,T+1}$ and $\hat{Y}_{i,T+1}$ are constant estimates, $\varepsilon_{i,T+1}$ is the only source for volatility of $R_{i,T+1}$. Consequently, there exist n distinct simulations for each LoB concerning the required reserve amount, where n is the number of simulation.

Proposition 4.2.2. *Given that $Z_{i,T+1}$ encompasses expenses, $\max(Z_{i,T+1} - 1, 0)$ constitutes the ratios of the excess values that must be covered by the SCR margin. This provides insight into the required provision of SCR Non-Life capital [17].*

Proposition 4.2.3. *Considering the excess values as per Proposition 4.2.2, the amount requiring coverage by the Non-Life Premium Solvency Capital Requirement (SCR NL) is expressed as follows:*

$$\begin{aligned} S_{i,T+1} &= \max(Z_{i,T+1} - 1; 0) \cdot \hat{V}_{i,T+1} \\ &= (\hat{X}_{i,T+1} + \hat{Y}_{i,T+1} + \varepsilon_{i,T+1} - 1; 0) \cdot \hat{V}_{i,T+1}. \end{aligned} \quad (4.6)$$

Excess values are computed by deducting "1", which corresponds to 100%, from the loss ratios, and negative values are substituted with "0" to mitigate any unnecessary diversification tendencies among the LoBs. Following summation holds for n simulation of $S_{i,T+1}$:

$$S_{total,T+1} = \sum_{i=1}^d S_{i,T+1}. \quad (4.7)$$

4.2.1 SCR: Non-Life Premium Risk

Ultimately, $\text{VaR}_{99.5}[S_{total,T+1}]$ attained from n simulations serves as a representative of the SCR attributable to Non-Life Premium risk. Furthermore, the calculation of $\text{TVaR}_{99.5}[S_{total,T+1}]$ is also conducted as a coherent risk measure.

4.2.2 Proposed Model Algorithm

The hybrid model algorithm is constructed by incorporating the previously discussed methodologies. The suggested framework accepts Loss Ratios and Volume measures as inputs and generates simulated capital requirement outcomes for numerous scenarios, covering a specified set of LoBs and alternative SCR values relating to the Non-Life premium risk. Algorithm 4, constructed for the proposed model is given as follows.

Algorithm 4: Proposed Model Algorithm

Let $Z_{i,t}$ be the Loss Ratio Matrix where i is LoB and t is time.

for $i \leftarrow 1$ **to** d **by** 1 **do**

 Apply Algorithm 1 to $Z_{i,t}$ where $t = 1, 2, \dots, T$ to assign the appropriate time series model for each LoB;

end

Let $\hat{X}_{i,t}$ be the fitted value for $Z_{i,t}$ and $w_{i,t}$ be the residuals such that

$$w_{i,t} = Z_{i,t} - \hat{X}_{i,t}.$$

Save $w_{i,t}$'s for $i = 1, 2, \dots, d$ and $t = 1, 2, \dots, T$.

for $i \leftarrow 1$ **to** d **by** 1 **do**

 Apply Autoregressive Neural Network (NNAR) model to the $w_{i,t}$'s where $t = 1, 2, \dots, T$;

if $w_{i,t} \sim N(0, \sigma_{w_i})$ **then**

if *Time series-ANN model performs better than Time series model*

then

 | Save the $\varepsilon_{i,t}$ where $t = 1, 2, \dots, T$

end

end

else if $w_{i,t} \approx N(0, \sigma_{w_i})$ **then**

 | Save the $\varepsilon_{i,t}$ where $t = 1, 2, \dots, T$

end

end

Let $\hat{Y}_{i,t}$ be the fitted value for $w_{i,t}$ and $\varepsilon_{i,t}$ be the residuals with the assumption that $\varepsilon_{i,t}$ follows a Normal distribution.

for $i \leftarrow 1$ **to** d **by** 1 **do**

 | Fit Normal distribution with zero mean to each $\varepsilon_{i,t}$ where $t = 1, 2, \dots, T$;

end

Apply Algorithm 3 to $\varepsilon_{i,t}$ matrix where i is LoB and t is time.

Save the tree structure and bivariate copula distributions assigned to the nodes.

Algorithm 5: Proposed Model Algorithm, continued

for $s \leftarrow 1$ **to** n **by** 1 **do**

 Apply Inverse Rosenblatt Transform 4.2.1 or Recursion formula 4.3 to
 create n simulations of $u_{i,s}$'s where $i = 1, 2, \dots, d$ under the chosen
 copula structure.

end

Save n simulated uniform margins $u_{i,s}$ where $i = 1, 2, \dots, d$ and

$s = 1, 2, \dots, n$.

for $i \leftarrow 1$ **to** d **by** 1 **do**

 Perform n simulations for the residuals $\varepsilon_{i,s}$ where $s = 1, 2, \dots, n$ by
 using Inverse Transfer Method on the fitted distribution.

end

Save n simulated residuals $\varepsilon_{i,s}$ where $i = 1, 2, \dots, d$ and $s = 1, 2, \dots, n$.

for $i \leftarrow 1$ **to** d **by** 1 **do**

for $l \leftarrow T + 1$ **to** $T + 4$ **by** 1 **do**

 Calculate the n simulations of $Z_{i,l}$ by $\hat{X}_{i,l} + \hat{Y}_{i,l} + \varepsilon_{i,s}$ where
 $s = 1, 2, \dots, n$;

 Calculate the n simulations of $S_{i,l}$ by $\max(Z_{i,l} - 1; 0) \cdot \hat{V}_{i,l}$;

end

end

Save the $n \cdot d$ simulations for $l = T + 1, T + 2, \dots, T + 4$

for $l \leftarrow T + 1$ **to** $T + 4$ **by** 1 **do**

for $s \leftarrow 1$ **to** n **by** 1 **do**

 Calculate the aggregate SCR, $S_{total,l}$ by $\sum_{i=1}^d S_{i,l}$

end

end

Save the $4 \cdot n$ simulated $S_{total,l}$ results.

Calculate the $\text{VaR}_{99.5}$ and $\text{TVaR}_{99.5}$ of aggregate values, $S_{total,l}$.

Moreover, a schematic diagram is presented to clarify the outlined model's algorithm, offering readers an enhanced visual comprehension of the proposed methodology.

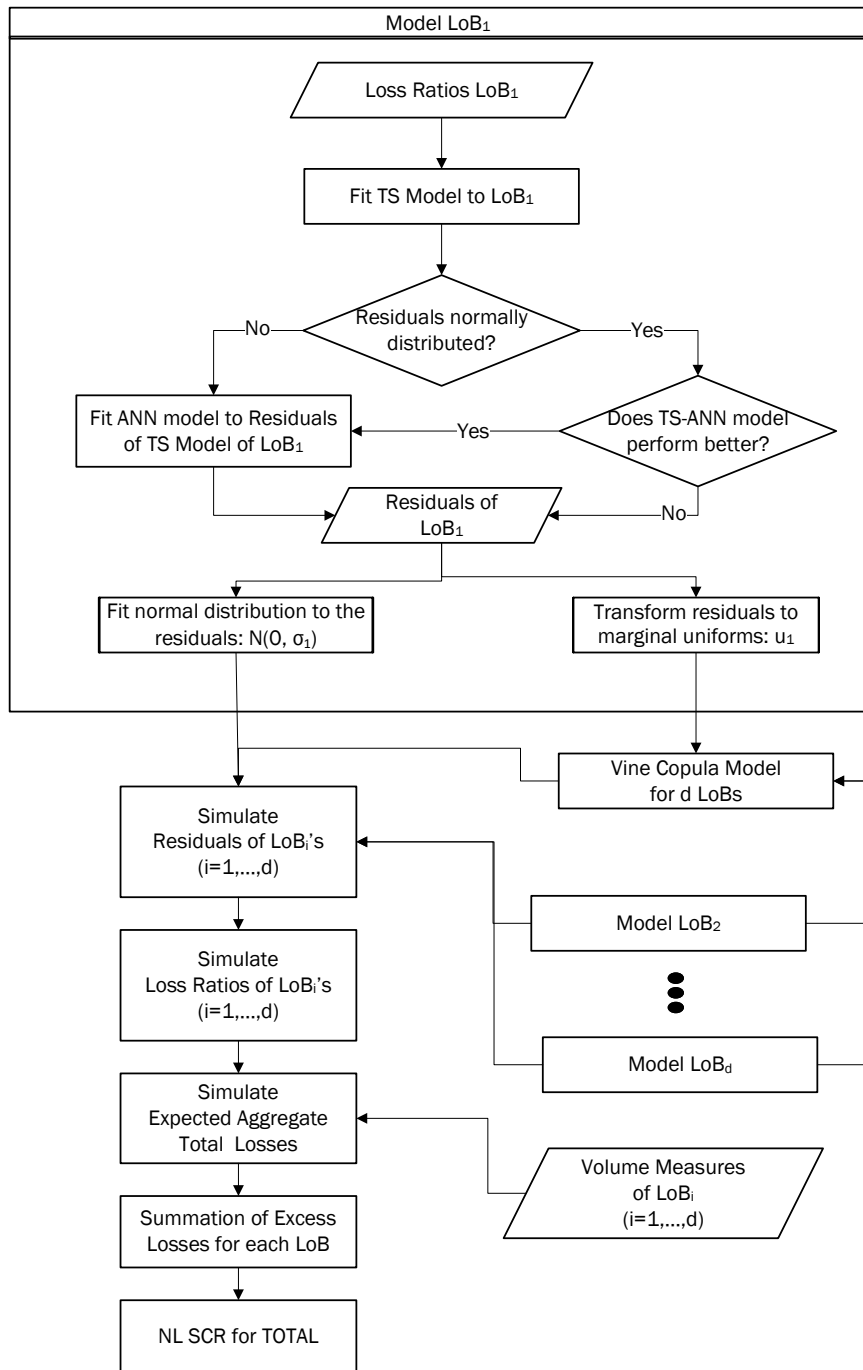


Figure 4.1: Internal Model Diagram

4.3 Limitations

A significant limitation of this study lies in the availability of the data. Given the absence of public quarterly data pertaining to a company or other countries, the scope of this case study is limited to data pertaining to the Turkish industry, which was obtained from the periodical reports published by the Insurance Association of Turkey [2].

Another limitation concerning the datasets pertains to recent significant occurrences, including the Covid-19 pandemic and the subsequent financial outcomes of the monetary measures taken during the pandemic. The initial impact of the pandemic manifests in the loss frequencies, which experience a significant reduction across various branches, consequently causing volatility in the rates. Subsequently, inflation due to the pandemic influences the severities. Thus, the time span of 2009 to 2019 (44 quarters) is employed for the loss ratios in order to mitigate the potential of abnormal volatility caused by the pandemic.

Furthermore, numerous methodological approaches can be hybridized, such as the utilization of Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNNs), Gated Recurrent Unit (GRU), or various other machine learning algorithms as Support Vector Machine (SVM) integrated into time series modeling. Nevertheless, it is essential to acknowledge that each of these methods comes with its distinct set of advantages and drawbacks. In this thesis, our choice is to leverage time series analysis in conjunction with the autoregressive neural network (NNAR) approach to make a novel contribution to the academic literature.

CHAPTER 5

CASE STUDY: NON-LIFE PREMIUM RISK

This chapter covers the practical implementation part of the thesis. As a general procedure, the application of the internal model to the chosen LoBs is done by using Algorithm 4. The procedure involving loss ratio calculations, the selection of appropriate time series models for individual LoBs, the application of ANNs to time series residuals, and the establishment of a suitable R-vine structure based on hybrid Time Series-ANN model residuals is presented. Conclusively, a comparison between the standard SCR model, the standard SCR model incorporating calculated standard deviations, and the outcomes of the proposed model is provided through tables and figures.

5.1 Data

Solvency II is a mandatory regulation for the (re)insurance companies operating in the European Union. However, this thesis centers its attention on the Turkish market. Although the obligatory implementation of Solvency II is absent in Türkiye, it remains imperative to exhibit the industry's performance with regard to premium risk to practitioners. Among many branches available in the Turkish Market, Accident, Fire, Health, Motor Third Party Liability (Traffic), Land Vehicle (Casco) and Miscellaneous have been chosen to be used in case study. These LoBs represent roughly 90% of the whole Turkish Non-Life market. Data for the study lies between 2009-2019, 44 quarters in total. Incurred Losses, Earned Premiums, Expenses and Written Premiums for the mentioned LoBs are retrieved from Insurance Association of Turkey's (TSB) public database [2]. The quarterly loss ratios for these branches

are calculated using Equation 3.2.

Let $Z_{i,t}$ represents the loss ratio for LoB i and for quarter t assuming that $i = 1, \dots, 6$ for Accident (Acc), Fire (Fir), Health (Hea), MTPL (Mtp), Land Vehicle (Lan) and Miscellaneous (Mis), respectively and $t = 1, \dots, 44$ starting from the first quarter of 2009 to the last quarter of 2019. Furthermore, 10% of the data is reserved as test set while remaining 90% is used as training set. The descriptive statistics of loss ratios for the chosen six LoBs is given in Table 5.1 and they are plotted in Figure 5.1. It is important to acknowledge that the provided descriptive statistics involves the entire dataset spanning from 2009 to 2019, while the modeling process is conducted using the training set.

Table 5.1: Descriptive Statistics: Loss Ratios

Stat/LoB	Acc	Fir	Hea	Mtp	Lan	Mis
Min	39%	56%	78%	91%	74%	56%
Median	54%	79%	95%	113%	91%	79%
Mean	53%	80%	95%	117%	93%	78%
Max	68%	131%	119%	190%	121%	109%
Std. Dev.	8%	13%	10%	18%	12%	12%
Skewness	0.06	1.22	0.17	1.77	0.54	0.31
Kurtosis	-0.85	3.11	-0.27	4.5	-0.39	-0.35

Drawing inference from Table 5.1, it is evident that MTPL exhibits the broadest range and highest standard deviation, thus positioning it as the branch with the highest premium risk. In contrast, Accident appears as the least risky one. Surpassing the threshold of 100% can be justified as an indication of requirement of an additional funding on the top of the regular reserves. It is essential to emphasize that, with the exception of "Accident", all listed LoBs experience instances where their loss ratios exceed 100% necessitate extra funding to ensure solvency within the context of these specific branches.

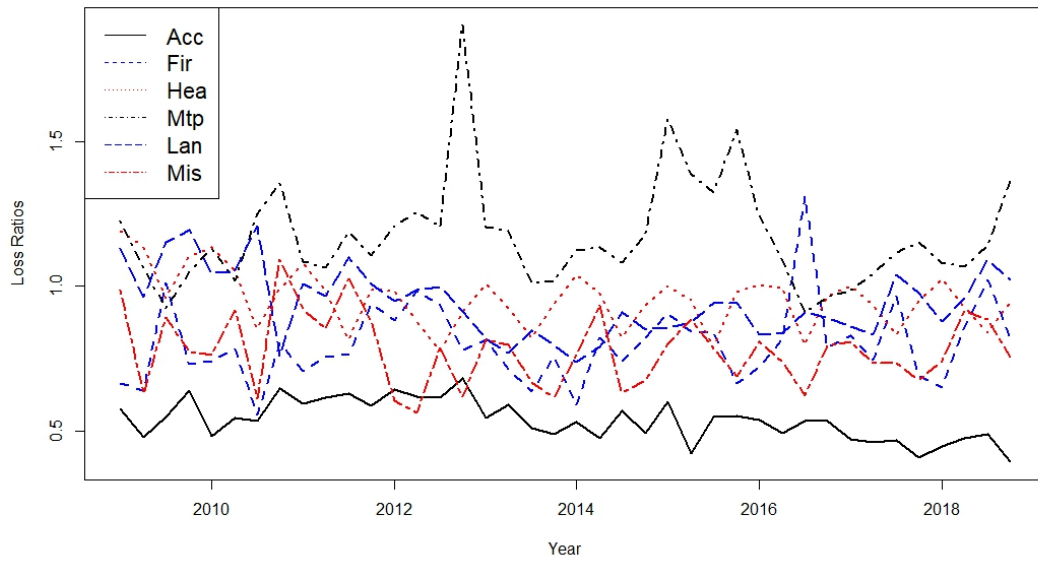


Figure 5.1: Loss Ratios of the chosen LoBs

MTPL branch displays a high sensitivity to inflation in spare part prices. Currency shocks make an immediate impact on spare part costs, subsequently leading to an elevation in MTPL loss ratios. Furthermore, substantial increases in minimum wage rates have a significant effect on the MTPL segment, as they directly impact bodily injury claims calculations in MTPL. Additionally, regulatory modifications such as the allocation of specific amounts to government entities result in temporary increases. The implementation of the new price cap system in 2017 led to a rise in MTPL loss ratios, too.

Similarly to MTPL, the augmentation in spare part costs initiates a rise in the loss ratio of Land Vehicles, especially when subjected to substantial currency fluctuations. Nonetheless, this impact is comparatively less noticeable than that observed in MTPL, primarily due to the absence of a price ceiling on Land Vehicles, a distinction from the MTPL framework.

The escalation of outstanding claims provisions within the Fire branch during the year 2016 can be attributed to the occurrence of terrorist incidents. This subsequently gives rise to a corresponding elevation in the loss ratio [2].

5.2 Time Series Modeling

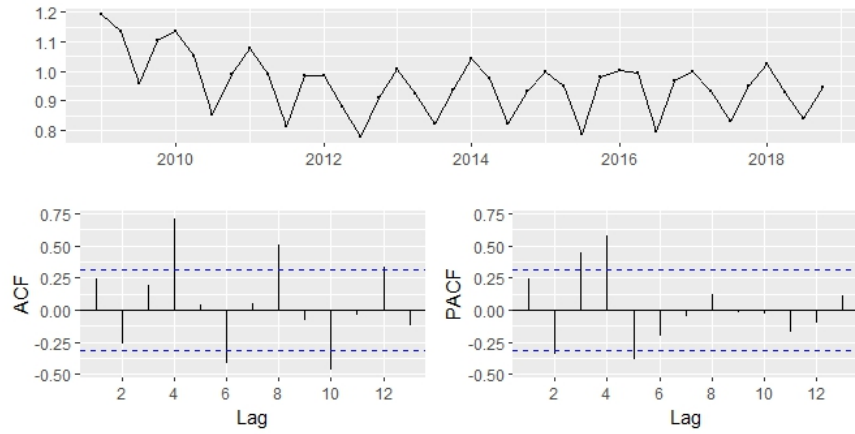
Following the computation of loss ratios, the time series modeling (Algorithm 1) is employed individually for each LoB. As an illustration, the health branch is presented as a whole analysis in this section to elucidate the procedural steps for the reader. However, all the results concerning the other LoBs are also presented in Appendix A.

The time series data, along with the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are depicted in the subsequent Figure 5.2. ACF function seems to decay to insignificant values as lags increase and there seems two spikes in PACF function. Taking into account the seasonal aspect as well, it is proposed that the first candidate model for analyzing the Health data is SARIMA(2,0,0)(1,1,0)[4].

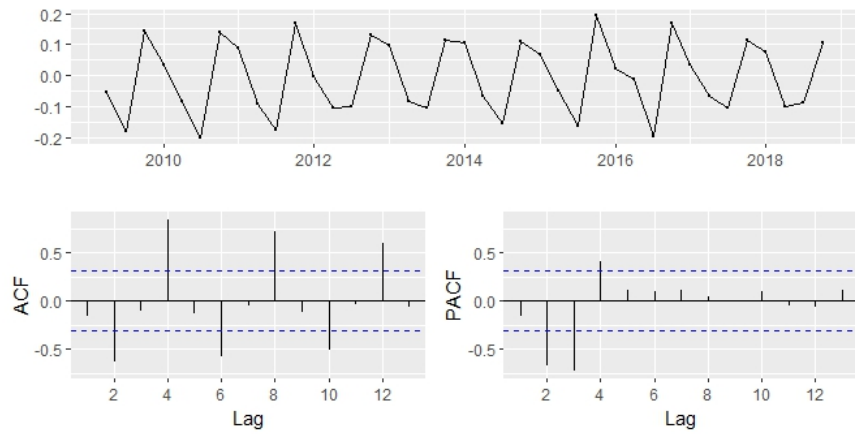
Three distinct types of tests were employed to assess the stationarity of the time series. Augmented Dickey-Fuller (ADF) and Kwiatkowski-Philips-Schmidt-Shin (KPSS) tests are utilized to detect regular stationary while Osborn-Chui-Smith-Birchenhall (OCSB) test is applied in order to detect seasonal stationarity.

The stationarity tests were conducted on the Health data, and the results (Table 5.3) indicate that for achieving stationarity, both regular and seasonal differences should be applied. Subsequent to applying both regular and seasonal differencing, the time series becomes stationary.

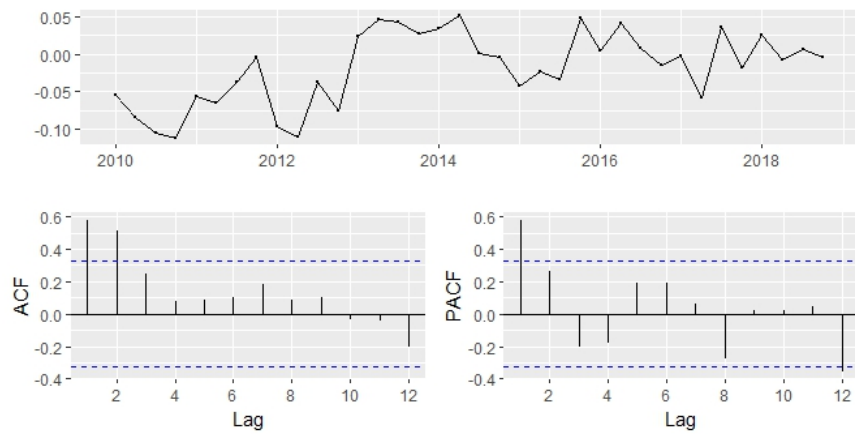
As observed, the PACF exhibits significant spikes while the ACF of Health loss ratio shows a decaying pattern. This observation leads us to consider an AR Model. However, given the outcomes of the regular and seasonal stationarity tests, it is logical to examine the ACF and PACF of the regular and seasonal lagged data.



(a) Original



(b) Regular Difference



(c) Seasonal Difference

Figure 5.2: Health Insurance loss ratios and its components, 2009-2018

Drawing inference from the ACF and PACF graphs of the regular lagged data, it can be observed that there are two significant spikes in the PACF while the ACF demonstrates a decaying pattern. On the other hand, concerning the seasonally lagged data, there is a single significant spike observed on the PACF while the ACF exhibits a tailing off pattern. Thus, the second candidate model according to the lagged graphs is: SARIMA(2,1,0)(1,1,0)[4].

Furthermore, "auto.arima" function in "forecast" package of R indicates a third candidate, which is: SARIMA(1,1,0)(0,1,1)[4]. A comparison is conducted among three candidate models to determine the most suitable model for the analysis as following.

Table 5.2: Comparison of Time Series Models: Health

Fitted Models	ADF	KPSS	OCSB	SW	BP	LB	AIC	BIC
(2,0,0)(1,1,0)[4]	0.070	0.043	-5.25	0.500	0.74	0.230	-129.6	-123.2
(2,1,0)(1,1,0)[4]	0.031	0.100	-3.49	0.310	0.70	0.071	-124.7	-118.4
(1,1,0)(0,1,1)[4]	0.031	0.100	-3.49	0.130	0.70	0.420	-131.4	-126.8
Criteria	<0.05	>0.05	<- 1.89	>0.05	>0.05	>0.05	Smallest	Smallest

It should be noted that criteria for ADF, KPSS, SW and LB tests represents p-values while decision on OCSB is made according to test statistics. The model selection based on AIC and BIC involves choosing the model with the smallest AIC or BIC value. Based on the comparison values on Table 5.2, the best-fitted model is identified as SARIMA(1,1,0)(0,1,1)[4].

The coefficient test results (in Figure 5.3) reveal that both the Autoregressive (AR) and Seasonal Moving Average (SMA) parameters in SARIMA(1,1,0)(0,1,1)[4] model are statistically significant.

z test of coefficients: SARIMA(1,1,0)(0,1,1)[4]				
	Estimate	Std. Err.	z value	Pr(> z)
ar1	-0.542	0.148	-3.67	0.00025 ***
sma1	-0.735	0.194	-3.78	0.00016 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Figure 5.3: Coefficient Test for Health Time Series Model

After completing the modeling phase, ACF and PACF functions of the model residuals are plotted, as illustrated in Figure 5.4.

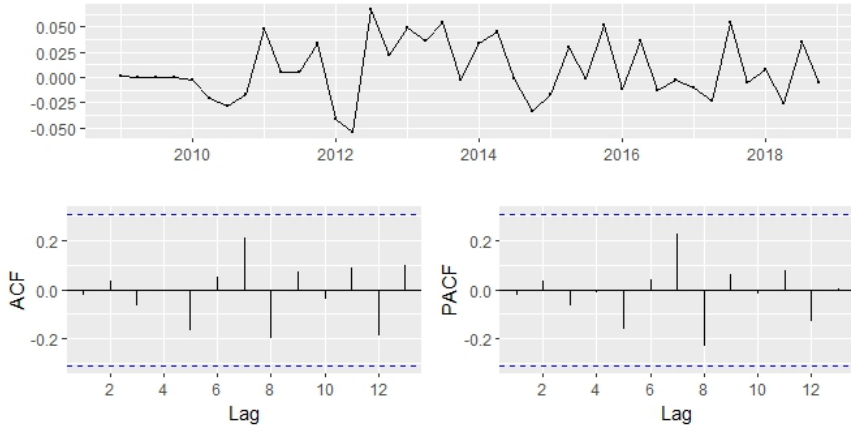


Figure 5.4: Health Residuals

It appears that there is no significant auto correlation or partial auto correlation observed between the lags of the model residuals.

The results of the stationary tests in Table 5.3 include original loss ratios, regularly differenced, seasonally differenced, and both regularly and seasonally differenced data sets across six LoBs. These results are presented only for the purpose of inferring trend detection.

Drawing inferences from Table 5.3, it can be concluded that Health exhibits non-stationarity with respect to seasonality since prior to the application of seasonal differencing, the test statistic is greater than the critical value. However, following the implementation of seasonal differencing, the time series becomes seasonally stationary.

The outcomes of the ADF tests indicates the presence of unit roots in all LoBs, necessitating transformations before modeling. In contrast, the results of the KPSS tests suggest non-stationarity for Accident and Health LoBs, while the remaining LoBs exhibit stationarity. Notably, there exist contradictions between the ADF and KPSS test results. However, the modeling results presented in Table 5.4 support the reliability of the KPSS test results within our datasets. This is substantiated by the fact that differences are not taken in Fire, MTPL and Miscellaneous, yet all assumptions

remain satisfied after modeling.

Table 5.3: Stationary Tests for LoBs' original loss ratios and differences

LoB/Tests	ADF	KPSS	OCSB
Acc	0.618	0.013	-3.948
Acc diff	0.022	0.100	-5.533
Acc r-s diff	0.020	0.100	-4.589
Fir	0.176	0.100	-4.048
Fir diff	0.014	0.100	-7.767
Fir r-s diff	0.016	0.100	-5.515
Hea	0.190	0.024	-1.040
Hea diff	0.077	0.025	-5.258
Hea r-s diff	0.031	0.100	-3.489
Mtp	0.086	0.100	-5.073
Mtp diff	0.069	0.100	-8.357
Mtp r-s diff	0.030	0.100	-5.331
Lan	0.946	0.064	-3.186
Lan diff	0.022	0.100	-11.559
Lan r-s diff	0.020	0.100	-7.949
Mis	0.273	0.100	-4.667
Mis diff	0.010	0.100	-7.215
Mis r-s diff	0.014	0.100	-4.731
Criteria	<0.05	>0.05	<1.8927

On the rows labeled with "diff" quote, regular differencing is applied to the LoBs before ADF and KPSS tests, while seasonal differencing is utilized before OCSB test. The criteria employed for ADF and KPSS tests involve the p-value, whereas this criteria is the test statistic for OCSB test.

The same process given in Algorithm 1 is followed to determine the most appropriate time series model for the other LoBs as well. Analyses concerning the remaining LoBs are available within the Appendix A. The best fitted time series models are given in the following Table 5.4.

Table 5.4: Best Fitted Time Series Models of LoBs

LoB	Best Fitted Model	Constant	Description
Accident	(1,1,1)(0,0,0)[4]	NO	ARIMA
Fire	(0,0,0)(0,0,0)[4]	YES	Random Walk With Drift
Health	(1,1,0)(0,1,1)[4]	NO	SARIMA
MTPL	(1,0,0)(0,0,0)[4]	YES	AR
Land Vehicle	(0,1,1)(0,0,0)[4]	NO	ARIMA
Miscellaneous	(0,0,0)(0,0,0)[4]	YES	Random Walk With Drift

The inclusion status of the constant term is also provided in Table 5.4. If a constant term is required in the model, it is denoted as "YES".

After completing the modeling phase for all LoBs, Breusch Pagan (BP) test is applied to the residuals to test if heteroskedasticity is present, Ljung-Box (LB) is applied to see if there is a serial autocorrelation and Shapiro-Wilk (SW) test is employed to check if normality assumption is satisfied. The results of these tests are given, as expressed in Table 5.5:

Table 5.5: Diagnostic Tests Results: P-values

Diagnostics Tests / LoB	Acc	Fir	Hea	Mtp	Lan	Mis
BP	0.94	0.31	0.69	0.15	0.71	0.07
LB	0.75	0.86	0.42	0.62	0.38	0.86
SW	0.63	0.01	0.13	0.01	0.02	0.37

It is important to observe that if the p-value in Table 5.5 exceeds 0.05, this implies the absence of heteroskedasticity according to the BP test, absence of serial autocorrelation based on LB test, and fulfillment of the normality assumption in relation to the SW test. Based on the BP tests results, it can be concluded that the variance in the data sets remains constant, which implies that there is no requirement for ARCH-GARCH modeling. According to the results of the Ljung-Box (LB) tests, it can be inferred that there are no significant serial autocorrelations present in the residuals obtained from the models. SW test results indicates that the residuals of Fire, MTPL and Land Vehicle LoBs do not satisfy the normality assumption. In such instances, it is customary to consider the possibility of non-linear patterns within the time series data. As a result, an Artificial Neural Network (ANN) namely, Autoregressive Neural Network (NNAR) is utilized to model the non-linear components in the time

series data. Despite the fulfillment of the normality assumption for certain residual sets from time series models, ANN models are still employed. The time series and Time series-ANN models are compared using several performance criteria, and the models that meet both the normality assumption and exhibit the best performance are chosen as the final models.

5.3 Artificial Neural Network - NNAR

NNAR models are chosen based on performance criteria, and the best-fitted models among them are selected as ANN models. Subsequently, a comparison is conducted between the Time Series (TS) models and the Time Series combined with Artificial Neural Network (TS+ANN) hybrid models.

The performance results regarding six LoBs are given in Table 5.6. Change column represents the performance improvement on each LoB. The "Change" column indicates the improvement in performance for each LoB. In cases where the value is negative, it signifies an improvement in performance compared to the TS model. Should enhancements be observed across all evaluation criteria, while simultaneously fulfilling the normality assumption, the hybrid model is selected for implementation.

According to the performance results, there are no improvements observed in two cases: Health and MTPL. However, for the remaining cases, the hybrid model demonstrated superior performance: Accident, Fire, Land Vehicle and Miscellaneous. Despite the fact that the performance of the Time Series model alone is better, the hybrid model is employed for the MTPL dataset due to the observed improvement in meeting the normality assumption. For the Health dataset, no improvement is observed in any perspective; therefore, the time series model is utilized for this case.

Table 5.6: Comparison of Time Series and TS+ANN Models

LoB	Criteria	TS	TS + ANN	Change (%)	Final Model
Accident	MAE	0.03342	0.02878	-14%	TS + ANN
	MAPE	0.07987	0.06791	-15%	TS + ANN
	MSE	0.00152	0.00092	-39%	TS + ANN
	RMSE	0.03893	0.03031	-22%	TS + ANN
Fire	MAE	0.06636	0.02735	-59%	TS+ ANN
	MAPE	0.08145	0.03424	-58%	TS + ANN
	MSE	0.00590	0.00082	-86%	TS + ANN
	RMSE	0.07681	0.02866	-63%	TS + ANN
Health	MAE	0.02202	0.03313	50%	TS
	MAPE	0.02384	0.03548	49%	TS
	MSE	0.00050	0.00180	260%	TS
	RMSE	0.02246	0.04245	89%	TS
MTPL	MAE	0.06511	0.06830	5%	TS
	MAPE	0.05693	0.05937	4%	TS
	MSE	0.00539	0.00564	5%	TS
	RMSE	0.07339	0.07513	2%	TS
Land Vehicle	MAE	0.15516	0.15391	-1%	TS + ANN
	MAPE	0.18492	0.18342	-1%	TS+ ANN
	MSE	0.02518	0.02477	-2%	TS + ANN
	RMSE	0.15869	0.15738	-1%	TS + ANN
Miscellaneous	MAE	0.06905	0.05955	-14%	TS + ANN
	MAPE	0.08599	0.07369	-14%	TS + ANN
	MSE	0.00618	0.00497	-20%	TS + ANN
	RMSE	0.07859	0.07050	-10%	TS + ANN

Furthermore, the normality tests are re-employed to the residuals of hybrid models whose results are given in Table 5.7.

Table 5.7: Normality Tests for TS and TS+ANN

LoBs	Acc	Fir	Hea	Mtp	Lan	Mis
TS	0.630*	0.012	0.130*	<0.01	0.020	0.370*
TS+ANN	0.084*	0.220*	0.860*	0.410*	0.570*	0.480*
*	Normally distributed if P-value>0.05					

Upon the application of the hybrid model to each LoB, the normality assumption is met for all of them. The ultimate model selection process prioritizes the fulfillment of the normality assumption as the initial criterion, followed by the performance improvement. The ultimate models, whether hybrid (TS+ANN) or time series only, are established and presented in Table 5.8.

Table 5.8: Best Performed Hybrid Models: TS+ANN

LoB	TS Model	ANN Model
Accident	(1,1,1)(0,0,0)[4]	NNAR(1,1,2)[4]
Fire	(0,0,0)(0,0,0)[4]	NNAR(2,1,4)[4]
Health	(1,1,0)(0,1,1)[4]	-
MTPL	(1,0,0)(0,0,0)[4]	NNAR(1,1,3)[4]
Land Vehicle	(0,1,1)(0,0,0)[4]	NNAR(1,3)
Miscellaneous	(0,0,0)(0,0,0)[4]	NNAR(1,6)

The residuals of the ultimate models, $\varepsilon_{i,t}$, are presented in Figure 5.5. The figure demonstrates a stationary pattern for each LoB, in concurrence with the numerical test results.

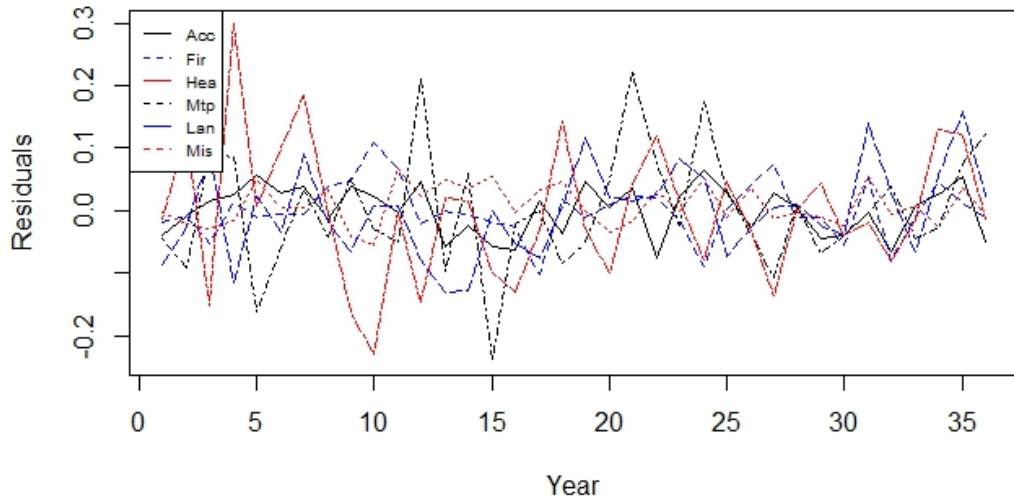


Figure 5.5: Residuals of the ultimate models

5.4 Copula Modeling

After obtaining the residuals from the final models, the proposed copula modeling process is followed (Algorithm 3). In line with the process, the correlation structure on a bi-variate basis and the contours are analyzed and illustrated in Figure 5.6. Contour plots provide insight into the dependency structure of a pair, enabling straightforward identification of negative or positive dependence, as well as upper tail, lower tail or symmetric dependencies. In Figure 5.6, some couples demonstrate negative correlations while most of the couples have positive dependence or no association at all. Certain pairs exhibit negative correlations, whereas the majority of pairs demonstrate positive dependencies or lack of association. Those with negative dependence are modeled with an appropriate bivariate copula function rotated by 90 or 270 degrees.

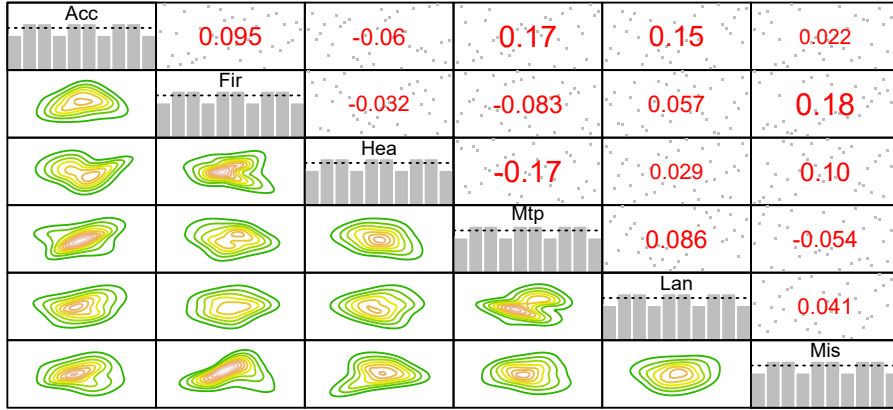


Figure 5.6: Copula pairs of the residuals of six LoBs

Initial substantial observations reveal that the Health-Miscellaneous pair exhibits lower tail dependence, whereas the Land Vehicle-Accident pair displays upper tail dependence. Furthermore, the Health-MTPL pair demonstrates negative dependence, whereas a noticeable positive dependence is observed in the Fire-Miscellaneous.

In accordance with the demand of copula modeling, a transformation is applied to the residuals. Let $\varepsilon_{i,t}$ be the residuals from time series models applied to six LoBs where $i = 1, \dots, 6$ and $t = 1, \dots, 40$. First, the domain of the residuals transferred to $[0,1]$ via the equation,

$$U_{i,k} \sim \hat{F}_n(k) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{x_j \leq k}, \quad (5.1)$$

and residuals become uniform marginals ($u_{i,t}$). These uniform marginals can be described as CDFs values of the residuals, $F_\varepsilon(\varepsilon_{i,t}) = u_{i,t}$.

Residuals are modeled using the methodology outlined in Algorithm [11]. This process includes identifying the optimal structure and selecting the best-fitting bivariate copulas. 6-dimensional R-vine structure is given by the tree representations in Figure 5.7:

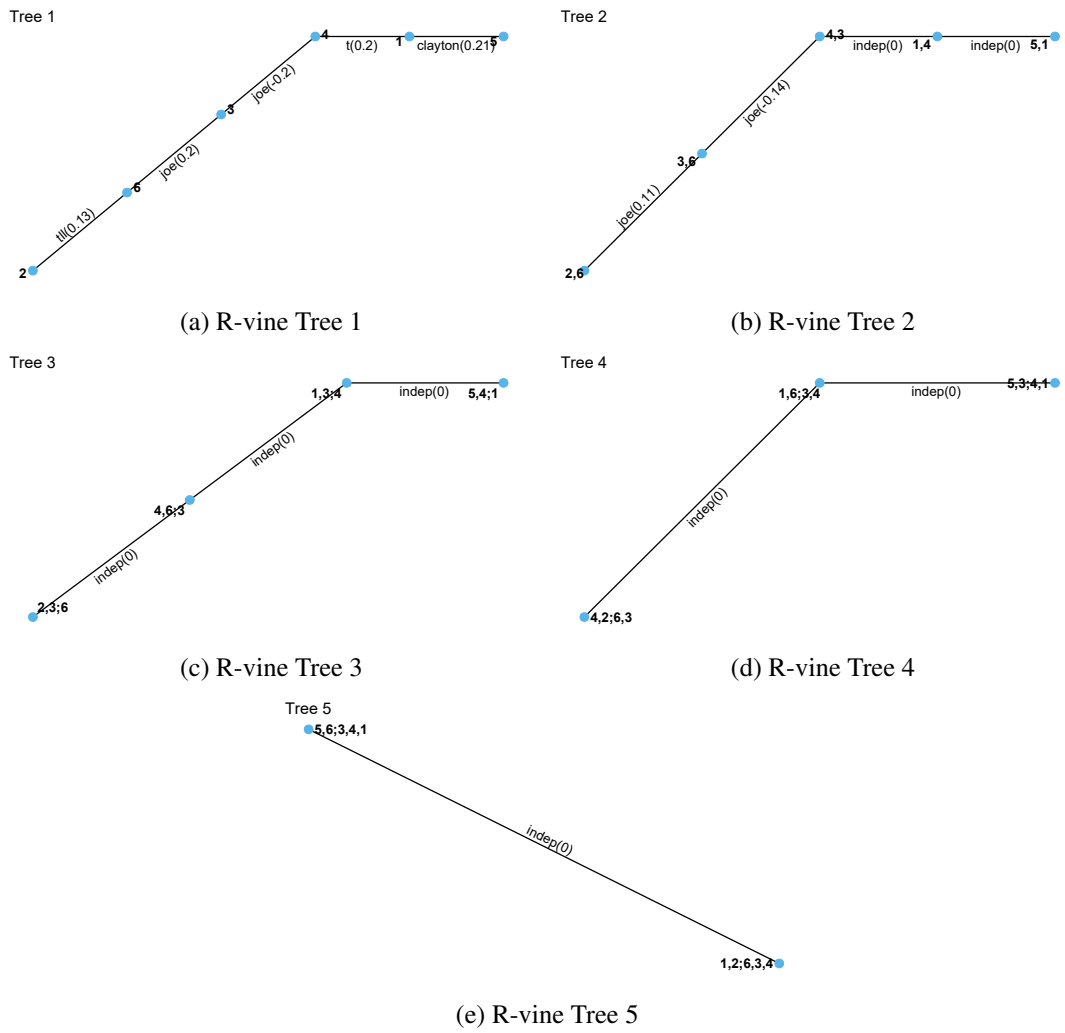


Figure 5.7: R-vine fitting for six dimensional structure

The specifications and the parameters of the bivariate copula models and R-vine trees are given in Table 5.9.

Table 5.9: R-vine structure: Bivariate distributions of edges

Tree	Edge	Cond-ed	Cond-ing	Family	Rotation	Param	df	Tau
1	1	5,1		Clayton	180	0.53	1	0.21
1	2	1,4		t	0	0.31, 2.00	2	0.20
1	3	4,3		Joe	90	1.4	1	-0.20
1	4	2,6		TLL	0	[30x30 g]	10	0.13
1	5	3,6		Joe	180	1.5	1	0.20
2	1	5,4	1	Indep	0		0	0
2	2	1,3	4	Indep	0		0	0
2	3	4,6	3	Joe	270	1.29	1	-0.14
2	4	2,3	6	Joe	0	1.22	1	0.11
3	1	5,3	4,1	Indep	0		0	0
3	2	6,1	4,3	Indep	0		0	0
3	3	2,4	3,6	Indep	0		0	0
4	1	1,2	3,4,6	Indep	0		0	0
4	2	5,6	1,3,4	Indep	0		0	0
5	1	2,5	1,3,4,6	Indep	0		0	0

The R-vine contour of the modeled variables is given in Figure (5.8). The contour plots of the initial five pairs under the first tree, namely Land Vehicle-Accident, Accident-MTPL, MTPL-Health, Fire-Miscellaneous, and Health-Miscellaneous, exhibit resemblances to the contour plots of the original values as shown in Figure 5.6. The resemblances between Figure 5.6 and Figure 5.8 are apparent in the contours of the first tree, indicating a good fit. The remaining contours within the upper-numbered trees describe the copula attributes of higher level pairings, exemplified by Fire-Health given Health and MTPL-Miscellaneous given Health.

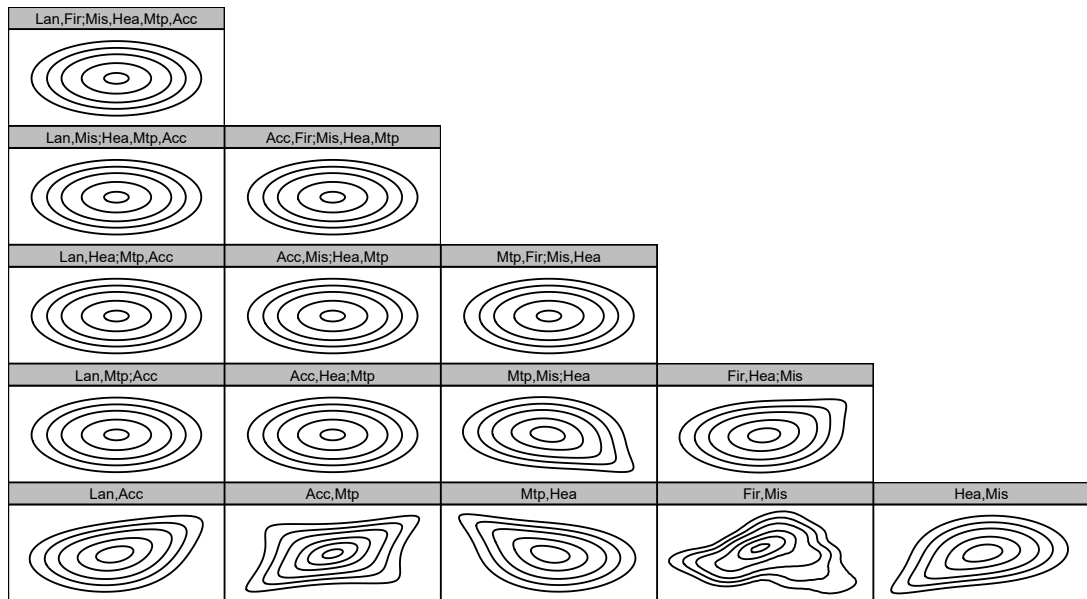


Figure 5.8: R-vine Contour

Recognizing that individual branches exhibit extreme tail values, with individual standard deviations addressing these tail observations, the proposed copula structure offers distinct advantages by collectively representing these extreme values. In our framework, values exceeding certain thresholds are considered in the calculation of the solvency capital requirement, a determination driven by the specified upon the dependence structure delineated by the R-vine. Consequently, the risk of inadequate diversification is mitigated.

5.5 Application Utilizing the Box-Cox Transformation

The utilization of the Box-Cox transformation to the response variable has the potential to enhance both its normality and, consequently, improve the accuracy and performance of the linear models [6]. Considering this information, the Box-Cox transformation is implemented on the loss ratios, and a similar methodology is adopted in the context of time series, artificial neural network and copula modeling.

A comparative analysis between time series and hybrid models has been conducted and the results are presented in Table 5.10 and Table 5.11, denoting the training and testing sets, respectively.

Table 5.10: Comparison of Time Series and Hybrid Models (training set)

LoB	Criteria	TS	TS+ANN	Change (%)	Final Model
Accident	MAE	0.04324	0.01917	-56%	TS+ANN
	MAPE	0.08310	0.03724	-55%	TS+ANN
	MSE	0.00290	0.00054	-81%	TS+ANN
	RMSE	0.05385	0.02325	-57%	TS+ANN
Fire	MAE	0.09978	0.08086	-19%	TS+ANN
	MAPE	0.12175	0.09609	-21%	TS+ANN
	MSE	0.01920	0.01380	-28%	TS+ANN
	RMSE	0.13856	0.11745	-15%	TS+ANN
Health	MAE	0.02613	0.02823	8%	TS
	MAPE	0.02811	0.03036	8%	TS
	MSE	0.00115	0.00115	0%	TS
	RMSE	0.03385	0.03387	0%	TS
MTPL	MAE	0.11329	0.10414	-8%	TS+ANN
	MAPE	0.09039	0.08150	-10%	TS+ANN
	MSE	0.02951	0.02803	-5%	TS+ANN
	RMSE	0.17178	0.16741	-3%	TS+ANN
Land Vehicle	MAE	0.07093	0.04463	-37%	TS+ANN
	MAPE	0.07682	0.04976	-35%	TS+ANN
	MSE	0.00965	0.00325	-66%	TS+ANN
	RMSE	0.09824	0.05704	-42%	TS+ANN
Miscellaneous	MAE	0.09994	0.04722	-53%	TS+ANN
	MAPE	0.12890	0.06118	-53%	TS+ANN
	MSE	0.01563	0.00378	-76%	TS+ANN
	RMSE	0.12501	0.06144	-51%	TS+ANN

Table 5.11: Comparison of Time Series and Hybrid Models (testing set)

LoB	Criteria	TS	TS+ANN	Change (%)	Final Model
Accident	MAE	0.03222	0.02484	-23%	TS+ANN
	MAPE	0.07730	0.05634	-27%	TS+ANN
	MSE	0.00137	0.00074	-46%	TS+ANN
	RMSE	0.03701	0.02728	-26%	TS+ANN
Fire	MAE	0.06636	0.04438	-33%	TS+ANN
	MAPE	0.07933	0.05239	-34%	TS+ANN
	MSE	0.00640	0.00372	-42%	TS+ANN
	RMSE	0.07998	0.06099	-24%	TS+ANN
Health	MAE	0.01396	0.01720	23%	TS
	MAPE	0.01483	0.01815	22%	TS
	MSE	0.00027	0.00038	43%	TS
	RMSE	0.01636	0.01958	20%	TS
MTPL	MAE	0.06002	0.05686	-5%	TS+ANN
	MAPE	0.05207	0.04884	-6%	TS+ANN
	MSE	0.00444	0.00404	-9%	TS+ANN
	RMSE	0.06665	0.06358	-5%	TS+ANN
Land Vehicle	MAE	0.15356	0.13394	-13%	TS+ANN
	MAPE	0.18302	0.15856	-13%	TS+ANN
	MSE	0.02469	0.02231	-10%	TS+ANN
	RMSE	0.15712	0.14936	-5%	TS+ANN
Miscellaneous	MAE	0.07737	0.06047	-22%	TS+ANN
	MAPE	0.09483	0.07133	-25%	TS+ANN
	MSE	0.00729	0.00528	-28%	TS+ANN
	RMSE	0.08539	0.07268	-15%	TS+ANN

The hybrid models which improve the performance in both training and testing sets are designated as the final models and listed in Table 5.12.

Table 5.12: Best Performed Hybrid Models for Transformed Data

LoB	TS Model	ANN Model
Accident	(0,1,2)(0,0,0)[4]	NNAR(1,1,5)[4]
Fire	(0,0,0)(0,0,0)[4]	NNAR(1,1,2)[4]
Health	(1,1,0)(1,1,0)[4]	-
MTPL	(1,0,0)(0,0,0)[4]	NNAR(1,1,1)[4]
Land Vehicle	(0,1,1)(0,0,0)[4]	NNAR(1,1,2)[4]
Miscellaneous	(0,0,0)(0,0,0)[4]	NNAR(1,1,6)[4]

The copula model structure after Box-Cox transformation is given in Table 5.13.

Table 5.13: R-vine structure: Bivariate distributions of edges (data with Box-Cox transformation)

Tree	Edge	Cond-ed	Cond-ing	Family	Rotation	Param	df	Tau
1	1	4,1		t	0	0.34, 2.00	2	0.224
1	2	1,6		Gumbel	90	1.4	1	-0.279
1	3	5,3		Indep	0		0	0
1	4	2,6		Indep	0		0	0
1	5	3,6		Clayton	0	0.59	1	0.228
2	1	4,6	1	Indep	0		0	0
2	2	1,2	6	Indep	0		0	0
2	3	5,6	3	Indep	0		0	0
2	4	2,3	6	t	0	0.12, 2.00	2	-0.014
3	1	4,2	6,1	Indep	0		0	0
3	2	3,1	6,2	Indep	0		0	0
3	3	5,2	6,3	Indep	0		0	0
4	1	4,3	6,2,1	Joe	90	1.3	0	-0.140
4	2	5,1	6,3,2	Indep	0		0	0
5	1	4,5	6,3,2,1	Indep	0		0	0

In the next chapter, the outcomes of two distinct analyses will be presented: one conducted without the Box-Cox transformation, and the other one incorporating the Box-Cox transformation. Subsequently, a comparative evaluation of these results will be undertaken.

5.6 Simulation

Based on the Algorithm 4, both the Inverse Rosenblatt Transform (1952) (Definition 4.2.1) and the recursive formula proposed by Czado (2022), (Equation (4.3)), can be employed to obtain simulations from the chosen structure.

The decision is made by comparing two methods using the absolute mean of the difference between the Kendall correlation coefficients matrix of the simulations and the modeled residuals. Despite the obtained results indicating similarity between the two methods, the recursive formula exhibited slightly superior performance (0.040 vs 0.042, as percentage of mean absolute deviance). Thus, the simulation of the residuals is carried out using the recursive formula.

The amounts derived from the standard formula are also provided for the purpose of comparison. The corresponding standard deviations utilized in the Standard Formula in Solvency II framework is given in Table 5.14. The rates used in the standard formula are directly sourced from the regulation [12], whereas the Turkish industry rates are computed as part of this study.

Table 5.14: Standard Deviations used in the Standard Formula

	Standard Formula	Turkish Market
Accident	4%	7%
Fire	8%	14%
Health	4%	10%
MTPL	10%	19%
Land Vehicle	8%	12%
Miscellaneous	13%	13%

The calculation of required capital involves two distinct approaches concerning volumes. In the first approach, volumes are derived from the average of the last four quarters and the weights in the industry are comprehensively represented. Conversely, the second approach employs equal weighting, with each LoB assigned a fixed value of 100 units. Having calculated Solvency Capital Requirements via Standard Model, Hybrid Internal Model and Proposed Model, results are shown in Figure (5.9a-5.9b)

Table 5.15: Calculations with the estimated volumes for 2021 and equally weighted volumes (each 100)

	Method	V_{total}	$SCR \cdot NL_p$	Ratio
Weighted	Standard Model	52 BLN	8.7 BLN	17%
	Standard Model TR	52 BLN	14.2 BLN	27%
	Internal Model (VaR)	52 BLN	9.2 BLN	18%
	Internal Model (TVaR)	52 BLN	10 BLN	20%
Equally weighted	Standard Model	600	91.7	15%
	Standard Model TR	600	136.5	23%
	Internal Model (VaR)	600	65.0	11%
	Internal Model (TVaR)	600	71.0	12%

In the estimated volumes case, Table 5.15 illustrates that the standard model suggests an SCR of 17% for the premium risk, whereas the Internal Model proposes a requirement of 18%. These figures exhibit significant similarity when considering the portfolio with proportions estimated from the industry data. In the case of equal weighting, Table 5.15 reveals that the standard model recommends an SCR of 15%, whereas the internal model suggests 11%. The notable difference can be attributed to the fact that the internal model possesses greater information regarding extreme events (tails), leading to more pronounced diversification effects among the LoBs.

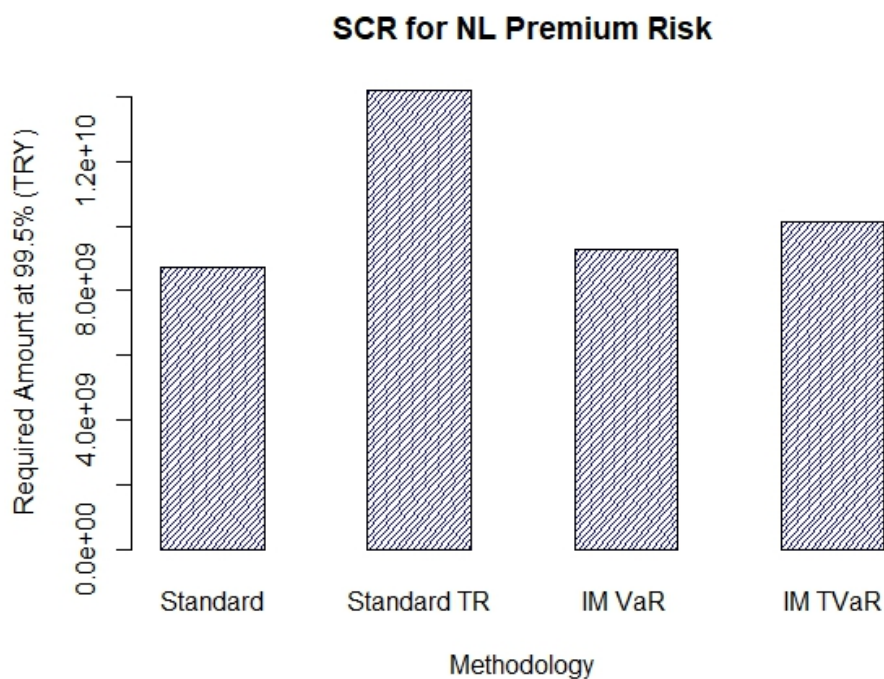
Furthermore, it is evident that the calculated amounts based on the Turkish insurance industry rates are substantially higher in both the estimated volumes and equally weighted volumes. This discrepancy arises due to the standard deviations being computed from time series data, where regular trends and seasonal variations contribute to elevated standard deviations. The absence of trend and seasonal modeling is expected to lead to overestimation of the required amounts.

Moreover, subsequent to the implementation of the Box-Cox transformation, the ratios are recalculated, and a comparative analysis of these ratios is presented in Table 5.16. Ratio 1 signifies the computations conducted without employing the Box-Cox transformation, while Ratio 2 corresponds to the calculations performed with the application of Box-Cox transformation.

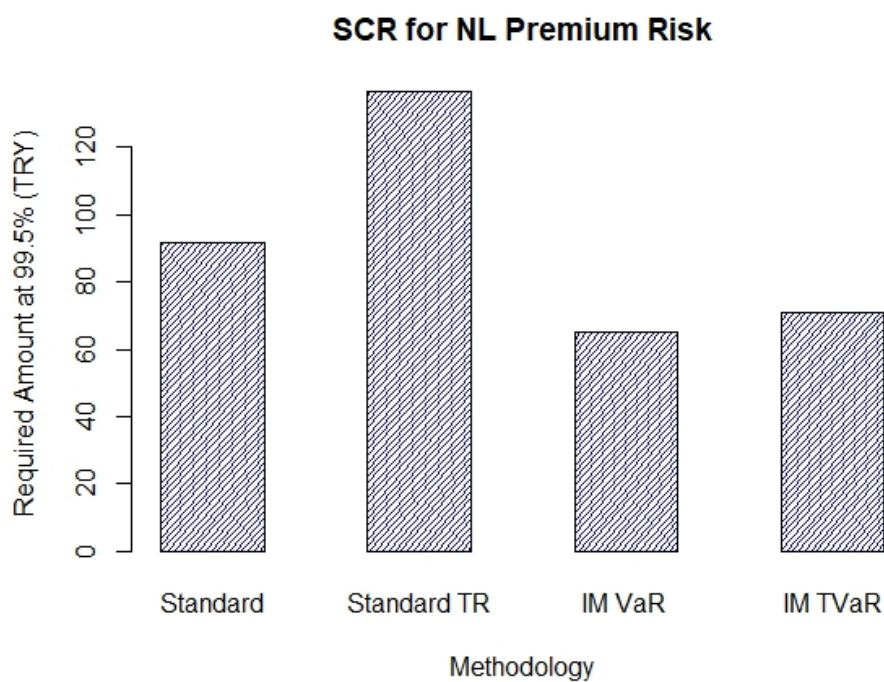
Table 5.16: SCR ratios with the estimated volumes for 2021 and equally weighted volumes

	Method	Ratio 1	Ratio 2: Box-Cox
Weighted	Standard Model	17%	17%
	Standard Model TR	27%	27%
	Internal Model (VaR)	18%	25%
	Internal Model (TVaR)	20%	29%
Equally weighted	Standard Model	15%	15%
	Standard Model TR	23%	23%
	Internal Model (VaR)	11%	15%
	Internal Model (TVaR)	12%	17%

Making inferences from the data presented in Table 5.16, it becomes evident that the use of the Box-Cox transformation leads to an increase in the SCR ratios of the internal models. This fact can be attributed to the enhanced capturing of tail values facilitated by the Box-Cox transformation. In contrast to the analyses conducted without the Box-Cox transformation, the analysis incorporating the Box-Cox transformation exhibits a larger difference between VaR and TVaR. This divergence can be explained by increased ability to generate more extreme tail values.

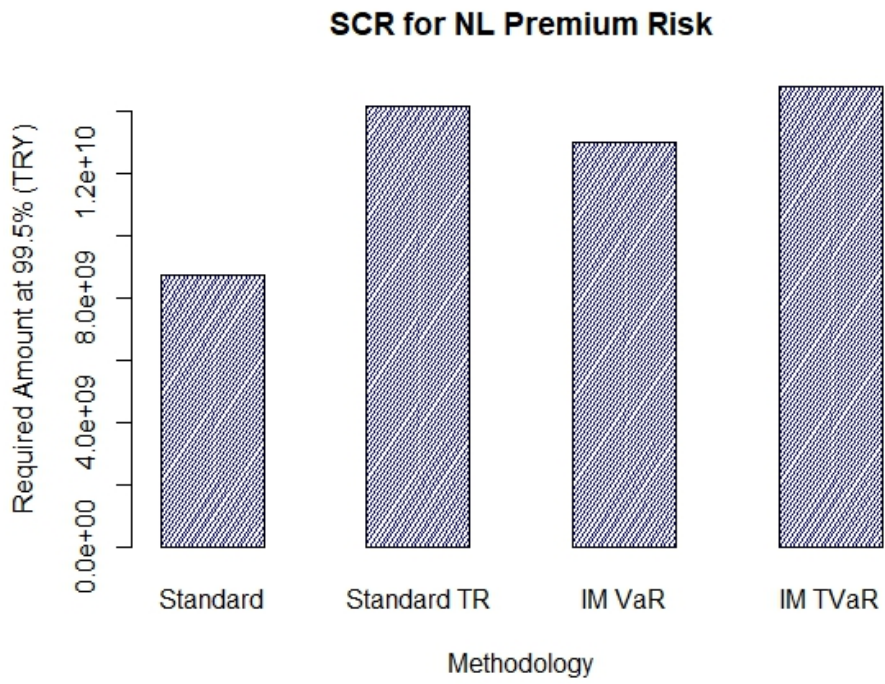


(a) The estimated volumes for 2021

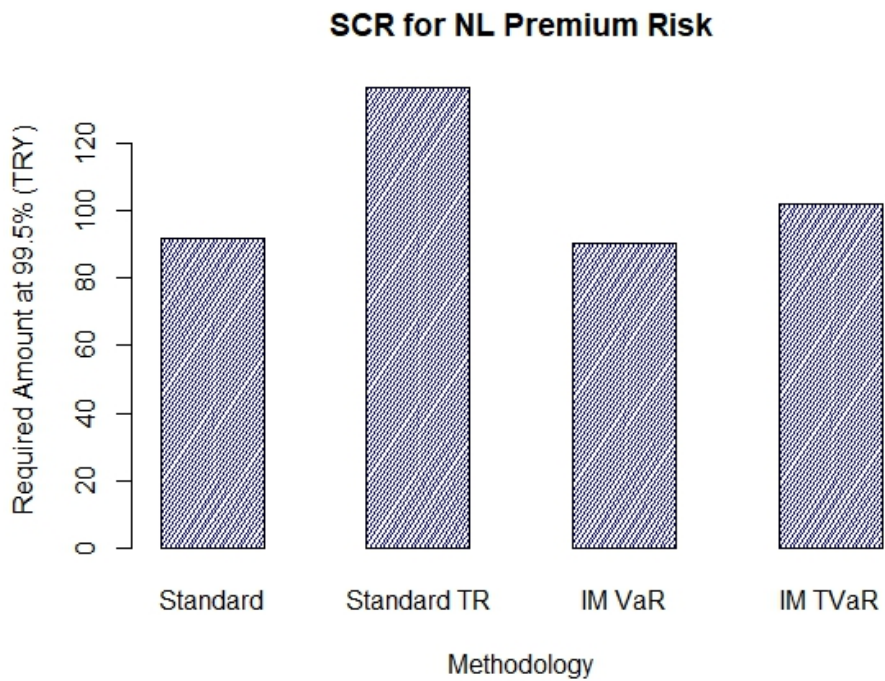


(b) Equally weighted volumes

Figure 5.9: Graphical representation of Solvency Capital Requirements



(a) The estimated volumes for 2021



(b) Equally weighted volumes

Figure 5.10: Graphical representation of Solvency Capital Requirements after the Box-Cox transformation is applied

In the context of sensitivity analysis, a 5% shock is imposed on each loss ratio, and the subsequent changes in the SCR ratios are investigated. In the case of a 5% escalation in historical loss ratios, the SCR values calculated by the proposed models exhibit incremental shifts of 2-3%. Conversely, there are no changes in the SCR values for standard models, as the volatility parameters used in standard models remain constant for all branches.

Table 5.17: Sensitivity analysis for 5% shock in each loss ratio

	Method	Original	After shock	Change
Weighted	Standard Model	17%	17%	0%
	Standard Model TR	27%	27%	0%
	Internal Model (VaR)	25%	28%	3%
	Internal Model (TVaR)	29%	32%	3%
Equally weighted	Standard Model	15%	15%	0%
	Standard Model TR	23%	23%	0%
	Internal Model (VaR)	15%	17%	2%
	Internal Model (TVaR)	17%	20%	3%

5.7 Ruin Probability

The regulatory framework sets the ruin probability at 0.5%. Within the simulation component of this thesis, a loss environment regarding chosen LoBs and spanning thousands of years is synthesized. While calculating SCR values under proposed model, the ruin probability remains fixed at 0.5% in accordance with regulatory guidelines. Additionally, for the purpose of establishing a comparative benchmark, we calculate the outcomes of the standard model within the framework of our artificially constructed losses environment. We determine the probability of exceeding SCR values, as computed by standard formulas, within the values in the synthesized data set of loss ratios and contrast this with a predefined fixed ruin probability of 0.5%. The calculated ruin probabilities are given in Table 5.18.

Table 5.18: Ruin probability comparison

	Type	Original	After shock
Weighted	Standard Model	5.94%	11.90%
	Standard Model TR	0.31%	0.62%
	Internal Model (VaR)	0.50%	0.50%
	Internal Model (TVaR)	0.20%	0.20%
Equally weighted	Standard Model	0.61%	1.30%
	Standard Model TR	0.04%	0.06%
	Internal Model (VaR)	0.50%	0.50%
	Internal Model (TVaR)	0.20%	0.20%

In the context of the weighted volumes case, the ruin probability associated with the standard model is notably high, primarily because it fails to adequately represent the volatility specific to the Turkish insurance industry. Conversely, when we calculate the standard formula using Turkish standard deviations, the resulting ruin probability is 0.31%, which exceeds the required threshold by approximately 0.2%. Consequently, this leads to the unnecessary retention of company's free cash flow.

In the scenario where a 5% shock is applied to all loss ratios, the ruin probability for the proposed model remains stable at the 0.5% level. However, the ruin probability for the standard model, when incorporating Turkish industry volatility, escalates to 0.62%. This increase in ruin probability signifies a shortage in the SCR and elevates the risk of bankruptcy.

Moreover, in the case of equal weighting among branches, the standard formula considering Turkish industry volatility yields ruin probabilities of 0.04% and 0.06%, corresponding to the original dataset and the dataset after applying a 5% shock, respectively. These rates fall significantly below the 0.5% threshold, thereby, this could result in using an excessive retention of the company's free cash flow.

When dealing with branches of varying weights, the outcomes may differ. Nevertheless, it can be inferred that employing our proposed model will provide users to calculate SCR estimates that align with the 0.5% ruin probability threshold in any case. This approach ensures that the company avoids both underestimation and overestimation of the SCR, thus accurately addressing its unique premium risk characteristics.

CHAPTER 6

CONCLUDING COMMENTS

The role of insurance as a pivotal component within the realm of financial services is of utmost significance. The insurance industry offers insurance coverage through various branches to diverse entities, including the individuals, real sector, the financial sector, among others. In this manner, companies can conduct their operations securely, while individuals can safeguard their future and maintain their present standing through the benefits of insurance coverage. To ensure the continuity of this advantageous environment, the utmost crucial aspect pertains to the insurance companies' capability to promptly fulfill the compensations arising from the policies they have issued. The Solvency II regulation serves as a foundation for mitigating potential disruptions that could emerge from the risks encountered by insurance companies such as market, life, non-life, health, default and intang risks. The regulation mandates that insurance companies maintain a solvency capital requirement (SCR) in addition to the regular reserves, ensuring the capacity to address unforeseen risks associated with the aforementioned ones. Determining the appropriate capital level is crucial for the financial health of the company. Although regulations offer standardized calculations, companies may utilize internal models, subject to approval by authorities.

In this thesis, an internal model is introduced to address Non-Life Premium risk, integrating a hybrid Time Series - Artificial Neural Network - Copula approach. In this methodology, the loss ratios regarding the non-life branches are initially characterized using suitable time series models. Subsequently, the residuals of these time series models are further captured using an autoregressive neural network (NNAR) model. While the linear component is addressed through time series modeling, the non-linear

aspect is handled using an artificial neural network model. This combined approach ensures that all residuals of the model conform to a normal distribution, which is a critical assumption. Additionally, the inclusion of ANN generally leads to improved performances in the majority of the cases. Finally, the suitable R-vine structure is devised to effectively accommodate the residuals of the Time Series-ANN model. After the aforementioned process, a Box-Cox transformation is applied to each Line of Business (LoB), and the same algorithm is then re-executed with the transformed variables. The results are preserved for the comparison purposes.

The copula structure is employed for the purpose of simulating residuals corresponding to the selected Line of Businesses (LoB). By aggregating the time series and ANN components with the residuals, the simulations for the subsequent time intervals are projected. Through deducting one unit from each projected ratio, the difference between the projected ratio and the affordable amount covered by regular reserves is determined, that is the additional amount that needs to be compensated through the SCR. As an additional necessary procedure, any negative values in the differences are transformed into zeros. Next, the projected loss ratios are combined with their corresponding volumes to calculate the total capital requirement with respect to the chosen LoBs.

After completing the simulation phase, the results for the Solvency Capital Requirement (SCR) are obtained using several approaches. These include the SCR calculated using the standard formula as proposed in the regulation [12], the SCR calculated using the same formula but with the standard deviations of Turkish data employed in the study, as well as the Value at Risk ($\text{VaR}_{99.5}$) and Tail Value at Risk ($\text{TVaR}_{99.5}$) of the simulated projections by employing proposed model. A comparative analysis is then performed to assess the differences and similarities among these different approaches.

- (i) Based on the outcomes derived from the simulations, it can be confidently stated that the hybrid models employed for the various LoBs have successfully met all underlying assumptions and establishing the credibility of the models. Thus, it can be affirmed that the hybrid approach facilitates the fulfillment of assumptions.
- (ii) Next, the utilization of the Box-Cox transformation offers advantages in terms

of obtaining more precisely fitted hybrid models and enhanced representation of extreme tail values.

- (iii) Moreover, in the scenario of equally weighted volumes, the analysis of SCR outcomes demonstrates a potential reduction in the SCR amount through the incorporation of a temporal component into the model. This arises due to the temporal parameter's simultaneous impact across all LoBs. This influence could be duly accounted for when determining regular reserves, rather than affecting the SCR.
- (iv) In addition, the copula structure facilitates a more profound representation of tail dependence. The utilization of advanced tail dependence methodologies, such as R-vine, enhances the realism of SCR estimates by effectively addressing extreme scenarios.

The methodological framework proposed in this thesis offers a robust alternative strategy for non-life insurance companies seeking to assess premium risk using the internal model option within the context of Solvency II. This approach can also be used to test the portfolios of LoBs to decide to invest in which LoBs. Moreover, this model can function as a tool for appraising portfolios, helping in the optimization of concentration across diverse LoBs to mitigate the allocation of high SCR.

APPENDIX A

TIME SERIES ANALYSES

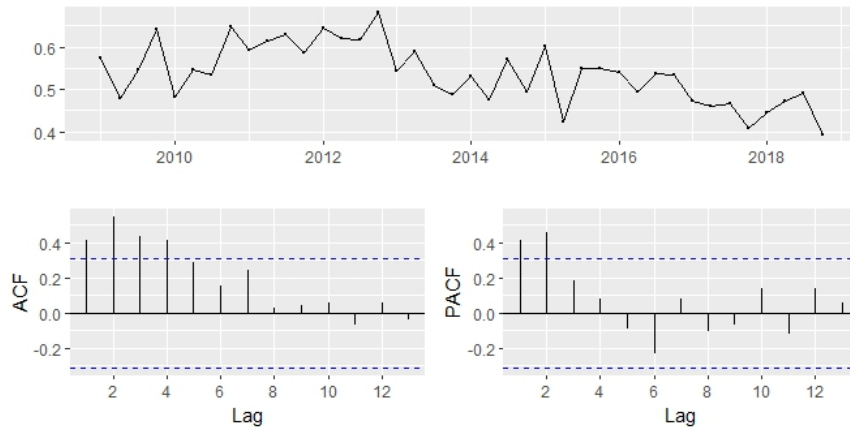
A.1 Accident

As indicated by the Stationarity Tests presented in Table 5.3, it is evident that the Accident loss ratios exhibit a consistent trend, thereby requiring the adoption of regular differencing.

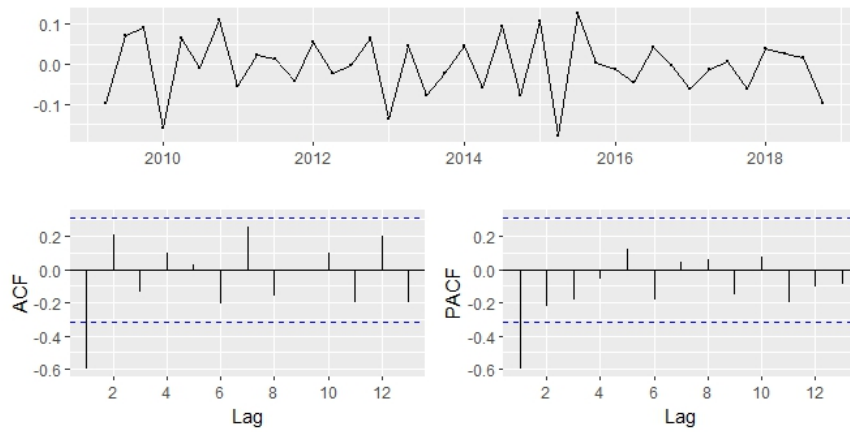
As evident from the analysis of Figure A.1, the analysis of the ACF and PACF plots for the lagged data indicates the presence of solitary spikes. Consequently, both AR and MA components can be contemplated within the model, as these spikes may signify either decaying or spiking patterns. Moreover, it is conceivable to consider an ARIMA model encompassing both AR and MA components. The potential model candidates are ARIMA(1,1,0), ARIMA(0,1,1), and ARIMA(1,1,1).

Table A.1: Comparison of Time Series Models: Accident

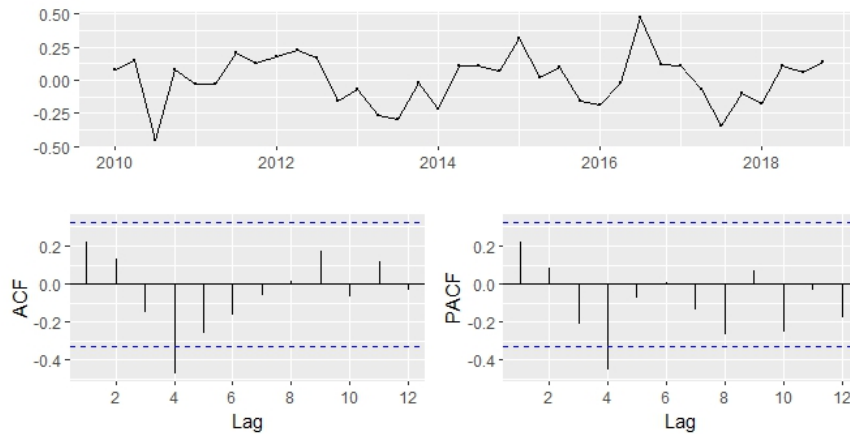
Fitted Models	ADF	KPSS	OCSB	SW	BP	LB	AIC	BIC
(1,1,0)(0,0,0)[4]	0.16	0.10	-3.68	0.87	0.91	0.30	-108.2	-105.0
(0,1,1)(0,0,0)[4]	0.37	0.10	-3.80	0.57	0.70	0.39	-108.5	-105.2
(1,1,1)(0,0,0)[4]	0.31	0.10	-4.13	0.63	0.94	0.75	-109.4	-104.4
Criteria	<0.05	>0.05	<- 1.89	>0.05	>0.05	>0.05	Smallest	Smallest



(a) Original



(b) Regular Difference



(c) Seasonal Difference

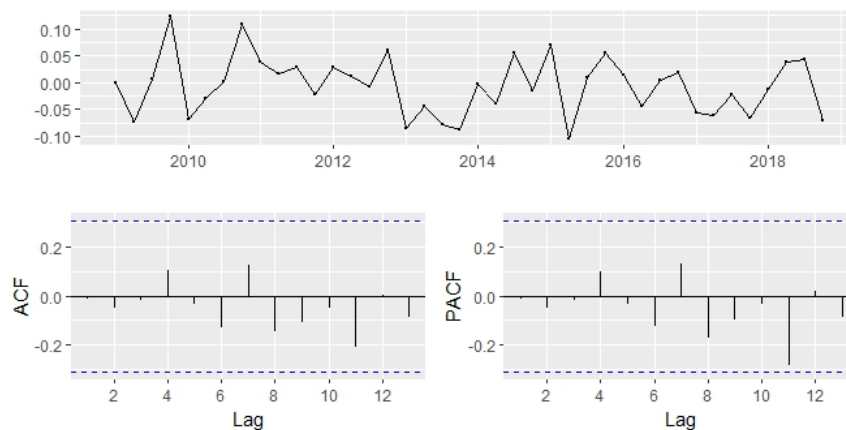
Figure A.1: Accident Insurance loss ratios and its components, 2009-2018

Based on the findings presented in Table A.1, the ARIMA(1,1,1) model is selected as the most suitable fitting model. However, the other models remain viable alternatives, as indicated by the relatively modest differences in the AIC and BIC scores. Upon model fitting, an examination of the parameters is conducted, and the results are illustrated in Figure A.2a. In addition, ACF and PACF plots of the model residuals are given in Figure A.2b.

z test of coefficients: ARIMA(1,1,1)				
	Estimate	Std. Error	z value	Pr(> z)
ar1	-0.401	0.210	-1.91	0.056 .
ma1	-0.425	0.197	-2.16	0.031 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

(a) Coefficient Test



(b) Residuals

Figure A.2: Diagnostic checks of Accident branch

In accordance with the outcomes, all assessed coefficients exhibit statistical significance, and the ACF and PACF plots display satisfactory characteristics. As a result, it can be inferred that the validity of ARIMA(1,1,1) model for the Accident dataset is affirmed.

A.2 Fire

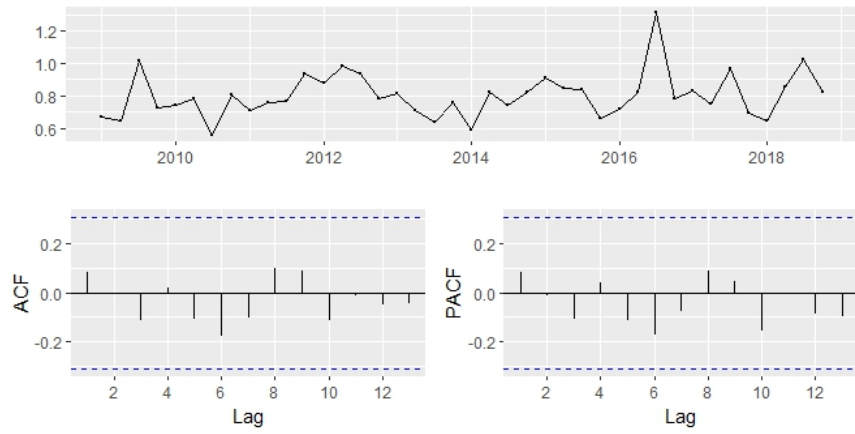
As evidenced from the Stationarity Tests outlined in Table 5.3, the ADF test proposes the need for differencing, whereas the KPSS test assumes the process to be stationary.

As observable through analysis in Figure A.1, the ACF and PACF plots for the lagged data indicates the presence of single spikes. Similarly in the instance of the Accident dataset, both AR and MA components can be employed in the model, as these spikes may indicate either decaying or spiking patterns. In Additionally, it is acceptable to consider an Random Walk With Drift (RWWD) model which does not include AR and MA components since the ACF and PACF plots of the original dataset show no indication of non-stationarity, in contrast to the ADF test. The model candidates to be considered are ARIMA(1,1,0), ARIMA(0,1,1), and ARIMA(0,0,0) model with a constant coefficient.

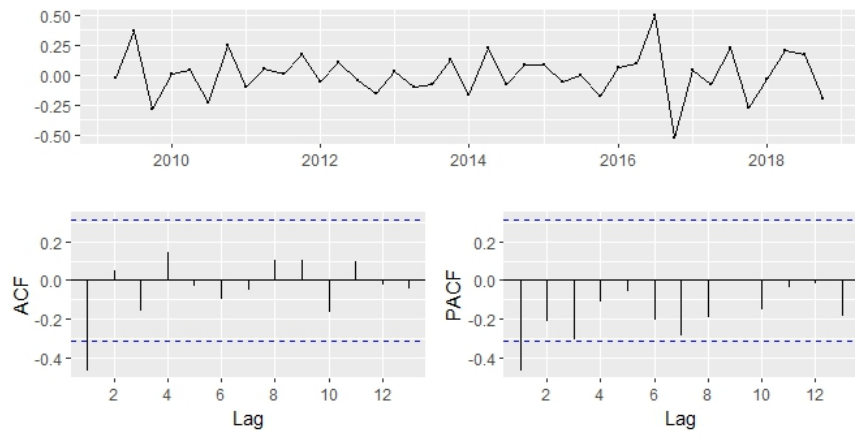
Table A.2: Comparison of Time Series Models: Fire

Fitted Models	ADF	KPSS	OCSB	SW	BP	LB	AIC	BIC
(1,1,0)(0,0,0)[4]	0.04	0.10	-3.68	0.097	0.56	0.28	-23.9	-18.9
(0,1,1)(0,0,0)[4]	0.18	0.10	-4.01	0.006	0.74	0.51	-35.6	-30.6
(0,0,0)(0,0,0)[4]	0.18	0.10	-4.04	0.013	0.31	0.86	-41.5	-38.1
Criteria	<0.05	>0.05	<- 1.89	>0.05	>0.05	>0.05	Smallest	Smallest

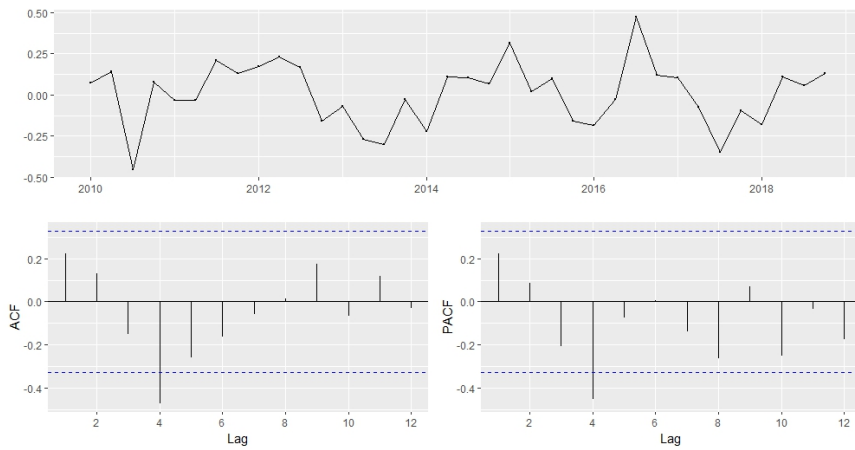
Based on the findings presented in Table A.2, the ARIMA (1,1,0) model fulfills all requisite criteria, yet its performance in terms of the AIC and BIC criteria is suboptimal when compared with the other alternatives. Considering the option to model the residuals through an ANN framework, the best performing model, ARIMA(0,0,0) with a constant coefficient (drift), is chosen as the best fitting one. This model is referred to as RWWD, to put it differently. After model fitting, an evaluation of the parameters is conducted, and the results are given in Figure A.4a. In addition, ACF and PACF plots of the model residuals are illustrated in Figure A.4b.



(a) Original



(b) Regular Difference



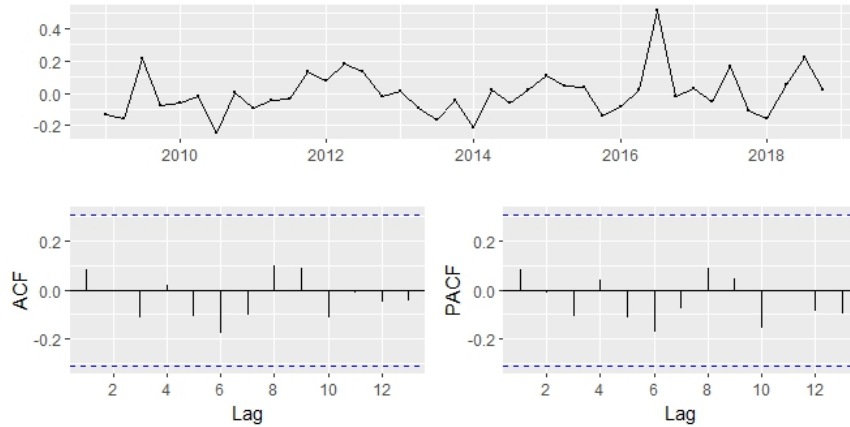
(c) Seasonal Difference

Figure A.3: Fire Insurance loss ratios and its components, 2009-2018

z test of coefficients: RWWD				
	Estimate	Std. Error	z value	Pr(> z)
intercept	0.8013	0.0217	37	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

(a) Coefficient Test



(b) Residuals

Figure A.4: Diagnostic checks of Fire branch

Consistent with the findings, the intercept is statistically significant and the ACF and PACF plots show satisfactory characteristics. Consequently, it can be affirmed that the RWWD model is valid for the Fire dataset.

A.3 MTPL

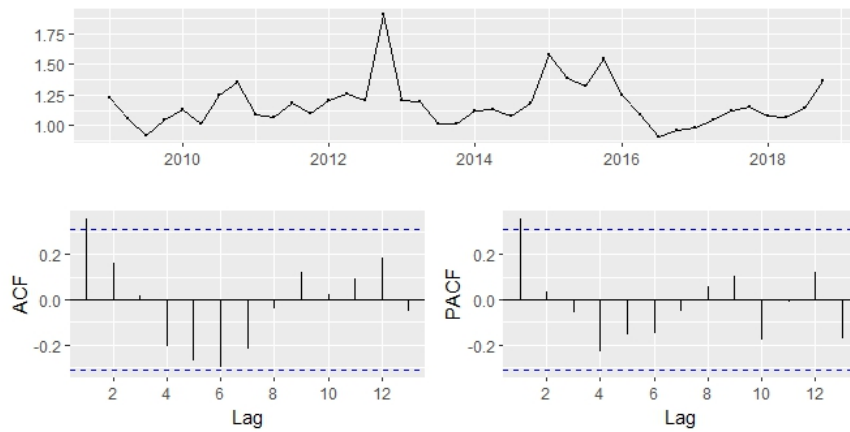
Corresponding to the results of the Stationarity Tests outlined in Table 5.3, the ADF test indicating a unit root, while the KPSS test proposes the process is stationary. In scenarios characterized by insufficient data size, the emergence of such contradictions is possible. Therefore, it is recommended to employ alternative models under such circumstances.

Observable through analysis in Figure A.5, the original dataset indicates one single spike on the ACF and PACF plots while having no significant spikes at the data subjected to regular differencing. In addition, both AR and MA components can be employed in the model, as these spikes may indicate either decaying or spiking patterns. In Additionally, it is acceptable to consider an integrated ARIMA model since ADF test results indicate a possible unit root. Finally, the model candidates to be considered are AR(1), MA(1), ARIMA(1,1,0) and ARIMA (0,1,1).

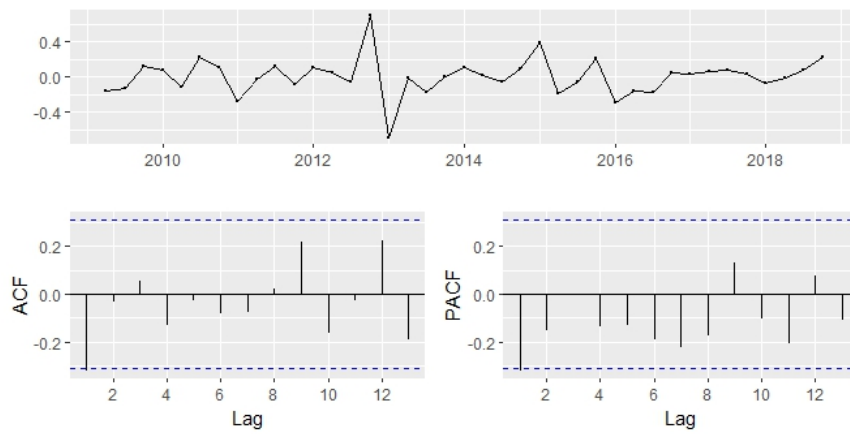
Table A.3: Comparison of Time Series Models: MTPL

Fitted Models	ADF	KPSS	OCSB	SW	BP	LB	AIC	BIC
(1,0,0)(0,0,0)[4]	0.11	0.10	-4.28	<0.01	0.15	0.62	-20.4	-15.3
(0,0,1)(0,0,0)[4]	0.12	0.10	-4.36	<0.01	0.24	0.39	-19.6	-14.5
(1,1,0)(0,0,0)[4]	0.05	0.10	-4.21	0.04	0.96	0.68	-8.9	-3.9
(0,1,1)(0,0,0)[4]	0.06	0.10	-4.60	0.01	0.71	0.35	-10.5	-5.5
Criteria	<0.05	>0.05	<- 1.89	>0.05	>0.05	>0.05	Smallest	Smallest

Following the findings presented in Table A.3, no model fulfills all required criteria. Thus, an ANN model is applied to the residuals of chosen model. However, it is imperative to select the most appropriate model given these circumstances. The AR(1) model is identified as the most suitable fitted model based on the comparison of AIC and BIC. After model fitting, an evaluation of the parameters is conducted, and the results are given in Figure A.6a. In addition, ACF and PACF plots of the model residuals are illustrated in Figure A.6b.



(a) Original



(b) Regular Difference



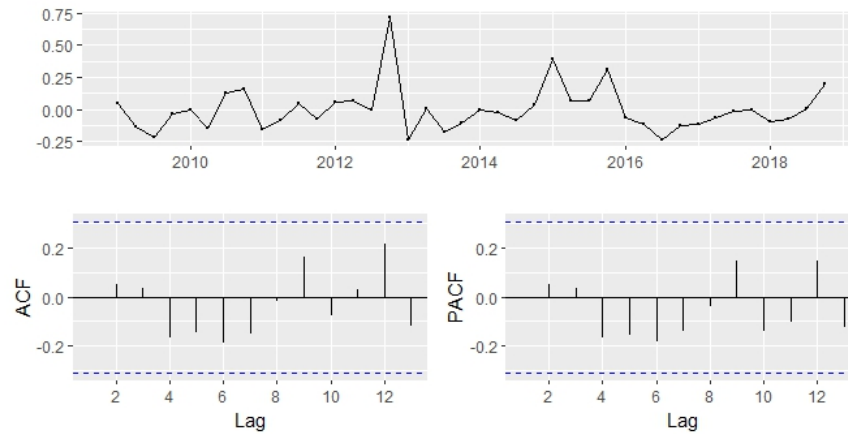
(c) Seasonal Difference

Figure A.5: MTPL Insurance loss ratios and its components, 2009-2018

z test of coefficients: AR(1)				
	Estimate	Std. Error	z value	Pr(> z)
ar1	0.3615	0.1470	2.46	0.014 *
intercept	1.1770	0.0425	27.72	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

(a) Coefficient Test



(b) Residuals

Figure A.6: Diagnostic checks of MTPL branch

Adhering to the outcomes, all evaluated coefficients display statistical significance, and the ACF and PACF plots proves the model has adequate characteristics. As a result, it can be said that AR(1) model is suitable for the MTPL dataset.

A.4 Land Vehicle

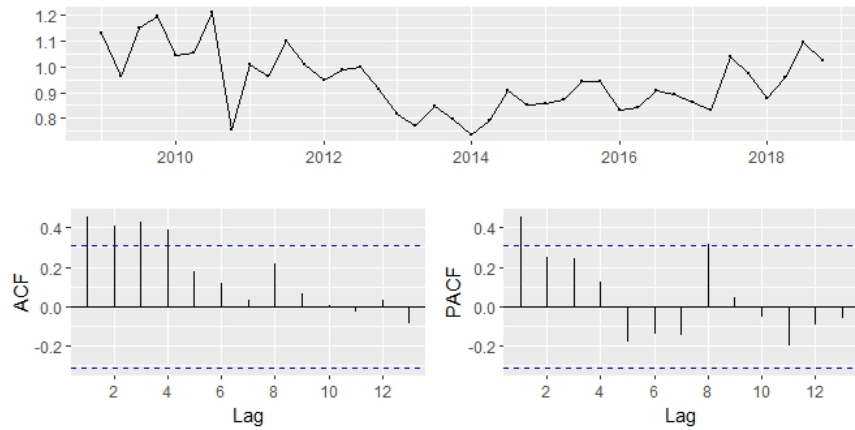
In accord with the findings of the Stationarity Tests stated in Table 5.3, the ADF test indicating a unit root, while the KPSS test proposes the process is stationary. In situations marked by limited data volume, such contradictions may occur as mentioned in other cases, too. Therefore, it is recommended to apply different models involving both integrated and non-integrated models.

Aligned with the observations presented in Figure A.7, the original dataset exhibits a singular spike in the PACF plot, while the ACF plot displays a decaying pattern. This observation suggests the suitability of an AR(1) model. In the event that the dataset undergoes regular differencing, the lagged data displays a one spike in the ACF function alongside a decaying pattern in the PACF, implying the applicability of an ARIMA(0,1,1). As alternative approaches, models such as MA(2) and ARIMA(0,1,2) can also be considered. To conclude, the potential model candidates for consideration are AR(1), ARIMA(0,1,1), MA(2), and ARIMA(0,1,2).

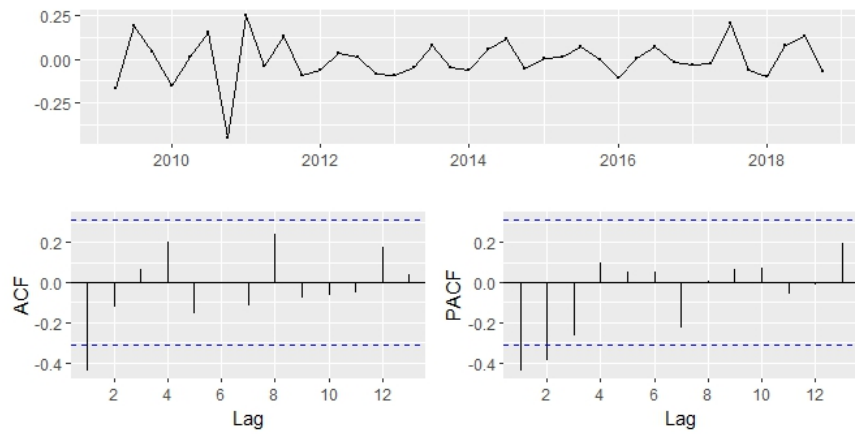
Table A.4: Comparison of Time Series Models: Land Vehicle

Fitted Models	ADF	KPSS	OCSB	SW	BP	LB	AIC	BIC
(1,0,0)(0,0,0)[4]	0.01	0.10	-1.79	0.01	0.59	0.01	-47.6	-44.2
(0,1,1)(0,0,0)[4]	0.25	0.08	-2.77	0.02	0.71	0.38	-65.0	-61.2
(0,0,2)(0,0,0)[4]	0.52	0.10	-4.20	0.52	0.74	<0.01	+39.0	+44.1
(0,1,2)(0,0,0)[4]	0.22	0.10	-2.93	0.03	0.55	0.22	-63.3	-58.3
Criteria	<0.05	>0.05	<- 1.89	>0.05	>0.05	>0.05	Smallest	Smallest

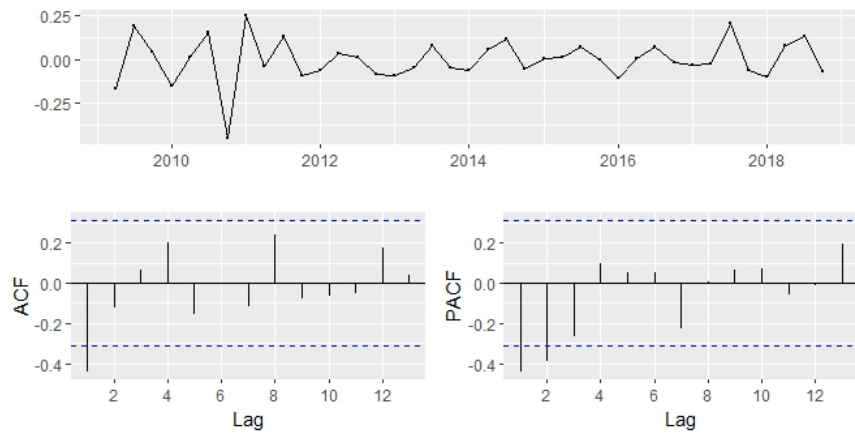
In agreement with the results showed in Table A.4, no model meets all the required conditions. Hence, an ANN model is employed for the residuals of chosen model. Nevertheless, it is important to determine the most suited model given these circumstances. The ARIMA(0,1,1) model is identified as the best fitted model based on the comparison of AIC and BIC. After model fitting, a review of the parameters is carried out, and the outcomes are displayed in Figure A.8a. Additionally, ACF and PACF plots of the model residuals are given in Figure A.8b.



(a) Original



(b) Regular Difference



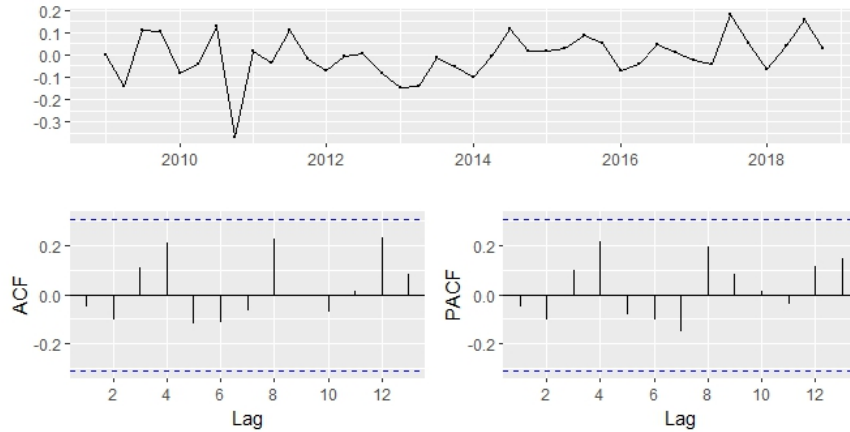
(c) Seasonal Difference

Figure A.7: Land Vehicle Insurance loss ratios and its components, 2009-2018

z test of coefficients: ARIMA(0,1,1)				
	Estimate	Std. Error	z value	Pr(> z)
ma1	-0.637	0.112	-5.67	1.5e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

(a) Coefficient Test



(b) Residuals

Figure A.8: Diagnostic checks of Land Vehicle branch

According to the results, evaluated coefficient is statistically significant, and the ACF and PACF plots shows that the model has the required characteristics. As a result, it is reasonable to assert that ARIMA(0,1,1) model is valid for the Land Vehicle dataset.

A.5 Miscellaneous

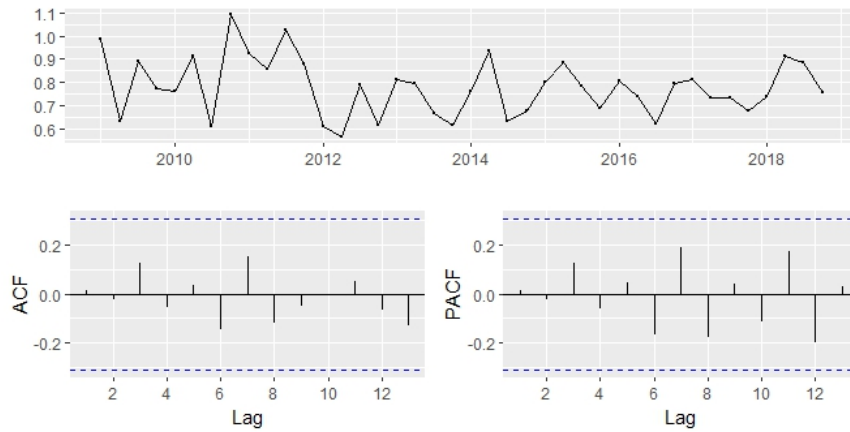
In line with the conclusions drawn from the Stationarity Tests stated in Table 5.3, the ADF test detects a unit root, while the KPSS test assumes the process is stationary. Such contradictions may occur as mentioned in other cases. Hence, it is recommended to perform different models such as integrated and non-integrated ones.

making inference from the observations presented in Figure A.9, the original dataset displays no spikes in either ACF or PACF plots. Thus, a random walk model can be an alternative for this LoB. On the contrary, if it is assumed that the process is not stationary and need differencing, then an ARIMA(0,1,1) model is proposed by investigating from the plots of regular differenced data. It has one spike in the ACF function while a decaying pattern occurring in the PACF. As an alternative, ARIMA(2,1,0) can be considered. To sum up, the possible models in consideration are ARIMA(0,0,0), ARIMA(0,1,1), and ARIMA(2,1,0).

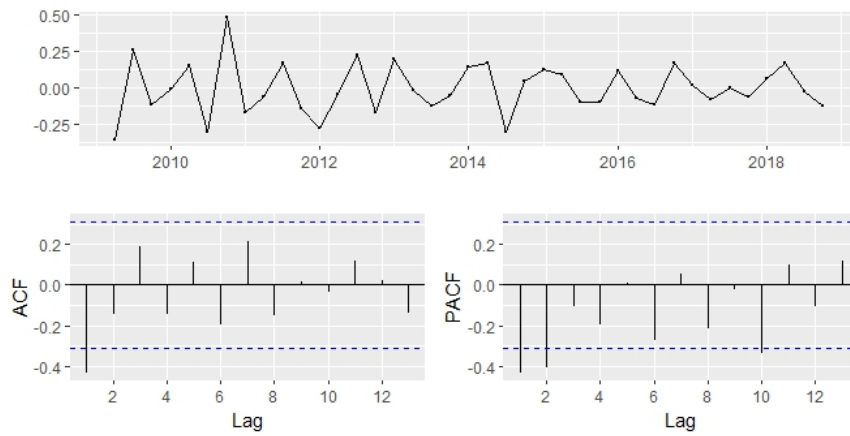
Table A.5: Comparison of Time Series Models: Miscellaneous

Fitted Models	ADF	KPSS	OCSB	SW	BP	LB	AIC	BIC
(0,0,0)(0,0,0)[4]	0.27	0.10	-4.67	0.37	0.07	0.86	-49.6	-46.2
(0,1,1)(0,0,0)[4]	0.23	0.08	-2.72	0.70	0.92	0.72	-43.7	-40.4
(2,1,0)(0,0,0)[4]	0.02	0.10	-5.34	0.47	0.81	0.34	-37.2	-32.2
Criteria	<0.05	>0.05	<- 1.89	>0.05	>0.05	>0.05	Smallest	Smallest

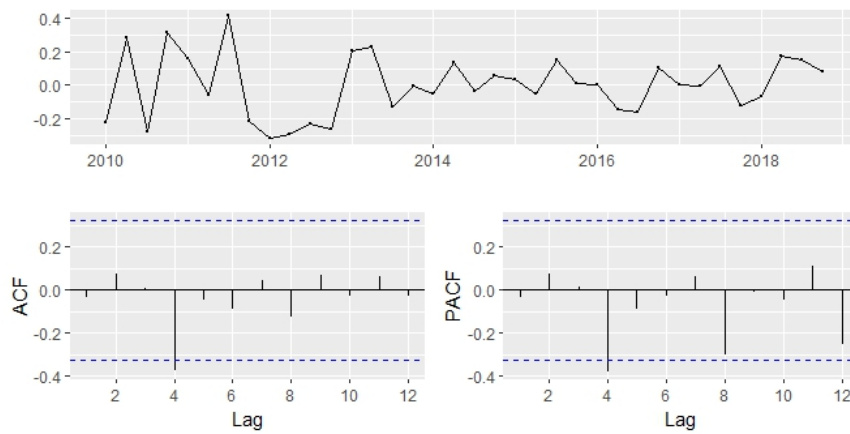
In alignment with the outcomes exhibited in Table A.5, the ARIMA (2,1,0) model fulfills all requisite criteria, however, its efficacy with regard to the AIC and BIC criteria is less than optimal when contrasted with the other alternatives. After deliberating the potential of employing an ANN model to the residuals, the most proficient model, ARIMA(0,0,0) with a constant coefficient (RWWD), is selected as the most suitable fitting choice. Subsequent to model fitting, an assessment for the parameters is conducted, and the outputs are given in Figure A.10a. Moreover, ACF and PACF plots of the model residuals are displayed in Figure A.10b.



(a) Original



(b) Regular Difference



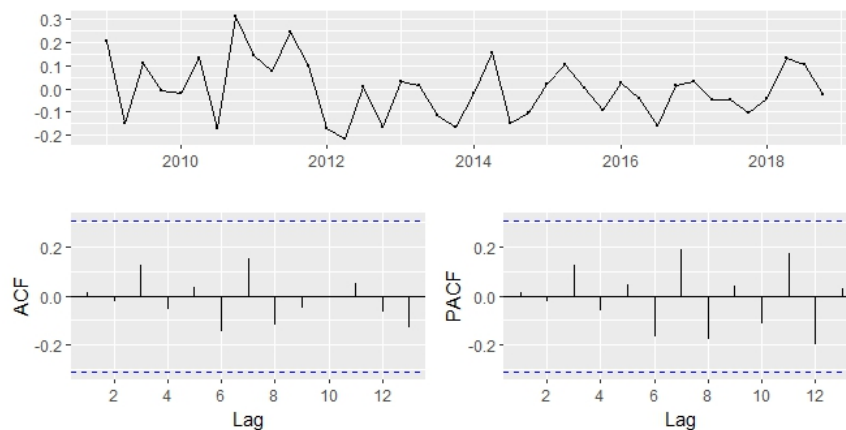
(c) Seasonal Difference

Figure A.9: Miscellaneous Insurance loss ratios and its components, 2009-2018

z test of coefficients: RWWD				
	Estimate	Std. Error	z value	Pr(> z)
intercept	0.7795	0.0196	39.8	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

(a) Coefficient Test



(b) Residuals

Figure A.10: Diagnostic checks of Miscellaneous branch

In conformity with the results, the intercept is statistically significant and the ACF and PACF plots indicates that the model has satisfactory characteristics. Accordingly, it is reasonable to deduce that ARIMA(0,0,0) is a valid model for the Miscellaneous dataset.

REFERENCES

- [1] TCMB Electronic data delivery system. URL: <https://evds2.tcmb.gov.tr> (visited on 01/08/2023).
- [2] TSB Financial and technical statements. URL: <https://tsb.org.tr> (visited on 01/08/2023).
- [3] S. Araichi, C. de Peretti, and L. Belkacem, Reserve modelling and the aggregation of risks using time varying copula models, *Economic Modelling*, 67, pp. 149–158, 2017.
- [4] T. Bedford and R. M. Cooke, Probability density decomposition for conditionally dependent random variables modeled by vines, *Annals of Mathematics and Artificial intelligence*, 32(1-4), pp. 245–268, 2001.
- [5] T. Bedford and R. M. Cooke, Vines—a new graphical model for dependent random variables, *The Annals of Statistics*, 30(4), pp. 1031–1068, 2002.
- [6] G. E. Box and D. R. Cox, An analysis of transformations, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 26(2), pp. 211–243, 1964.
- [7] E. C. Brechmann, C. Czado, and K. Aas, Truncated regular vines in high dimensions with application to financial data, *Canadian journal of statistics*, 40(1), pp. 68–85, 2012.
- [8] CEIOPS, QIS3 calibration of the underwriting risk, market risk and mcr, 2007.
- [9] R. Cont and P. Tankov, *Financial modelling with jump processes*, Chapman & Hall, 2004.
- [10] C. Czado and T. Nagler, Vine copula based modeling, *Annual Review of Statistics and Its Application*, 9, pp. 453–477, 2022.
- [11] J. Dissmann, E. C. Brechmann, C. Czado, and D. Kurowicka, Selecting and estimating regular vine copulae and application to financial returns, *Computational Statistics & Data Analysis*, 59, pp. 52–69, 2013.
- [12] EIOPA, Solvency II directive (directive 138/2009/ec), 2009.
- [13] M. Eling and K. Jung, Risk aggregation in non-life insurance: Standard models vs. internal models, *Insurance: Mathematics and Economics*, 95, pp. 183–198, 2020.

- [14] P. Embrechts, F. Lindskog, and A. McNeil, Modelling dependence with copulas, Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich, 2001.
- [15] E. W. Frees and E. A. Valdez, Understanding relationships using copulas, North American actuarial journal, 2(1), pp. 1–25, 1998.
- [16] S. Haug, C. Klüppelberg, and L. Peng, Statistical models and methods for dependence in insurance data, Journal of the Korean Statistical Society, 40(2), pp. 125–139, 2011.
- [17] M. Höbek, *Assessment of Solvency II requirements for Turkish insurance market*, Master’s thesis, Middle East Technical University, 2016.
- [18] R. J. Hyndman and G. Caceress, Neural network time series forecasts, nnetar {forecast}. URL: <https://search.r-project.org/cran/refmans/forecast/html/nnetar.html> (visited on 01/08/2023).
- [19] C. İbiş and N. Çoban Çelikdemir, Sigorta şirketlerinde risk ve mali yeterlilik: Avrupa birliği düzenlemesi solvency ii1, 2017.
- [20] I. Ismail, *Dependence Structures and Risk Aggregation Using Copulas*, Ph.D. thesis, University of Leicester, 2018.
- [21] H. Joe, Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters, Lecture notes-monograph series, pp. 120–141, 1996.
- [22] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit, *Modern actuarial risk theory: using R*, volume 128, Springer Science & Business Media, 2008.
- [23] D. Kurowicka and R. Cooke, Completion problem with partial correlation vines, Linear algebra and its applications, 418(1), pp. 188–200, 2006.
- [24] T. Nguyen, R. D. Molinari, et al., Risk aggregation by using copulas in internal models, Journal of Mathematical Finance, 1(03), p. 50, 2011.
- [25] A. J. Patton, Modelling time-varying exchange rate dependence using the conditional copula, 2001.
- [26] A. J. Patton, A review of copula models for economic time series, Journal of Multivariate Analysis, 110, pp. 4–18, 2012.
- [27] M. Pellicchia and G. Perciaccante, The calculation of solvency capital requirement using copulas, Available at SSRN 3382556, 2019.
- [28] M. Rosenblatt, Remarks on a multivariate transformation, The annals of mathematical statistics, 23(3), pp. 470–472, 1952.

- [29] A. Sandström, Solvency II: Calibration for skewness, *Scandinavian Actuarial Journal*, 2007(2), pp. 126–134, 2007.
- [30] N. Savelli and G. P. Clemente, Hierarchical structures in the aggregation of premium risk for insurance underwriting, *Scandinavian Actuarial Journal*, 2011(3), pp. 193–213, 2011.
- [31] T. Schmidt, Coping with copulas, *Copulas-From theory to application in finance*, 3, pp. 1–34, 2007.
- [32] R. H. Shumway and D. S. Stoffer, *Time series and its applications: With R examples*, 2011.
- [33] A. Sklar, *Random variables, Joint Distribution Functions and Copulas*, 1959.
- [34] A. Sklar, Random variables, joint distribution functions, and copulas, *Kybernetika*, 9(6), pp. 449–460, 1973.
- [35] O. Sokolinskiy and D. van Dijk, Forecasting volatility with copula-based time series models, Technical report, Tinbergen Institute Discussion Paper, 2011.
- [36] E. Usta, *Risk premium estimation in MTPL insurance using copula: Turkey case*, Master’s thesis, Middle East Technical University, 2016.
- [37] T. Vatter and T. Nagler, rvinecopulib: high performance algorithms for vine copula modeling. r package, version 0.5.5.1.1. URL: <https://cran.r-project.org/package=rvinecopulib> (visited on 01/08/2023), 2020.
- [38] S. Wang, Aggregation of correlated risk portfolios: models and algorithms, in *Proceedings of the Casualty Actuarial society*, volume 85, pp. 848–939, Cite-seer, 1998.
- [39] G. P. Zhang, Time series forecasting using a hybrid arima and neural network model, *Neurocomputing*, 50, pp. 159–175, 2003.
- [40] Z. Zhao, P. Shi, and Z. Zhang, Modeling multivariate time series with copula-linked univariate d-vines, *Journal of Business & Economic Statistics*, pp. 1–15, 2021.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Hasgül, Etkin

Nationality: Turkish (TC)

Date and Place of Birth: 13.07.1988, İzmir

Marital Status: Married

EDUCATION

Degree	Institution	Year of Graduation
M.S.	Institute of Applied Mathematics, METU	2015
B.S.	Department of Statistics, METU	2012
High School	Buca Anatolian High School	2006

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
6 years	Middle East Technical University	Research Assistant
2 years	Insurance Association of Turkey	Researcher/Specialist
3 years	MarshMclennan Turkey	Analytics Solutions Leader

PUBLICATIONS

Journal Publications

- **Hasgül, E.**, Selcuk-Kestel, A. S. & Yolcu-Okur, Y. (2020). Forecasting mortality rates with a general stochastic mortality trend model . Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics, 69 (1), 910-928. DOI: 10.31801/cfsuasmas.478265
- **Hasgül, E.** (2015). Modeling future mortality rates using both deterministic and stochastic approaches (Master's thesis, Middle East Technical University).

International Conferences

- **Hasgül, E.** & Selcuk-Kestel, A. S. An alternative stochastic mortality trend model. Perspectives on Actuarial Risks in Talks of Young researchers, 2017, Ascona-Switzerland.
- **Hasgül, E.** & Selcuk-Kestel, A. S. Determination of nonlinear dependency on Solvency II requirements using copulas. Dependence Modeling in Finance, Insurance, and Environmental Science, 2016, Technical University of Munich, Germany.
- **Hasgül, E.**, Selcuk-Kestel, A. S. & Yıldırak Ş. K. How effective is SCR when the association is measured with Copulas. International 9. Statistics Congress, 2015, Antalya, Türkiye.
- Basbuğ-Erkan, B.B., **Hasgül, E.** & Usman, M. Turkish Disaster Data Bank (TABB). Van International Earthquake Symposium, 2013, Van-Türkiye.

National Conferences

- **Hasgül, E.** & Selcuk-Kestel, A. S. A Useful Dependence Structure for Solvency Capital Requirements using Copulas. 3rd Ankara-Istanbul Workshop on Stochastic Processes, 2016, Middle East Technical University, Türkiye.