



UNIVERSITY OF THE WITWATERSRAND

Essays on Bankruptcy Prediction and Systemic Risk in Financial Institutions

Author:
Senyo CUDJOE

Supervisor:
Prof. Jones ODEI-MENSAH

*A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy*

at the

WITS Business School
Faculty of Commerce, Law and Management

Abstract of thesis entitled

Essays on Bankruptcy Prediction and Systemic Risk in Financial Institutions

Submitted by

Senyo CUDJOE

for the degree of Doctor of Philosophy

at The University of the Witwatersrand

in October, 2025

Financial institutions play an important role in the economic environment through economic stability, economic growth, management of risk and general development of the economy. These activities of financial institutions often expose them to risks and sometimes lead to their collapse. Financial bankruptcy causes major disruptions which usually result in a rippling effect on the economy. Accurate bankruptcy prediction is important to financial institutions, regulators, investors and central banks as it guides investment decisions and help mitigate financial losses. Prediction of bankruptcy is significant for the whole economy to help prevent financial crisis. In an interconnected global financial system, the need to understand and manage systemic risk is crucial to prevent financial crisis and maintain the economic stability economies across the world. Following the recent financial crisis of 2007-2008, systemic risk has gained prominence. The study focuses on systemic risk within the South African financial system using econometric models to identify and measure the risks within the financial system. The research provides a framework for detecting early warning signals of systemic risk. The findings from this study offer insights for policymakers, financial institutions, investors and other stakeholders to develop strategies which mitigate risks and enhance the resilience in the financial system.

The first empirical study investigates the application of state of the art ML models to predict bankruptcy in the US financial system leveraging on their ability to handle large dataset, deal with non-linear relationships and have higher predictive accuracy. We investigate the effects of different feature selection and different parameter optimization techniques on the performance of the ML models. We also implemented the SMOTE technique to mitigate the adverse effect of class imbalance on the predictive ability of the models. The thesis employs multiple models including support vector machine (SVM), online sequential extreme learning machine (OSELN), extreme gradient boosting machine (XGBoost) and light gradient boosting machine (LightGBM) using the CAMELS features. We performed experiments using financial ratios of both

failed and solvent US banks to demonstrate the effectiveness of our proposed models for bankruptcy prediction. The experimental results indicate that LightGBM outperform the rest of the state-of-the-art machine learning approaches for bankruptcy prediction although the other ML algorithms performed well on both the training and testing data. The first empirical study investigates the application of state of the art ML models to predict bankruptcy in the US financial system leveraging on their ability to handle large dataset, deal with non-linear relationships and have higher predictive accuracy. We investigate the effects of different feature selection and different parameter optimization techniques on the performance of the ML models. We also implemented the SMOTE technique to mitigate the adverse effect of class imbalance on the predictive ability of the models. The thesis employs multiple models including support vector machine (SVM), online sequential extreme learning machine (OSLM), extreme gradient boosting machine (XGBoost) and light gradient boosting machine (LightGBM) using the CAMELS features. We performed experiments using financial ratios of both failed and solvent US banks to demonstrate the effectiveness of our proposed models for bankruptcy prediction. The experimental results indicate that LightGBM outperform the rest of the state-of-the-art machine learning approaches for bankruptcy prediction although the other ML algorithms performed well on both the training and testing data.

The second empirical chapter examines the systemic risk contribution of financial institutions in South Africa to the overall risk in the financial system. The chapter utilised the VaR and conditional value at risk method based on GARCH models. This study investigates the VaR and CoVaR using SGARCH, EGARCH and GJRGARCH and compares the forecasting results of each volatility model. The distributions used in the analysis are the Student-t, Skewed Student-t and generalised error distribution. We use the forecast volatilities corresponding to each model to predict the VaR and CoVaR of each of the indices. We test the predictive performance of the estimated models using backtesting procedures. For each estimated model, we computed the VaR, CoVaR and the ΔCoVaR . Using a sample of daily equity prices spanning the period from February 19 2002 to May 31 2023, our results reveal that in South Africa, Nedbank is the most systemically important financial institution followed by Standard bank while Absa contributes little to the systemic risk in the South African financial system. There is the need for financial institutions, regulators, central banks and other stakeholders to implement regulatory measures such as higher capital adequacy requirements for systemically important institutions. The findings show that Nedbank is the systemically important financial institution. In addition, the CoVaR indicator provides a more comprehensive view of risk and provides valuable information for risk management and investment decisions.

The third empirical chapter evaluates and compares the performance of univariate GAS framework model combined with heavy-tailed asymmetrical distribution in estimating

time-varying joint VaR and ES to estimate one-day-ahead tail risk forecast of daily equity returns of financial institutions. We used the MAE and RMSE to rank the predictive abilities of the models. Also, we employed backtesting procedures and loss functions to evaluate the predictive accuracy of the different models. The results indicate that although all the models were adequate for tail risk forecasting, their predictive abilities were different. The findings suggest that GAS model with heavy-tailed distributions are appropriate for modelling financial returns. The best model across all financial institutions is the model with the Ald distribution at 1% and 2.5% risk levels. Results suggest that Absa was the least risky financial institution and required the least capital to absorb its losses.

The fourth and fifth chapters use a variety of GARCH and GAS models with different distributions to predict risk and uses the Model Confidence Set approach to evaluate the forecasting performance. We compare the Value-at-Risk (VaR) forecasts delivered by alternative model specifications using the Model Confidence Set (MCS). We analyse the predictive and the forecasting ability of various GAS and GARCH frameworks for the daily returns of South African financial institutions for the period May 2003 to May 2023. We further examine the impact of different model re-estimation strategies, different forecasting periods and different quantile levels on out-of-sample VaR risk prediction performance of the selected GARCH models using 12 GARCH and 7 GAS models. The findings suggest that the quantile levels, different forecasting periods and different model re-estimation strategies had little impact on the forecasting accuracy of the GARCH models. Also, the findings from the joint VaR and ES models indicate that the risk models developed using the GAS framework exhibited homogeneity, are time-invariant and percentile-independent.

Declaration

I, Senyo CUDJOE, declare that this thesis titled, Machine learning, which is submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy, represents my own work except where due acknowledgement have been made. I further declared that it has not been previously included in a thesis, dissertation, or report submitted to this University or to any other institution for a degree, diploma or other qualifications.

Signed: _____

Date: _____ October 14, 2025

This thesis is dedicated to my daughter Seyram Peace Cudjoe who brings me so much joy. I thank God for your life and surviving cancer at an early age. To my daughter, whose enduring spirit to conquer cancer is a source of strength and motivation to me to persevere against all odds. To my wife Dr. Mrs. Monica Cudjoe for your persistent patience; God richly blesses you. To my mother, my solid backbone, my shelter, and my prayer warrior, God bless you immensely. The thesis is also dedicated to the memory of my beloved grandmother who brought me up. God bless and keep you. Rest well with the Lord. Finally, this thesis is dedicated to the memory of my beloved daughter, Selikem Love Abui. Rest well with the Lord. Xede nyuie

Acknowledgements

First of all, I want to thank God Almighty for His protection, gift of life, presence in my life and guidance throughout my research journey. I am eternally grateful to the Almighty God, my sustainer, and tower of refuge, for bringing me this far and constantly reminding me of his steadfast and unconditional love and mercy. I am also grateful to my immediate family for their tolerance, love, support, and encouragement.

I am forever grateful to my supervisors and mentors, Professor Jones Odei-Mensah and Professor Gregory Farrell, for agreeing to supervise my work. I am grateful to my supervisors for their guidance and helping me become more independent. Without your continuous support and encouragement, I would not have been able to overcome all the difficulties. I have been very fortunate to work with Professor Jones Odei-Mensah and Professor Greg Farrell.

I am also indebted to Professor Imhotep Alagidede for his wise counsel, frequent visits, words of encouragement and the opportunity to distress by hosting me in his home.

It has been my great privilege to study at the Business School of the University of the Witwatersrand, as it provides an excellent study and research environment built by its dedicated faculty and staff members, to whom I want to express my sincere gratitude. I want to particularly thank Veli Mongwe, Mmabatho Leeuw and Jennifer Mgolodela for their patience, service, and dedication to duty. I would also like to thank the graduate students at the Business school for their assistance.

I also want to thank Professor Harrison Coffie for his support. A special thanks is extended to Mr. Nelson Kpogli, Mr. Mawuli Amedoda, Dr. Julius Afetsi Akaba and Mr. Ernest Addo Addotey for the constant support, and encouragement through each step of this journey.

To my colleagues and friends, Dr. Ellis Ayayee, Professor Sam Azasu, Dr. Edem Klenam, Professor Michael Bondurin, Professor Christmal Christmals, Dr. Edem Ameyibor, Dr. Bertin Zinsou, Mr. Leonard Nii Adom, Khomo, Dr. Damilola Bondurin, Dr. James Ayodele Oke, Dr. Emmanuel Ademola Ayodele, Dr. Patrick Adadzi, Mr David otoo Arthur, Dr. Herbert Abekah, Mr Haruna Alhassan and Mr Michael Dankwa osafo may the good Lord reward all of you abundantly.

I would like to take this opportunity to show my gratitude to individuals who kindly gave their time to give detailed and constructive feedback during the proposal and final design of the research.

Lastly, my family deserves endless gratitude especially my mother Margaret Dzakpasu for her unconditional love and support, Samuel Dzakpasu, Joel Dzokoto, Kofi Dzokoto.

I would like to thank my mom for her support during difficult times. I want to say thank you to my lovely daughter; Seyram Peace Cudjoe who has been so patient and understanding whilst I have been on this journey. To my lovely sister, Evelyn Cudjoe who has always been there to encourage me, I say Mawu nayra wo.

Finally I want to give special thanks to my loving wife and best friend, Dr. Mrs Monica Cudjoe for being my rock and support.

Senyo CUDJOE
University of the Witwatersrand
October 14, 2025

Contents

Abstract	i
Declaration	i
Acknowledgements	ii
List of Figures	xi
List of Tables	xv
List of Algorithms	xvii
List of Abbreviations	xix
1 Introduction	1
1.1 Background of the study	1
1.2 Problem Statement	5
1.3 Research Questions	7
1.4 Research Aims and Objectives	8
1.5 Justification and Significance of the study	8
1.6 Contribution to knowledge	9
1.7 Theoretical Framework	11
1.7.1 Statistical Learning Theory	11
1.7.2 Financial Contagion Theory	12
1.8 Delineation, Limitation and Assumptions	12
1.9 Structure of the thesis	12
2 Machine learning approach in bankruptcy prediction	15
2.1 Introduction	15
2.2 Literature review	25
2.2.1 Introduction	25
2.2.2 Definition of bankruptcy	25
2.2.3 Empirical review	26
2.2.4 Application of Statistical methods in bankruptcy prediction . . .	26

2.2.5	Application of Artificial intelligence methods in bankruptcy prediction	28
2.2.6	Gaps in literature	32
2.3	Methodology	34
2.3.1	Introduction	34
2.3.2	Data Acquisition	34
2.3.2.1	Features	35
2.3.3	Data Pre-processing	38
2.3.3.1	Handling missing data	38
2.3.3.2	Cross Validation	39
2.3.3.3	Data Partitioning	40
2.3.3.4	Fixing data imbalances	41
2.3.3.5	Synthetic Minority Oversampling Technique (SMOTE)	41
2.3.3.6	Feature Selection	42
2.3.3.7	Feature Importance with SHAP	45
2.3.4	Model hyperparameter optimization	46
2.3.4.1	Grid search	47
2.3.4.2	Bayesian Optimisation	48
2.3.5	Model Evaluation metrics	49
2.3.5.1	Confusion Matrix	49
2.3.5.2	Accuracy	51
2.3.5.3	Sensitivity	51
2.3.5.4	Specificity	51
2.3.5.5	Precision	52
2.3.5.6	F1 score	52
2.3.5.7	Mathews Correlation Coefficient	52
2.3.5.8	Cohen's Kappa	54
2.3.5.9	Receiver Operating Characteristics	54
2.3.6	Extreme Gradient Boosting (XGBoost) Machine	55
2.3.6.1	XGBoost hyperparameters	59
2.3.7	Light Gradient Boosting (LightGBM) Machine	60
2.3.7.1	LightGBM hyperparameters	64
2.3.8	Review of Extreme Learning Machine	65
2.3.9	Online Sequential Extreme Learning Machine (OSELN)	68
2.3.10	Support vector Machine	70
2.3.10.1	Optimizing parameters of SVM	72
2.4	Result and Discussion	73
2.4.1	Introduction	73
2.4.2	Empirical results	73
2.4.2.1	Descriptive statistics	73

2.4.2.2	Pairwise comparison of the financial ratios for bankrupt and solvent banks	78
2.4.3	Support Vector Machine	79
2.4.3.1	SVM-Linear kernel	80
2.4.3.2	SVM-Sigmoid Kernel	84
2.4.3.3	SVM-Polynomial kernel	88
2.4.4	Extreme Gradient Boosting Machine (XGBoost)	92
2.4.4.1	Feature Importance	96
2.4.4.2	Feature importance and SHAP Value	97
2.4.5	Light Gradient Boosting Machine (LightGBM)	101
2.4.5.1	Feature importance	105
2.4.5.2	Feature importance and SHAP Value	106
2.4.6	Online Sequential Extreme Learning Machine	111
2.4.6.1	OSELM-All variables	111
2.4.6.2	OSELM-Simulated Annealing	112
2.4.6.3	OSELM-Genetic algorithm	113
2.4.6.4	OSELM-Tabu Search	114
2.4.6.5	OSELM-Recursive Feature Elimination	114
2.4.6.6	OSELM-Stepwise regression	115
2.4.6.7	OSELM-LVQ	116
2.4.7	All model comparison	117
2.5	Discussion	117
3	Measuring Systemic Risk across banks in South Africa	123
3.1	Introduction	123
3.2	Literature Review	127
3.2.1	Introduction	127
3.2.2	Definition of systemic risk	127
3.2.3	Review of systemic risk measures	128
3.2.3.1	Stress testing measures	128
3.2.3.2	Forward looking risk measures	130
3.2.3.3	Measures of illiquidity and insolvency	131
3.2.3.4	Cross sectional measures	132
3.3	Methodology	132
3.3.1	Introduction	132
3.3.2	Data Source and description	133
3.3.3	Univariate Volatility GARCH Models	133
3.3.3.1	Standard GARCH	133
3.3.3.2	Exponential GARCH	135
3.3.3.3	GJR GARCH	136
3.3.3.4	Persistence	136

3.3.3.5	Half life	137
3.3.4	Conditional distribution	137
3.3.4.1	Normal distribution	137
3.3.4.2	Student distribution	137
3.3.4.3	The generalised error distribution	138
3.3.5	Model selection criteria	138
3.3.5.1	Akaike Information Criterion (AIC)	139
3.3.5.2	Bayesian Information Criterion (BIC)	139
3.3.5.3	Shibata Information Criterion (SIC)	139
3.3.5.4	Hannan-Quinn Information Criterion (HQIC)	140
3.3.6	Measures of Systemic Risk	140
3.3.6.1	CoVaR	140
3.3.6.2	Δ CoVaR and systemic risk contribution	140
3.4	Econometric Analysis and Results	141
3.4.1	Exploratory data analysis	141
3.4.2	Results of Marginal models	147
3.4.2.1	Model selection using Information criteria	147
3.4.2.2	Backtesting of GARCH Models	155
3.4.3	Systemic Risk Contribution	159
3.5	Conclusion and policy implication	161
4	Time-Varying Risk Analysis in South African Equity Markets	163
4.1	Introduction	163
4.2	Generalised Autoregressive Score Models	169
4.2.1	Selected distributions	171
4.2.1.1	STD, SSTD, SNORM	171
4.2.1.2	AST and AST1	173
4.2.1.3	Asymmetric Laplace distributions ALD	174
4.2.2	Backtesting methods	174
4.2.2.1	Unconditional coverage	175
4.2.2.2	Conditional coverage	176
4.2.2.3	Dynamic Quantile Test	178
4.3	Empirical Results	179
4.3.1	Univariate GAS FZL functions	179
4.3.2	MAE and RMSE ranking of univariate GAS (VaR, ES) model forecasts per distributional innovation	180
4.3.3	Backtesting and model ranking of joint (VaR, ES) models	180
4.3.4	Characteristic (VAR, ES) estimate for 1% univariate GAS models	181
4.4	Conclusion	188
5	Elicitability and risk model comparison of financial institutions	191

5.1	Introduction	191
5.2	Model Confidence Set Procedure	194
5.3	Evaluation of VaR prediction performance using GARCH	195
5.3.1	Evaluation of VaR prediction performance under different different quantiles	195
5.3.2	Evaluation of VaR prediction performance under different forecasting periods	210
5.3.3	Evaluation of VaR prediction performance under different model re-estimation strategies	223
5.3.4	GAS Empirical Results	236
5.4	Conclusion and policy implication	241
6	Summary, Conclusions and Recommendations	243
6.1	Introduction	243
6.2	Summary of findings and policy implications	243
6.2.1	Machine learning approach in bankruptcy prediction	243
6.2.2	Measuring Systemic Risk across banks in South Africa	245
6.2.3	Time-varying risk analysis in South African Equity market	246
6.2.4	Elicitability and risk model comparison of financial institutions	246
6.3	Limitations	247
6.4	Recommendations	248
6.5	Areas for further research	249
A	Graphs	251
B	Tables	255
	Bibliography	261

List of Figures

2.1	Model training and testing using K-Fold CV (Source: Devar, 2019)	40
2.2	Level wise tree growth (Source: Park et al., 2021).	61
2.3	Leaf wise tree growth (Source: Park et al., 2021).	61
2.4	ROC curves for SVM-linear on training data for different classifiers . . .	84
2.5	ROC curves for SVM-linear on test data for different classifiers	84
2.6	ROC curves for SVM-sigmoid on training data for different classifiers . .	85
2.7	ROC curves for SVM-sigmoid on test data for different classifiers	85
2.8	ROC curves for SVM-Polynomial on training data for different classifiers	89
2.9	ROC curves for SVM-Polynomial on test data for different classifiers . .	89
2.10	ROC curves for XGBoost on test data for different classifiers	95
2.11	ROC curves for XGBoost-Tabu on training and test data	95
2.12	ROC curves for XGBoost-GA on training and test data	96
2.13	ROC curves for XGBoost-LVQ on training and test data	96
2.14	Feature importance using XGBoost-Stepwise	97
2.15	Feature importance using XGBoost-Tabu	97
2.16	Feature importance using XGBoost-GA	97
2.17	Feature importance using XGBoost-LVQ	97
2.18	Mean absolute SHAP values of XGBoost - All features	98
2.19	Mean absolute SHAP values of XGBoost - Tabu	98
2.20	Mean absolute SHAP values of XGBoost - GA	99
2.21	Mean absolute SHAP values of XGBoost - LVQ	99
2.22	SHAP summary plot of XGBoost-Stepwise	99
2.23	SHAP summary plot of XGBoost-Tabu	99
2.24	SHAP summary plot of XGBoost-GA	100
2.25	SHAP summary plot of XGBoost-LVQ	100
2.26	SHAP Force plot using XGBoost-Stepwise	100
2.27	SHAP Force plot using XGboost-Tabu	100
2.28	SHAP Force plot using XGboost-GA	101
2.29	SHAP Force plot using XGboost-LVQ	101
2.30	ROC curves for LightGBM models on test data	102
2.31	ROC curves for LightGBM using all features for training and test data .	103
2.32	ROC curves for LightGBM-RFE on training and test data	103

2.33	ROC curves for LightGBM-GA on training and test data	103
2.34	ROC curves for LightGBM-LVQ on training and test data	103
2.35	Feature Importance using all variables	106
2.36	Feature Importance using RFE selected variables	106
2.37	Tree diagram of LightGBM using all features	106
2.38	SHAP Bar Graph	107
2.39	Aggregated SHAP values for top 10 features of LightGBM	108
2.40	Aggregated SHAP Plot Graph	109
2.41	SHAP values of individual samples	110
2.42	SHAP values of individual samples	110
2.43	Force Plot for LightGBM with all features	110
2.44	Force Plot for LightGBM-RFE	110
3.1	Daily stock prices of each series	143
3.2	Quantile Quantile plots of daily stock returns	143
3.3	Histogram and density plot of daily stock returns	144
3.4	Daily stock returns	146
3.5	GARCH model diagnostics for JSE	153
3.6	GARCH model diagnostics for FNB	153
3.7	GARCH model diagnostics for Standard bank	153
3.8	GARCH model diagnostics for Absa bank	153
3.9	GARCH model diagnostics for Capitec bank	154
3.10	GARCH model diagnostics for Nedbank	154
3.11	GARCH model diagnostics for Investec	154
3.12	Quantile quantile plot of standardised residuals	154
3.13	VaR backtesting plot for JSE	157
3.14	VaR backtesting plot for FNB	157
3.15	VaR backtesting plot for Standard bank	158
3.16	VaR backtesting plot for ABSA	158
3.17	VaR backtesting plot for Capitec bank	158
3.18	VaR backtesting plot for Nedbank	158
3.19	VaR backtesting plot for Investec	158
A.1	ROC curve for XGBoost with all features	251
A.2	ROC curve for XGBoost-Stepwise	251
A.3	ROC curve for XGBoost-RFE	251
A.4	ROC curve for XGBoost-SA	251
A.5	ROC curve for LightGBM-SA	252
A.6	ROC curve for LightGBM-Tabu	252
A.7	ROC curve for LightGBM-Stepwise	252
A.8	RFE	252

A.9 XGBoost-SA	252
A.10 Tree diagram of LightGBM using all features	252
A.11 Tree diagram of LightGBM-RFE	252
A.12 Tree diagram of LightGBM-SA	253
A.13 SHAP Bar Graph XGBoost-RFE	253
A.14 SHAP Bar Graph XGBoost-SA	253
A.15 SHAP Bar Graph XGBoost-Stepwise	253
A.16 Force plot using XGboost-RFE	253
A.17 Force plot using XGboost-SA	253
A.18 Force plot XGboost using all features	253
A.19 Feature Importance using GA selected features	254
A.20 Feature Importance using SA selected variables	254
A.21 Feature Importance using stepwise selected features	254
A.22 Feature Importance using Tabu selected features	254
A.23 Feature Importance using LVQ selected variables	254
A.24 Tree ensemble model	254

List of Tables

2.1	Number of failed U.S banks in the dataset	35
2.2	Financial ratios with categories	37
2.3	Features selected for each selection technique	45
2.4	Confusion matrix	50
2.5	Descriptive statistics for the financial ratios for solvent and failed banks	77
2.6	Mann-Whitney U test for failed banks years before failure	79
2.7	Predictive performance of SVM-Linear models using different feature selection methods	82
2.8	Performance of SVM-linear classifiers on training and testing datasets .	83
2.9	McNemar test for pairwise performance comparison for SVM-Linear Kernel	84
2.10	Predictive performance of SVM-Sigmoid models using different feature selection methods	86
2.11	Performance of SVM-sigmoid classifiers on training and testing datasets	87
2.12	McNemar test for pairwise performance comparison for SVM-Sigmoid Kernel	87
2.13	Predictive performance of SVM-Polynomial models using different feature selection methods	90
2.14	Performance of SVM-polynomial classifiers on training and testing datasets	91
2.15	McNemar test for pairwise performance comparison for SVM-Polynomial Kernel	91
2.16	Optimised hyperparameters for XGBoost model	93
2.17	Predictive performance of XGBoost models using different feature selection methods	94
2.18	Performance of XGBoost classifiers on training and testing datasets . . .	95
2.19	McNemar test for pairwise performance comparison using XGBoost model	101
2.20	Optimised hyperparameters for LightGBM model	101
2.21	Predictive performance of LightGBM models using different feature selection methods	104
2.22	Performance of LightGBM classifiers on training and testing datasets . .	105
2.23	McNemar test for pairwise performance comparison using LightGBM model	111

2.24	Predictive performance of OSELM using all features	112
2.25	Predictive performance of OSELM-SA	113
2.26	Predictive performance of OSELM-GA	113
2.27	Predictive performance of OSELM-Tabu search	114
2.28	Predictive performance of OSELM-RFE	115
2.29	Predictive performance of OSELM-Stepwise regression	115
2.30	Predictive performance of OSELM-LVQ	116
2.31	McNemar test for pairwise performance comparison for OSELM	117
2.32	McNemar test for pairwise performance comparison of best models . . .	117
3.1	Descriptive statistics of the daily stock prices	141
3.2	Summary statistics of daily stock returns	145
3.3	Kendall Correlation between daily stock returns	146
3.4	GARCH models using information criteria	147
3.5	GARCH models using information criteria	148
3.6	GARCH models using information criteria	149
3.7	Estimated parameters of GARCH (1,1) model results	151
3.8	Persistence and half-life volatility of GARCH models	152
3.9	Kupiec's POF and Joint test of Christoffersen backtesting results	156
3.10	Backtesting results at 1%, 2.5% and 5%	157
3.11	Systemic Risk Contributions at 5%	160
3.12	Systemic Risk Contributions at 1%	160
4.1	Christoffersen's Markov Chain representation	177
4.2	Multivariate Diebold-Mariano test of equal predictive accuracy	179
4.3	MAE and RMSE ranking of univariate GAS (VaR, ES) model forecasts per distributional innovations	183
4.4	Backtesting results of selected univariate GAS (VaR, ES) models	184
4.5	Characteristic FZL forecast values for selected distributional innovations	184
4.6	Estimated information criteria and Log-Likelihood	185
4.7	Estimated information criteria and Log-Likelihood	186
4.8	Backtest results of univariate GAS (VAR, ES) models	187
4.9	Backtest results of univariate GAS (VAR, ES) models	189
5.1	Comparison of Value-at-Risk model using the MCS approach	198
5.2	Comparison of Value-at-Risk model using the MCS approach	199
5.3	Comparison of Value-at-Risk model using the MCS approach	200
5.4	Comparison of Value-at-Risk model using the MCS approach	201
5.5	Comparison of Value-at-Risk model using the MCS approach	202
5.6	Comparison of Value-at-Risk model using the MCS approach	203
5.7	Comparison of Value-at-Risk model using the MCS approach	204
5.8	Comparison of Value-at-Risk model using the MCS approach	205

5.9	Comparison of Value-at-Risk model using the MCS approach	206
5.10	Comparison of Value-at-Risk model using the MCS approach	207
5.11	Comparison of Value-at-Risk model using the MCS approach	208
5.12	Summary of SSM composition for different quantile levels	209
5.13	Comparison of Value-at-Risk model using the MCS approach	211
5.14	Comparison of Value-at-Risk model using the MCS approach	212
5.15	Comparison of Value-at-Risk model using the MCS approach	213
5.16	Comparison of Value-at-Risk model using the MCS approach	214
5.17	Comparison of Value-at-Risk model using the MCS approach	215
5.18	Comparison of Value-at-Risk model using the MCS approach	216
5.19	Comparison of Value-at-Risk model using the MCS approach	217
5.20	Comparison of Value-at-Risk model using the MCS approach	218
5.21	Comparison of Value-at-Risk model using the MCS approach	219
5.22	Comparison of Value-at-Risk model using the MCS approach	220
5.23	Comparison of Value-at-Risk model using the MCS approach	221
5.24	Summary of SSM composition for different forecasting periods	222
5.25	Comparison of Value-at-Risk model using the MCS approach	224
5.26	Comparison of Value-at-Risk model using the MCS approach	225
5.27	Comparison of Value-at-Risk model using the MCS approach	226
5.28	Comparison of Value-at-Risk model using the MCS approach	227
5.29	Comparison of Value-at-Risk model using the MCS approach	228
5.30	Comparison of Value-at-Risk model using the MCS approach	229
5.31	Comparison of Value-at-Risk model using the MCS approach	230
5.32	Comparison of Value-at-Risk model using the MCS approach	231
5.33	Comparison of Value-at-Risk model using the MCS approach	232
5.34	Comparison of Value-at-Risk model using the MCS approach	233
5.35	Comparison of Value-at-Risk model using the MCS approach	234
5.36	Summary of SSM composition for different model re-estimation strategies	235
5.37	SSM of univariate GAS (VaR, ES) model forecasts	238
5.38	SSM of univariate GAS (VaR, ES) model forecasts	239
5.39	SSM of univariate GAS (VaR, ES) model forecasts	240
B.1	Backtest Results of Univariate GAS (VAR, ES) models at 1%	256
B.2	Backtest Results of Univariate GAS (VAR, ES) models at 1%	257
B.3	Backtest results of univariate GAS (VAR, ES) models testing it at 2.5% . .	258
B.4	Backtest results of univariate GAS (VAR, ES) models testing it at 2.5% . .	259

List of Abbreviations

<u>ABBREVIATION</u>	<u>MEANING</u>
ACD	Autoregressive Conditional Density
AI	Artificial Intelligence
AIC	Akaike Information Criterion
ALD	Asymmetric Laplace Distribution
AUC	Area Under the Curve
AR	Auto Regressive
ARCH	Autoregressive Conditional Heteroscedasticity
AIG	American International Group
ANN	Artificial Neural Network
AST	Asymmetric Student-t with two tail decay parameters
AST1	Asymmetric Student-t with one tail decay parameters
BCBS	Basel Committee on Banking Supervision
BIC	Bayesian Information Criterion
BO	Bayesian Optimization
BPNN	Back Propagation Neural Network
CAMEL	Capital adequacy Asset quality Management competency Earning Liquidity
CARTS	Classification And Regression Trees
CC	Conditional Coverage
CCA	Contingent Claims Analysis
CDF	Cumulative Distribution Function
CFZL	Characteristic Fissler and Ziegel Loss
CNN	Convolutional Neural Networks
CoVaR	Conditional Value at Risk
CV	Cross Validation
DA	Discriminant Analysis
DCC	Dynamic Conditional Correlation
DCS	Dynamic Conditional Scoring
DF	Dickey Fuller
DIP	Distress Insurance Premium
DT	Decision Trees
DQ	Dynamic Quantile
EFB	Exclusive Feature Bundling

EI	Expected Improvement
ELM	Extreme Learning Machine
EPA	Equal Predictive Accuracy
ES	Expected Shortfall
EWS	Early Warning Systems
EWR	Early Warning Ratio
FDIC	Federal Deposit Insurance Corporation
FN	False Negative
FNB	First National Bank
FP	False Positive
FPR	False Positive Rate
FZL	Fissler Ziegel Loss
GA	Genetic Algorithm
GARCH	Generalised Autoregressive Conditional Heteroscedasticity
GAS	Generalised Autoregressive Score
GBDT	Gradient Boosting Decision Tree
GBM	Gradient Boosting Machine
GDP	Gross Domestic Product
GED	Generalised Error Distribution
GFC	Global Financial Crisis
GOSS	Gradient-based One Side Sampling
GS	Grid Search
HQIC	Hannan Quinnne Information Criterion
IEWS	Integrated Early Warning System
IMF	International Monetary Fund
JB	Jarque Bera
JSE	Johannesburg Stock Exchange
KPSS	Kwiatkowski Phillips Schmidt Shin
LCFI	Large Complex Financial Institutions
LM	Lagrange Multiplier
LR	Likelihood Ratio
LRMES	Long Run Marginal Expected Shortfall
LVQ	Learning Vector Quantization
LightGBM	Light Gradient Boosting Machine
MAE	Mean Absolute Error
MDA	Multiple Discriminant Analysis
MCC	Mathews Correlation Coefficient
MCS	Model Confidence Set
MES	Marginal Eexpected Shortfall
MICE	Multivariate Imputation Chained Equations
ML	Machine Learning

MLP	Multilayer Perceptron
MDM	Modified Diebold Mariano
MSE	Mean Squared Error
OPT	Option Pricing Theory
OSELM	Online Sequential Extreme Learning Machine
PCA	Principal Component Analysis
PDF	Probability Density Function
PCNN	Principal Component Neural Network
POF	Proportion Of Failures
PPV	Positive Predictive Value
PVC	Parallel Voting
PSO	Particle Swarm Optimisation
QL	Quantile Loss
QP	Quadratic Programming
QQ	Quantile Quantile
RBF	Radial Basis Function
RE	Retained Earnings
RFE	Recursive Feature Elimination
RMSE	Root Mean Squared Error
ROA	Return On Average
ROC	Receiver Operating Characteristics
RS	Random Search
SA	South Africa
SA	Simulated Annealing
SES	Systemic Expected Shortfall
SHAP	Shapley Additive ExPlanation
SIC	Schwarz Information Criterion
SIFI	Systemically Important Financial Institution
SLT	Statistical Learning Theory
SLFN	Single Layer Feedforward Network
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SQL	Structured Query Language
SMOTE	Synthetic Minority Oversampling TEchnique
SNORM	Skewed Gaussian
STD	Student-t Distribution
SSTD	Skewed Student-t Distribution
SSM	Superior Set Model
SRM	Structural Risk Minimisation
TA	Total Assets
TPE	Tree Parzen Estimator

TSE	Tehran Stock Exchange
TP	True Positive
TN	True Negative
TPR	True Positive Rate
TNR	True Negative Rate
UC	Unconditional Coverage
US	United States
VaR	Value-at- Risk
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting

Chapter 1

Introduction

1.1 Background of the study

Financial organisations play a significant role in the economy of any country, acting as mediators between parties that need to borrow and parties willing to lend or invest (Acharya et al., 2017). These financial organisations include commercial banks, investment banks, insurance institutions, and in some cases, the non-banking sector, such as savings and loans corporations, shadow banks, or non-bank financial intermediaries. It is therefore without doubt that these financial institutions are central to the economic performance of any country. According to Boukhatem and Moussa (2018), banks are essential to the smooth operation of any economy because they provide the markets with the necessary liquidity and support economic growth. Advances in technology and computing have enabled innovations in finance and effective integration between local and international financial markets (Silva et al., 2017). It is essential to emphasise that the integration of financial technology and the financial sector in the current wave of technological and industrial revolution increases the complexity of the financial interconnectedness (Barroso and Laborda, 2022; Chen, Wu, and Yang, 2019), made the financial system more interconnected and complex (Barroso and Laborda, 2022; Claessens et al., 2018). The development of advanced technologies such as blockchain (Martino et al., 2020), artificial intelligence (Buchanan, 2019; Helbing, 2018) and big data (Barr et al., 2018) led to further integration of the financial system.

The links within this network may be made through indirect routes like overlapping portfolios, derivatives markets, portfolio correlation or direct channels like interbank debt interconnections. This integration of financial institutions leads to economic growth. Financial integration can enhance economic growth through several mechanisms such as capital flows, improved financial efficiency, technology transfer and better risk sharing between developed and developing countries. The growth of economies largely depends on financial integration, especially portfolio investments. The integration of financial markets globally has delivered gains (better governance, risk sharing, higher economic growth and efficient capital allocation) in the allocation of real resources and

the rate of economic growth (Acharya et al., 2009). Integrated financial markets support economic stability, resource allocation, and risk reduction (Nassani et al., 2025). Financial integration facilitates capital flows, which in turn improve investment and economic growth in developing countries (Qamruzzaman, 2023). The integration of monetary systems accelerates the transmission of new ideas and technologies, stimulates investment, and boosts innovation and production (Zhou et al., 2023). This integration cushions the financial institutions against financial instability, resulting in economic and financial resilience.

However, these gains do not come without significant risks. While the nexus between financial integration and economic growth enhances economic growth, it can also expose the vulnerabilities in the system, leading to a financial crisis. Financial integration distorts growth as countries with less developed institutions are susceptible to economic shocks. These interdependencies among financial markets can lead to cascading defaults and failures, and also come at the cost of increased fragility in the system. This integration has, therefore, made the financial system more connected than before, sophisticated and complex, leading to increased risk over time (Silva et al., 2017). These linkages have the potential to introduce channels for contagion. These channels of contagion become especially important when early losses at one bank spread via these connections, impacting other economic sectors and culminating in a financial crisis, like the 2008 financial crisis. The connectedness between financial institutions enables the propagation of risk from one institution to another. The global financial crisis that occurred between 2007 and 2009 provided evidence of how rapidly distress may propagate across the financial system and jeopardise stability (Brunnermeier, 2009). Consequently, shocks experienced by a given financial institution can propagate to other institutions within the network of institutions, potentially increasing rapidly into a systemic crisis. The failures of these interconnected financial institutions affect the economy as a whole. The integration of the financial markets poses risks to the fragile system and institutions with weak supervision and regulatory structures. Consequently, the difficult task of regulating and controlling the threat arising from this industry falls on policy-makers and regulators. The threat has grown over the past 20 years due to the deep integration and liberalisation of financial markets, which have created a highly linked and complex global financial system. The decline in the Hong Kong stock market in 1997 had effects on South and North America, Europe and African markets as a result of interconnectedness (Forbes and Rigobon, 2002). Also, the crash of the US market in 1987 had a significant impact on major stock markets globally. The empirical literature on shock propagation is extensive (Claessens, Dornbusch, and Park, 2001; Claessens and Forbes, 2001). There is still no agreement among researchers and regulators on a definition of systemic risk. Different researchers have their own definitions of systemic risk. According to Acharya et al. (2017), systemic risk is widespread failures of financial institutions or freezing up of capital markets that can reduce the supply of

critical intermediaries. According to Billio et al. (2012), systemic risk arises as a result of circumstances that endanger the trust and stability in the financial system. Financial institutions are referred to as systemically important whenever their failure to meet their responsibilities to customers and creditors has serious adverse consequences for the financial system (Brownlees and Engle, 2017). Thus, much of the recent research regarding systemic risk has been focused on these institutions, which are referred to as Systemically Important Financial Institutions (SIFIs).

The financial crisis of 2007 to 2009 saw the collapse of large financial institutions, which in turn challenged the stability of the financial market and the economy. Losses in the US home market signalled the start of the 2008 financial crisis. These losses sparked a subprime mortgage market crisis and caused financial instruments like mortgage-backed securities to lose value rapidly. An acute liquidity crisis struck banks heavily exposed to these assets, and the financial industry as a whole was rocked by the September 15, 2008, failure of investment bank Lehman Brothers. People suffered greatly as a result, and enormous losses were sustained in all areas of the economy, including employment, income, stock prices, and property values. American International Group (AIG), Bear Stearns, Citigroup, Fannie Mae, and Freddie Mac were just a few of the failed financial institutions that required large-scale intervention from the United States (US) government. These failures had an enormous impact on the global economy. From late 2008 through 2009, stock markets fell by 42% in the US, 49% in Europe, 35% in Japan, and 50% in Latin America on a dollar-adjusted basis (Acharya et al., 2012). The global gross domestic product (GDP) fell by 0.8% with a sharp decline in advanced economies of 3.2%. Also, international trade fell by almost 12% (Acharya et al., 2012).

Given the negative impact of systemic risk on the global economy, central banks, financial institutions, regulators, and academia must monitor any underlying risks to avoid the spread of a financial crisis to other sectors of the economy. The Basel II was published by the Basel Committee on Banking Supervision (BCBS) to advance international standards, which seek to maintain economic stability from 2001 to 2004. After the recent financial crisis, Basel III was introduced to develop preventive policies, enhance microprudential regulation of Basel II and also seek to preserve financial stability and regulate systemic risk (Board, 2012). Numerous crises have struck the global financial system in recent decades, each having a different impact on the actual economy of developed and emerging nations as well as the financial sector. Two crucial realities have been brought to light by the crisis, which led to the downfall of major financial institutions. First, it demonstrated the global interconnectedness of financial institutions, demonstrating how shocks from one institution may quickly spread to others and hurt the financial system. Secondly, it has brought attention to how susceptible financial institutions are to shocks. The 2007-2009 financial crisis has drawn a sharp increase in the academic research which has used network models to study financial systems and to

propose ways to preserve financial stability (Chen, Zhang, and Li, 2016; Elliott, Golub, and Jackson, 2014; Gai and Kapadia, 2010; Nier et al., 2007). The 2008 financial crisis, often attributed in part to contagion, has rekindled interest in the role of complex economic, financial or social interlinkages as channels for propagation and amplification of shocks (Acemoglu, Ozdaglar, and Tahbaz-Salehi, 2015). This contagion gives rise to a phenomenon referred to as systemic risk. After the financial crisis that occurred from 2007 to 2009, researchers and financial regulators are trying to understand and regulate the banking system (International Settlements, 2011; Callaghan et al., 2014; Gola and Roselli, 2009). There has been an increase in academic and government research in systemic risk as the models used to capture the effects of significant disturbances failed before the 2007-2009 crisis. In the past, regulators were much more focused on managing the risk of individual banks to safeguard themselves from the inherent risk of their investments by keeping sufficient reserves. This traditional approach of regulation neglected the links between banks. Therefore, a failure in one bank could quickly spread to the other banks connected to the failed bank and eventually affect the whole financial system. It has, therefore, become necessary to focus on systemic risk in a network of banks rather than on individual financial institutions. The need to manage systemic risk has persuaded financial regulators to think of new regulatory approaches that recognise the interconnected nature of financial institutions. The 2008 GFC has prompted the need to understand the channels which propagate financial contagion and to identify such contagion beforehand, develop measures for effective prevention.

Rose and Spiegel (2010) and Frankel and Saravelos (2012) show that the impact of the global financial crisis on the real economy was substantially less in emerging markets than in advanced economies. According to these authors, developing economies fared better throughout the same period when it came to systemic risk's effects—one of these economies, which was affected but not as much as developed countries, is South Africa. The financial sector in SA managed to withstand the effects of the GFC and the spread of systemic risk from developed countries. Financial crisis is a significant factor contributing to job loss, economic hardship, which can adversely affect the economy in general. The 2007–2009 global financial crisis provided evidence of the negative repercussions of such a system, as the failure of failing institutions sent ripple effects across the world economy that hampered international trade and stunted economic development for years to come. The failure of financial institutions has negative economic and social consequences. The effects of bank failure had underlined the importance of bankruptcy prediction models both in industry and academia. Many stakeholders, including employees, creditors, regulators and policymakers, are severely affected by bank failure. Accurate prediction of bankruptcy is critical to all stakeholders. The bankruptcy prediction models help to keep track of the financial institution's financial performance over a number of years. One of the major objectives of the Basel III regulations is to minimise risk. The early detection of risk gives financial institutions and

other stakeholders enough time to implement strategies to mitigate the risks to avoid bankruptcy. An accurate bankruptcy prediction model helps to understand the reasons behind the collapse of financial institutions, as this enables timely and preventive measures to be taken. The application of machine learning models for bankruptcy prediction has the potential to greatly reduce economic and job losses. The detection of financial crisis includes large amount of data. This dissertation focuses on the application of different machine learning models for bankruptcy prediction.

The thesis focuses on developing early warning systems using machine learning algorithms such as online extreme learning machine, support vector machine, extreme gradient boosting machine and light gradient boosting machine. In addition, the thesis uses the value-at-risk (VaR), conditional value-at-risk (CoVaR) and ΔCoVaR econometric model to rank systemically important institutions.

1.2 Problem Statement

The recent global financial crisis of 2007-2009 has renewed interest in systemic risk analysis and the need to develop early warning systems, which will be important for financial crisis prediction. The increasing interest from financial institutions, central banks and regulatory bodies has prompted researchers to investigate the key drivers of financial crisis. Numerous studies have been done on systemic risk analysis using warning systems. Some of these studies focused on currency crises, the use of panel logits models, Bayesian Model Averaging (Kaminsky, Lizondo, and Reinhart, 1998; Kumar, Moorthy, and Perraudin, 2003). Kumar and Ravi (2007) indicated the weakness in using statistical techniques due to their low predictive accuracy. This shifted the focus from statistical techniques to machine learning techniques over the last decade. Few studies have used both statistical and machine learning techniques to predict financial crises (Canbas, Cabuk, and Kilic, 2005; Le and Viviani, 2018; Martin, 1977; Sinkey Jr, 1975; Tam, 1991) among others. However, prior studies that used statistical and machine learning techniques in financial crisis and currency crisis prediction produced inconclusive results. Also, the statistical techniques often rely on linear assumptions and are unable to capture the complexity of modern financial data fully. The existing Early Warning Systems (EWS) could not predict the 2007-2009 financial crisis, thereby questioning their usefulness (Rose and Spiegel, 2010). Despite the growing research in the area of systemic risk and financial crisis prediction, the application of machine learning and deep learning techniques remains limited. The financial system, which is complex, diverse and changing regularly, requires advanced capabilities to detect, analyse and understand threats to the financial system (Flood et al., 2016). With these big data issues facing the financial sector, the use of machine learning and big data techniques to address challenges such as financial crises and bank collapses is needed. The thesis used

machine learning techniques such as online sequential Extreme Learning Machine (OS-ELM), Support Vector Machine (SVM), extreme gradient boosting machine (XGBoost) and Light gradient boosting machine (LightGBM) to predict financial crisis. The thesis also aims to provide financial institutions, regulators, central banks and academia with a comprehensive understanding of the use of machine learning algorithms in predicting financial crises.

Previous studies used a modelling approach based on generalised autoregressive conditional heteroskedasticity-dynamic conditional correlation (GARCH-DCC) (Engle, 2002; Engle, 2009) to estimate the marginal expected shortfall (MES). Even though the GARCH-DCC can capture the stylised facts of the data, it does not allow for time-varying nonlinear dependence. The generalised autoregressive score (GAS) dynamics of Creal, Koopman, and Lucas (2013) was used to estimate joint VaR and ES in this thesis. The GAS model goes beyond the first and second moments, hence can capture all the vital information needed (Mensah and Alagidede, 2017; Manguzvane and Muteba Mwamba, 2020). Also, the use of score to update the time-varying parameters improves the model fit (Creal, Koopman, and Lucas, 2013). The financial crisis has emphasised the need to predict financial crises accurately to mitigate the risk. The ability of economic models to accurately predict systemic risk depends on their ability to capture the asymmetric and nonlinear dependence of financial returns (Durante and Sempì, 2015; Nelsen, 2007).

The growth in trading activities and trading losses of financial institutions has led to banking supervisors and financial regulators proposing a technique which can anticipate the losses financial institutions are likely to incur. The VaR and ES are two of the most widely used risk measures in financial risk management. Value at Risk (VaR) has become the standard measure used by financial analysts to quantify the market risk of an asset or a portfolio (Hotta, Lucas, and Palaro, 2008). Besides, one of the main difficulties in estimating VaR is to model the dependence structure, mainly because VaR is concerned with the tail of the distribution (Hotta, Lucas, and Palaro, 2008). Artzner et al. (1999) established that VaR is not a coherent risk metric since it does not satisfy the axioms of subadditivity and convexity. The financial crisis of 2007 led to a significant number of banks becoming undercapitalised, revealing the shortfalls of the VaR measure as defined by Basel II. As a result, the Basel Committee on Banking Supervision suggested the application of Expected Shortfall (ES). However, according to Fissler and Ziegel (2016) and Nolde and Ziegel (2017), the ES is not elicitable. This implies there is no scoring function for which the expectation is minimised by the actual ES (Gneiting, 2011). Thus, both VaR and ES suffer from practical and coherent risk weakness (Fissler and Ziegel, 2016; Nolde and Ziegel, 2017). While VaR models are elicitable, ES lacks this property. These two measures have been estimated separately, which could lead to inefficiency, especially when used for risk assessment and decision-making under regulatory frameworks. However, the joint VaR and ES is elicitable and supports the agenda of Basel III to reduce regulatory arbitrage. The joint estimation of VaR and ES

is a more coherent and efficient approach. This study models tail risks in the South African equity markets with the joint (VaR, ES) model based on the FZL function. It conforms to the current Basel III regulatory framework. The joint VaR and ES offers a more accurate and coherent risk assessment and has the potential to improve the reliability of the risk forecasts.

Accurate risk assessment is fundamental to the stability of the financial system. Financial institutions, investors and other stakeholders are confronted with selecting the most appropriate risk model from a wide range of competing models. This selection of the most appropriate model is crucial because the choice of sub-optimal risk models could result in inaccurate risk assessment and potentially lead to financial loss. Several studies have used the AIC, BIC, SIC and HQIC to select or rank risk models. Other selection techniques include root mean squared error (RMSE), log-likelihood and mean absolute error (Blazsek and Hernández, 2018). These techniques do not rank all competing models concurrently as a set. Also, these measures lack a consistent scoring function to rank the models and do not provide any information about the statistically significant differences among the models. Recent testing procedures such as Reality Check White (2000), Superior Predictive Ability Hansen and Lunde (2005), Stepwise Multiple Testing Romano and Wolf (2005) and Conditional Predictive Ability Giacomini and White (2006) lack the FZL consistent scoring function. Also, Hansen (2005) proposed the SPA test, which can be used to compare the VaR forecasting accuracy of different models but requires the selection of a benchmark model. In this study, we employ the Model Confidence Set (MCS) technique of Hansen, Lunde, and Nason (2011) to construct a superior set of competing models and rank them. The MCS procedure examines equal predictive ability among the models through a sequential bootstrap process. The MCS technique helps to choose a model that best fits the data from a list of possible candidate models. The MCS technique is utilised to test whether there is a significant difference between different models in terms of producing accurate forecasts.

1.3 Research Questions

- How effective are optimised machine learning models in predicting bankruptcy?
- What is the individual risk contribution of each financial institution?
- How effective are tail risk time-varying volatility models in estimating financial risk?
- How do risk managers model, select and rank risks?

1.4 Research Aims and Objectives

There have been many studies on systemic risk in the financial sector. However, the use of machine learning techniques and big data techniques is scant. There is a need to explore the use of big data analytics and data management systems. This research will address the existing gap in the literature on this topic. The research looks at the application of machine learning techniques and big data analytics and their impact on measuring systemic risk. The objectives of this research include

- To develop and provide an evaluation of the effectiveness of optimised machine learning models in predicting bankruptcy.
- Examine financial institutions and rank them according to risk contribution using econometric models.
- To assess the performance and precision of tail risk and time varying volatility models in estimating financial risk
- To model, select and rank risk models.

1.5 Justification and Significance of the study

The financial crisis has significant social and economic costs. For these costs to be minimised or prevented, prompt measures are required to serve as early warning signals to policymakers, central banks and financial institutions. This thesis adds to the growing number of contributions on the analysis of financial crisis and the measurement of systemic risk. It uses machine learning techniques to serve as an early warning system for financial crisis prediction. The application of big data analysis and machine learning techniques in systemic risk and financial crisis prediction has become a vital issue. This has become important because of the large volume of financial data that is generated and its complexities.

The thesis provides insights for academics, financial institutions and central bank regulators on how to detect a financial crisis so it is prevented or the effects are minimised. The motivation is to obtain an efficient approach to collect financial data and extract relevant information. The extracted information was used to develop an optimal predictive model for financial crisis prediction and provide helpful information to financial institutions and regulators for enhanced supervision. Financial stability is beneficial for the economy and helps mitigate potential social losses. It also enhances economic efficiency growth. Over the years, big data and machine learning techniques have rapidly increased solutions to diverse problems in different fields. Machine learning techniques have become indispensable in most financial applications, including risk and financial crisis prediction. The latest challenge systemic risk analysts face involves working with a large volume of data from a variety of sources and types that were not previously

available. The abundance of data could help solve financial crisis prediction problems, which were previously challenging. New sets of tools and techniques are required to face the challenges posed by such large data sets. The understanding of systemic risk and its causes is evolving, and the relevance of data is gradually being recognised.

1.6 Contribution to knowledge

Although several studies have been conducted in bankruptcy prediction, the debate regarding the best or superior bankruptcy prediction model continues as new machine learning models evolve and the volume of data keeps increasing. This thesis evaluates the performance of several state-of-the-art machine learning classification models to predict bankruptcy in the US financial sector. These ML models include support vector machine (SVM), online sequential extreme learning machine (OSELM), Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM). Our study contributes to the growing literature on bankruptcy prediction in several ways. The sampling period covers the failures of banks before, during, and after the recent Global Financial Crisis (GFC). The study uses machine learning algorithms to analyse a large US financial dataset from 2007 to 2019, which is unusual in bankruptcy prediction, resulting in an extensive dataset to be used for modelling. We highlight that, to the best of our knowledge, there is no literature in the machine learning studies of bankruptcy prediction that used large volumes of US financial data with the technique used in the study. As long as economic conditions change continuously, so do the likely causes of bankruptcy, requiring the latest extensive financial data for empirical analysis.

Few studies have concentrated on the impact of the input variables (features) on prediction performance. To fill this gap, the study examines the discriminatory power of feature selection on bankruptcy prediction. To determine the ideal feature combination for bankruptcy prediction, we used both linear and nonlinear feature selection methods. Specifically, we used feature selection techniques, including simulated annealing (SA), genetic algorithm (GA), Tabu search, recursive feature elimination (RFE), step-wise selection, and learning vector quantisation (LVQ), to select the best features for bankruptcy prediction. Through the use of these feature selection methods, we were able to pinpoint the essential features that could increase model precision.

Most researchers have ignored the imbalance of property in real datasets used for bankruptcy prediction. We propose a pre-processing data approach referred to as the Synthetic Minority Oversampling Technique (SMOTE). This approach eliminates any potential bias and skewness of the class labels. The study investigates the effectiveness of the SVM approach in predicting bankruptcy using linear, sigmoid and polynomial kernels. We compared the prediction performance of the three kernels in bankruptcy prediction. Also, we applied an online sequential extreme learning machine algorithm

using single hidden-layer feedforward networks with additive or radial basis function hidden nodes to predict bankruptcy. This technique can learn data one-by-one or chunk-by-chunk with fixed or varying chunk size. Furthermore, it can discard data for which training has already been done. The thesis also applied the XGBoost algorithm to predict bankruptcy. XGBoost has the advantage of parallel processing. It is highly scalable and allows the use of a wide variety of computing environments, such as parallelisation, out-of-core computing, distributed computing for handling large models, and Cache optimisation for the efficient use of hardware. The XGBoost model can make a weak learner into a strong learner through its optimisation step for every new tree that is added. Regularisation is an important aspect of the XGBoost algorithm, as it helps in avoiding data overfitting problems. Most machine learning models suffer from low learning rates, slow computational time, and low prediction accuracy. To overcome these, a new and popular method, LightGBM, is considered in this thesis. We implemented an improved LightGBM. To find the ideal combination of the hyperparameters for LightGBM, we applied Bayesian optimisation with TPE in this research. Unlike the grid search and random search techniques, which do not learn from previous iterations to have a better search space, the BO-TPE chooses the hyperparameters using an optimisation approach based on previously obtained results. Thus, grid search and random search methods are uninformed, while BO-TPE is an informed search method. It is possible to miss better hyperparameter combinations outside the predefined search space for the uninformed search methods. The suggested method has an advantage over previous machine learning models regarding prediction speed and accuracy due to gradient-based one-sided sampling (GOSS) and Exclusive Feature bundling (EFB) optimisation.

The thesis contributes to the growing literature on systemic risk analysis by exploring the risk of the South African financial sector and ranking financial institutions according to their risk contribution to the entire financial system (Leukes and Mensah, 2019). The study leveraged GARCH models to estimate VaR, CoVaR and ΔCoVaR to identify systemically important financial institutions in South Africa to advance risk management. The South African economy has been significantly impacted in recent times during the COVID-19 pandemic. It is therefore imperative to conduct a systemic risk analysis due to the COVID-19 pandemic. The study included financial distress periods induced by the COVID-19 pandemic, unlike other studies that focused on pre-COVID-19 periods. This study helps to understand the interconnectedness and potential spillover effects within the South African financial system and offers valuable insights to regulators, financial institutions and policymakers. The findings of the study offer practical tools to stakeholders to improve financial stability.

In this thesis, we use the generalised autoregressive score (GAS) dynamics of Creal, Koopman, and Lucas (2013) to estimate the joint VaR and ES. The GAS model uses the complete density structure, which goes beyond the first and second moments and

can capture all the vital information. The use of the Fissler and Ziegel joint loss model for VaR and ES captures all the stylised facts exhibited by financial returns. The study develops rankings and determines which bank poses the biggest threat to the South African banking system.

The study adds to the growing literature on risk modelling by using the Model Confidence Set to rank a set of GARCH and GAS risk competing models. It offers risk managers the ease and ability to select a single optimal model from a set of competing models. The thesis advances model selection in financial econometrics. This technique provides a more robust approach than the traditional AIC, BIC, SIC and HQIC. The ranking of the models helps to identify a superior set of models, which is critical for managing risk and making better investment decisions. The study contributes to the literature on model selection by demonstrating the effectiveness of the MCS in ranking financial risk models.

1.7 Theoretical Framework

1.7.1 Statistical Learning Theory

The Statistical Learning Theory (SLT) by Vapnik (2013) provides the basis for machine learning models. Statistical learning theory (SLT) is an ML framework that integrates principles from both statistics and functional analysis (Alnuaimi and Albaldawi, 2024). It is primarily concerned with the challenge of deriving predictive functions from data. Statistical learning methods serve as the foundation for developing machine intelligence and hold a pivotal position in a wide range of fields. The main goal of statistical learning theory (SLT) (Bousquet, Boucheron, and Lugosi, 2003) is to provide a framework for studying the problem of inference that is of gaining knowledge, making predictions, making decisions or constructing models from a set of data.

SLT provides the theoretical basis for machine learning algorithms and serves as one of the developed branches of artificial intelligence (Von Luxburg and Schölkopf, 2011). The basic principle of Statistical Learning Theory (SLT) essentially coincides with the aims of machine learning in general, in that an algorithm is defined which learns a relationship between two sets of data. The statistical theory explains how models generalise from using training data to testing data by balancing bias and variance. SLT is related to a supervised learning framework for pattern recognition. Supervised learning is machine learning for learning a function from training data that consists of pairs of input objects and desired outputs. The theoretical underpinning of machine learning provides insight into a model's ability to learn complex patterns from large amounts of data. Machine learning models generalise from these patterns and make accurate predictions on unseen data. Learning patterns from data to predict the state of unseen data is the basis of statistical learning theory (Hastie et al., 2009). While each

model has its mathematical foundation, the core idea is to adjust the model's parameters to minimise the discrepancy between its predictions and actual outcomes, often measured through a loss function. According to Mello and Ponti (2018), other theories, such as Optimisation theory and Bayesian theory, provide theoretical underpinning for machine learning models. However, the machine learning model utilised in the study focuses on statistical theory, which deals with learning, analysis, and optimisation of machine learning algorithms Hastie et al. (2009). This framework underpins the choice of the ML models for the study.

1.7.2 Financial Contagion Theory

The financial contagion theory focuses on the spread of shocks or financial distress among interconnected financial institutions. According to Engle and Ruan (2019), investor behaviour and network play a role in amplifying the shock within the financial interconnected network. A network of interconnected financial institutions creates channels through which contagion can spread quickly (Brownlees, Hans, and Nualart, 2021). Distress in one institution can quickly affect other financial institutions. Institutions with weak systems and high debt are often more prone to contagion. Thus, the theory highlights that financial institutions are interconnected, where shocks can travel in the network, leading to contagion effects. The study relies on the contagion theory to model the risk of each financial institution and its overall contribution to the risk in the financial system.

1.8 Delineation, Limitation and Assumptions

The proposed approach does not model all the real bank's activities, thus dividends to shareholders are not considered. It also does not consider how investment decisions are made. The network structure of the banking system is assumed to be constant, although in the real world, it changes with time. Further research needs to be conducted for the non-banking financial sector.

1.9 Structure of the thesis

The completed thesis contains six chapters. Chapter 1 introduces the thesis with the research questions, significance of the study, problem statement and contributions. Chapter 2 uses state-of-the-art machine learning models to predict bankruptcy in the US financial sector using SVM, OSELM, XGBoost and LightGBM. Chapter 3 measures the systemic risk of South African financial institutions using VaR, CoVaR and Δ CoVaR. It provides the risk contribution of each financial institution to the whole financial sector and identifies the systemically important financial institutions. Chapter 4 utilises the

GAS framework to estimate the joint VaR and ES forecasts. Chapter 5 uses the MCS to rank GARCH and GAS risk models and identify a superior set of models. Lastly, chapter 6 covers the conclusion, limitations, recommendations and policy implications based on the findings.

Chapter 2

Machine learning approach in bankruptcy prediction

2.1 Introduction

Financial institutions play a significant role in the economy, mediating between parties that need to borrow and parties willing to lend or invest (Acharya et al., 2017). Undoubtedly, these financial institutions make significant contributions to the economic growth and development of any economy, primarily driven by the expansion of business activities and the provision of various services to the broader economy. Financial institutions play a crucial role in driving innovative financial developments within a country's economy by funding business and economic activities. These institutions provide services and products to the economy and are of key systemic importance. The financial sector's contribution to economic growth is crucial. Advances in technology have enabled the integration of the financial industry. Financial integration occurs when the financial markets of neighbouring, regional, or global countries are closely linked. Financial integration involves eliminating restrictions on cross-border financial operations, allowing financial institutions to operate freely, borrow directly, raise funds, and invest directly across countries without limitations (Ehigiamusoe and Lean, 2019). Financial integration can promote capital allocation, productivity, risk sharing of global consumption, and economic growth (Saafi, Mohamed, and Doudou, 2016). Additionally, it facilitates the mobilisation of resources (goods, capital) across borders, resulting in increased integration of stock markets (Song et al., 2021). According to Song et al. (2021), the integration of the financial market helps mobilise resources across borders, leading to increased integration of stock markets.

However, this interdependence among the financial markets can lead to cascading defaults and failures. According to Silva, Souza, and Tabak (2017), this integration has connected the financial system, leading to increased risk over time. Thus, the failure of one bank due to interdependence among financial institutions could cause a knock-on

effect, which can spread quickly and negatively impact other banks connected directly or indirectly to the bankrupt bank (Le and Viviani, 2018). The failure of a financial institution can have far-reaching effects, potentially disrupting the entire financial system and affecting financial services. It is, therefore, essential to detect bank failure before it occurs to minimise its impact on the economy or mitigate its effects. Any disruptions to the financial sector due to exposure to risks are detrimental to economic growth, as they reduce employment, income, and social welfare. This level of risk in the financial system can eventually lead to bankruptcy. According to Kirkos (2015), bankruptcy can be triggered by several factors, including financial mismanagement, regulatory changes, external shocks and deteriorating market conditions. The 2007-2009 global financial crisis threatened the stability of major economies and monetary markets, underscoring the need for bankruptcy prediction to ensure the financial stability of economies. The impetus of the recent global financial and economic crisis, which started in the US, was a result of the failure of a few systemically important financial institutions. According to Cox and Wang (2014), the GFC was associated with an increase in the number of bank failures, from 3 in 2007 to 25 in 2008, 140 in 2009, and 157 in 2020. The Federal Deposit Insurance Corporation (FDIC) declared more than 500 banks as failures in the United States between 2008 and 2014. According to Le and Viviani (2018), the bankruptcy of one bank can have a knock-on effect, which can spread quickly and negatively impact other banks that are directly or indirectly connected to the bankrupt bank. Bankruptcy has far-reaching consequences, including job loss, financial instability, and economic losses. It is therefore essential to detect bank failure before it occurs to mitigate its impact on the economy or reduce its effects. Thus, timely and accurate bankruptcy prediction is crucial for risk management, financial decision-making, and the overall health of any economy (Shi and Li, 2019). It has become necessary to prevent this financial crisis as early as possible through the use of early warning tools, such as bankruptcy prediction models (Pilch, 2021). Recognising these warning indicators and predicting these adverse events in advance can enable stakeholders to take preventative actions that could reduce the risk or possibly avoid bankruptcy.

The prediction of bankruptcy using machine learning techniques has been a topic of interest for over three decades. Bankruptcy prediction has been widely studied mainly after the 2008 GFC because of the impact on the economy. However, accurately classifying or predicting bankruptcy remains a significant problem. Although financial modelling has been on the rise, accurate bankruptcy prediction remains a challenge due to a lack of financial data and the influence of external factors, such as COVID-19, which led to the unexpected collapse of companies. Also, the assumption of linearity between financial ratios and the health of the financial sector by many models is another reason. Additionally, some machine learning models perform well on the training data but poorly on the test data. There is a need for changes in financial supervision to control and detect systemic risk early enough, thereby preventing the spread of risk

within the economy. Stress testing has become an integral part of banking regulation, assessing the soundness of the financial system and banks, including capital, liquidity, and other types of risks. There has been extensive research by academics and practitioners on models for bankruptcy prediction, aiming to understand the challenges and risk factors responsible for the collapse of financial institutions and to suggest policies and remedies for banking operations using statistical or machine learning techniques. Several academics and practitioners have extensively researched models for predicting bankruptcy using statistical or machine learning techniques (Balcaen and Ooghe, 2006; Kumar and Ravi, 2007; Lin, Hu, and Tsai, 2011), with machine learning methods having the best predictive power. These models aim to understand the challenges and risk factors responsible for the collapse of financial institutions and suggest policies and remedies for banking operations, utilising statistical or machine learning techniques.

Since the work of Beaver (1966) and Altman (1968), bankruptcy prediction has garnered considerable attention from academics, policymakers, and central banks worldwide. Many models have been proposed and empirically tested using various datasets, sometimes yielding conflicting results. The earliest techniques for predicting bankruptcy were primarily statistical, using logistic regression (Ohlson, 1980) or multiple discriminant analysis (Altman, 1968; Altman, Haldeman, and Narayanan, 1977). Statistical models, such as discriminant analysis (Deakin, 1972), logit and probit models (Wiginton, 1980), logistic regression (Jabeur, 2017), and others, have been used extensively in predicting bankruptcy. Altman (1968) applied discriminant analysis to bankruptcy prediction using five financial ratios. Some of these prominent techniques include logistic regression (Ohlson, 1980), neural networks (Odom and Sharda, 1990), support vector machines (Li and Sun, 2009), and multivariate discriminant analysis (Altman, Haldeman, and Narayanan, 1977; Cox and Wang, 2014; Sinkey Jr, 1975). Other methods have also been proposed and applied. For instance, Altman (1968) applied multivariate statistical techniques to classify solvent and insolvent companies using financial statement data. Beaver (1966) and Altman (1968) employed statistical models to predict firm failure using accounting ratios. Ravi et al. (2008b) presented a comprehensive review of the applications of prediction techniques to solve bankruptcy predictions of banks and firms. However, Begley, Ming, and Watts (1996) argued that the models built by Altman (1968) and Ohlson (1980) were inaccurate and suggested the need for enhancements in the modelling of default risk. They argue that using the original models on current data may introduce measurement error in the analysis and do not perform well, leading to an increase in Type I and Type II errors.

Machine learning models have been introduced to resolve the abovementioned issues using statistical models. Machine learning models overcome the shortcomings of statistical models by handling the complex nature of features and capturing the non-linear relationships between features. The machine learning models incorporate cross-validation and regularisation to reduce overfitting. The machine learning models do

not assume specific data distribution and can handle large, dimensionless datasets. The use of computational intelligence techniques (Kumar and Ravi, 2006), which include decision trees, ANN, SVM, convolutional neural networks (CNN), Adaboost ensemble, and others, has been documented in the literature as well. Artificial intelligence techniques such as ANN have been used extensively as a tool for bankruptcy prediction (Atiya, 2001; Baek and Cho, 2003; Becerra, Galvão, and Abou-Seada, 2005). The forecasts of the Back Propagation Neural Network (BPNN), one of the early techniques applied to bankruptcy prediction, were compared with other results (Kay, Titterington, Titterington, et al., 1999; Shin, Lee, and Kim, 2005). Support vector machine which is based on statistical learning theory has been used as a tool for bankruptcy prediction in literature (Kay, Titterington, Titterington, et al., 1999; Kumar and Ravi, 2006; Min, Lee, and Han, 2006; Ravi et al., 2008a; Ravi et al., 2008b; Shin, Lee, and Kim, 2005). These machine learning methods have proven to outperform traditional statistical methods, such as logistic regression, regarding prediction accuracy and error (Tsai, 2008; Tsai and Wu, 2008). Additionally, combining several classification techniques (ensemble classifiers) yields better results than single classification methods (Lin, Hu, and Tsai, 2011). According to Ravi et al. (2008b), statistical techniques are no longer preferred because of their relatively low accuracy. Statistical models have limitations in significantly improving predictive power despite being widely employed in academia and industry (Begley, Ming, and Watts, 1996). Statistical models assume a linear relationship between the input features and the target variable, although the financial ratios used for bankruptcy prediction are non-linear and complex in nature. In the real world, economic data is often skewed and exhibits leptokurtic behaviour. Violating the distributional assumptions can bias estimates and make the predictions unreliable. Thus, bankruptcy prediction involves complex, nonlinear relationships that statistical models often struggle to capture. This assumption has the potential to reduce the predictive accuracy of the models. There is a possibility of multicollinearity among the financial ratios used for bankruptcy prediction, which can decrease the reliability of the models. Also, statistical models assume specific data distributions. Multivariate discriminant analysis is based on several assumptions, including normally distributed independent variables and equal variance-covariance matrices across solvent and bankrupt banks (Balcaen and Ooghe, 2006; Charitou, Neophytou, and Charalambous, 2004). The predictive performance of statistical models deteriorates due to missing data and lack of normality of the data (Jones, Johnstone, and Wilson, 2017). Additionally, statistical models struggle to handle imbalanced data and often bias predictions towards the majority class, resulting in poor accuracy. These models create a simple linear decision boundary to classify bankrupt and solvent banks, which may not be sufficient to separate the two classes of banks, leading to Type I and Type II errors. These limitations could affect their performance and reliability. This has led to a significant increase in the attention and application of machine learning techniques over the past decade.

Machine learning (ML) models were introduced to resolve the issues above using statistical models. The application of ML for bankruptcy prediction has gained significant attention over the years and has been acknowledged in literature (Ahmed et al., 2022; Lahmiri et al., 2020; Shi and Li, 2019). Machine learning models overcome the shortcomings of statistical models by handling the complex nature of features and capturing the non-linear relationships between features. Additionally, these ML models do not assume a specific data distribution, can handle large datasets with high dimensions, and incorporate CV and regularisation to reduce overfitting. Machine learning models are not affected by multicollinearity or heteroskedasticity, which affect the performance of statistical models (Jones, Johnstone, and Wilson, 2017). These ML models are known to achieve better predictive performance than statistical models (Hsieh and Hung, 2010). Support vector machines (SVM), artificial neural networks (ANN), decision trees (DT), and ensemble approaches have all been utilised to overcome the limitations of statistical models and are more accurate than conventional statistical techniques. The use of computational intelligence techniques (Kumar and Ravi, 2006), which include decision trees, ANN, SVM, convolutional neural networks, Adaboost ensemble, and others, has been documented in the literature as well. Artificial intelligence techniques, such as ANN, have been used extensively as a tool for bankruptcy prediction (Atiya, 2001; Baek and Cho, 2003; Becerra, Galvão, and Abou-Seada, 2005). The predictions of the Back Propagation Neural Network (BPNN), one of the early techniques applied to bankruptcy prediction, were compared with other results (Kay, Titterington, Titterington, et al., 1999; Shin, Lee, and Kim, 2005). Support vector machine which is based on statistical learning theory has been used as a tool for bankruptcy prediction in literature (Kay, Titterington, Titterington, et al., 1999; Kumar and Ravi, 2006; Min, Lee, and Han, 2006; Ravi et al., 2008a; Ravi et al., 2008b; Shin, Lee, and Kim, 2005). These machine learning methods have proven to outperform traditional statistical methods, such as logistic regression, regarding prediction accuracy and error (Tsai, 2008; Tsai and Wu, 2008). Many studies have produced empirical results showing the superiority of NN over linear classifiers (Altman, Marco, and Varetto, 1994; Tam and Kiang, 1992). Additionally, combining multiple classification techniques (ensemble classifiers) yields better results than single classification methods (Lin, Hu, and Tsai, 2011).

Despite the wide use of these ML methods, they suffer from low learning rates, slow computational time, and low prediction accuracy. Although most ML models, including NN, have been used extensively in bankruptcy prediction, they act as black boxes, computationally intensive and lack interpretability (Iturriaga and Sanz, 2015). In addition, the ML models that account for nonlinear relationships fail to indicate the contribution of individual features to the overall prediction (Lohmann, Möllenhoff, and Ohliger, 2023). Additionally, many machine learning models suffer from overfitting and lack interpretability, which can be significant drawbacks. An effective

bankruptcy prediction model must capture the actual effects of the most important explanatory variables on the probability of bankruptcy while remaining transparent and interpretable. These ML models should have high accuracy and offer interpretability (Son et al., 2019). Addressing the challenges confronting machine learning models is crucial to further enhancing bankruptcy performance. Addressing these challenges also enhances the capabilities and practical applications of machine learning models. Due to the limitations of both statistical and some machine learning models applied in previous studies, this chapter investigates the application of state-of-the-art machine learning techniques to classify banks. A crucial issue in bankruptcy prediction modelling concerns the selection of variables to be included in the model. The variable selection process is essential for achieving model parsimony, accuracy, and generalisation (Veganzones and Severin, 2021). To create the most accurate model for bankruptcy prediction, it is crucial to select the optimal features to ensure optimal model performance and interpretation. According to Alaka et al. (2018), identifying the best features that can accurately detect bankruptcy requires further research to facilitate the seamless interpretation of financial statements and produce reliable results. Several studies have been conducted on dimension reduction (Brédart, Séverin, and Veganzones, 2021; Jabeur et al., 2021; Tsai et al., 2021). Previous studies focused on filter and wrapper feature selection methods. Existing methods largely excluded a combination of different feature selection methods applied simultaneously. A comparison of both linear and non-linear feature selection methods has not been used in US bankruptcy prediction models. Few studies have focused on the impact of input variables (features) on prediction performance, despite numerous studies attempting to suggest recent machine learning techniques that improve the models' prediction performance.

Although most ML models, including NN, have been used extensively in bankruptcy prediction, stakeholders are careful with their adoption (Park et al., 2021) because they are difficult to explain. According to Iturriaga and Sanz (2015), these ML models are computationally intensive, lack interpretability and act as black boxes. In addition, the ML models that account for nonlinear relationships fail to indicate the contribution of individual features to the overall prediction (Lohmann, Möllenhoff, and Ohliger, 2023). Despite the remarkable success of these ML models, they lack transparency and interpretability. Most ML models have overlooked the importance of data visualisation and the interpretability of results. An effective bankruptcy prediction model must capture the actual effects of the most significant explanatory variables on the probability of bankruptcy while remaining transparent and interpretable. These ML models should have high accuracy and offer interpretability (Son et al., 2019). A combination of ML models and data visualisation tools such as SHAP can help address this gap.

Machine learning models struggle to handle highly imbalanced datasets, with their predictive ability reducing if not resolved (Kim, Kang, and Kim, 2015). Most researchers have ignored the imbalanced data distribution in real datasets used for bankruptcy

prediction. However, the prediction performance of these models often suffers due to an imbalanced training dataset (Chowdhury, Kaisar, and Naha, 2023). The imbalanced distribution of bankruptcy datasets tends to impact the performance of the prediction models, because models trained on unbalanced data tend to be biased towards the majority class to increase the overall accuracy (Safi, Castillo, and Faris, 2022; Faris et al., 2020). When machine learning models are trained on such an imbalanced dataset (Kaisar and Chowdhury, 2022), their performance suffers during the testing phase, particularly when detecting samples from the minority class. Consequently, classification algorithms tend not to perform well when the data are heavily skewed to one class. Misclassifying a bankrupt company as non-bankrupt may result in severe financial consequences, while misclassifying a non-bankrupt company as bankrupt may lead to inconvenience, but not necessarily to significant economic losses. This cost imbalance highlights the importance of addressing imbalanced data in bankruptcy prediction (Liu, Chen, and Wang, 2022; Liu et al., 2022). Addressing the issue of unbalanced data leads to more reliable, accurate, and practical models in real-world applications, ultimately assisting stakeholders in making informed decisions and managing financial risks effectively (Papíková and Papík, 2022).

The US has recorded the highest number of bankrupt banks. As of 2023, 2456 banks were declared bankrupt by the FDIC. During the 2008-2009 GFC, the US was the net transmitter of spillovers due to their interconnectedness with the global markets. The 2008-2009 GFC exposed vulnerabilities in the US financial system, particularly in risk management. The crisis had cascading effects on economies that were connected directly and indirectly to the US financial sector. It is essential to comprehend the dynamics and factors that contributed to the widespread failure of banks to prevent similar crises in the future through improved risk assessment and enhanced regulatory frameworks. In the aftermath of the GFC, the Dodd-Frank Act and stress testing of financial institutions were introduced to ensure the stability of the financial system. Conducting research within the current regulatory framework enables the assessment of the effectiveness of these measures. Hence, the role of the US financial sector in the cross-country and cross-market transmission of shocks to other countries and the real economy is of interest to policymakers, regulators, practitioners and other stakeholders.

Although many studies have been conducted on bankruptcy prediction using machine learning techniques (Danenas and Garsva, 2015; Du Jardin, 2010), it is necessary to have new studies that utilise updated datasets and different models for bankruptcy prediction. The debate over the best and most accurate machine learning models for bankruptcy prediction will continue as new techniques are developed that incorporate additional variables. A comparison of new machine learning techniques with established ones is worthwhile, as it will enable us to make comparisons regarding their predictive abilities. Research in bankruptcy prediction has become necessary due to

ongoing economic uncertainties, financial market volatility, and advancements in AI and predictive analytics techniques. The aftermath of COVID-19 has left many financial institutions with debts and in a slow recovery mood. The cost of borrowing is increasing and could lead to bankruptcy for weak financial institutions. The growth in AI and Big Data in financial risk management enables the efficient analysis of complex, non-linear financial patterns. With the global financial distress and evolving market conditions, it is more imperative to develop more accurate, real-time, and explainable bankruptcy prediction models for better risk assessment and the prevention of future financial crises.

The motivation for the study stems from the demand for accurate, interpretable, and robust models that can predict bankruptcy in extreme and uncertain economic conditions. Over the years, financial institutions, regulators, investors, and other stakeholders have consistently faced growing challenges in identifying financially distressed institutions early enough to take proactive measures. It has become necessary and crucial to prevent this financial crisis as early as possible through the use of early warning tools, such as bankruptcy prediction models (Pilch, 2021). Bankruptcy poses a risk to individual banks, the financial system, investors, employees, creditors and the broader economy. Bankruptcy has far-reaching consequences, including job loss, financial instability, and economic losses. Therefore, timely and accurate bankruptcy prediction is crucial for risk management, financial decision-making, and the overall health of any economy (Shi and Li, 2019). The ability to predict bankruptcy using ML models serves as an early warning system, enabling stakeholders to take proactive decisions and mitigate risk. These ML models assist various stakeholders in assessing risk and making informed decisions. It also helps regulators and policymakers monitor the risk within the financial system. The ability to predict bankruptcy using ML models serves as an early warning system, enabling stakeholders to take proactive decisions and mitigate risk. These ML models assist various stakeholders in assessing risk and making informed decisions. It also helps regulators and policymakers monitor the risk within the financial system.

Recognising these warning indicators and predicting these adverse events in advance can enable stakeholders to take preventative actions that could reduce the risk or possibly avoid bankruptcy. Statistical models are not able to capture the complex and nonlinear patterns in financial data. This study used machine learning techniques to predict bankruptcy. Accurate prediction of bankruptcy is crucial for investors, regulators, creditors, and all relevant stakeholders. The ramifications of bankruptcy include economic losses, financial instability, and job losses. The bankruptcy of financial institutions has significant consequences for individuals who deposit money in the affected institutions and for the economy as a whole. The collapse of banks is detrimental to the economy and society. Bankruptcy hurts not only firms but also the personal lives of entrepreneurs (Kliestik et al., 2018) and employees (Ellul and Pagano, 2019; Graham

et al., 2023). The collapse of financial institutions can have a domino effect on the economy, generating panic in the financial market and leading to a broader financial crisis that could impact economic growth and investment (Erer and Erer, 2024; Yadav et al., 2023). Additionally, bankruptcy can lead to the loss of employment, which may pose a security risk. The ability to accurately and timely predict bankruptcy is essential for risk management and the general health of the economy (Shi and Li, 2019). Additionally, the study compares the predictive ability of SVM, online sequential extreme learning machine, extreme gradient boosting machine, and light gradient boosting machine models in bankruptcy prediction for effective financial decision-making.

Against this background, this chapter predicts bankruptcy using state-of-the-art ML models. The main objective of the study is to predict whether a financial institution will be bankrupt or solvent using ML models. This chapter develops predictive machine-learning algorithms to classify bankrupt and solvent banks, identifying the dominant features that can lead to financial institutions becoming bankrupt. The study bridges the gap by providing feature engineering and parameter optimisation techniques and tests different feature selection techniques. Our proposed methods significantly improve existing models by incorporating the importance of features using SHAP.

This chapter makes several notable contributions. The first contribution is the use of the LightGBM technique to predict bankruptcy efficiently. LightGBM is a recent machine learning technique that offers state-of-the-art performance in various fields. The suggested method has an advantage over previous machine learning models in terms of prediction speed and accuracy due to gradient-based one-sided sampling (GOSS) and Exclusive Feature bundling (EFB) optimisation. LightGBM has the potential to capture the multidimensionality of factors and build robust predictions that can identify financial institutions at risk, allowing preemptive actions to be taken and mitigating the risk. Secondly, we applied the Bayesian optimisation with TPE in this research to find the ideal combination of the hyperparameters for LightGBM. Unlike the grid search and random search techniques, which do not learn from previous iterations to improve the search space, the BO-TPE chooses hyperparameters using an optimisation approach based on previously obtained results. Thus, grid search and random search methods are uninformed, while BO-TPE is an informed search method. It is possible to miss better hyperparameter combinations outside the predefined search space for the uninformed search methods.

The third contribution is the use of Shapley additive explanations (SHAP) model from the field of Explainable Artificial Intelligence (XAI) to capture the contribution of each feature to the overall prediction, as well as the local and global behaviour of the model, to obtain the end user's trust in adopting the model. Applying SHAP to interpret bankruptcy prediction models is of great interest as it allows users to understand the decision-making process of the ML models. The integration of machine learning with

visualisation techniques, such as SHAP (Shapley Additive Explanations), offers a new approach to understanding and interpreting the factors that lead to bankruptcy. It also provides an interpretable framework that provides visual representations of feature importance. SHAP, combined with machine learning models, enhances prediction accuracy and provides visual representations of feature importance, facilitating more precise bankruptcy predictions. The contribution of each feature to the model output is captured using the Shapley value. The study explains the behaviour of ML models in predicting bankruptcy, comprehensively elucidating how input features affect the prediction.

The fourth contribution is the use of the Synthetic Minority Oversampling Technique (SMOTE) with oversampling to mitigate the issue of unbalanced samples, which can lead to biased results. To address the issue of imbalanced datasets, Chawla et al. (2002) proposed a technique known as SMOTE. SMOTE is a data oversampling technique that generates samples from the minority class to achieve a more balanced or nearly balanced training dataset (Liu, Chen, and Wang, 2022; Liu et al., 2022). This approach eliminates any potential bias and skewness of the class labels. Additionally, the study examines the discriminatory power of feature selection in bankruptcy prediction. We employed linear and non-linear feature selection methods for bankruptcy prediction to identify the optimal feature combination. Specifically, we used simulated annealing (SA), genetic algorithm (GA), Tabu search, recursive feature elimination (RFE), step-wise selection, and learning vector quantisation (LVQ) to select the best features that could increase bankruptcy prediction accuracy.

The sixth contribution of the chapter is the use of extensive data. We focus on the financial data of 327 US banks from 2007 to 2019, encompassing a diverse range of banks, resulting in a comprehensive dataset for modelling. The sampling period covers the failures of banks before, during, and after the recent Global Financial Crisis (GFC). We highlight that, to the best of our knowledge, there is no existing literature in the machine learning studies of bankruptcy prediction that utilises large volumes of US financial data with the technique employed in this study. As long as economic conditions continue to change, so do the likely causes of bankruptcy, necessitating the use of the latest extensive financial data for empirical analysis.

The study's findings will further enhance risk management practices in the financial sector by facilitating the development of early warning systems. The findings enable financial institutions and other stakeholders to detect financial crises earlier than traditional models for proactive risk management.

The findings also provide investors with the opportunity to assess the financial health of financial institutions before making any investment decisions. Early and timely bankruptcy predictions conserve scarce resources by enabling lenders, investors, and

other stakeholders to take corrective measures, thereby limiting the magnitude and impact on the broader economy. Accurate bankruptcy prediction enables legislators and policymakers to promptly and effectively address any systemic issues that may contribute to these bankruptcies. It will lead to regulatory frameworks based on forward-looking data rather than historical data. Financial regulators could also utilise the study's findings to monitor risk within the financial system and prevent the occurrence of cascading financial crises.

The rest of the study is organised as follows: Section 2 outlines the relevant literature while Section 3 provides details on methodology and data. Section 4 presents the results, analysis and discussion while Section 5 provides the conclusion.

2.2 Literature review

2.2.1 Introduction

There has been substantial work done on bankruptcy prediction for the past 50 years. The existing research focuses on statistical and machine learning techniques to detect bankruptcy. This section discusses a few of the pioneering research studies that have been conducted over the years. The study compared the performance of several statistical and machine learning techniques. Also, a critical analysis of the different statistical and machine learning models was done, evaluating the weaknesses and strengths of existing studies. This chapter detailed the various gaps in existing literature and proposed how this study filled these gaps and made a new contribution to the field of bankruptcy prediction using machine learning models. The literature review section is divided into three sections. The first section covers the definition of bankruptcy in the study. The second part dealt with the application of statistical methods in bankruptcy prediction, while the last section dealt with the application of machine learning models. Several models have been used to predict the bankruptcy of banks. Notable among them include multiple discriminant analysis (MDA) (Altman, 1968; Zhang et al., 2016), logit models (Martin, 1977; Ohlson, 1980), neural network (NN) models (Ecer, 2013; Martínez, 1996; Vellido, Lisboa, and Vaughan, 1999).

2.2.2 Definition of bankruptcy

Altman (1968) and Beaver (1966) were the first to define bankruptcy. According to Altman (1968), bankruptcy is an instance where the rate of return on the capital invested is less than the rates on similar investments. Beaver (1966) defined bankruptcy as an instance where a company defaulted on a loan obligation or missed a preferred dividend payment. There are many definitions of bankruptcy. According to Elloumi and Gueyié (2001), bankruptcy is the deterioration of a company's condition whereby it is unable to meet financial obligations.

2.2.3 Empirical review

The global financial crisis of 2008 had a significant impact on the health of banks around the world and on the soundness of individual banks. The development of an effective and accurate bankruptcy prediction model is a crucial yet challenging task, as it requires producing low prediction error rates and high prediction accuracy. Financial contagion, distress prediction and bankruptcy prediction have become a popular topic among academic researchers, central banks and policy makers around the world, especially after the impact of the 2008 GFC on the world economies. Forecasting bankruptcy has been one of the most researched areas in academic literature and also in the financial industry. Various types of techniques have been applied in the literature in an attempt to forecast or predict bankruptcies. Several methods have been used, including discriminant analysis, Cox proportional hazards models, logistic regression and other statistical methods. Recently, several other studies have attempted to use machine learning techniques to predict bankruptcy in the financial sector. The use of early warning systems in systemic risk is reviewed, and the application of machine learning techniques is thoroughly discussed. The section provides a systematic review of the existing literature on systemic risk and financial stability using statistical methods and machine learning techniques. The use of machine learning techniques for bankruptcy prediction has become popular among researchers and practitioners over the years.

2.2.4 Application of Statistical methods in bankruptcy prediction

Most of the prediction models are either statistical techniques or machine learning (Artificial intelligence) methods (Balcaen and Ooghe, 2006). The most widely used statistical techniques in bankruptcy prediction are the multiple discriminant analysis, which was first used by Altman (1968) to develop a Z model and logistic regression. Since the work of Altman (1968), who suggested using the Z-score to predict firms' default, hundreds of research articles have studied this issue. Beaver (1966) and Altman (1968) built statistical models to predict firm failure based on accounting ratios. Altman (1968) used discriminant analysis to predict the failure of firms. Sinkey Jr (1975) employed multiple discriminant analysis to predict the failure of a bank. He used MDA to identify financial characteristics which distinguish problem and non-problem banks for the years 1969 to 1972. Martin (1977) and Ohlson (1980) employed logistic regression to predict failures of firms and banks. According to Martin (1977), logistic regression has a higher percentage of correct classification than linear discriminant analysis. Frankel and Rose (1996) used a probit model to evaluate the predictive power of several indicators for studying the determinants of currency crashes. They used annual panel data from 1971 to 1992 for over a hundred developing countries and found output growth, rate of change of domestic credit, low ratio of foreign direct investment, and foreign interest rates to be associated with currency crashes. Kaminsky and Reinhart (1999) extended the work of

Frankel and Rose to include a wider set of crises, such as the signal approach. They analysed the connection between currency and banking crises and found that banking crises precede currency crises.

Jordan et al. (2010) examined 225 failed banks and 225 solvent banks for the period 2007 to 2010 using discriminant analysis and achieved an out-of-sample forecasting accuracy of 78.10%. Mayes and Stremmel (2014) examined a large number of 16188 U.S banks from 1992 to 2012 using a logit model and achieved an out-of-sample forecasting accuracy of 83%. Cleary and Hebb (2016) examined failed and solvent US banks from 2002 to 2009 using discriminant analysis. They used variables related to bank capital, loan quality and profitability to forecast bank failures in out-of-sample data. They found that loan loss reserves per loans, equity/total assets (EQ/TA), retained earnings/total assets (RE/TA), return on average assets (ROA), and cash/total assets (cash/TA) were the most important variables in predicting bank failure. The model achieved a forecasting accuracy in the 90-95% range for an out-of-sample of 191 banks during 2010 to 2011. Cox and Wang (2014) examined US bank failures using discriminant analysis and univariate t-test of mean differences during the period from 2007 to 2010. The authors developed different models with different financial variables. The best model had real estate loans, growth rate of loans, equity capital to assets, size of bank, return on assets, loan loss allowance, nonperforming loans, net charge-offs and foreclosures as predictors of failure or survival of a bank. They found that the proportion of illiquid loans and exposure to interbank funding markets were the main drivers of bank failures.

Altman (1968) used the original Z-score formula and showed its advantages by analysing five main financial and economic aspects of a firm: the liquidity, size dimensions, operating efficiency and profitability of the assets, financial leverage, as well as considering the capability of management in dealing with competitive conditions. Discriminant analysis was also used to predict bank failure (Sinkey Jr, 1975). The findings indicated that measures of banking factors such as asset composition, loan characteristics, capital adequacy, sources and uses of revenue, efficiency, and profitability are good discriminators between the two groups (problem-bank or the non-problem bank group). Martin (1977) predicted the failure of US commercial banks using 25 financial ratios of asset risk, liquidity, capital adequacy and earnings. He found out that logistic regression has a higher percentage of correctly classified than linear discriminant analysis. The studies used univariate comparisons for failed and non-failed banks and showed that higher capital/risk asset ratios made banks less likely to fail. Jordan et al. (2010) examined 225 failed banks and 225 solvent banks for the period covering 2007 to 2010 using discriminant analysis. The model achieved an out-of-sample forecasting accuracy of 78.10%. Mayes and Stremmel (2014) examined 16188 banks from the US during the period 1992 to 2012 using a logit model. Their model achieved an out-of-sample forecasting accuracy of 83%.

2.2.5 Application of Artificial intelligence methods in bankruptcy prediction

Financial crisis and distress prediction is an important tool which is utilised in the banking sector, finance and other sectors. Financial crisis prediction has become very popular in academic research, especially after the impact of the 2008 global financial crisis. The use of advanced techniques to forecast financial crises has become one of the most widely researched topics in academic literature. Models based on Altman (1968) and Ohlson (1980) have become inaccurate and are unable to predict bankruptcy in the financial sector. This has led to the use of artificial intelligence and machine learning techniques for predicting bankruptcy due to advances in technology. Machine learning techniques, which assess patterns in observations and identify features which differentiate observations into distinct groups, have been used in applied mathematics and other disciplines such as medicine (De Menezes et al., 2017; Subasi and Gursoy, 2010), education (Bernard et al., 2017) and finance (Gerlein et al., 2016). Neural networks and multi-layer perceptron neural networks have been widely used to predict bankruptcy since the early 1990s (Atiya, 2001). This technique, unlike other statistical techniques such as discriminant analysis, can capture the non-linear relationship between variables and the likelihood of bankruptcy. Artificial intelligence (AI) techniques such as ANN are mostly used for bankruptcy prediction, especially using the back-propagation algorithm of ANN (Aziz and Dar, 2006; Tseng and Hu, 2010).

The use of ANN in bankruptcy prediction has been on the rise. Various studies have demonstrated that Artificial Intelligence techniques such as Artificial neural networks (ANN) (Atiya, 2001; Baek and Cho, 2003; Becerra, Galvão, and Abou-Seada, 2005; Kay, Titterington, Titterington, et al., 1999; Koh and Tan, 1999; Odom and Sharda, 1990; Pramodh and Ravi, 2007; Purvinis, Šukys, and Virbickaitė, 2005; Ravi et al., 2008a; Shah and Murtaza, 2000; Tam and Kiang, 1992; Tsakonas et al., 2006) can serve as a useful technique for bankruptcy prediction. ANN has become a dominant modelling technique for bankruptcy prediction over the last decade. Odom and Sharda (1990) used ANN for bankruptcy for the first time, using five input variables similar to Altman's works (Altman, 1968) with one hidden layer and five nodes. The study used 65 bankrupt companies and 64 non-bankrupt companies sampled between 1975 to 1982. An MDA was used as a benchmark model and compared to the performance of the ANN. The study found that ANN performed better than MDA with an 81.81% accuracy compared to 74.28% by the MDA model. Tam (1991) used a back propagation neural network (BPNN) for bankruptcy prediction. The BPNN is more accurate in predicting bankruptcy than discriminant analysis. Sevim et al. (2014) developed three early warning systems using decision trees, neural networks and logistic regression using Turkish data. The results indicated that the artificial neural network outperformed the other methods. Also, Du Jardin (2010) used more than 500 ratios taken from approximately

200 previous papers, showing that an NN-based model that uses a set of variables selected with a criterion specifically adapted to the network leads to better results than a set chosen with criteria used in the financial literature. Pramodh and Ravi (2007) proposed a principal component neural network (PCNN) architecture for bankruptcy prediction in commercial banks. They tested it on Spanish and Turkish banks' data. Papadimitriou et al. (2013) used a sample of 300 US banks for forecasting US bank failures with SVM and obtained an out-of-sample forecasting accuracy of 76.40% using variables that refer to banks' efficiency, leverage and market appreciation in terms of goodwill and other tangibles. Gogas, Papadimitriou, and Agrapetidou (2018) used a linear decision boundary that separates failed and solvent banks using a sample of 1443 US banks. A set of feature selection techniques was used to select variables and later modelled using a support vector machine. The model achieved a 99.22% forecasting accuracy and performed better than Ohlson's score.

The use of different classifiers has proven to increase the accuracy of machine learning models. Several studies which used the ensemble approach reported high accuracy in prediction compared with single models. Specifically, combining multiple classification techniques or classifier ensembles performs far better than single classification techniques (Lin, Hu, and Tsai, 2011). A hybrid of intelligent models which combine the advantages of single models has been developed recently (Sánchez-Lasheras et al., 2012; Tsai, Hsu, and Yen, 2014; Xu et al., 2014). The result of hybrid models might be more accurate than either of the techniques used separately (Kainulainen et al., 2011). West, Dellana, and Qian (2005) applied bagging and boosting with a Multilayer Perceptron network (MLP) as the base classifier and found that ensembles were a better classifier than single models. The Adaboost ensemble is a better predictor than a neural network (Alfaro and Drehmann, 2009). Canbas, Cabuk, and Kilic (2005) proposed an integrated early warning system (IEWS) that combines DA, logit, probit and principal component analysis (PCA), which can help predict bank failure. They used PCA to detect financial components that significantly explain the changes in the financial conditions of banks and then used discriminant analysis (DA), logit and probit regression models. They combined these to form IEWS. They showed that discriminant analysis had slightly better results than probit and tobit models. They argued that this combination improves the prediction accuracy. Backpropagation neural network was the most popular technique used by most researchers in credit risk modelling in the 90s (Vellido, Lisboa, and Vaughan, 1999) and for finance (Wong, Lai, and Lam, 2000). Several other hybrid ANN models, such as the fuzzy ANN with particle swarm optimisation for parameter selection (Huang, 2008), neurofuzzy systems (Chen, Huang, and Lin, 2009), are just a few of the examples of hybrid methods adopted in literature.

Artificial intelligence techniques such as rule-based expert systems and machine learning techniques such as neural networks have been used in predicting bankruptcy successfully (Angelini, Di Tollo, and Roli, 2008; Atiya, 2001; Paliwal and Kumar, 2009).

There have been so many comparisons between traditional statistical models and AI techniques, such as NN models, resulting in mixed results (Fulmer et al., 1984; Ohlson, 1980). The choice of method is sometimes based on the availability of quality data. Related literature and other studies have shown that machine learning methods such as neural networks, support vector machines outperform traditional statistical methods such as logistic regression, discriminant analysis (Tsai, 2008). Several studies (Miguel et al., 1993; Olmeda and Fernandez, 1997; Swicegood and Clark, 2001; Wu and Wang, 2000) compared ANN and statistical techniques to predict bank failure and concluded that the ANN is a better predictor compared to the statistical techniques. The superiority of NN relative to other techniques in terms of predictive accuracy has been demonstrated (Hol, 2007; Lee and Choi, 2013; Levy-Yeyati, Martinez Peria, and Schmukler, 2010). Martínez (1996) compares back propagation methods with discriminant analysis, logit analysis and the k-nearest neighbour for a sample of Texas banks and concludes that the first set of methods outperforms. Similarly, the results of Vellido, Lisboa, and Vaughan (1999) suggest that NNs are more able to predict commercial bank failures than the logit model. According to Wu and Wang (2000), the back propagation method outperforms discriminant analysis and human judgment in predicting bank failures. Also, De Miguel et al. (1993) shows that the MLP is more successful than multivariate discriminant analysis, k-means cluster analysis, and logistic regression analysis in predicting the financial failure of Turkish banks. Boyacioglu, Kara, and Baykan (2009) evaluate four different neural network models, support vector machines and three multivariate statistical methods to the problem of predicting bank failures. Although SVMs provide satisfying prediction performance, these authors corroborate the superiority of MLP in prediction problems, given the difficulties in the selection of the kernel and the slow behaviour in the test phases. These assertions are consistent with Lee, Booth, and Alam (2005).

Le and Viviani (2018) compared traditional statistical techniques (discriminant analysis, logistic regression) to machine learning techniques (artificial neural network, support vector machines and k-nearest neighbours) on predicting the failure of 3000 US banks for 5 years. The study used 31 financial ratios and found that artificial neural network and k-nearest neighbour methods are the most accurate techniques. This suggests that machine learning techniques performed better in prediction than traditional statistical methods used in the study. Tam (1991) implemented a neural network on 118 banks (59 failed and 59 non-failed) during 1985 to 1987 and indicated that the neural network performed more effectively than discriminant analysis, K-NN and decision trees. Lee and Choi (2013) compared the prediction accuracy of neural networks to linear discriminant analysis on a sample of Korean companies and concluded that the accuracy using neural networks is greater than that of linear discriminant analysis. Vellido, Lisboa, and Vaughan (1999) indicated that ANN is a better predictor than

the logit model, while Lee and Choi (2013) compared the prediction accuracy of neural networks and linear discriminant analysis on a sample of Korean companies. The results suggest that the prediction accuracy using a neural network is greater than statistical techniques. Tam and Kiang (1992) compares discriminant analysis, Logit, k-nearest neighbour and artificial neural networks on bank failure prediction and finds that ANN outperforms the other techniques. Boyacioglu, Kara, and Baykan (2009) compared MDA, logistic regression, SVM, k-means cluster analysis and four neural network (NN) architectures for the period 1997 to 2003 to forecast Turkish bank failures. The study used 21 failed banks and 44 solvent banks, and 20 explanatory variables for the analysis. Although NN yielded the best results, SVM outperformed the majority of the other techniques with a 90.90% forecasting accuracy. Davis and Karim (2008) evaluated statistical and intelligence techniques in their analysis of banking crises by comparing logistic regression and signal extraction EWS methods. They stated that logit models perform better as a global EWS, and signal extraction is preferable as a country-specific EWS. Zhao, Sinha, and Ge (2009) compares Logit, ANN and k-NN and found ANN to be a better predictor than Logit and k-NN, while Logit performed better than k-NN.

Serrano-Cinca and Gutiérrez-Nieto (2013) compared nine different methods to predict the bankruptcy of USA banks during the financial crisis. These methods included logistic regression, linear discriminant analysis, support vector machine, k-nearest neighbour and neural network. The authors found that no technique is better than others, as performance depends on the measures chosen. Recent studies have demonstrated that other machine learning techniques, such as support vector machines (Kim and Sohn, 2010; Min and Lee, 2005; Shin, Lee, and Kim, 2005) and generalised additive models (Berg, 2007), have better predictive power than neural networks. Rafiei, Manzari, and Bostanian (2011) predicted the financial health of 180 Iranian manufacturing companies listed in the Tehran Stock Exchange (TSE) using an artificial neural network, genetic algorithm and multiple discriminant analysis using financial ratios. The study found that ANN had the highest accuracy rates for both training and holdout samples compared to the other techniques, such as the genetic algorithm and multiple discriminant analysis. Also, GA had a better accuracy rate for both training and hold-out data than MDA. Chen and Du (2009) found that artificial intelligence techniques were superior in predicting the financial distress of companies than traditional statistical models. The study found that the factor analysis approach increased classification error. Also, the ANN obtained a better prediction accuracy than the clustering approach. Ecer (2013) also compared the performance of NN and SVM for forecasting bank failures for the period 1994 to 2001. The NN yielded the optimum out-of-sample result prediction. Iturriaga and Sanz (2015) compared the performance of NN using a combination of multilayer perceptron and self-organising maps to SVM, random forest, discriminant analysis and logistic regression for forecasting US bank failures over the period 2002 to

2012. They used 386 failed and 386 solvent US banks and found that NN outperformed the SVM for the short-term (one-year) horizon, but the SVM outperformed NN for the medium and long-term horizons (two and three years before failure). The study found that the most important variables are provisions, risk concentration in the construction industry and equity support to loans. Sevim et al. (2014) developed early warning systems using artificial neural network (ANN), decision trees and logistic regression using Turkish data. The findings indicated that ANN yielded the best performance among the three models. Bankruptcy prediction often involves developing an effective model with a higher prediction accuracy. Tsai, Hsu, and Yen (2014) conducted a study to compare three classifier ensembles using a multilayer perceptron (MLP) neural network, support vector machine (SVM), and decision trees (DT) based on bagging and boosting and different classifiers. Results indicated that DT ensembles using the boosting approach performed best. Chen (2011) compared traditional statistical methods to machine learning techniques for predicting financial distress and posited that traditional techniques handled large datasets without sacrificing prediction performance, while intelligent techniques achieved better results with smaller data sets. The empirical results indicated that particle swarm optimisation (PSO) integrated with SVM was considered for predicting financial distress.

2.2.6 Gaps in literature

The literature survey presented in the earlier sections detailed two distinct methods used for bankruptcy prediction. These two distinct methods include statistical and machine learning algorithms. The early methods for bankruptcy prediction relied on traditional statistical methods such as logistic regression, discriminant analysis and probit models. For instance, Altman (1968) used the Z-score discriminant analysis to generate a score which classifies observations. One limitation of the statistical methods is the use of a relatively small sample size for the models. This could be due to the high cost of data storage in the early years of bankruptcy prediction. Also, the use of computers was not widespread during the early 1960s through to the early 1990s. The statistical methods have become inefficient due to the growing number of features needed for bankruptcy prediction. Begley, Ming, and Watts (1996) indicated the shortcomings of the statistical methods and suggested the need to develop improved models. Many statistical models assume a linear relationship between the variables. However, the financial ratios used in bankruptcy prediction exhibit non-linear characteristics. In addition, the statistical model fails to handle high-dimensional data. According to Jones, Johnstone, and Wilson (2017), the machine learning models perform better than statistical models in handling complex data. Also, machine learning models perform better than statistical models in cross-validation data (Sun et al., 2014). This study addressed the shortcomings of the statistical models by employing machine learning models to predict bankruptcy.

Machine learning moved from the use of traditional financial ratios to a broader set of features that could capture the complexities of the financial system. However, these may often result in the curse of dimensionality. This could be resolved through feature selection, feature extraction and regularisation. Although most of the machine learning models achieved higher accuracies, the use of rigorous variable selection techniques is scanty. Most of the literature failed to focus on feature selection, which is usually utilised when there is a large number of features. The integration of variable selection techniques (linear and non-linear) and machine learning models can lead to more robust bankruptcy prediction models. The use of linear and non-linear variable selection techniques, together with machine learning models, which could help uncover hidden patterns, is limited in the literature. A size-fit-all approach to variable selection is not sufficient to capture the non-linear and multifaceted nature of the risk inherent in financial data. While the machine learning models have used a large set of financial ratios, there is no standard set of ratios for bankruptcy prediction, although most studies recommend the CAMEL features.

Barboza, Kimura, and Altman (2017) in their work compared the performance of traditional statistical models to machine learning models for bankruptcy prediction. The findings indicated that machine learning models outperform statistical models. However, their study did not consider feature selection techniques. Also, the study did not consider different classification costs. The use of accuracy alone as a performance metric is not enough. This thesis addressed these limitations by using a wide range of performance metrics such as kappa, precision, F1 score and MCC. The thesis also used linear and non-linear feature selection methods. Although machine learning models attain high predictive accuracy, the models are too complex to be interpreted. One limitation of the machine learning models is their lack of interpretability. Most of the models fail to identify the contribution of each feature to the overall accuracy of the model. It is impossible to track the important features in the model. According to Son et al. (2019), the use of feature importance techniques for boosting tree models overcomes the lack of interpretability of machine learning methods. The thesis leveraged the advantage of SHAP to develop highly accurate, interpretable bankruptcy prediction models. The model meets both limitations of high accuracy and interpretability. This study explores ways of improving the existing models using machine learning techniques. Machine learning methods are considered to be among the most important of the recent advances in applied mathematics, with significant implications for classification problems (Tian, Shi, and Liu, 2012). Predicting the default risks for banks is a classic and timely issue. The availability of large amounts of historical data increases each day, and these data could be used to improve forecast accuracy.

2.3 Methodology

2.3.1 Introduction

This section describes the machine learning techniques employed in the study. We applied five supervised machine learning techniques, namely Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine, Support Vector Machine, and Online-Sequential Extreme Learning Machine. We utilised financial data from U.S. banks from 2007 to 2019 and discussed the issue of class imbalance. The variables used in the study are thoroughly discussed. Different feature selection techniques used to select the best combination of predictors of bankruptcy were discussed, as well as the performance metrics. We detail the general experimental design regarding model evaluation.

2.3.2 Data Acquisition

The data, which was used for this study, consists of failed and solvent U.S. banks from 2007 to 2019 and includes all types of banking institutions, such as commercial banks, savings associations and national banks, as categorised by the FDIC. We collected financial data on all banks from 2007 to 2019 from Banksfocus. Information on bank insolvency was gathered from the FDIC. The sample of failed banks was obtained from the failed bank list of the FDIC (<http://www.fdic.gov/bank/individual/failed/banklist.html>). The financial data for both failed and solvent banks were obtained from the BankFocus database. The data period coincides with the 2008 financial crisis. There were a total of 559 US banks that failed during the selected period of analysis, but the sample was restricted to banks with data available for a minimum of four years before their collapse. For example, if a bank failed in 2019, we included data from its financial statements for the years 2018, 2017, 2016 and 2015. Thus, for each bank, 4 observations of 33 financial ratios were collected from their financial statements over a period of four years, resulting in a total of 132 observations per bank.

We collected data from failed banks four years prior to their year of bankruptcy to identify early warning signs. Examining data four years before bank failure enabled the study to detect risk early enough for intervention and also capture financial distress indicators that gradually lead to bankruptcy. This data enables us to detect any gradual deterioration in the financial health of the financial institutions, and also observe the trends in liquidity, asset quality, profitability and capital adequacy of the financial institutions over time. Also, the use of data four years before bank failure allowed for comparison between solvent and bankrupt banks and helped to identify financial ratios which deteriorated over time. The analysis of the years prior to bankruptcy helps regulators, central banks, financial institutions and other stakeholders to fully understand what went wrong. This will assist with better supervision, risk assessment and strategic decision making. The choice of the four years before bankruptcy aligns with that of

supervisory authorities that often monitor banks over multi-year periods making the findings more relevant for policy. A robust bankruptcy prediction model should not only predict imminent bank failure but also forecast risk over a long period. This provides regulators and other stakeholders with time to mitigate the risks. Failed banks without data four years before their collapse were excluded from the sample. The solvent banks were randomly selected with the criteria of being active still and being a U.S bank during the period 2007 to 2019. Active banks were also restricted to banks with at least 3 consecutive years of financial data. Following Altman (1968), we selected the same number of solvent and insolvent banks during the period of study for the sample to have a balanced set. The final dataset includes 376 failed and 376 solvent banks, giving a total of 752 banks. Table 2.1 shows the number of failed banks in the US from 2007 to 2019 and the number included in the sample for the study.

Table 2.1: Number of failed U.S banks in the dataset

Year	Number of failed banks	Number of failed banks in our dataset
2007	3	2
2008	25	12
2009	140	89
2010	157	115
2011	92	71
2012	51	34
2013	24	17
2014	18	13
2015	8	7
2016	5	5
2017	8	7
2019	4	4
Total	535	376

2.3.2.1 Features

Some of the studies on bankruptcy prediction rely on accounting data, which are augmented in some cases with macroeconomic and market data (Agarwal and Taffler, 2008; Berger and Bouwman, 2013). However, the use of high-level constructs referred to as financial ratios instead of low-level accounting variables for bankruptcy prediction has been on the rise. According to Zhao, Sinha, and Ge (2009), the use of financial ratios in bankruptcy models rather than raw accounting variables improves the predictive performance of the models. Bankruptcy prediction models are usually designed using financial ratios from balance sheets and income statements of financial institutions. These financial ratios are usually constructed based on publicly available balance and income data that commercial banks are required to report to regulatory authorities regularly. The financial ratios, which are calculated in the same way within a regulatory

framework, usually allow for good discrimination between failed and non-failed financial institutions and are easily available (Altman, 1968). They are designed to reflect the soundness of a commercial bank in several aspects. These financial ratios are believed to be more effective explanatory variables than the raw accounting data in predicting and explaining bank failures.

A common financial indicator used by past researchers to measure bank performance is the Capital Adequacy, Asset Quality, Management Competency, Earnings Quality, and Liquidity (CAMEL) framework. According to Wirnkar and Tanko (2008), the CAMEL framework was developed by US Federal regulators in the early 1970s to help structure the bank examination process. Since then, the use of CAMEL factors in evaluating a bank's financial health has become widespread among regulators. The CAMEL rating system is a useful tool for examining the safety and soundness of banks, and for helping to mitigate the potential risk of bank failure (Dang, 2011). The CAMEL analysis is another approach for researchers and policy analysts to measure bank financial performance (Douglas, Lont, and Scott, 2014). The CAMEL analysis approach is beneficial as it is an internationally standardised rating and provides flexibility between on-site and off-site examination; hence, it is the main model for assessing bank performance.

For this study, the research instrument used is the CAMEL Model, which is a recent innovation in the area of financial performance evaluation of banks. The financial ratios used in the study are classified into the categories of the CAMEL rating framework used by the FDIC. CAMEL is an acronym for the five major characteristics of a bank's financial and operational conditions: Capital adequacy, Asset quality, Management quality/efficiency, Earnings ability/quality, and Liquidity position. According to the empirical evidence in the literature, the CAMEL framework is widely used to evaluate the financial performance of banks (Kumar et al., 2012). Capital adequacy is a measure of the level and quality of a bank's capital base. A bank's capital base is critical since it is the last line of defence against losses to uninsured depositors and general creditors. Asset quality measures the level of risk of a bank's assets. This is related to the quality and diversity of loan borrowers and their ability to repay the loans. Management quality is a measure of the quality of a bank's officers and the efficiency of its management structure. Earnings ability is a measure of the performance of a bank and the stability of its earnings stream. Liquidity measures a bank's ability to meet unforeseen deposit outflows in a short time. We deliberately chose the simple financial ratios in order to demonstrate that these ratios may significantly improve the performance of classifiers. As a result, 33 financial ratios (presented in Table 2.2) were selected as input features for bank failure prediction. This set of financial ratios can be interpreted as a full set of measures of the five CAMEL components. While the 33 financial ratios are all in relatively simple forms, they constitute important domain knowledge, which is not explicitly captured in, and cannot be automatically learned from, the raw accounting data. The identified variables were fed into the proposed models using the training

and testing data. The financial ratios are listed in Table 2.2 with their categories. Each financial ratio was labelled by X from X1 to X29 to shorten the name. The final data set for the model consists of financial ratios of 752 financial institutions in the USA. The financial ratios were standardised into z-values.

Table 2.2: Financial ratios with categories

Category	Variable	Label
Capital quality	Equity/total assets	X9
	Equity/net loans	X10
	Equity/deposit and short term funding	X11
	Equity/liabilities	X12
	Capital funds/total assets	X13
	Capital funds/net loans	X14
	Capital funds/deposit and short term funding	X15
	Capital funds/liabilities	X16
Asset quality ratio	Loan loss reserve/gross loan	X1
	Loan loss provision/net interest revenue	X2
	Loan loss reserve/non-performing loans	X3
	Non-performing loans/gross loans	X4
	Net charge off/average gross loan	X5
	Net charge off/net income before loan loss provision	X6
	Impaired loans/equity	X7
	Unreserved impaired loans/equity	X8
Management	Other operating income/average assets	X19
	Non-interest expense/average assets	X20
	Non-operating items and taxes/average assets	X22
Earnings	Net interest margin	X17
	Net interest revenue/average assets	X18
	Pre-tax operating income/average assets	X21
	Return on average assets (ROAA)	X23
	Return on average equity (ROAE)	X24
	Income net of dist/average equity	X25
	Cost to income ratio	X26
	Recurring earning power	X27
	Dividend pay out	X33
Liquidity ratios	Net loans/total asset	X28
	Net loans/deposit and short term funding	X29
	Net loans/total deposit and borrowing	X30
	Liquid assets/total deposit and borrowing	X31
	Liquid assets/deposit and short term funding	X32

2.3.3 Data Pre-processing

Data pre-processing, which includes data transformation, data cleaning and data selection, is a useful step before applying a classification or regression algorithm (Han and Kamber, 2001). These data pre-processing techniques are applied to the raw data to clean the data and make it consistent by converting data values into a specific range (Patel and Mehta, 2011). The process of scaling features into a dimensionless unit in machine learning is referred to as feature standardisation or scaling. This can be done by computing the mean and standard deviation of all features in the training data, and using z-score standardisation to transform features into a space with zero mean and unit variance. The transformed value is obtained by subtracting the mean from the original value and dividing by the standard deviation. It is done to change the feature vector into a format which is suitable for machine learning techniques. The training data was standardised (zero mean and unit variance), and the test data set was also standardised using the same standardisation parameters from the training data.

2.3.3.1 Handling missing data

Real world data usually contain missing values which adversely affect the performance of machine learning models. Missing data are values not captured for certain variables. It is a general problem in most datasets especially financial data where records could be incomplete or missing totally. Missing values occur due to refusal by respondents to answer certain questions, merging unrelated data, human error, machine error, data collection errors, incomplete information and unavailable information (Suthar, Patel, and Goswami, 2012; Houari et al., 2014; Hasan et al., 2021).

According to Kang (2013), the impact of the missing data on data analysis depends on how much data is missing, the type of missingness and the pattern of missing data. It is common to find missing data in financial data. Since the accuracy and efficiency of machine learning depends on quality data, it is important to use relevant techniques to address missing data. Handling missing values help the machine learning models to perform optimally. Handling missing values is crucial in machine learning models, especially when the variables with missingness contribute to the performance of the model. The two ways of handling missing data include deletion and imputation of values (Langkamp, Lehman, and Lemeshow, 2010). Deletion of missing data from the data set is normally adopted when the ratio of missing data is significantly small. The other approach is the process of imputation which involves replacing missing values with predicted values. Different statistical and machine learning techniques have been proposed and used to handle missing data (Alabadla et al., 2022). Hybrid methods have also been used to resolve missing data (Aydilek and Arslan, 2013; Lin et al., 2020; Choudhury and Kosorok, 2020; Al-Helali et al., 2021). In this study, we adopted the multivariate imputation by chained equations (MICE) for missing data imputation.

According to Lyngdoh et al. (2022), MICE is an iterative method where missing data is imputed based on error-minimisation. This method utilises multiple imputations and measures the uncertainty of the missing values systematically (Van Buuren and Groothuis-Oudshoorn, 2011). The MICE has the ability to handle different data types including binary and continuous data.

2.3.3.2 Cross Validation

Evaluating the performance of machine learning models requires testing against unseen data. Cross-validation is a technique used to determine the effectiveness of machine learning models, evaluate how the model performs in training and ensure that the model is not over-fitted or under-fitted (Jung et al., 2020). According to Rao and Wu (2005), CV assesses the model's capacity to classify new data that was not used in training the model, identify selection bias and overfitting. Cross-validation is a statistical technique of comparing and evaluating models by dividing the data into segments; one segment is used to train the model (training set) and the other is used to validate the model (test set) (Refaeilzadeh, Tang, Liu, et al., 2009). Of all the available CV techniques, the K-fold cross-validation has been used to assess the performance of machine learning techniques (Kumar et al., 2015). Hence we used the K-fold CV in this study because it increases model accuracy (Kohavi et al., 1995), prevents overfitting (Hastie et al., 2009), determines ideal values of hyperparameters (Bergstra and Bengio, 2012) and enhances the overall model performance. The K-CV is a commonly used method for training sets. It divides the training set into K equal-sized parts. Each part of the training set is regarded as the validation set, and the remaining $k - 1$ parts are considered as new training sets. Then, k models will be established, trained and validated with $k - 1$ training sets and k validation sets. The accuracy of K models can be obtained, and the performance of the K-CV classifier model is evaluated based on the average accuracy of the K models. Commonly used Partitions are divided into 10 partitions or 10-fold cross-validation (Gaspar, Carbonell, and Oliveira, 2012; Syarif, Prugel-Bennett, and Wills, 2016).

To choose parameters using k-fold CV, we first split the available data into k subsets (in most experiments, we set $k=10$). One subset was used as a testing dataset and then evaluated using the remaining $k-1$ training subsets. Then we calculated the CV error using this split error for parameter combinations. This process was repeated k number of times, and the model performance values reported are the average of the performance values observed in each loop. The process is thus termed k-fold cross-validation and illustrated in Figure 2.1.

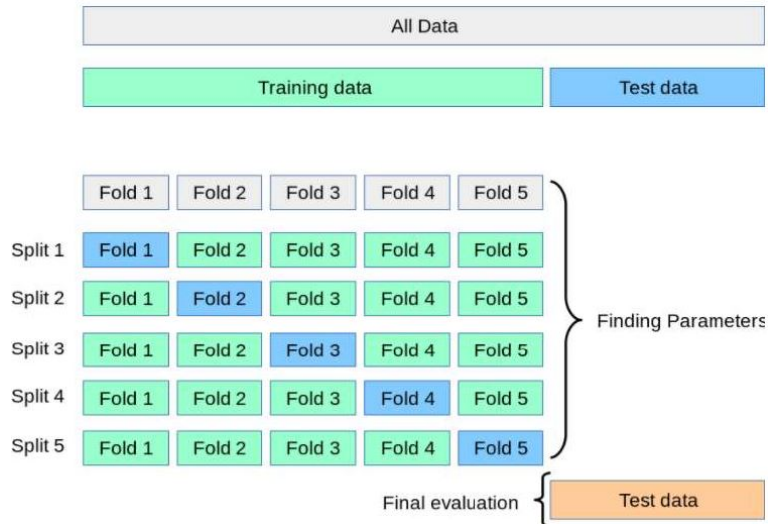


Figure 2.1: Model training and testing using K-Fold CV (Source: Devar, 2019)

2.3.3.3 Data Partitioning

After collecting data from both solvent and insolvent banks, we shuffled the data to prevent bias during the modelling process. We used data for all banks that failed between 2007 and 2011 for the in-sample analysis (training data) and data for all banks that failed between 2012 and 2019 for the out-of-sample analysis (test data). The data was split into two subdivisions: a training set of 80% and a testing set of 20%. It is important to test the efficiency of machine learning models used for bankruptcy on new samples (Taffler, 1983). The first set was used to train the models, and the test set was used to evaluate the performance of the models and to also evaluate the feature selection and classifier methods. The validation set was used to assess the performance of the model (Hastie et al., 2009). The accuracy and misclassification rates of each model using the different groups of variables were computed using a holdout validation dataset. The hold-out data was used to quantify the generalisation ability of the classifier by measuring the model error on the test data set. Data partitioning prevents overfitting of the model and ensures that the model learns general patterns from the training data. The test data provides an unbiased evaluation of the trained model on an unseen data. Also, partitioning the data enables the use of cross-validation methods which assess the stability of the model across several random data splits and also fine tune the model parameters. During model evaluation, it is important to minimise test error rather than minimise training error, as the test error provides a more realistic assessment of the model's performance.

2.3.3.4 Fixing data imbalances

The ability of a machine learning model to discriminate between bankrupt and solvent banks is highly dependent on the inherent patterns in the data and data labels. In most cases, the number of bankrupt banks is small in relation to solvent banks. This imbalance between the bankrupt and solvent bank labels impacts the performance of supervised machine learning techniques. This data imbalance is known as class imbalance and can have adverse effects on a machine learning model's (or learner's) ability to detect bankrupt banks. The use of data sampling techniques to address the issue of class imbalance improves the performance of machine learning techniques. Classification of data with an imbalanced class distribution has a significant drawback on the performance of classification algorithms. Class imbalance refers to the condition where classes are not represented equally within a dataset (Van Hulse, Khoshgoftaar, and Napolitano, 2007). An imbalanced dataset is characterised as having more classes than others. For instance, there are more observations for solvent banks than for failed banks. Classification algorithms tend to perform poorly on imbalanced data sets in which one class heavily outnumbers the other class. The performance of bankruptcy models is not only dependent on the features selected and the machine learning algorithms used, but also depends on the samples selected for fitting the models.

The minority class (bankrupt cases) are usually of much interest but tends to be misclassified. This severe imbalance in our dataset may have adverse repercussions for the performance of the machine learning models. Specifically, it is well-known that fitting machine learning classifiers to imbalanced datasets renders them biased against the minority class. This imbalance creates a bias towards the majority class during model training. The effects of class imbalance on machine learning performance have been widely researched (Ali, Shamsuddin, and Ralescu, 2013). The use of machine learning algorithms on imbalanced datasets has received much attention in recent research due to its wide real applications (García, Sánchez, and Mollineda, 2012; Pérez-Godoy et al., 2010; Sun et al., 2007; Van Hulse and Khoshgoftaar, 2009). In this study, we used the Synthetic Minority Oversampling Technique (SMOTE) to fix the data imbalance by randomly duplicating observations in the minority class. SMOTE generates new instances from the minority group in the training data before training the model. We used a data sampling technique because not all classifiers have algorithm-based implementation, and also the misclassification cost of using the algorithmic approach is not known (Seiffert et al., 2007; Weiss, McCarthy, and Zabar, 2007).

2.3.3.5 Synthetic Minority Oversampling Technique (SMOTE)

Training a machine learning model with imbalanced data will result in a low recall rate for the minority class. SMOTE is a heuristic oversampling technique designed to solve the category imbalance and significantly improve the overfitting caused by

non-heuristic random oversampling methods. This method generates artificial minority observations, which increases the size of the minority class using sampling with replacement and k-nearest neighbours (Han, Wang, and Mao, 2005). SMOTE interpolates from original minority examples rather than just duplicating the original observations. The core idea is to insert randomly generated new samples between a sample of minority categories and their neighbouring classes, which can increase the number of minority categories and improve the unbalanced category distribution of the dataset. This technique first finds the k-nearest neighbours of the minority class for each minority observation. The new, artificial examples are then generated in the direction of some or all of the nearest neighbours, depending on the amount of oversampling desired, by calculating the difference between the original minority example and its nearest neighbours and multiplying this difference by a random number (between 0 and 1).

The samples of a minority category in the training set are set as S_{min} . The steps in SMOTE are as follows

1. A sample X_i is randomly selected from S_{min} . The Euclidean distance is used as the standard to calculate its distance to other samples in S_{min} , and its k nearest neighbours are obtained
2. A sample X_j is randomly selected from k neighbours, and a new sample X_{new} is generated using Eq. 2.1

$$X_{new} = X_i + rand(0,1) \times |x_i - X_j| \quad (2.1)$$

3. The first two steps are repeated until the number of samples for the category in the training set is equal to the number of samples for the category with the largest number of samples

2.3.3.6 Feature Selection

Real-world datasets often contain a large number of features, some of which are either redundant or irrelevant to the given tasks (Fayyad, Piatetsky-Shapiro, and Smyth, 1996). This happens when it is unknown which features are relevant to a target concept, especially when domain knowledge is unavailable or incomplete. Classifying banks and financial institutions as healthy or bankrupt often requires a large number of features. However, some of the features are redundant or irrelevant, resulting in noise in the dataset, thereby reducing the performance of the classification algorithm (Dash and Liu, 1997). The presence of irrelevant and redundant features may mask or obscure the distribution of truly relevant features for a target concept and hence harm the performance of classification models (John, Kohavi, and Pfleger, 1994; Koller and Sahami, 1996; Ragavan et al., 1993). In addition, increasing the dimensionality of the feature space will generally result in increased complexity of interactions among the

features and increased degree of noise, making it difficult for classification methods to learn high-level concepts effectively and efficiently (Piramuthu, Ragavan, and Shaw, 1998).

One of the challenges of bankruptcy prediction models involves choosing the right set of predictors. Quite a several different financial ratios have been used in modelling. Therefore, there is a need to carefully choose variables from the many sets of reliable predictors used in previous literature. Reducing the number of redundant or irrelevant features reduces the running time of the algorithm and enhances its accuracy of the algorithm. The selection of relevant features and the elimination of redundant features is an important part of machine learning modelling. It is very important to decide on which features to use in the model and which ones to ignore to avoid increasing the search space, which leads to an increase in the time for classification and makes generalisation more difficult. A number of studies have combined feature-preprocessing algorithms with classification methods to improve their performance. For example, Piramuthu, Ragavan, and Shaw (1998) used feature construction algorithms to construct new feature sets and improved the performance of a neural network for predicting bankruptcy and default loan payment. West (1985) extended a logit model by preprocessing features with factor analysis. The combined factor-logistic approach was shown to be more effective in predicting bank bankruptcy.

In machine learning, feature selection, which is an important aspect of model building, is the process of choosing relevant variables that are useful in developing a predictive model. It is desirable to reduce the number of input variables to both reduce the computational cost of modelling and, in some cases, to improve the performance of the model. Feature selection is the process of selecting a set of relevant features or attributes to build a predictive model. Too many features, including redundant ones, add noise to the result and cause the algorithm to take a lot of time and reduce the accuracy of the prediction model. In most cases, high-dimensional features cause high computational cost and could result in overfitting the model. It is therefore important to reduce the dimensionality of the features using feature selection techniques (Zhuo et al., 2008). Feature selection techniques pick a subset of features that are relevant to the target variable in a classification or regression algorithm from the data. These features are the best combination of the original subsets that contribute most to the classification algorithm. Feature selection aims to choose a subset of M features from a set of N features, $M < N$, to improve the prediction accuracy of the algorithm or decrease the size of the structure without a decrease in prediction accuracy of the algorithm using only the selected features (Koller and Sahami, 1996). According to Koller and Sahami (1996), feature selection picks a small subset such that the distribution of the class, given only the value of the selected features, is as close as possible to the original class distribution given all the features. Feature selection enables the possibility of using features that are important in fitting a predictive model with reduced model complexity and better time

efficiency. However, due to the fact that feature selection is an NP-complete problem, there is no single efficient solution (Albrecht, 2006). Feature selection techniques enable machine learning algorithms to train faster, reduce the complexity of a model and make it easier to interpret, improve the accuracy of the model and reduce overfitting.

Feature selection methods search through the subset of features to select the best one among the competing 2^N candidate subsets according to some evaluation function. There are four basic steps in the feature selection method, namely generation procedure, evaluation function, stopping criterion and validation procedure (Dash and Liu, 1997). The generation procedure generates the next candidate subset of features to be evaluated. It usually starts with no features (features are added iteratively) or all features (features are iteratively removed) or with a random subset of features (features are iteratively added or removed) (Langley, 1994). The evaluation function measures the goodness of the subset and compares it with the previous best subset. When the present subset is found to be better, it replaces the previous subset. However, without a stopping criterion, the feature selection process may run exhaustively through the space of subsets. Stopping criteria based on a generation procedure indicate whether a predefined number of features was selected or whether a predefined number of iterations was reached. Stopping criteria based on an evaluation function indicate whether the addition (or deletion) of any feature does not produce a better subset, and whether an optimal subset according to some evaluation function is obtained. The loop continues until some stopping criterion is satisfied. The feature selection process halts by outputting a selected subset of features to a validation procedure. The validation procedure tests the validity of the selected subset of features using different tests and compares the results with previous results or competing selected features. The best feature subset contains the fewest number of features with a higher accuracy.

Variables are sometimes selected based on their usefulness in predicting bank failure in the past or by the use of statistical or optimisation methods. However, the statistical approach of stepwise selection of variables assumes that input and output dependence is linear and, as such, may not be appropriate for non-linear models such as extreme learning machine and others. In order to overcome this setback, different variable selection techniques such as genetic algorithm, simulated annealing and Boruta algorithm were employed in this study. We used genetic algorithm, simulated annealing, tabu search, recursive feature elimination, stepwise regression, learning vector quantisation and Boruta algorithm combined with SVM-linear, SVM-RBF, OSELM, XGboost and LightGBM classifiers. The performance of all features and each feature selection technique was compared using each of the classifiers. Also, we compared the performance of each of the classifiers with each other. The dataset consists of 33 financial ratios for each failed and solvent bank, as shown in Table 2.2. We identified the most important financial ratios for bankruptcy prediction by employing different feature selection techniques. The selected features are shown in Table 2.3. Group 1 represents

all financial ratios used in the study. Groups 2 and 3 represent all financial ratios selected by stepwise logistic regression and simulated annealing algorithm, respectively. Features selected by genetic algorithm technique, Tabu search and recursive feature elimination methods were represented by groups 4, 5 and 6, respectively.

Table 2.3: Features selected for each selection technique

Group	Description/technique	Features
Group 1	All variables from the data	X1-X33
Group 2	Stepwise logistic regression	X2, X3, X4, X5, X6, X7, X8, X9, X10,X11 X12, X13, X14, X15, X16, X17,X18,X19, X20, X22, X23, X24, X25,X26, X27,X28, X29, X30, X31, X32,X33
Group 3	Simulated annealing	X3, X5, X9, X12, X14, X16, X20, X22, X23,X25, X26, X27, X32, X33
Group 4	Genetic algorithm	X1, X4, X6, X7, X9, X10, X11, X13, X14, X15, X16, X17,X18, X21, X22, X24, X26, X27,X29, X30, X32, X33
Group 5	Tabu search	X1, X3, X4, X7, X9, X11, X13, X14, X17, X19,X21, X23,X24, X26, X29, X30, X31, X32
Group 6	Recursive feature elimination	X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11,X12, X13,X14, X16, X17, X18, X19, X20, X21,X22, X23, X24, X25,X26, X27, X31, X32
Group 7	Learning vector quantization	X1 ,X2 ,X3, X4, X5, X7, X9, X10, X11,X12, X13, X14, X15, X16, X20, X21, X23, X24, X25, X27

2.3.3.7 Feature Importance with SHAP

Traditional feature importance only tells which feature is important, but it does not reflect how the feature affects the prediction result. SHAP value not only reflects the influence of the features in each sample, but also shows the positive and negative influence. Coalitional game theory serves as the basis for the additive explanation model that Shapley Additive exPlanation (SHAP) develops. The SHAP method applies mathematical functions using the Shapley values method from game theory to determine the weight of each feature. The SHAP feature ranking method was developed by Shapley (1953) in contributing to Game Theory and was recommended by Lundberg and Lee (2017). SHapley Additive exPlanation (SHAP) constructs an additive explanation model which is inspired by coalitional game theory (Shapley et al., 1953). The SHAP framework is based on the calculation of the Shapley value for measuring how features affect the dependent variable. In the feature selection task, the SHAP value can be used to calculate the marginal contribution value of each feature and measure the importance of the feature. SHapley Additive exPlanations is a unified approach to explain

the output of any machine learning model. It connects game theory with local explanations, uniting several previous methods and representing the only possible consistent and locally accurate additive feature attribution method based on expectations (Lundberg and Lee, 2017).

In SHAP, each feature in the sample is regarded as a "contributor" to the machine learning model's prediction (Lundberg and Lee, 2017). For each prediction sample, SHAP produces an evaluation value assigned to each feature in that sample, which is referred to as the SHAP value. In order to estimate the target function f , SHAP creates a linear function g with binary variables based on the additive feature attribution approaches. The following is the explanation model:

$$f(x) \approx g(Z^*) = \phi_0 + \sum_{i=1}^M \phi_i Z_i^* \quad (2.2)$$

where $Z_i^* = (0, 1)$ means that M is the number of features if feature i participates in the prediction process, and ϕ_i represents the contribution of feature i . The contribution of feature i is introduced as follows:

$$\phi_i = \sum_{S \subseteq N \setminus i} \frac{|S|!(M - |S| - 1)!}{M!} [f_x(S \cup \{i\}) - f_x(S)] \quad (2.3)$$

where N is the set of all features, S is a subset of different features combination without feature i , and $f_x(C)$ is the prediction model conditioned on input instance x with the subset C of feature combination. After calculating the feature importance through SHAP value, visualisations are obtained. The SHAP dependence plot is displayed to show important features and how they affect the prediction.

2.3.4 Model hyperparameter optimization

Generally, most machine learning algorithms will not achieve optimal results if their parameters are not tuned properly. To build a high-accuracy classification model, it is very important to choose a powerful machine learning algorithm as well as adjust its parameters. Parameter optimisation can be very time-consuming if done manually, especially when the algorithm has many parameters (Friedrichs and Igel, 2005; Rossi and Carvalho, 2008). In addition to feature selection, the choice of optimal model hyperparameters can improve the performance of machine learning models. Machine learning models use parameters and hyperparameters. Parameters are generated by the model during the training stage. These values are estimated from the data. On the other hand, hyperparameters are provided before the training phase of the model. The machine learning algorithms utilised in this thesis have various hyperparameters that control the performance of the model. Hyperparameters determine the complexity of

a machine learning model. Therefore, an optimal choice of hyperparameters ensures that there is no overfitting or underfitting of the model. In order to optimise the performance of each model, it is important to perform hyperparameter tuning using a grid search approach.

Hyperparameter tuning is a process of selecting the best hyperparameters from a set of all possible values of a machine learning model in order to optimise the performance of the model. In most cases, the default hyperparameter settings do not guarantee optimal performance of the machine learning algorithm. The objective of hyperparameter tuning is to obtain the maximum evaluation score on the training and testing datasets without causing overfitting issues. The hyperparameter searching process can be done using exhaustive search, heuristic search, Bayesian optimisation (BO) or multi-fidelity optimisation. Finding the hyperparameters of a certain machine or deep learning algorithm that perform the best when evaluated against a validation set is the aim of hyperparameter optimisation in machine learning or deep learning. The hyperparameters have an impact on the models' ability to predict outcomes. It is essential to adjust these hyperparameters sensibly. Hyperparameter optimisation, on the other hand, is a combinatorial optimisation problem that cannot be resolved using the gradient descent method. We used the grid search and Bayesian optimisation with 10-fold cross-validation to select the best possible hyperparameters.

2.3.4.1 Grid search

The grid search cross-validation is an exhaustive search method based on a defined subset of hyperparameter space to optimise the model's performance by traversing a given combination of parameters (Erdogan Erten, Bozkurt Keser, and Yavuz, 2021). The general idea of using grid search is to determine the best parameter based on the given parameter grid value to build the machine learning model. With a proper tuning of the parameter, a good, balanced and well-generalised model can be obtained. The wider the parameter range is, the more possibilities and effectiveness the grid search method with 10-fold CV has of finding the best combination of parameters to significantly increase the accuracy value. Chang and Lin (2011) suggests choosing an initial set of possible input parameters and performing grid search cross-validation to find optimal parameters for the model. The grid-search cross-validation produces cross-validation estimates of performance statistics (for example, error rate) for each point in the grid. This parameter tuning method cycles through all the candidate parameter choices for each possibility, with the final result being the best-performing parameter. The grid search algorithm trains the model using each set of hyperparameters and picks the hyperparameter combination with the smallest validation error. According to Lin et al. (2008), the cross-validation technique can prevent the over-fitting problem. The performance of every combination is evaluated using some performance metrics.

2.3.4.2 Bayesian Optimisation

Bayesian optimisation uses the results of the past evaluations to form a probabilistic model of the objective function and uses this model to choose the next set of hyperparameter values. The probabilistic model, which is otherwise referred to as a surrogate model, is represented by $p(y|x)$; y being the performance metric for the model and x being the hyperparameter values. The objective function maps the hyperparameter values to the performance metric (AUC, accuracy, etc) on the data. In contrast to grid search (GS) and random search (RS), BO bases its determination of future evaluation points on results from the past. BO relies on an acquisition function and a surrogate model as its two main building blocks to decide the subsequent hyper-parameter setting (Salo et al., 2018). All of the currently observed points are intended to fit into the objective function using the surrogate model. After getting the probabilistic surrogate model's prediction distribution, the acquisition function weighs the trade-off between exploration and exploitation to decide which points to use. While exploitation entails sampling in the presently promising regions where the global optimum is most likely to occur, based on the posterior distribution, exploration entails sampling the instances in the areas that have not yet been sampled. In order to identify the currently most optimal locations and prevent missing better configurations in the unknown areas, BO models balance the exploration and exploitation processes. The steps involved in Bayesian optimisation are as follows (Snoek, Larochelle, and Adams, 2012):

1. Build a surrogate model of the objective function using results of previous evaluations
2. Find the hyperparameters that perform best on the surrogate model
3. Evaluate the objective function with the hyperparameter values chosen in step 2
4. Update the surrogate model by adding results obtained in step 3
5. Repeat steps 2 to 4 until a stopping criterion is reached
6. Select the best performing hyperparameters from all the trials

As a result, BO operates by modifying the surrogate model following each assessment of the objective function. Due to its ability to identify the best hyper-parameter combinations by examining the results of prior tests and the fact that running a surrogate model is frequently significantly less expensive than running the complete objective function, BO is more effective than GS and RS. Bayesian optimisation, which depends on previously tested values, belongs to sequential methods which are difficult to parallelise but are able to find optimal hyperparameter combinations in a small number of iterations. A number of variants of Bayesian optimisation methods are available for hyperparameter optimisation. The methods differ by the type of surrogate model and optimisation criteria used in step 2 above. Common surrogate models for BO

include Gaussian process (Seeger, 2004), random forest (Hutter, Hoos, and Leyton-Brown, 2011) and the Tree Parzen estimator (Bergstra et al., 2011).

We use the Tree Parzen Estimator (TPE) Bergstra et al. (2011), a variant of Bayesian optimisation as the surrogate model and Expected Improvement (EI) as the criterion to optimise the surrogate model. BO-TPE creates two density functions, $l(x)$ and $g(x)$, to act as the generative models for all domain variables(3). To apply TPE, the observation results are divided into good results and poor results by a pre-defined percentile y^* , and the two sets of results are modelled by simple Parzen windows (Bergstra et al., 2011). Bayesian Optimisation using TPE is mathematically expressed as:

$$p(x|y, D) = \begin{cases} l(x) & \text{if } y < y^* \\ g(x) & \text{if } y \geq y^* \end{cases} \quad (2.4)$$

The expected improvement in the acquisition function, which is reflected by the ratio between the two density functions, is used to determine the new configurations for evaluations. The Parzen estimators are organised in a tree structure, whereby the specified conditional dependencies are retained. TPE naturally supports specified conditional hyper-parameters.

2.3.5 Model Evaluation metrics

To evaluate the prediction performance of the different classifiers for bankruptcy detection, we compute different performance metrics using a cross-validation test. The K-fold cross-validation has been used to assess the performance of machine learning techniques (Kumar et al., 2015). We conduct a 10-fold CV to assess the performance of the proposed methods. We used the CV to train and test the models in each of the two data sets, and the metrics are calculated over the data set. The efficiency of the different classifiers were assessed and compared using confusion matrix, accuracy, F1 score, kappa, MCC, sensitivity (recall or true positive rate), specificity (true negative rate), precision (positive predictive value), Type I error, Type II error, receiver operating characteristics (ROC) curve and area under the curve for both training and testing dataset. The area under the curve (AUC) was calculated for each of the models to determine the accuracy. The other metrics are defined based on the confusion matrix.

2.3.5.1 Confusion Matrix

In a binary classification problem in bankruptcy prediction, a classifier labels a financial institution as either solvent or bankrupt. The decision made by the classifier can be represented in a structure known as a confusion matrix or contingency table. A confusion matrix is a tabular layout that contains information about the actual and predicted

classifications done by a classification algorithm. A confusion matrix for a binary classification classifier is a 2×2 table displaying the observed labels against the predicted labels for a data set. It describes the performance of a classifier on the training and testing data where the actual classes are known. The confusion matrix is shown below in Table 2.4. In general, given a classification machine learning model, the information regarding its performance is summarised into a confusion matrix. For a binary classification problem, there are four possible outcomes when the model is applied to an observation.

True positive (TP) is the number of bankrupt banks that were correctly predicted. True negative (TN) refers to the number of predictions where the classifier correctly predicts a solvent bank as solvent. False positive (FP) refers to the number of predictions where the classifier incorrectly predicts the solvent bank as a bankrupt bank. It is a Type I error made in classification. False negative (FN) refers to the number of predictions where the classifier incorrectly predicts a bankrupt bank as a solvent bank. It is a Type II error made during classification. When the model output matches the label of the observation, we either have a true positive or a true negative. However, if the model output does not match the label, we have either a false positive or a false negative, both of which are error outcomes. These two error outcomes are usually distinguished in classification problems. For a given classifier and an appropriate labelled data set, one of the intuitive metrics for model performance is the error of a classifier. The error is the number of mistakes the model makes on the data set divided by the number of observations in the data set. Thus, the model error is computed by summing the number of errors made by the classifier on the data set and dividing by the size of the data set.

For a model predicting bankruptcy, two possible errors can be committed. The model can commit a false positive error; it predicts that a bank is bankrupt when it is solvent. It can also commit a false negative error where the model predicts that a bank is solvent when it is bankrupt. In financial crisis prediction and analysis, false negative error is much serious than false positive. A false negative, the bank is assumed to be solvent and necessary steps may not be taken by regulators to salvage the bank. A false positive result will prompt more tests and analysis to be conducted on the bank, which may eventually prove that it is a false positive error. A representation of the model performance using a confusion matrix distinguishes between the two types of errors and is often used as a tool of choice for model performance in classification problems.

Table 2.4: Confusion matrix

		Predicted class	
		Bankrupt	Solvent
True class	Bankrupt	TP	FN
	Solvent	FP	TN

2.3.5.2 Accuracy

The performance of a classifier can be characterised in terms of accuracy. The accuracy of a classifier is given as the number of correct predictions divided by the total number of observations. It is the most frequently used metric in classification (Géron, 2019; Raschka, 2015). It is often used as a singular performance metric when there is a balanced data set. For an imbalanced dataset, accuracy is very sensitive and very misleading to rely on this metric alone (Fitzpatrick, Bull, and Dupriez, 2018; Géron, 2019). This metric is therefore often used in conjunction with other performance metrics. Accuracy in bankruptcy is defined as the proportion of total banks that were correctly classified by the model. It is the number of correct classifications from the total data. It provides the overall accuracy of the model. It is expressed mathematically as below:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (2.5)$$

2.3.5.3 Sensitivity

In addition to accuracy and error, the sensitivity of the classifier is also used as a measure of model performance. The sensitivity of a model is defined as the true positive predictions divided by the sum of all positive observations. Sensitivity (recall) is the fraction of bankrupt banks which were correctly predicted as bankrupt by the classifier. Thus, it is the fraction of positive predictions by the classifier which were positive. Sensitivity is the ratio between correctly assigned data and all the data points in a class. It is also referred to as the true positive rate (TPR). A high value of recall is indicative of the classifier's ability to correctly classify. A sensitivity of 1 for a model indicates that the model predicts all positive observations correctly. Thus, the model does not commit any false negative predictions. It is a good metric to use in situations where the cost of false negatives is high.

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (2.6)$$

2.3.5.4 Specificity

The specificity of a model is defined as the true negative predictions divided by the sum of all negative observations. Specificity is the fraction of solvent banks that are correctly predicted as solvent by the classifier. Specificity is the fraction of all negative samples that are correctly predicted as negative by the classifier. It is also known as the true negative rate (TNR). A specificity of 1 for a model indicates the model predicts all negative observations correctly and does not commit any false positive predictions.

$$Specificity = \frac{TN}{FP + TN} \times 100\% \quad (2.7)$$

2.3.5.5 Precision

Precision or positive predictive value (PPV) is defined as the number of true positives divided by the number of true positives plus the number of false positives. Precision is the fraction of bankrupt banks that were predicted accurately. Thus, Precision is the fraction of positive predictions by the classifier that are positive. Precision shows what proportion of observations classified as non-bankrupt by a model is really non-bankrupt. A high precision is indicative of a high number of correctly assigned classes out of the total classification assignments by the classifier. Precision is a good metric to use in situations where the cost of false positives is high.

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2.8)$$

2.3.5.6 F1 score

F1-score is the combination of precision and recall into a single measure. It is a harmonic mean of precision and recall. It is a good measure when there is an uneven class distribution. A good F1 score indicates low false negatives and low false positives. The classifier with the highest F1 score provides the best performance. F1 score ranges from zero to one. A classifier gets a high F1 score if both recall and precision are high. F1 score provides the weighted average of precision and recall, and is given.

$$F1 \text{ score} = \frac{2TP}{2TP + FP + FN} \quad (2.9)$$

or

$$F1 \text{ score} = 2 \times \frac{precision \times recall}{precision + recall} \quad (2.10)$$

2.3.5.7 Matthews Correlation Coefficient

To evaluate binary classifications and their confusion matrices, several statistical rates are employed. Accuracy and F1 score computed on confusion matrices have been (and still are) among the most popular adopted metrics in binary classification tasks. However, when the dataset is unbalanced (the number of samples in one class is much larger than the number of samples in the other classes), accuracy cannot be considered a reliable measure anymore, because it provides an overoptimistic estimation of the classifier's ability on the majority class (Bekkar, Djemaa, and Alitouche, 2013; Gu, Zhu, and Cai, 2009; Sokolova, Japkowicz, and Szpakowicz, 2006). An effective solution overcoming the class imbalance issue comes from the Matthews correlation coefficient (MCC). As an alternative measure unaffected by the unbalanced datasets issue, the Matthews correlation coefficient is a contingency matrix method of calculating the Pearson product-moment correlation coefficient (Powers, 2020) between actual and predicted values.

This measure of classification performance was originally developed in 1975 for comparing chemical structures (Chicco and Jurman, 2020). Although it was first introduced in biochemistry (Matthews, 1975), the MCC gained popularity in several scientific fields, including software detection prediction (Yao and Shepperd, 2020), recommender systems (Liu et al., 2015), pattern recognition (Zhu, 2020) and medicine (Saqlain et al., 2019). MCC was re-proposed by Baldi et al. (2000) as a standard performance metric for datasets that have unbalanced target labels for machine learning. Matthew's correlation coefficient (MCC) is a measure of the quality of binary classifications. It is similar to the Pearson correlation coefficient. This method treats the true class and predicted class of the confusion matrix as two binary variables, and it calculates their correlation coefficient. The lower the correlation between the true and predicted values, the lower the classification performance. Computing the MCC involves solving the phi coefficient (ϕ).

The Matthews correlation coefficient (MCC) is a more reliable statistical measure which produces a high score only if the prediction obtained good results in all the four confusion matrix categories (true positives, false negatives, true negatives, and false positives). Thus, the MCC is the only rate that maximises all four basic rates (Chicco and Jurman, 2020; Chicco, Tötsch, and Jurman, 2021). MCC is the only binary classification rate that generates a high score only if the binary predictor was able to correctly predict the majority of positive data instances and the majority of negative data instances (Jurman, Riccadonna, and Furlanello, 2012). The MCC score represents the correlation between the predicted positives and the true positives. The MCC value ranges from -1 to +1. A +1 value corresponds to a perfect classification; a value close to 0 corresponds to a prediction made by chance; and a -1 value corresponds to a perfectly opposite prediction where all the negative samples were predicted as positive and vice versa (Matthews, 1975). A negative number represents a negative correlation between predicted and true negatives. Some previous studies have argued that the MCC is more informative than the F1 score (Chicco and Jurman, 2020), accuracy (Chicco and Jurman, 2020), and Cohen's Kappa (Chicco, Tötsch, and Jurman, 2021). The key asset of the MCC is that its high value always corresponds to high values for each of the four confusion matrix basic rates: sensitivity, specificity, precision, and negative predictive value (Chicco, Tötsch, and Jurman, 2021). The MCC or phi coefficient can be derived from Equation 2.3.5.7.

$$MCC = \frac{TP \times TN - FP \times FN \times recall}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (2.11)$$

2.3.5.8 Cohen's Kappa

Cohen's kappa is a metric that captures the accuracy of a model with a single value. Cohen's Kappa measures the influence of chance on a model's accuracy. It does this by comparing the accuracy achieved and expected accuracy (Cohen, 1968). According to Fitzpatrick and Mues (2016), expected accuracy is the accuracy rate that a random classifier would be expected to achieve based on the confusion matrix of predictions. A Kappa score of 1 is the perfect score, representing complete agreement between the model's performance score and the ground truth, whereas a negative score indicates chance agreement. A kappa of zero means all the accuracy achieved was by chance.

$$Kappa = \frac{2 \times (TP \times TN - FN \times FP)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)} \quad (2.12)$$

2.3.5.9 Receiver Operating Characteristics

In machine learning, current research has shifted away from simply presenting accuracy results when performing an empirical validation of new algorithms. This is especially true when evaluating algorithms that output probabilities of class values. Provost, Fawcett, Kohavi, et al. (1998) have argued that simply using accuracy results can be misleading. They recommended that when evaluating binary decision problems, use Receiver Operator Characteristic (ROC) curves, which show how the number of correctly classified positive examples varies with the number of incorrectly classified negative examples. ROC curves are commonly used to present results for binary decision problems in machine learning. It is among the most used tools to gauge the performance of ML classifiers, especially visually (Davis and Goadrich, 2006).

The ROC curve is a graphical plot of the sensitivity (proportion of true positives) of the predictive model versus the complement of the specificity (proportion of false positives), in a series of thresholds for a positive result (Hanley et al., 1989). Thus, in ROC space, one plots the False Positive Rate (FPR) on the x-axis and the True Positive Rate (TPR) on the y-axis. The FPR measures the fraction of negative examples that are misclassified as positive. The TPR measures the fraction of positive examples that are correctly labelled. The ROC curve is a performance metric representable in graphic form, which displays the relationship between the TPR and the FPR classes of prediction. The ROC curve provides a visual illustration of the performance of classifiers on binary datasets, where a classifier corresponds to a point (Fawcett, 2006). An ideal curve climbs towards the top left, indicating that the model accurately predicts cases. The area under the ROC curve (AUC), which shows the quality of the classifier, is a performance metric for evaluating classifiers. The AUC is one of the most common statistics used in scientific research to assess binary classifications, and can range from 0 (worst result) to 1 (perfect result) (Hanley and McNeil, 1982). A random classifier will have an AUC of 0.5 while a perfect classifier will have an AUC of 1. The larger the AUC

(AUC closer to 1), the better the model performance. A high AUC value represents a high TPR and low FPR. Often, the area under the curve is used as a simple metric to define how an algorithm performs over the whole space in the ROC curve (Bradley, 1997; Davis et al., 2005; McSkassy and Provost, 2005). The AUC can be calculated by using the trapezoidal areas created between each ROC point (Cortes and Mohri, 2003). We used the area under the curve (AUC) and receiver operating characteristic (ROC) curve as general performance in addition to the measures above.

2.3.6 Extreme Gradient Boosting (XGBoost) Machine

The Extreme Gradient Boosting (XGBoost) algorithm proposed by Chen and Guestrin (2016) is an improved version of Gradient Boosting Machine (GBM) for both regression and classification problems. XGBoost is an optimised version of GBM. According to Chen et al. (2015), XGBOOST is an efficient and scalable implementation of the gradient boosting framework originally proposed by (Friedman, 2001). The traditional GBM deals with the first derivative in learning, while XGBoost improves the loss function with Taylor expansion. The main improvement on GBM is the normalisation of the loss function to reduce model variance. Although both GBM and XGBoost use the gradient boosting principle, XGBOOST introduces a regularised model formalisation to control over-fitting and obtain better performance results.

Gradient boosting methods use an ensemble learning approach, which combines the predictive power of several weaker models (trees or classifiers) in order to obtain a superior predictive model (Chambers and Dinsmore, 2014). One important feature of boosting is that each new classifier is dependent on the errors from previous fitted classifiers, and each decision tree is adjusted to the residuals from the model. These individuals' classifiers are referred to as base learners or weak learners. Boosting algorithms were developed to combine a number of weak classifiers to produce an ensemble classifier with a superior generalised misclassification error rate (Kuhn, Johnson, et al., 2013). The combination of the weak learners will produce better predictive performance than any individual base learner. The tree ensemble model is a linear combination of several trees that can be considered as a regression model where each term is a tree (Elith, Leathwick, and Hastie, 2008). XGBoost is an algorithm that is based on boosting; thus, it combines predictions of a set of weak learners to develop a strong learner through additive training strategies.

The scalability of XGBoost is due to important systems and algorithm optimisation. It can handle sparse data and implements distributed and parallel computing. The parallel and distributed computing makes learning faster, which enables quicker model exploration. XGBoost has new design features which give it an edge over other ensemble algorithms. These include approximate split-finding algorithm, sparsity-aware split-finding, parallel computing, cache-aware access, block compression and sharing.

This technique, unlike the original GBM, controls overfitting by introducing a regularised model formalisation. The additive learning process in XGBoost is explained below:

The first learner is fitted to the training data, while the second model is fitted to the residuals in order to address the shortfalls of the weak learner. The process continues iteratively several times until a stopping criterion is met. The final prediction of the model is the sum of the predictions of each learner.

The general model for the prediction at step t is given by:

$$f_i^t = \sum_{k=1}^t f_k(x_i) = f_i^{t-1} + f_t(x_i) \quad (2.13)$$

where $f_t(x_i)$ is the learner at step t , f_i^t and f_i^{t-1} are predictions at steps t and $t - 1$ respectively, x_i is the input variables/features.

The objective function of XGBoost, which measures how well the model fits the training data, consists of two parts: the loss function and the regularisation term. The objective function is defined as

$$Obj(\theta) = L(\theta) + \Omega(\theta) \quad (2.14)$$

where L is the training loss function (mean squared error, MSE for regression or log loss for classification) and Ω is the regularisation term (penalty term to prevent overfitting). The training loss measures how predictive the model is with respect to the training data. Thus, it measures the model's ability to accurately predict the training data. The regularisation term controls the complexity of the model to help avoid overfitting.

A common choice of L is the mean squared error given by $L(\theta) = \sum (y_i - \hat{y}_i)^2$ where y_i is the target value for the i th row, $\hat{y}_i$ is the predicted value by the model for the i th row. Another commonly used loss function is the logistic loss for classification algorithms given as $L(\theta) = \sum_i [y_i \ln(1 + e^{-\hat{y}_i}) + (1 - y_i) \ln(1 + e^{\hat{y}_i})]$.

The prediction scores of each tree are summed up to get the final score. Thus, the function that determines the prediction includes the sum of all previous functions. This is written mathematically as

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (2.15)$$

where K is the number of boosted trees, f_K is a function in the functional space F and F is the set of all possible CARTS. The objective function to be optimised is written as

$$Obj = \sum_{i=1}^n \ell(y_i, \hat{y}_i^t) + \sum_{i=1}^t \Omega(f_i) \quad (2.16)$$

$$\hat{y}_i^t = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{t-1} + f_t(x_i) \quad (2.17)$$

Substituting equation 2.17 into 2.16

$$Obj^t = \sum_{i=1}^n \ell(y_i, \hat{y}_i^{t-1} + f_t(x_i)) + \Omega(f_t) + constant \quad (2.18)$$

Taking the Taylor expansion of the loss function up to the second order gives

$$Obj^t = \sum_{i=1}^n [\ell(y_i, \hat{y}_i^{t-1}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) + constant \quad (2.19)$$

where g_i and h_i are defined as

$$g_i = \partial_{\hat{y}_i^{t-1}} \ell(y_i, \hat{y}_i^{t-1})$$

$$h_i = \partial_{\hat{y}_i^{t-1}}^2 \ell(y_i, \hat{y}_i^{t-1})$$

. The objective function at step t becomes

$$Obj^t = \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \quad (2.20)$$

The objective function depends on g_i and h_i only. Let ω be the vector of scores on leaves. The function f , which maps the tree root to the leaves, is given as

$$f_t(x) = \omega_{q(x)}, \omega \in R, q : R^d \rightarrow 1, 2, \dots, T$$

q is the function assigning each data point to the corresponding leaf, and T is the number of leaves. The regularisation function (model complexity) in XGBoost is defined as

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2$$

Combining the loss function and the regularisation function, the objective function becomes

$$Obj^t = \sum_{i=1}^n [g_i \omega_{q(x_i)} + \frac{1}{2} h_i \omega_{q(x_i)}^2] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (2.21)$$

Let $I_j = \{i | q(x_i) = j\}$ be the set of indices of data points assigned to the j th leaf.

$$Obj^t = \sum_{j=1}^T [(\sum_{i \in I_j} g_i) \omega_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) \omega_j^2] + \gamma T \quad (2.22)$$

$$Obj^t = \sum_{j=1}^T [G_j \omega_j + \frac{1}{2} (H_j + \lambda) \omega_j^2] + \gamma T \quad (2.23)$$

where

$$G_j = \sum_{i \in I_j} g_i$$

and

$$H_j = \sum_{i \in I_j} h_i$$

. Minimising the objective function by taking the derivative with respect to ω_j and set left side equal to zero, we obtain

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \quad (2.24)$$

Substituting 2.24 into the objective function gives

$$Obj = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (2.25)$$

Equation 2.25 can be used as a scoring function to measure the quality of the tree structure, q . It is impossible to enumerate all the possible tree structures of q . A greedy algorithm, which starts from a single leaf and adds branches iteratively to the tree, is used. Assume I_L and I_R are the instance sets of left and right nodes after the split. Letting $I = I_L \cup I_R$, then the loss reduction after the split is given by

$$L_{split} = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (2.26)$$

The split candidates are evaluated using Equation 2.26.

2.3.6.1 XGBoost hyperparameters

Generally, hyperparameter tuning is needed/required to obtain the best performance. XGBoost has a number of hyperparameters that are manually set when building a model. The hyper-parameter consists of general parameters, booster parameters and task parameters.

Learning rate: It is also referred to as *eta*. It limits the variance by reducing the weight of each tree to a given percentage. It determines the contribution of each tree to the growing model. This parameter is used to control overfitting by shrinking the feature weights, thereby making the boosting process more conservative.

Maximum depth: It is the number of splits in each tree. It is used to specify the maximum depth of the tree. It controls the complexity of the boosted structure and helps to control overfitting. XGBoost makes splits up to the maximum depth specified and then starts pruning the tree backwards.

N-estimators: It provides the number of trees in the ensemble. Thus, it is the number of trees trained on the residuals.

Gamma: It is the minimum loss reduction required to make a further partition on a leaf node of the tree. It provides a threshold that nodes must surpass before making further splits according to the loss function. A node is split only when the resulting split gives a positive reduction in this loss function.

Sub-sample: It specifies the fraction of data to be selected for each tree. It increases the training speed (computation) of the algorithm and helps to prevent overfitting. Sub-sampling occurs once in every boosting iteration.

Minimum child weight: It refers to the minimum sum of weight required for a node to split into a child. If the sum of weights during the tree partition step is less than the value assigned to this parameter, no further splits are made. It reduces overfitting.

Col sample-by tree: Specifies the fraction of columns that are selected randomly for each tree. It increases the training speed of the algorithm and helps prevent overfitting. It limits the percentage of training columns for each boosting ratio.

Col sample-by level: This parameter limits the percentage of columns for each depth level of the tree. It is the subsample ratio of columns for each level. The subsampling occurs once for every new depth level reached in a tree. The columns are subsampled from the set of columns chosen for the current tree.

Col sample-by node: It is the subsample ratio of columns for each node (split). The subsampling occurs once every time a new split is evaluated. The columns are sampled from the set of columns chosen for the current level. It limits the percentage of columns to evaluate splits.

Lambda: It is the L2 regularisation (Ridge regression) term on weights. It helps to control the complexity of the model, thus avoiding overfitting.

Alpha: It is the L1 regularisation (Least Absolute Shrinkage and Selection Operator) term on weights. It also assists in controlling the complexity of the model (makes the model more conservative). It helps to avoid overfitting.

2.3.7 Light Gradient Boosting (LightGBM) Machine

This section explains the proposed intelligent approach for predicting bankruptcy among financial institutions using an optimised light gradient boosting framework based on tree learning algorithms. In the proposed approach, a Bayesian-based hyperparameter optimisation algorithm is intelligently integrated to tune the parameters of the LightGBM algorithm. Light Gradient Boosting Machine (LightGBM) (Meng et al., 2016), an open-source GBDT algorithm by Microsoft, is a gradient lifting framework based on tree learning, which was proposed by Microsoft. LightGBM is an improved version of the Gradient Boosting Decision Tree (GBDT) framework based on the decision tree algorithm and the idea of weak learners proposed using gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB). Since this technique has been developed (Ke et al., 2017), LightGBM has been applied in other fields such as breast cancer identification (Wang, Zhang, and Zhao, 2017), P2P network loan default (Ma et al., 2018), music recommendation (Zhang et al., 2018; Zhang, Quan, and Srinivasan, 2018), acoustic scene classification (Fonseca et al., 2017) and smart grid load forecasting (Zhang et al., 2018) as a result of high prediction accuracy, fast computational speed and excellent capability of minimising over-fitting issues.

It uses the histogram-based algorithm (Jin and Agrawal, 2003; Ranka and Singh, 1998) to speed up the training process, improve training efficiency, reduce memory consumption, achieve higher accuracy, implement large-scale data processing and combine advanced network communication to optimise parallel learning, called the parallel voting (PV) decision tree algorithm. The PV divides the training data into multiple machines, and the local voting decision to select the top-k attributes and the global voting decision to receive the top-2k attributes in each iteration are performed. The basic idea of the histogram algorithm is to discretise continuous floating-point eigenvalues into k bins and construct a histogram with a width of k. The histogram algorithm does not need extra storage of pre-sorted results, and also can save the value after the discretisation of features, which is generally enough to store with an 8-bit integer, and the memory consumption can be reduced to $\frac{1}{8}$ of the original. Other algorithms, such as XGBoost, use the level-wise algorithm, an inefficient algorithm to grow the decision trees. The level-wise approach treats the leaves of the same layer, resulting in a large amount of unnecessary memory consumption.

LightGBM uses a leaf-wise algorithm to grow trees vertically. A leaf that reduces the loss the most is chosen to split and grow the tree. LightGBM uses a histogram-based method to find the best split candidates. This method uses the leaf-wise strategy with depth constraints, unlike other gradient boosting decision tree algorithms, which use the level-wise decision tree strategy. This approach results in higher accuracy and reduced errors using the same number of splitting times. Leaf-wise is a more effective strategy, which finds the leaves with the highest branching gain each time from all the leaves, and then goes through the branching cycle. Therefore, compared with the horizontal direction, the blade can reduce more errors and obtain better precision with the same number of times of segmentation. However, the leaf-wise technique may result in over-fitting as a result of growing deeper decision trees. To control for over-fitting, a maximum depth limit is added to the top of the leaf to ensure high efficiency while preventing over-fitting. The schematic diagrams of level-wise and leaf-wise tree growth strategies are shown in Figures 2.2 and 2.3.

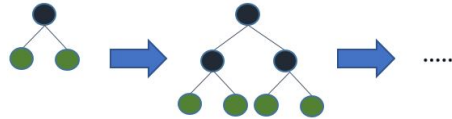


Figure 2.2: Level wise tree growth (Source: Park et al., 2021).

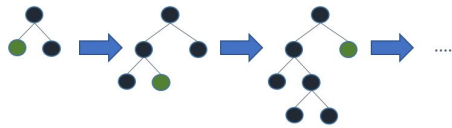


Figure 2.3: Leaf wise tree growth (Source: Park et al., 2021).

To improve the training, LightGBM uses a sampling algorithm, Gradient-based One-Side Sampling (GOSS), to indicate the importance of data instances. The GOSS chooses data with high gradients (e.g., larger than a predefined threshold or those among the top percentiles) for training and samples small portions of the data with low gradients as a supplement (Ke et al., 2017). This method speeds up the training process and improves the model's generalisation performance. The assumption is that the data with small gradients has lower errors; thus, they are already well trained. Therefore, GOSS proposed to ignore these less-informative data points and use the rest to compute the information gain when finding the best splits. However, this will result in a bias problem towards the sample with larger gradients, and will change the original distribution of the data. To solve this issue, GOSS performs a random sampling on the data with small gradients while keeping all the samples with large gradients. Because the sample would still be biased towards the data with large gradients, GOSS increases the weights (i.e. adding a constant multiplier) of the data instances with small gradients when computing the information gain.

In addition, LightGBM uses the Exclusive Feature Bundling algorithm to handle sparsity in datasets. High-dimensional data are usually sparse. Many of these sparse features in a sparse feature space do not take nonzero values simultaneously. It then becomes necessary to bundle these exclusive features into a single feature using the exclusive feature bundle technique. It combines mutually exclusive features in a nearly lossless way, resulting in reducing the number of features while keeping the most informative ones. The EFB is used to reduce the dimension of the features without necessarily losing any information. These optimisation techniques make LightGBM perform better in terms of speed and accuracy than other machine learning techniques. The features are merged in the same bundle to reduce training complexity by letting exclusive features reside in different bins using the histogram-based algorithm. The EFB algorithm bundles many exclusive features into fewer dense features, which effectively avoids unnecessary computation for zero feature values.

Given a supervised learning dataset $X = (x_i, y_i)$, LightGBM minimizes the objective function

$$Obj = \sum \ell(y_i, \hat{y}_i) + \sum \Omega(f_k) \quad (2.27)$$

The logistic loss function measures the difference between predictions $\hat{y}_i$ and the target y_i .

$$\ell(y_i, \hat{y}_i) = y_i \ln(1 + e^{-\hat{y}_i}) + (1 - y_i) \ln(1 + e^{\hat{y}_i})$$

Given the supervised learning data set $X = [(x_i, y_i)]_{i=1}^n$, LightGBM finds an approximation $\hat{f}(x)$ to a function $f^*(x)$ for minimising expected values of specific loss function $L(y, f(x))$:

$$\hat{f} = \arg \min_f E_{y,x} L(y, f(x)) \quad (2.28)$$

LightGBM minimises the regularised objective function

$$Obj = \sum_i \ell(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (2.29)$$

In this algorithm, the logistic function is used to measure the difference between the prediction $\hat{y}_i$ and the target y_i .

$$\ell(y_i, \hat{y}_i) = y_i \ln(1 + e^{-\hat{y}_i}) + (1 - y_i) \ln(1 + e^{\hat{y}_i})$$

LightGBM integrates a number of T regression trees $\sum_{t=1}^T f_t(x)$ to approximate the final model given as

$$f(T)(X) = \sum_{t=1}^T f_t(x) \quad (2.30)$$

The regression trees can be expressed in another form, $w_{q(x)}, q \in (1, 2, \dots, J)$ where J is the number of leaf nodes, q is the decision rule of the tree, and w is a vector denoting the sample weights of leaf nodes. The model is trained in the additive form at step t ; the objective function can be expressed as

$$\Gamma_t = \sum_{i=1}^n (y_i, f_{t-1}(x_i) + f_t(x_i)) \quad (2.31)$$

In LightGBM, the objective function is approximated using Newton's method, unlike traditional GBDT, which uses the steepest descent method. After simplification, the objective function can be written or transformed as

$$\Gamma_t \cong \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] \quad (2.32)$$

where g_i and h_i represent first and second order gradient statistical results of the loss function, respectively.

$$g_i = \partial_{F_{t-1}}(x_i) \Psi(y_i, F_{t-1}(x_i))$$

$$h_i = \partial_{F_{t-1}}^2(x_i) \Psi(y_i, F_{t-1}(x_i))$$

Let I_j denote the sample set of leaf j , then 2.32 can be transformed to

$$\Gamma_t = \sum_{j=1}^J [(\sum_{i \in I_j} g_i) \omega_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) \omega_j^2] \quad (2.33)$$

For the structure of a tree $q(x)$, the optimal weight of each leaf node w_j^* and the extreme values of Γ_k can be obtained as follows:

$$\omega_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (2.34)$$

$$\Gamma_T^* = - \frac{1}{2} \sum_{j=1}^J \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} \quad (2.35)$$

where Γ_T^* is the function which measures the quality of the tree structure q . The objective function after integrating the split is

$$G = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] \quad (2.36)$$

where I_L and I_R are sample sets of the left and right branches, respectively. Unlike traditional GBDT techniques, LightGBM grows the tree vertically, whereas other algorithms grow trees horizontally. This makes LightGBM an effective technique in processing large-scale data and features.

2.3.7.1 LightGBM hyperparameters

The LightGBM algorithm includes several parameters, termed hyperparameters. The hyperparameters have a significant impact on the performance of the LightGBM algorithm. These hyperparameters include:

Max depth: It describes the maximum depth of the tree. It limits the max depth for the tree model and is used to deal with model overfitting. Thus, it prevents trees from growing too deep, as shallow trees have less likelihood of overfitting. It is important to set this parameter if the dataset is small. The default value is -1.

Num leaves: It is the maximum number of leaves in one tree. It controls the complexity of the tree model. Larger values of *num_leaves* increase accuracy at the risk of a higher chance of overfitting. Setting small values helps to prevent overfitting. The default value is 31.

Num iterations: It sets the number of trees, which represent the number of training boosting iterations. This parameter influences the training speed of the tree model. Small values of this parameter lead to underfitting, while too large values lead to overfitting. The default value is 100.

Learning rate: This controls the speed of iteration. It is a shrinkage rate. The default value is 0.1.

Min data in leaf: It is the minimal number of data in one leaf. Setting this parameter to a large value can cause overfitting. It is a very important parameter for controlling overfitting. The default value is 20.

Bagging fraction: It randomly selects part of the data without resampling to be used for each iteration. It increases the training speed and helps to prevent overfitting. The default value is 1.

Feature fraction: This parameter enables feature subsampling. It randomly selects a subset of features on each iteration if the value is less than 1. This parameter increases training speed and helps to prevent overfitting. The default value is 1.

Feature fraction bynode: This parameter randomly selects a subset of features on each tree if the value is less than 1. It is used to deal with overfitting. The default value is 1.

Early stopping round: This parameter stops the training of the tree model if one metric of validation does not improve in the previous *early_stopping_round*.

Max bin: It is the maximum number of bins that feature values will be bucketed in. A small number of bins may reduce training accuracy, but control overfitting and improve training speed. Large value improves accuracy. The default value is 255.

2.3.8 Review of Extreme Learning Machine

This section briefly reviews the batch ELM development by Huang, Zhu, and Siew (2006) to provide the necessary background for the development of OS-ELM. A brief mathematical description of SLFN incorporating both additive and RBF hidden nodes in a verified way is given first.

For N arbitrary distinct samples (x_i, t_i) where $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$, the output of a standard SLFNs with $\tilde{N}$ hidden nodes and activation function $g(x)$ can be represented by

$$\sum_{i=1}^{\tilde{N}} \beta_i G_i(x_j) = \sum_{i=1}^{\tilde{N}} \beta_i G_i(a_i, b_i, x) = o_j, \quad a_i \in R^n, x \in R^n, j = 1, \dots, N \quad (2.37)$$

where $a_i = [a_{i1}, a_{i2}, \dots, a_{in}]^T$ and b_i are the learning parameters randomly generated for the hidden nodes, respectively. $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ is the weight vector connecting the i th hidden node and the output nodes, o_j , are the output vector of the SLFN concerning the input sample X_i .

For additive hidden nodes with the activation function $g(x) : R \rightarrow R$, $G(a_i, b_i, x)$ is given by

$$G(a_i, b_i, x) = g(a_i \cdot x + b_i) \quad b_i \in R \quad (2.38)$$

Where a_i is the input weight vector connecting the input layer to the i th hidden node, b_i is the bias of the hidden node, $a_i \cdot x$ denotes the inner product of vectors a_i and X in R^n .

For RBF hidden node with activation function $g(x) : R \rightarrow R$, $G(a_i, b_i, x)$ is given as

$$G(a_i, b_i, x) = g(b_i \| x - a_i \|), \quad b_i \in R^+ \quad (2.39)$$

where a_i and b_i are the center and impact factor of i th RBF mode. R^+ is the set of all positive real values.

The standard SLFNs with $\tilde{N}$ hidden nodes with activation function $g(x)$ can approximate these N samples with zero error, implying that $\sum_{j=1}^{\tilde{N}} \|O_j - t_j\| = 0$, i.e., there exist β_i, a_i and b_i such that

$$\sum_{i=1}^{\tilde{N}} \beta_i G(a_i \cdot X_j + b_i) = t_j, \quad j = 1, \dots, N \quad (2.40)$$

Equation 2.40 can be written compactly as

$$H\beta = T \quad (2.41)$$

where

$$H(a_1, \dots, a_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}, X_1, \dots, X_N) = \begin{bmatrix} G(a_1 \cdot X_1 + b_1) & \dots & G(a_{\tilde{N}} \cdot X_1 + b_{\tilde{N}}) \\ \vdots & \dots & \vdots \\ G(a_1 \cdot X_N + b_1) & \dots & G(a_{\tilde{N}} \cdot X_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}} \quad (2.42)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}$$

H is the hidden layer output matrix of the neural network; the i th column of H is the i th hidden node's output with respect to inputs $x_1, x_2, \dots, x_N$.

Theorem 1.1

Given a standard SLFN with N hidden nodes and activation function, $g : R \rightarrow R$ which is infinitely differentiable in any interval, for N arbitrary distinct samples (x_i, t_i) , where $x_i \in R^n$ and $t_i \in R^n$ for any a_i and b_i randomly chosen from any intervals of R^n and R , respectively, according to any continuous probability distribution, then with probability one, the hidden layer output matrix H of the SLFN is invertible and $\|H\beta - T\| = 0$.

Theorem 1.2

Given any small positive value $\epsilon > 0$ and activation function $g : R \rightarrow R$ which is infinitely differentiable in any interval, there exists $\tilde{N} \leq N$ such that for N arbitrary distinct samples (x_i, t_i) where $x_i \in R^n$ and $t_i \in R^m$, for any a_i and b_i randomly chosen

from any intervals of R'' and R , respectively, according to any continuous probability distribution, then with probability one, $\|H_{N \times \tilde{N}}\beta_{\tilde{N} \times m} - T_{N \times m}\| < \epsilon$.

From Theorems 1.1 and 1.2, the hidden node parameters a_i and b_i (input weights and hidden layer biases) of SLFNs can be randomly assigned during training with random values if only the activation function is infinitely differentiable.

If the number of N distinct training samples is equal to the number $\tilde{N}$ of hidden nodes, $N = \tilde{N}$, matrix H is square and invertible when the input weight vectors a_i and the hidden biases b_i are randomly chosen, and SLFNs can approximate these training samples with zero error.

When the number of hidden nodes is much less than number of distinct training samples $\tilde{N} < N$, H is a nonsquare matrix and there may not exist $a_i, b_i, \beta_i (i = 1, \dots, \tilde{N})$ such that $H\beta = T$.

The smallest norm least-squares solution of Equation 2.41, which is a linear system is

$$\hat{\beta} = H^\dagger T \quad (2.43)$$

where $H^\dagger$ is the Moore-Penrose generalised inverse of matrix H (Pollard, 1973; Serre, 2000). There are several methods of calculating the Moore-Penrose generalised inverse of H , including orthogonal projection, orthogonalisation method, iterative method and singular value decomposition (SVD) (Ortega and Ortega, 1987; Pollard, 1973). The orthogonal projection method can be used when $H^T H$ is nonsingular and $H^\dagger = (H^T H)^{-1} H^T$. However, $H^T H$ may tend to be singular in some applications. Therefore, the orthogonal projection method may not perform well in all applications.

The SVD can be used generally to calculate the Moore-Penrose generalised inverse of H in all cases. The minimum norm least-squares solution of a general linear system has the following properties.

1. Minimum training error

The special solution $\hat{\beta} = H^\dagger T$ is one of the least-squares solutions of a general linear system $H\beta = T$. This implies that the smallest training error can be reached by this special Solution: $\|H\hat{\beta} - T\| = \|HH^\dagger T - T\| = \min_{\beta} \|H\beta - T\|$

2. Smallest norm of weights

The special solution $\hat{\beta} = H^\dagger T$ has the smallest norm among all the least-squares solutions of $H\beta = T$

$$\|\hat{\beta}\| = \|H^\dagger T\| \leq \|\beta\|,$$

$$\forall \beta \in (\beta : \|HH^\dagger T - T\| \leq \|H\beta - T\|, \forall z \in R^{\tilde{N} \times N})$$

3. The minimum norm least-squares solution of $H\beta = T$ is unique, which is $\hat{\beta} = H^\dagger T$

The simple learning method for SLFNs referred to as extreme learning machine can be summarised as follows:

Given a training set $\aleph = ((x_i, t_i) | x_i \in R^n, t_i \in R^m, i = 1, \dots, N)$, activation function $g(x)$ and hidden node number $\tilde{N}$,

- Step 1: Randomly assign the parameters of hidden nodes, input weight a_i and bias $b_i, i = 1, \dots, N$
- Step 2: Calculate the hidden layer output matrix H
- Step 3: Calculate the output weight $\beta = H^\dagger T$

2.3.9 Online Sequential Extreme Learning Machine (OSELM)

The batch ELM assumes that all the training data (N samples) is a variable for training. However, in real applications, the training data may arrive chunk by chunk or one by one. For example, data from bankrupt banks is not available at the same time. Due to the financial crisis and other factors such as COVID-19, data may be available for other banks that are likely to collapse. Hence, the batch ELM algorithm has to be modified to accommodate additional data, which arrive later. This led to the introduction of a sequential learning algorithm referred to as online sequential extreme learning machine (OS-ELM).

The OSELM develops SLFNs with both addition and radial basis function (RBF) hidden nodes. The output weight matrix $\hat{\beta}(\hat{\beta} = H^\dagger T)$ given in Equation 2.43 is a least squares solution of Equation 2.41. Considering the case where $\text{rank}(H) = \tilde{N}$, the number of hidden nodes, $H^\dagger$ of Equation 2.43 is given by

$$H^\dagger = (H^T H)^{-1} H^T \quad (2.44)$$

This is also called the left pseudoinverse of H from the fact that $H^\dagger H = I_{\tilde{N}}$. Substituting Equation 2.44 into Equation 2.43, the estimation of β becomes

$$\hat{\beta} = (H^T H)^{-1} H^T T \quad (2.45)$$

Equation 2.45 is called the least-squares solution to $H\beta = T$. The sequential implementation of the least-squares solution of Equation 2.45 results in OSELM. OSELM consists of two steps, namely the initialisation step and the sequential learning phase.

Initialization phase:

From the given training set $\aleph = (x_i, t_i) | x_i \in R^n, t_i \in R^m, i = 1, \dots, N$, a small chunk of initial training data $N_0 = (x_i, t_i)_{i=1}^{N_0}$ is used to initialise the learning where $N_0 \geq \tilde{N}$.

- assign random input weight a_i and bias b_i (for additive hidden nodes) or center a_i and impact factor b_i (for RBF hidden nodes), where $i = 1, \dots, \tilde{N}$
- Calculate the initial layer output matrix H_0

$$\begin{bmatrix} G(a_1, b_1, X_1) & \dots & G(a_{\tilde{N}}, b_{\tilde{N}}, X_1) \\ \vdots & \dots & \vdots \\ G(a_1, b_1, X_{N_0}) & \dots & G(a_{\tilde{N}}, b_{\tilde{N}}, X_{N_0}) \end{bmatrix}_{N_0 \times \tilde{N}}$$
- estimate the initial output weight β^0
 $\beta^0 = P_0 H_0^T T_0$ where $P_0 = (H_0^T H_0)^{-1}$ and $T_0 = [t_1, t_2, \dots, t_{N_0}]^T$
- Set $k = 0$, where k is the number of chunks that is trianed currently

Sequential learning phase:

Present the $(k + 1)th$ chunk of new data/observation $N_{k+1} = (x_i, t_i)_{i=(\sum_{j=0}^k N_j)+1}^{\sum_{j=0}^{k+1} N_j}$

- calculate the partial hidden layer output matrix H_{k+1} for the $(k + 1)th$ chunk of data N_{k+1}

$$\begin{bmatrix} G(a_1, b_1, X_{(\sum_{j=0}^k N_j)+1}) & \dots & G(a_{\tilde{N}}, b_{\tilde{N}}, X_{(\sum_{j=0}^k N_j)+1}) \\ \vdots & \dots & \vdots \\ G(a_1, b_1, X_{\sum_{j=0}^{k+1} N_j}) & \dots & G(a_{\tilde{N}}, b_{\tilde{N}}, X_{\sum_{j=0}^{k+1} N_j}) \end{bmatrix}_{N_{k+1} \times \tilde{N}}$$

where H_{k+1} is calculated by the $(k + 1)th$ chunk of data $\aleph_{k+1}$

$$\aleph_{k+1} = (x_i, t_i)_{i=(\sum_{j=0}^k N_j)+1}^{\sum_{j=0}^{k+1} N_j}$$

- set
 $T_{k+1} = [t_{(\sum_{j=0}^k N_j)+1}, \dots, t_{\sum_{j=0}^{k+1} N_j}]^T$
- calculate the output weight matrix β^{k+1}

$$P_{k+1} = P_k - P_k H_{k+1}^T (I + H_{k+1} P_k H_{k+1}^T)^{-1} H_{k+1} P_k$$

$$\beta^{k+1} = \beta^k + P_{k+1} H_{k+1}^T (T_{k+1} - H_{k+1} \beta^k) \quad (2.46)$$

- Set $k = k + 1$. Go to sequential learning phase

The chunk size does not necessarily have to be constant(the same). The number N_{k+1} of training observations in the $(k + 1)th$ chunk does not need to be the same as N_k . Whenever the training data is received one-by-one instead of chunk-by-chunk, $N_{k+1} \equiv$

1. Equation 2.46 has the following simple format (Sherman-Morrison formula) (Golub and Van Loan, 2013).

$$P_{k+1} = P_k - \frac{P_k h_{k+1} h_{k+1}^T P_k}{1 + h_{k+1}^T P_k h_{k+1}}$$

$$\beta^{k+1} = \beta^k + P_{k+1} h_{k+1} (t_{k+1}^T - h_{k+1}^T \beta^k)$$

where $h_{k+1} = [G(a_1, b_1, X_{k+1}), \dots, G(a_N, b_N, X_{k+1})]$

2.3.10 Support vector Machine

Support vector machine (SVM), which is a classification and regression technique, is based on statistical learning (SL) theory and Vapnik-Chervonenkis structural risk minimisation (SRM) principle (Vapnik, 1999). It tries to find the trade-off between minimising the training errors and maximising the margin in order to achieve the best generalisation ability and remain resistant to over-fitting. The mapping function in SVM can either be a classification or a regression. SVM maps data from input features into some high-dimensional feature space through a non-linear mapping. SVM is based on mapping the sample points into a high-dimensional feature space in order to search and obtain an optimal separating hyperplane, which maximises the sum of the distances between two classes in this space. The data is modified from the input space into a high-dimensional feature space by using a kernel function. The kernel function decides the complexity of the classification function set, therefore affecting the performance of the SVM (Yuan and Chu, 2007). SVM is an optimisation technique in which prediction error and model complexities are simultaneously minimised. SVM uses convex quadratic programming that provides global optimal solutions and avoids being trapped in local minima. SVM is used to find an optimal hyperplane which maximises the margin between itself and the nearest training examples in the new high-dimensional space and minimises the expected generalisation error. The training examples that are closest to the hyperplane are called support vectors (Lee, 2007).

A simple description of the SVM algorithm is provided as follows. Given a training dataset $D = [(x_i, y_i)]_{i=1}^N$ with input vectors $x_i = (x_i^1, \dots, x_i^n)^T \in R^n$ and target labels $y_i \in -1, +1$, here x_i denotes attributes of the financial institution and y_i is the observed result of whether the financial institution is bankrupt or solvent. If a financial institution is bankrupt, then $y_i = -1$, else $y_i = 1$ (Altman and Hotchkiss, 1993). The SVM classifier, according to Vapnik's original formulation, satisfies the following conditions

$$\begin{cases} W^T \phi(x_i) + b \geq +1 & \text{if } y_i = +1, \\ W^T \phi(x_i) + b \leq -1 & \text{if } y_i = -1 \end{cases} \quad (2.47)$$

which is equivalent to

$$y_i[W^T\phi(x_i) + b] \geq 1, \quad i = 1, \dots, N \quad (2.48)$$

where W represents the weight vector and b the bias. Nonlinear function $\phi(\cdot) : R^n \rightarrow R^{n_k}$ maps input to a high-dimensional and possibly infinite-dimensional feature space. Equation 2.48 comes down to the construction of two parallel bounding hyperplanes at opposite sides of a separating hyperplane $W^T\phi(x) + b$ in the feature space with the margin width between both hyperplanes equal to $\frac{2}{||w||^2}$. In primal weight space, the classifier then takes the decision function form.

$$\text{sgn}(W^T\phi(x) + b) \quad (2.49)$$

In the real world, a training set is usually not linearly separable even when mapped into high-dimensional feature space, which implies we cannot find a perfect separating hyperplane that makes each x_i satisfy equation 2.47. Most classification problems are linearly non-separable. It is therefore general to find the weight vector using a slack variable (ξ_i) to permit misclassification. The optimisation problem is defined as

$$\text{minimize} \quad \frac{1}{2}W^TW + c \sum_{i=1}^N \xi_i \quad (2.50)$$

subject to

$$\begin{cases} y_i[W^T\phi(x_i) + b] \geq 1 - \xi_i & i = 1, \dots, N, \\ \xi_i \geq 0 & i = 1, \dots, N \end{cases} \quad (2.51)$$

where ξ_i 's are slack variables needed to allow misclassification in the set of inequalities/margin of error related to classification cost C , y_i are the classifications in the training set, $C \in R^+$ is a tuning hyperparameter or a regularization parameter, $\phi(x)$ transforms space R^N .

The solution to the primal problem is obtained after constructing the Lagrangian. From the conditions of optimality, we obtain a quadratic programming (QP) problem with Lagrange multipliers α_i 's. By applying the Lagrangian technique to Equation 2.51, we obtain

$$\text{maximise} \quad \frac{1}{2}\alpha^T Q \alpha - e^T \alpha \quad (2.52)$$

subject to

$$\begin{cases} 0 \leq \alpha_i \leq C & i = 1, \dots, N \\ y^T \alpha = 0 \end{cases}$$

e^T is the vector of all ones, Q is a $N \times N$ positive semi-definite matrix, α_i is the Lagrange multiplier, $Q_{ij} = y_i y_j K(x_i, x_j)$ where $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ is the kernel function. The training vectors x_i 's are mapped into a higher dimensional space by the function ϕ . The W is not calculated but made possible due to Mercer's condition which relates mapping function $\phi(x)$ to kernel function $K(\cdot, \cdot)$ as follows:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j). \quad (2.53)$$

The kernel function has different types including linear kernel of $K(x_i, x_j) = x_i^T x_j$, polynomial kernel of degree d of $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, r > 0$, radial basis function (RBF) kernel of $K(x_i, x_j) = \exp[-\gamma \|x_i - x_j\|^2], r > 0$ and sigma kernel of $K(x_i, x_j) = \tanh[\gamma x_i^T x_j + r]$ where $d, r \in N$ and $\gamma \in R^+$ are constants. The final SVM classifier is constructed as

$$\text{sgn}\left(\sum_i^N \alpha_i y_i K(x, x_i) + b\right) \quad (2.54)$$

2.3.10.1 Optimizing parameters of SVM

In addition to feature selection, model parameters can improve the SVM classification accuracy. Parameter values for the SVM algorithm are chosen in advance. These parameters include (1) the regularisation parameter, C , which determines the tradeoff cost between minimising the training error and the complexity of the model, (2) gamma, which defines the non-linear mapping from the input space to some high-dimensional feature space, and (3) degree. In this study, we use the grid search technique (Hsu, Chang, Lin, et al., 2003) using 10-fold cross-validation to determine the optimal parameter values of linear, sigmoid and polynomial kernel functions. We performed 10-fold cross-validation to choose the parameters of C (0.01325, 0.125, 0.5, 1, 3), γ =(1,2,3,4) and degree=(1,2,3). Different values of C are tried for the linear kernel, pairs of C and gamma for the sigmoid kernel and C , gamma and degree for the polynomial kernel function and the one with the best accuracy is chosen.

2.4 Result and Discussion

2.4.1 Introduction

In this section, the results of the data analysis, as described in the methodology section, are discussed. The results of the experimental analysis with all machine learning algorithms discussed in the previous section are presented. We also compared the performance of each machine learning technique. A set of feature selection methods, such as simulated annealing, genetic algorithm, tabu search, recursive feature elimination (RFE), Boruta algorithm and learning vector quantisation (LVQ) were applied to select the most important features in predicting bankruptcy. The features selected by each of the feature selection techniques are presented in Table 2.3. We used K-fold (10) cross-validation for training each of the classifiers after re-sampling the data with the SMOTE technique to solve class imbalance.

The first part presents descriptive statistics of the financial ratios used. We compared the means of the financial ratios between bankrupt and solvent banks. We also used Mann Mann-Whitney test to determine the significant differences in financial ratios between bankrupt and solvent banks. The second part discussed the findings of the computational machine learning models and also determined whether there are any significant differences in terms of predictive accuracy amongst the machine learning models using different feature selection methods. The last part of this chapter compared the different machine learning models in order to determine the overall effective model for bankruptcy prediction for the US financial industry.

2.4.2 Empirical results

2.4.2.1 Descriptive statistics

We checked the normality of the features using the Shapiro-Wilk test. The results indicated that the financial ratios were not normally distributed; therefore, we used the non-parametric Mann-Whitney U test to analyse the mean differences between failed and solvent banks. Table 2.5 presents the mean, standard deviation and Mann-Whitney U test for the financial ratios for solvent and failed banks. The last two columns present the Mann-Whitney test results used to compare the differences between failed and solvent banks, along with their corresponding significance levels. The mean of the financial ratios was calculated for the bankrupt and solvent banks. A comparison of the summary statistics for failed banks and solvent banks provided results that are expected. We find remarkable differences between the mean ratios for failed and solvent banks.

Results indicated that 32 out of the 33 financial ratios showed a statistically significant difference between failed and solvent banks at a 5% level. The 32 financial ratios serve

as early warning indicators that can distinguish between bankrupt and solvent banks. Therefore, these 32 financial ratios can be described as early warning ratios (EWRs). As shown in Table 2.5, the mean of the EWR ratios differ between bankrupt and solvent banks. There was a statistically significant difference in all the capital ratios. There was no statistically significant difference in the average net loan-to-total assets ratio for both failed and solvent banks. The lack of statistical significance in the average net loan-to-total assets ratio for solvent and bankrupt banks in the US may result from a combination of regulatory and industry practices, as well as the inherent nature of bank failures themselves. The net loan-to-total assets ratio measures the quantity of loans relative to the total assets of the banks, but not the quality of the loans. Bankrupt banks may have a higher proportion of non-performing loans than solvent banks, although the total number of loans is similar. A bankrupt and a solvent company may have the same or similar net loan-to-asset ratios but differ in their management of liquidity risk, thereby yielding different results. The similarity in the net loans-to-assets ratio between bankrupt and solvent banks may be due to regulatory and risk management frameworks in the US.

Bankrupt banks have lower capital adequacy ratios, which makes them vulnerable to economic shocks, while the solvent banks often maintain higher capital adequacy. The lower capital quality values for the bankrupt banks may be a result of structural weaknesses in their finances. The solvent banks most likely have lower exposure to non-performing loans or do not invest in more risky assets compared to bankrupt banks. The solvent banks are more likely to comply with Basel III capital requirements, resulting in a healthier capital adequacy ratio compared to bankrupt banks. It is possible that bankrupt banks ignored the deteriorating levels of their capital quality, hoping to recover in the future, while solvent banks act early when stress is detected. The asset quality mean ratios for solvent banks were much lower than those of bankrupt banks, suggesting that solvent banks have more effective credit risk management than bankrupt banks. Additionally, solvent banks may have lower exposure to high-risk lending as a result of stricter policies, leading to fewer non-performing loans, unlike bankrupt banks, which may have an aggressive lending practice. It is most likely that solvent banks have more diversified asset portfolios.

Additionally, the failed banks have a high percentage of loan loss reserve ratio, loan loss provision to net interest revenue, non-performing loans to gross loans, and net charge-offs to average gross loans, relative to banks classified as solvent (Table 2.5). For example, the average loan loss reserve ratio for failed banks was 4.165% relative to 1.131% for solvent banks. In comparison, the average loan loss provision to net interest revenue for failed banks was 151.764% relative to 5.337% for solvent banks. These findings confirm similar results obtained by (Cleary and Hebb, 2016). The lower values of loan loss reserve to gross loan indicate that the solvent banks were safer to invest in, while the higher loan loss reserve to gross loan values for the bankrupt banks suggest

that the banks have a higher risk profile. It is therefore not suitable for investors to invest in banks with a higher loan loss reserve to gross loan ratio. Bankrupt banks not only have a higher loan loss provision to net interest revenue ratio than solvent banks, but also have a smaller reservoir of assets. The proportion of debt owed by bankrupt banks that is unlikely to be repaid is higher than that of solvent banks. Comparatively, the net charge-off-to-average-gross-loan ratio for bankrupt banks is significantly higher than that of solvent banks. Although bankrupt banks have less capacity to meet their obligations, they tend to incur more debt than solvent banks.

The values of the financial ratios indicate that solvent banks continue to grow year after year, while bankrupt banks decline year after year, ultimately leading to failure. This is demonstrated in the higher capital ratios (equity/total assets, equity/net loans, equity/deposit and short term funding, equity/liabilities, capital funds/total assets, capital funds/net loans, capital funds/deposit and short term funding and capital funds/liabilities) for solvent banks compared to lower values for bankrupt banks. The mean capital ratios for solvent banks are consistently higher than those for bankrupt banks. All capital ratios for failed banks were lower than those of solvent banks. Results indicated that for one year before failure, all the capital quality ratios (equity/total assets, equity/net loans, equity/deposit and short term funding, equity/liabilities, capital funds/total assets, capital funds/net loans, capital funds/deposit and short term funding and capital funds/liabilities) for solvent banks were significantly higher than that of failed banks. For instance, the equity-to-total-assets ratio for solvent banks was 11.648% relative to 3.037% for failed banks. The equity ratio measures the amount of leverage used by a bank. The low equity ratio for the failed banks suggests that these banks typically used debt to acquire assets, indicating a higher level of financial risk. The higher equity ratios for the solvent banks suggest that these banks effectively funded their assets with a minimal amount of debt. The higher equity ratio for solvent banks indicates that they are financially stronger and tend to have a longer-term solvency compared to banks with lower equity ratios. The equity/net loans for solvent banks was higher compared to failed banks, indicating that banks with more capitalisation are less likely to fail, unlike banks with less capitalisation. Capital ratios have proven to be significant determinants of bankruptcy prediction.

Some of the financial ratios (non-interest expense/average assets, non-operating items and taxes/average assets and Cost/income ratio) were significantly lower for solvent banks. The failed banks have negative ROAA and ROAE, indicating a weak balance sheet, while the solvent banks have positive ROAA and ROAE. This is indicative of lower profitability. The cost-to-income ratio for solvent banks was 60.878% and that for failed banks was 196.771%. The cost-to-income ratio measures the efficiency of the bank. It is a ratio which displays the bank's cost as a proportion of its income and compares the operating expenses of the bank to the income generated. A lower cost-to-income ratio is indicative of better performance, profitability, and competitiveness for

the bank. The lower cost-to-income ratio for solvent banks is an indication that these banks were more profitable and performed better than failed banks. The dividend pay-out ratio for solvent banks was 46.369% while that for failed banks was -1.740%. The dividend pay-out ratio determines the amount of dividends paid to shareholders of the bank relative to the income generated by the bank. The solvent banks, on average, paid 46.369% of their net income to their shareholders. Banks with a negative dividend pay-out ratio may need to utilise existing cash or raise funds to pay their dividends.

Some of the liquidity ratios (net loans/total assets, liquid assets/total deposit and borrowing, and liquid assets/deposit and short-term funding) were higher for the failed banks and others (net loans/deposit and short-term funding and net loans/total deposit and borrowing) were higher for solvent banks. The net loan-to-total asset ratio (risk ratio) for solvent banks was 66.255%, while the risk ratio for failed banks was 68.466%. The higher risk ratio for the failed banks indicates that these banks are loaned up and have low liquidity. The higher ratio indicates that the bank is riskier and has a higher tendency to default. Generally, the failed banks had lower values for loan loss reserve/non-performing loans, equity/total assets, equity/net loans, equity/deposit and short term funding, equity/liabilities, capital funds/total assets, capital funds/net loans, capital funds/deposit and short term funding, capital funds/liabilities, net interest margin, net interest revenue/average assets, other operating income/average assets, pre-tax operating income/average assets, ROAA, ROAE, income net of distribution/average equity, recurring earning power, net loans/deposit and short term funding, net loans/total deposit and borrowing and dividend pay out. The failed banks have higher values for loan loss reserve/gross loan, loan loss provision/net interest revenue, non-performing loans/gross loans, net charge off/average gross loan, Net charge off/net income before loan loss provision, impaired loans/equity, unreserved impaired loans/equity, non-interest expense/average assets, non-operating items and taxes/average assets, Cost/income, net loans/total asset, liquid assets/total deposit and borrowing and liquid assets/deposit and short term funding.

The values of the ratios loan loss reserve to gross loan, loan loss provision to net interest revenue, non-performing loans to gross loans, and net charge-off to average gross loan indicate that solvent banks continue to grow year after year. In contrast, bankrupt banks decline year after year before ultimately failing. This is demonstrated in the higher capital ratios (equity/total assets, equity/net loans, equity/deposit and short term funding, equity/liabilities, capital funds/total assets, capital funds/net loans, capital funds/deposit and short term funding and capital funds/liabilities) for solvent banks compared to lower values for the bankrupt banks. The means of loan loss reserve to gross loan, loan loss provision to net interest revenue, non-performing loans to gross loans, and net charge-off to average gross loan decreased over the years, indicating a

strong financial statement for the banks. The mean of all capital ratios (equity/total assets, equity/net loans, equity/deposit and short term funding, equity/liabilities, capital funds/total assets, capital funds/net loans, capital funds/deposit and short term funding and capital funds/liabilities) increased year after year for the solvent banks. These capital ratios are valuable for bank regulators, supervisors, and bank analysts. These ratios serve as a guideline to assess the adequacy of the banks' levels. A significant capital-to-total assets ratio provides a measure of the bank's appropriate capital level.

Table 2.5: Descriptive statistics for the financial ratios for solvent and failed banks

Ratio	Solvent		Failed		U	Sig.
	Mean	SD	Mean	SD		
X1	1.131	0.738	4.165	2.413	27791	0.000
X2	5.337	7.766	151.764	340.084	42602	0.000
X3	579.985	4919.982	42.63	59.217	508078	0.000
X4	0.899	1.311	14.409	9.163	15363	0.000
X5	0.268	0.684	3.769	3.245	34235	0.000
X6	9.319	17.405	117.562	6299.118	415737	0.000
X7	5.194	8.918	1498.516	21471.003	44421	0.000
X8	-1.562	9.296	1206.668	18126.799	58032	0.000
X9	11.648	3.037	3.491	2.903	545984	0.000
X10	20.467	20.375	5.107	4.295	544107	0.000
X11	14.252	5.389	3.854	3.299	545944	0.000
X12	13.369	4.462	3.714	3.158	546055	0.000
X13	11.866	3.087	3.586	2.895	546174	0.000
X14	20.898	20.703	5.238	4.261	544433	0.000
X15	14.544	5.569	3.959	3.301	545899	0.000
X16	13.62	4.521	3.814	3.155	546174	0.000
X17	3.722	1.875	2.984	0.986	375337	0.000
X18	3.365	1.592	2.633	0.898	393538	0.000
X19	1.292	1.868	0.301	2.269	436817	0.000
X20	3.082	1.977	7.412	3.659	40444	0.000
X21	1.575	0.769	-4.478	3.261	547941	0.000
X22	-0.419	0.334	-0.162	1.809	85880	0.000
X23	1.183	0.594	-4.575	3.388	546131	0.000
X24	10.557	5.334	-97.059	141.409	540604	0.000
X25	5.847	5.215	-97.948	141.028	542687	0.000
X26	60.878	12.389	196.771	487.07	39690	0.000
X27	1.808	1.178	-1.109	1.896	533338	0.000
X28	66.255	16.033	68.466	10.802	276119	0.818
X29	79.837	25.117	73.913	13.111	353200	0.000
X30	76.538	18.485	71.675	12.029	360889	0.000
X31	9.968	19.524	12.49	7.892	155426	0.000
X32	8.987	13.697	12.172	7.803	155688	0.000
X33	46.369	84.239	-1.74	59.266	504075.5	0.000

The details regarding the interpretations of X1-X33 is found in [Table 2.2](#)

2.4.2.2 Pairwise comparison of the financial ratios for bankrupt and solvent banks

Table 2.6 presents the Mann-Whitney U test for the bankrupt banks. The analysis compared the financial ratios during the year of failure with those for the three years preceding bankruptcy. For each of the financial ratios, there was a statistically significant difference between the ratios in the year of bankruptcy declaration and those in each of the four years preceding bankruptcy. This suggests that the financial ratios of the bankrupt banks were deteriorating at least four years before the year the banks collapsed. The deterioration in specific financial ratios for bankrupt banks is pronounced over the three years preceding failure (Table 2.6). The difference in means is evident for at least three years preceding failure, with the difference increasing as the year of failure approaches, for loan loss reserve/gross loan, loan loss provision/net interest revenue, non-performing loans/gross loans, and net charge off/average gross loan. The mean values for loan loss reserve/non-performing loans, equity/total assets, equity/net loans, equity/deposit and short term funding, equity/liabilities, capital funds/total assets, capital funds/net loans, capital funds/deposit and short term funding and capital funds/liabilities decreased over the years until failure of the banks. The bankrupt banks displayed signs of failure or weakness in their financial statement as early as three years before bankruptcy. The results provide a deeper understanding of the financial position of financial institutions at least four years before failure. A careful examination of this trend could lead to the identification of early warning signs of financial distress. This finding suggests that rather than focusing solely on the current state of financial institutions, it is essential to evaluate and monitor the trajectory of their financial status.

Table 2.6: Mann-Whitney U test for failed banks years before failure

Ratio	Failed	Lag 1	P-value	Lag 2	P-value	Lag 3	P-value
X1	4.165	2.823	0.000	1.924	0.000	1.543	0.000
X2	151.764	74.981	0.000	39.076	0.000	19.191	0.000
X3	42.63	238.912	0.000	314.436	0.000	1880.482	0.000
X4	14.409	8.605	0.000	4.813	0.000	2.42	0.000
X5	3.769	1.825	0.000	1.023	0.000	0.448	0.000
X6	117.562	-41.128	0.000	159.666	0.000	-21.621	0.000
X7	1498.516	98.659	0.000	59.078	0.000	22.992	0.000
X8	1206.668	71.089	0.000	36.819	0.000	9.511	0.000
X9	3.491	7.469	0.000	9.488	0.000	10.699	0.000
X10	5.107	10.506	0.000	13.147	0.000	15.754	0.000
X11	3.854	8.599	0.000	11.271	0.000	13.178	0.000
X12	3.714	8.225	0.000	10.777	0.000	12.673	0.000
X13	3.586	7.553	0.000	9.579	0.000	10.771	0.000
X14	5.238	10.619	0.000	13.263	0.000	15.844	0.000
X15	3.959	8.698	0.000	11.38	0.000	13.264	0.000
X16	3.814	8.316	0.000	10.879	0.000	12.753	0.000
X17	2.984	3.46	0.000	3.897	0.000	4.156	0.000
X18	2.633	3.151	0.000	3.603	0.000	3.858	0.000
X19	0.301	0.677	0.001	0.727	0.000	0.871	0.000
X20	7.412	5.66	0.000	4.665	0.000	4.211	0.000
X21	-4.478	-1.832	0.000	-0.334	0.000	0.518	0.000
X22	-0.162	-0.068	0.000	-0.207	0.000	-0.315	0.000
X23	-4.575	-1.884	0.000	-0.497	0.000	0.274	0.000
X24	-97.059	-32.655	0.000	-15.155	0.000	3.141	0.000
X25	-97.948	-35.144	0.000	-18.377	0.000	-0.01	0.000
X26	196.771	99.109	0.000	81.243	0.000	83.364	0.000
X27	-1.109	0.138	0.000	0.785	0.000	1.127	0.000
X28	68.466	72.822	0.000	74.331	0.000	73.738	0.000
X29	73.913	82.535	0.000	86.235	0.000	86.484	0.000
X30	71.675	79.596	0.000	83.103	0.000	83.476	0.000
X31	12.49	9.454	0.000	8.858	0.000	9.894	0.000
X32	12.172	9.193	0.000	8.64	0.000	9.681	0.000
X33	-1.74	7.744	0.004	17.111	0.000	23.328	0.000

The details regarding the interpretations of X1-X33 is found in Table 2.2

2.4.3 Support Vector Machine

We trained benchmark SVM models using linear, sigmoid and polynomial kernels on all the variables available and also on variables selected by feature selection techniques (simulated annealing, genetic algorithm, tabu search, recursive feature elimination, stepwise logistic regression, LVQ and Boruta algorithm). Different variable selection techniques were employed to identify the most significant variables for predicting bankruptcy. Table 2.3 displays the variables selected by each of the variable selection techniques. The benchmark models were trained using default parameters for each of the kernels.

2.4.3.1 SVM-Linear kernel

We compared the performance of the SVM model using the linear kernels with different feature selection techniques. We used various performance metrics, including accuracy, F1 score, kappa, AUC, and ROC curve, to compare the models. The classification of banks using SVM models is performed on both training and test datasets. The accuracy of all the SVM models with different features on the training dataset exceeded 98%, indicating the practical and strong predictive abilities of the SVM models. Table 2.7 shows the classification accuracies, sensitivity, specificity, kappa, precision, F1 score, MCC and AUC of various feature selection techniques using the SVM-linear kernel model for both in-sample and out-of-sample data.

It is evident from Table 2.7 that the SVM-linear model with all features and the SVM-linear model using the Boruta feature selection technique yield the best-performing outcome in terms of prediction power, with an accuracy rate of 98.86% and 27 misclassified banks on the training data. This is followed by an SVM-linear kernel model, which uses a stepwise regression feature selection technique with an accuracy of 98.74%. The lowest-performing SVM-linear kernel model utilises the SA feature selection technique, achieving an accuracy of 98.02% and an F1 score of 0.9609. Although all the models performed well on the training data, it is evident that SVM-Boruta and SVM-All variables outperformed the other models in terms of accuracy, F1 score, kappa, MCC, and AUC. It is therefore concluded that SVM-boruta and SVM-all variables yield the best performance in terms of accuracy, F1 score, MCC, kappa and AUC for the linear kernel. The type I and type II errors for the SVM-boruta and SVM using all variables are 1.83% and 0.43%, respectively. The best-performing model on the training data had an overall misclassification rate of 1.14%. That is, our optimal forecasting model for the linear kernel is achieved using SVM with Boruta and SVM with all variables. The in-sample forecasting accuracy for the solvent banks is 98.17% (1181 out of 1203), while that for the failed banks is 99.57% (1164 out of 1169). The overall in-sample forecasting accuracy of the SVM-Boruta model is 98.86% (2345 out of 2372). The model produced 27 errors in the training data: 22 solvent banks were misclassified as failed, and five failed banks were misclassified as solvent. The SVM-LVQ model performed better on the test data, even though all the other models performed well, too. For the SVM-LVQ model, the accuracy is 95.14%, with an F1 score of 0.8742 and an AUC of 0.9603. The type I error and type II error for the SVM-LVQ on the test data are 5.85% and 0.00%, respectively. Additionally, the SVM-LVQ model, which is the most predictive model on the test data, has an overall misclassification rate of 4.86%. The optimal model minimises the sum of both Type I and Type II errors.

Summarising the results across all metrics for both training and testing data, it is evident that SVM-LVQ exhibits higher discriminatory power compared to the other models. The SVM-LVQ classifier yielded the best classification results compared with the

different classifiers. The out-of-sample forecasting accuracy for the solvent bank is 94.15% (306 out of 325), while that for the failed banks is 100.00% (66 out of 66). The overall out-of-sample forecasting accuracy of the SVM-LVQ model is 95.14% (372 out of 391). The model produced 19 errors in the test data: 19 solvent banks were misclassified as failed banks. While the SVM-LVQ had a higher Type I error than SVM-GA for the test data, it had a lower Type II error than SVM-GA. For banks, a false negative (Type II error) is considered to be more costly than a false positive. Although SVM-GA, SVM-RFE, SVM-stepwise regression, and SVM-Boruta had a slightly lower Type I error rate than SVM-LVQ for the training data, their performance on the test data was significantly poor. The confusion matrix for each of the SVM-linear classifiers is presented in Table 2.8. As shown in Table 2.8, the SVM-linear models with the different feature selection techniques did quite well in classifying solvent and failed banks. The misclassification rates of solvent banks in the training data range from 1.83% to 3.59%, while those of failed banks range from 0.26% to 0.43%. Additionally, the misclassification rates of solvent banks in the test data range from 4.55% to 21.24%, while those of failed banks range from 0.00% to 40.0%.

Table 2.7: Predictive performance of SVM-Linear models using different feature selection methods

	Accuracy	Sensitivity	Specificity	Kappa	Precision	F1 score	MCC	AUC
Training model								
ALL variables	0.9886	0.9957	0.9817	0.9772	0.9815	0.9885	0.9773	0.9983
SA	0.9802	0.9974	0.9641	0.9604	0.9629	0.9798	0.9609	0.9969
GA	0.9865	0.9966	0.9769	0.973	0.9764	0.9864	0.9732	0.9973
Tabu Search	0.984	0.9957	0.9728	0.968	0.9722	0.9838	0.9682	0.9973
RFE	0.9861	0.9957	0.9768	0.9722	0.9764	0.986	0.9724	0.9982
Stepwise	0.9874	0.9957	0.9793	0.9747	0.9789	0.9872	0.9748	0.998
LVQ	0.9848	0.9957	0.9744	0.9696	0.9739	0.9847	0.9699	0.9974
Boruta	0.9886	0.9957	0.9817	0.9772	0.9815	0.9885	0.9773	0.9983
Testing model								
ALL variables	0.9003	1.000	0.887	0.6486	0.5412	0.7023	0.6928	0.9353
SA	0.8082	1.000	0.8032	0.1727	0.1177	0.2105	0.3074	0.935
GA	0.9335	0.8554	0.9545	0.8029	0.8353	0.8452	0.803	0.8911
Tabu Search	0.7852	0.6000	0.7876	0.0436	0.0353	0.0667	0.1056	0.781
RFE	0.8849	1.000	0.8718	0.5818	0.4706	0.64	0.6405	0.9356
Stepwise	0.9258	1.000	0.9134	0.7514	0.6588	0.7943	0.7758	0.9418
LVQ	0.9514	1.000	0.9415	0.8446	0.7765	0.8742	0.855	0.9603
Boruta	0.9003	1.000	0.887	0.6486	0.5412	0.7023	0.6928	0.9353

Note: SA-Simulated annealing, GA-genetic algorithm, RBE-recursive feature elimination, LVQ- learning vector quantization, MCC- Matthews Correlation Coefficient, AUC-Area under curve

Table 2.8: Performance of SVM-linear classifiers on training and testing datasets

			Training data		Test data	
Features selection techniques			Predicted		Predicted	
			Bankrupt	Solvent	Bankrupt	Solvent
Actual	All variables	Bankrupt	1164	5	46	0
		Solvent	22	1181	39	306
	SA	Bankrupt	1142	3	10	0
		Solvent	44	1183	75	306
	GA	Bankrupt	1158	4	71	12
		Solvent	28	1182	14	294
	Tabu search	Bankrupt	1153	5	3	2
		Solvent	33	1181	82	304
	RFE	Bankrupt	1158	5	40	0
		Solvent	28	1181	45	306
	Stepwise	Bankrupt	1161	5	56	0
		Solvent	25	1181	29	306
	LVQ	Bankrupt	1155	5	66	0
		Solvent	31	1181	19	306

Note: SA-Simulated annealing, GA-genetic algorithm, RBE-recursive feature elimination, LVQ- learning vector quantization

Figures 2.4 and 2.5 show the receiver-operating characteristic (ROC) curves for the linear kernel on the training and testing data, respectively, for each of the feature selection techniques. The ROC curve plots the true positive rate against the false positive rate. The true positive rate represents the fraction of positive cases that the model correctly classifies. In contrast, the false positive rate represents the fraction of negative instances that are incorrectly classified as positive. These curves are created by plotting the sensitivity (true positive rate) against 1-specificity (false positive rate) at various thresholds. The curve provides a visual tool to evaluate the performance of machine learning models. The ROC is an efficient method of expressing the diagnostic abilities of models. It provides diagnostic information as well as the area under the curve. We used the AUC of the ROC curve to assess the performance of the models. The larger the area under the ROC curve, the better the classifier. For the ROC curve (Figure 2.4) of the training data, it is difficult to confirm a single preferred technique, as all the models produced very high AUC values and identical curves that overlap. However, SVM-Boruta, SVM-all variables, SVM-RFE, SVM-stepwise and SVM-LVQ were the most promising classifiers. The ROC curve (Figure 2.5) demonstrates that the SVM-linear model with LVQ feature selection technique performs better on the test data than the other models. The classifier relies less on chance, achieving an AUC score of 0.9603, which is acceptable.

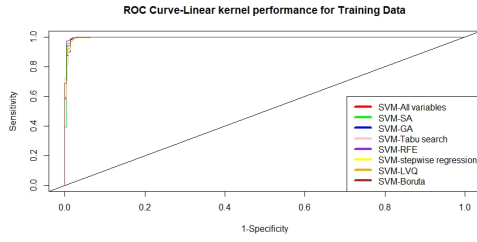


Figure 2.4: ROC curves for SVM-linear on training data for different classifiers

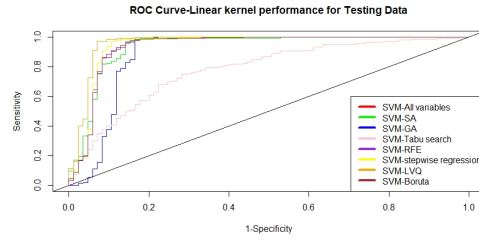


Figure 2.5: ROC curves for SVM-linear on test data for different classifiers

We conducted a McNemar test to test the hypothesis that none of the models performs better than the other. The performance of each model is compared using a statistical method. It determines whether there is a statistically significant difference between two machine learning classifiers. It is a non-parametric test for two related samples, used for before-and-after measurements of the same subjects (Kim, 2003). Table 2.9 presents the results of the McNemar test, which statistically compares the classification accuracy of different variable selection techniques for the SVM-linear model on the hold-out data. As shown in Table 2.9, SVM-LVQ outperforms SVM-GA, SVM-Stepwise, SVM-ALL, SVM-RFE, SVM-SA and SVM-tabu at a 5% statistical level of significance. Results also indicate that SVM-GA outperforms all the other models, except SVM-LVQ, at a 5% significance level. The prediction performance between SVM-SA and SVM-Tab does not differ significantly.

Table 2.9: McNemar test for pairwise performance comparison for SVM-Linear Kernel

	SA	GA	Tabu	RFE	Stepwise	LVQ
All variables	0.00	0.00	0.00	0.04	0.00	0.00
SA		0.00	0.18	0.00	0.00	0.00
GA			0.00	0.00	0.00	0.01
Tabu				0.00	0.00	0.00
RFE					0.00	0.00
Stepwise						0.02

Note: SA-Simulated annealing, GA-genetic algorithm, RBE-recursive feature elimination, LVQ- learning vector quantization

2.4.3.2 SVM-Sigmoid Kernel

Table 2.10 presents the classification accuracies, sensitivity, specificity, kappa, precision, F1 score, MCC, and AUC of SVM with a sigmoid kernel using various feature selection techniques. All the models performed well on the training models as seen in Table 2.10. From the results, SVM-LVQ with a sigmoid kernel performed better than the other models in terms of accuracy, sensitivity, specificity, kappa, precision, F1 score, MCC, and AUC. The most optimal model had Type I and Type II errors of 3.84% and 2.81%,

respectively, for the training data. The overall misclassification rate for SVM-LVQ for the training data is 3.33%. The in-sample forecasting accuracy for the solvent banks is 96.16% (1153 out of 1199), while that for the failed banks is 97.19% (1140 out of 1173). The overall in-sample forecasting accuracy of the SVM-LVQ model is 96.67% (2293 out of 2372) with a total of 79 errors in the training data. The worst-performing models on the test data for the SVM-sigmoid kernel are SVM-LVQ and SVM-GA, with accuracies of 34.27% and 20.97%, respectively. Although SVM-Tabu, SVM-RFE, SVM-stepwise, SVM-Boruta, and SVM-all variable performed well on the test data, SVM-SA achieved the best performance with higher accuracy, AUC, F1 score, MCC, and kappa. The SVM-SA had Type I and Type II errors of 2.08% and 2.02%, respectively, on the test data, resulting in an overall misclassification rate of 7.42%. The out-of-sample forecasting accuracy of the SVM-SA sigmoid kernel is 92.58% (362 out of 391). The out-of-sample accuracy for the solvent banks is 97.92% (283 out of 289), and that for the failed banks is 77.45% (79 out of 102). The optimal forecasting model (for both training and test data) for the sigmoid kernel function is achieved when the feature selection technique is SA. The SVM-SA has the highest discriminatory power among the other models. SVM-SA did well in classifying solvent and failed banks for the out-of-sample data.

Figures 2.6 and 2.7 show the ROC curves for the SVM-sigmoid kernel on the training and test data, respectively, for each of the different features. From figure 2.6, it is evident that SVM-LVQ had the greatest AUC compared to the rest of the models for the training data. On the other hand, SVM-SA is the best model for the out-of-sample data, as indicated by the ROC curve and AUC. According to the ROC curve for the out-of-sample data, SVM-LVQ is the worst model with an AUC of 0.2961, followed by SVM-GA with an AUC of 0.5957.

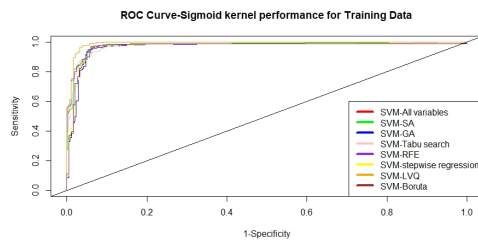


Figure 2.6: ROC curves for SVM-sigmoid on training data for different classifiers

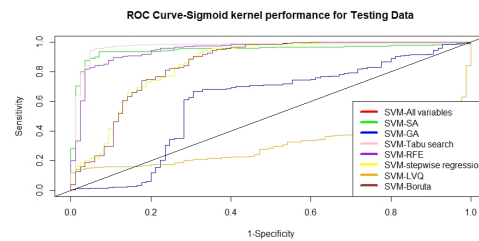


Figure 2.7: ROC curves for SVM-sigmoid on test data for different classifiers

Table 2.10: Predictive performance of SVM-Sigmoid models using different feature selection methods

	Accuracy	Sensitivity	Specificity	Kappa	Precision	F1 score	MCC	AUC
Training model								
ALL variables	0.9503	0.9588	0.9421	0.9005	0.941	0.9498	0.9007	0.969
SA	0.9473	0.9546	0.9402	0.8946	0.9393	0.9469	0.8947	0.9797
GA	0.9427	0.9442	0.9412	0.8853	0.941	0.9426	0.8853	0.9646
Tabu	0.9317	0.9384	0.9252	0.8634	0.9241	0.9312	0.8635	0.9748
RFE	0.9494	0.9517	0.9471	0.8988	0.9469	0.9493	0.8988	0.9755
Stepwise	0.9507	0.9596	0.9421	0.9013	0.941	0.9502	0.9015	0.9686
LVQ	0.9667	0.9719	0.9616	0.9334	0.9612	0.9665	0.9334	0.9921
Boruta	0.9503	0.9588	0.9421	0.9005	0.941	0.9498	0.9007	0.969
Testing model								
ALL variables	0.757	0.4667	0.9378	0.4405	0.8235	0.5957	0.4728	0.851
SA	0.9258	0.7745	0.9792	0.7967	0.9294	0.8449	0.8024	0.9468
GA	0.2097	0.2005	0.4118	-0.0425	0.8824	0.3268	-0.1917	0.5957
Tabu	0.8824	0.9333	0.8757	0.5835	0.4941	0.6462	0.626	0.967
RFE	0.8645	1	0.8524	0.4859	0.3765	0.547	0.5665	0.9493
Stepwise	0.7545	0.4615	0.9234	0.4211	0.7765	0.5789	0.4495	0.8528
LVQ	0.3427	0.1587	0.6763	-0.13	0.4706	0.2374	-0.1915	0.2961
Boruta	0.757	0.4667	0.9378	0.4405	0.8235	0.5957	0.4728	0.851

Note: SA-Simulated annealing, GA-genetic algorithm, RBE-recursive feature elimination, LVQ- learning vector quantization

Table 2.11: Performance of SVM-sigmoid classifiers on training and testing datasets

			Training data		Test data	
Features selection techniques			Predicted		Predicted	
			Bankrupt	Solvent	Bankrupt	Solvent
Actual	All variables	Bankrupt	1116	48	70	80
		Solvent	70	1138	15	226
	SA	Bankrupt	1114	53	79	23
		Solvent	72	1133	6	283
	GA	Bankrupt	1116	66	75	299
		Solvent	70	1120	10	7
	Tabu search	Bankrupt	1096	72	42	3
		Solvent	90	1114	43	303
	RFE	Bankrupt	1123	57	32	0
		Solvent	63	1129	53	306
	Stepwise	Bankrupt	1116	47	66	77
		Solvent	70	1139	19	229
	LVQ	Bankrupt	1140	33	40	212
		Solvent	46	1153	45	94

Note: SA-Simulated annealing, GA-genetic algorithm, RBE-recursive feature elimination, LVQ- learning vector quantization

Table 2.12 shows the results of the McNemar test to statistically compare the performance of SVM using the sigmoid activation on the different features selected on the hold-out data. Results, as shown in Table 2.12, indicate that SVM-SA outperforms all other models at a 5% significance level. Additionally, SVM-tabu search outperforms the different models, except for SVM-SA. The SVM-ALL model outperforms SVM-LVQ and SVM-GA at a 5% significance level and outperforms SVM-Stepwise at a 10% significance level. The classification accuracy between SVM-All and SVM-stepwise does not differ significantly at a 5% significance level.

Table 2.12: McNemar test for pairwise performance comparison for SVM-Sigmoid Kernel

	SA	GA	Tabu	RFE	Stepwise	LVQ
All variables	0.00	0.00	0.00	0.00	0.05	0.00
SA		0.00	0.00	0.00	0.00	0.00
GA			0.00	0.00	0.00	0.00
Tabu				0.02	0.00	0.00
RFE					0.00	0.00
Stepwise						0.00

Note: SA-Simulated annealing, GA-genetic algorithm, RBE-recursive feature elimination, LVQ- learning vector quantization

2.4.3.3 SVM-Polynomial kernel

The performance of the models using the polynomial kernel on the training data, in terms of predictive accuracy, ranges from 97.43% to 98.02%. Table 2.13 compares the classification performance of SVM-SA, SVM-GA, SVM-tabu, SVM-RFE, SVM-Stepwise, SVM-LVQ and SVM-Boruta in both training and test data and shows that SVM-tabu outperforms SVM-SA, SVM-GA, SVM-RFE, SVM-Stepwise regression, SVM-LVQ and SVM-Boruta by 0.59%, 0.51%, 0.25%, 0.34%, 0.25% and 0.34%, respectively. The forecasting accuracy for the solvent banks is 96.87% (1177 out of 1215), and that for the failed banks is 99.22% (1148 out of 1157). Thus, the model misclassifies 38 solvent banks and nine failed banks. Table 5 shows that when the feature selection method is the tabu search algorithm, the SVM model with a polynomial kernel achieves an accuracy of 98.02%, an overall misclassification error of 1.98%, and an AUC of 0.9949. The Type I and Type II errors for the SVM-tabu search algorithm are 3.13% and 0.78%, respectively. The most accurate classification model for the test data, using the polynomial kernel function, was obtained through the tabu search algorithm feature selection technique. The SVM-tabu search for the test data achieves an accuracy of 97.44%, with an F1 score of 0.9419 and an AUC of 0.9686. The forecasting accuracy for the solvent banks is 98.68% (300 out of 304), and that for the failed banks is 93.10% (81 out of 87). The overall misclassification rate for the test data is 2.56%. The Type I and Type II errors for the test data are 1.32% and 0.52%, respectively. Our optimal forecasting model (for both training and test data) for the polynomial kernel function is achieved when the feature selection technique is tabu search. The SVM-tabu search has the highest discriminatory power among the models. Even though the ROC curves for SVM-SA, SVM-RFE and SVM-LVQ for the training data were marginally better than that of SVM-tabu search, other performance metrics (accuracy, kappa, F1 score, precision and MCC) from SVM-tabu search are better compared to SVM-SA, SVM-RFE and SVM-LVQ. The AUCs of SVM-SA, SVM-RFE and SVM-LVQ are 0.9952, 0.9962 and 0.9957, respectively, while that of SVM-tabu search is 0.9949. For the test data, the ROC curve and AUC for the SVM-tabu search are superior to those of the other models. According to the ROC curve for the out-of-sample data, SVM-GA is the worst model with an AUC of 0.8836, followed by SVM-RFE with an AUC of 0.8948.

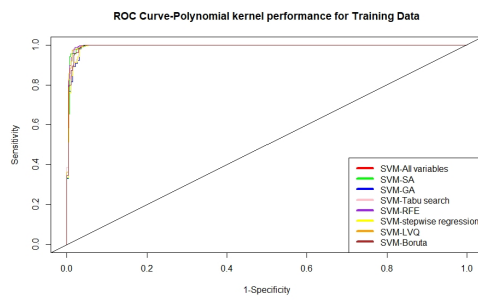


Figure 2.8: ROC curves for SVM-Polynomial on training data for different classifiers

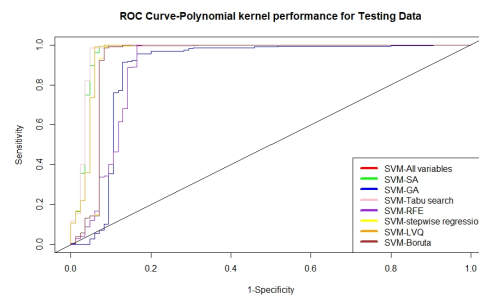


Figure 2.9: ROC curves for SVM-Polynomial on test data for different classifiers

Table 2.13: Predictive performance of SVM-Polynomial models using different feature selection methods

	Accuracy	Sensitivity	Specificity	Kappa	Precision	F1 score	MCC	AUC
Training model								
ALL variables	0.9768	0.9974	0.9579	0.9536	0.9562	0.9763	0.9544	0.9949
SA	0.9743	0.9947	0.9555	0.9486	0.9536	0.9737	0.9494	0.9952
GA	0.9751	0.9982	0.9541	0.9503	0.9519	0.9745	0.9513	0.9933
Tabu	0.9802	0.9922	0.9687	0.9604	0.968	0.9799	0.9607	0.9949
RFE	0.9777	0.9965	0.9602	0.9553	0.9587	0.9772	0.956	0.9962
Stepwise	0.9768	0.9974	0.9579	0.9536	0.9562	0.9763	0.9544	0.9934
LVQ	0.9777	0.9913	0.9647	0.9553	0.9637	0.9773	0.9557	0.9957
Boruta	0.9768	0.9974	0.9579	0.9536	0.9562	0.9763	0.9544	0.9949
Testing model								
ALL variables	0.9642	0.961	0.965	0.8911	0.8706	0.9136	0.8928	0.9341
SA	0.9668	1	0.9592	0.8966	0.8471	0.9172	0.9014	0.9673
GA	0.6905	0.4043	0.9557	0.3674	0.8941	0.5568	0.436	0.8836
Tabu	0.9744	0.931	0.9868	0.9255	0.9529	0.9419	0.9256	0.9686
RFE	0.9591	1	0.9503	0.871	0.8118	0.8961	0.8783	0.8948
Stepwise	0.9693	0.962	0.9712	0.9074	0.8941	0.9268	0.9084	0.9343
LVQ	0.9565	1	0.9474	0.8623	0.8	0.8889	0.8706	0.9597
Boruta	0.9642	0.961	0.965	0.8911	0.8706	0.9136	0.8928	0.9341

Note: SA-Simulated annealing, GA-genetic algorithm, RBE-recursive feature elimination, LVQ- learning vector quantization, MCC- Matthews Correlation Coefficient, AUC-Area under curve

Table 2.14: Performance of SVM-polynomial classifiers on training and testing datasets

			Training data		Test data	
Features selection techniques			Predicted		Predicted	
			Bankrupt	Solvent	Bankrupt	Solvent
Actual	All variables	Bankrupt	1134	3	74	3
		Solvent	52	1183	11	303
	SA	Bankrupt	1131	6	72	0
		Solvent	55	1180	13	306
	GA	Bankrupt	1129	2	76	112
		Solvent	57	1184	9	194
	Tabu search	Bankrupt	1148	9	81	6
		Solvent	38	1177	4	300
	RFE	Bankrupt	1137	4	69	0
		Solvent	49	1182	16	306
	Stepwise	Bankrupt	1134	3	76	3
		Solvent	52	1183	9	303
	LVQ	Bankrupt	1143	10	68	0
		Solvent	43	1176	17	306

Note: SA-Simulated annealing, GA-genetic algorithm, RBE-recursive feature elimination, LVQ- learning vector quantization

Table 2.15 shows the results of the McNemar test to statistically compare the performance of SVM using the polynomial activation on the different features selected on the hold-out data. SVM-tabu outperforms all the other classifiers at 5% significance level. Additionally, SVM-All outperforms SVM-RFE and SVM-LVQ at a 5% significance level. However, there was no statistically significant difference in performance among SVM-All variables, SVM-SA and SVM-Stepwise. Additionally, the prediction performance of SVM-SA, SVM-RFE, and SVM-LVQ, all of which use a polynomial activation function, does not differ significantly from one another.

Table 2.15: McNemar test for pairwise performance comparison for SVM-Polynomial Kernel

	SA	GA	Tabu	RFE	Stepwise	LVQ
ALL variables	0.23	0.00	0.01	0.01	0.48	0.04
SA		0.00	0.00	0.55	0.10	0.22
GA			0.00	0.00	0.00	0.00
Tabu				0.00	0.03	0.00
RFE					0.00	1.00
Stepwise						0.02

Note: SA-Simulated annealing, GA-genetic algorithm, RBE-recursive feature elimination, LVQ- learning vector quantization

2.4.4 Extreme Gradient Boosting Machine (XGBoost)

This section compares the predictive accuracy of XGBoost using different feature selection techniques. The hyperparameters of the XGBoost model were determined using a grid search method with 10-fold cross-validation based on the AUC score. Table 2.16 presents the optimised hyperparameters for the XGBoost model. The results of the tuned XGBoost models are presented in Table 2.17 for both training and test data, using 10-fold cross-validation for all features and those selected by different feature selection techniques. Table 2.18 presents the metric based on the confusion matrix for the different XGBoost models. The XGBoost technique, combined with the stepwise feature selection technique, achieved higher accuracy for the training data compared to the other techniques. The accuracy of the model with stepwise feature selection is 99.96% (2371 out of 2372), with a sensitivity of 100% and a specificity of 99.92%. The AUC yields a value of 0.9996, which is relatively close to the maximum value of 1. This suggests that most bankruptcy predictions are correct. The model using the stepwise feature selection technique achieved a better F-measure than the other models. Generally, the XGBoost technique with a stepwise feature selection approach performed better than all other methods, as measured by various performance metrics (accuracy, AUC, F1-score, kappa, MCC, sensitivity, Type I and II errors). The overall accuracy of XGBoost-stepwise is higher than that of other techniques. The in-sample prediction for solvent banks is 99.92% (1185 out of 1186), and that for failed banks is 100% (1186 out of 1186), with a 0.08% Type I error rate and 0.00% Type II error rate. The XGBoost technique, combined with a stepwise selection approach, achieved precision scores of 99.92% and an F1-score of 99.96%. As shown in Table 2.17, XGBoost-stepwise achieved the highest AUC score of 99.96%.

The XGBoost model with all features and the XGBoost model with the tabu search technique achieved the same results. The overall in-sample forecasting accuracy is 99.75% for both the XGBoost model, which uses all variables, and the model that uses features selected through the tabu search feature technique. For these two classifiers, the in-sample prediction accuracy for solvent banks is 99.92% (1185 out of 1186) and 99.58% for failed banks (1181 out of 1186). These two models yielded Type I error rates of 0.08% and Type II error rates of 0.42%. Additionally, XGBoost-SA, XGBoost-GA, XGBoost-RFE, and XGBoost-LVQ techniques yielded the same results with the training data. The in-sample forecast for solvent banks is 99.92% (1185 out of 1186), and that of failed banks is 99.24% (1177 out of 1186). The overall in-sample forecast is 99.58% for the XGBoost-SA, XGBoost-GA, and XGBoost-RFE classifiers, with Type I errors of 0.0843 and Type II errors of 0.76%. The XGBoost classifier, combined with genetic algorithm, tabu search, and LVQ feature selection techniques, yielded the highest prediction accuracy for out-of-sample data. The overall out-of-sample forecasting accuracy for XGBoost-GA, XGBoost-Tabu, and XGBoost-LVQ is 98.21%, with 384 banks correctly predicted out of a total of 391. The out-of-sample forecasting accuracy for

solvent banks is 99.35% (304 out of 306), and that of failed banks is 94.12% (80 out of 85). The models had a misclassification error of 0.08% for Type I and 0.76% for Type II. The XGBoost-GA, XGBoost-Tabu and XGBoost-LVQ classifiers misclassified two solvent banks and five failed banks. XGBoost-RFE achieved an accuracy of 97.95% with an F-score of 0.9524 and a kappa of 0.9394. This technique misclassified three solvent banks and five failed banks. It has a precision of 0.9639, with Type I and Type II error rates of 0.9804 and 5.8824, respectively.

XGBoost with all variables, XGBoost-SA, and XGBoost-stepwise regression techniques yield the exact prediction for the out-of-sample data. The overall accuracy is 97.70% with an F-score of 0.9467 and MCC of 0.9321. The out-of-sample prediction accuracy for failed banks is 94.12% (80 out of 85) and 98.69% (302 out of 306) for solvent banks. The sensitivity and specificity of these models are 0.9412 and 0.9869, respectively. For the XGBoost classifier with all variables, the misclassification ratio for bankrupt banks in the test data is 5.88% and that for solvent banks is 1.31%. The misclassification rate for bankrupt banks is higher than that of solvent banks. In a similar vein, the misclassification ratios for bankrupt banks in the test data using XGBoost-GA, XGBoost-Tabu search, and XGBoost-LVQ classifiers are 5.88% and 0.65% for solvent banks. The error in classifying bankrupt banks is higher compared to the error in classifying solvent banks when using XGBoost-GA, XGBoost-Tabu search and XGBoost-LVQ.

Table 2.16: Optimised hyperparameters for XGBoost model

Name of hyperparameter	Value
learning rate	0.02
Number of estimators	200
Subsample	1
Maximum depth	7
Gamma	0
Minimum child weight	1
Colsample by tree	0.5
Alpha	10

Table 2.17: Predictive performance of XGBoost models using different feature selection methods

	Accuracy	Sensitivity	Specificity	Precision	F1 score	MCC	Kappa	AUC
Training model								
XGBoost	99.75	0.9958	0.9992	0.9992	0.9975	0.9949	0.9949	0.9975
XGBoost-SA	99.58	0.9924	0.9992	0.9992	0.9958	0.9916	0.9916	0.9958
XGBoost-GA	99.58	0.9924	0.9992	0.9992	0.9958	0.9916	0.9916	0.9958
XGBoost-rfe	99.58	0.9924	0.9992	0.9992	0.9958	0.9916	0.9916	0.9958
XGBoost-tabu	99.75	0.9958	0.9992	0.9992	0.9975	0.9949	0.9949	0.9975
XGBoost-stepwise	99.96	1	0.9992	0.9992	0.9996	0.9992	0.9992	0.9996
XGBoost-LVQ	99.58	0.9924	0.9992	0.9992	0.9958	0.9916	0.9916	0.9958
Testing model								
XGBoost	97.7	0.9412	0.9869	0.9524	0.9467	0.9321	0.9321	0.9768
XGBoost-SA	97.7	0.9412	0.9869	0.9524	0.9467	0.9321	0.9321	0.9658
XGBoost-GA	98.21	0.9412	0.9935	0.9756	0.9581	0.947	0.9467	0.9648
XGBoost-RFE	97.95	0.9412	0.9902	0.9639	0.9524	0.9395	0.9394	0.9756
XGBoost-Tabu	98.21	0.9412	0.9935	0.9756	0.9581	0.947	0.9467	0.9700
XGBoost-Stepwise	97.7	0.9412	0.9869	0.9524	0.9467	0.9321	0.9321	0.9753
XGBoost-LVQ	98.21	0.9412	0.9935	0.9756	0.9581	0.947	0.9467	0.9723

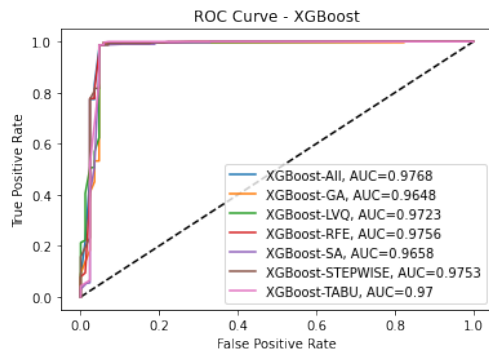
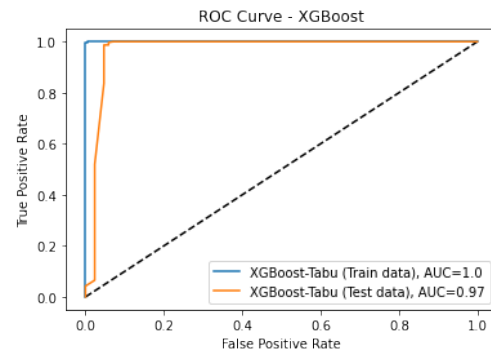
Note: SA-Simulated annealing, GA-genetic algorithm, RBE-recursive feature elimination, LVQ- learning vector quantization

Table 2.18: Performance of XGBoost classifiers on training and testing datasets

			Training data		Test data	
Features selection techniques			Predicted		Predicted	
			Bankrupt	Solvent	Bankrupt	Solvent
Actual	All variables	Bankrupt	1181	5	80	5
		Solvent	1	1185	4	302
	SA	Bankrupt	1177	9	80	5
		Solvent	1	1185	4	302
	GA	Bankrupt	1177	9	80	5
		Solvent	1	1185	2	304
	Tabu search	Bankrupt	1181	5	80	5
		Solvent	1	1185	2	304
	RFE	Bankrupt	1177	9	80	5
		Solvent	1	1185	3	303
	Stepwise	Bankrupt	1186	0	80	5
		Solvent	1	1185	4	302
	LVQ	Bankrupt	1177	9	80	5
		Solvent	1	1185	2	304

Note: SA-Simulated annealing, GA-genetic algorithm, RBE-recursive feature elimination, LVQ- learning vector quantization

From Figure 2.10, which displays the ROC curve for each of the models using the test data, it isn't easy to select a single classifier as the most preferred model since all the ROC curves overlap. Additionally, the AUC for each model is well above 0.95, indicating that all models performed well on the test data. XGBoost with all features had the highest AUC as shown in Figure 2.10. The ROC curves for XGBoost using all features, XGBoost-Stepwise, XGBoost-RFE, and XGBoost-SA are shown in Appendix A.

**Figure 2.10:** ROC curves for XGBoost on test data for different classifiers**Figure 2.11:** ROC curves for XGBoost-Tabu on training and test data

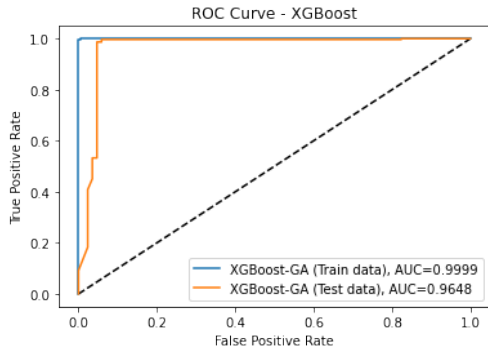


Figure 2.12: ROC curves for XGBoost-GA on training and test data

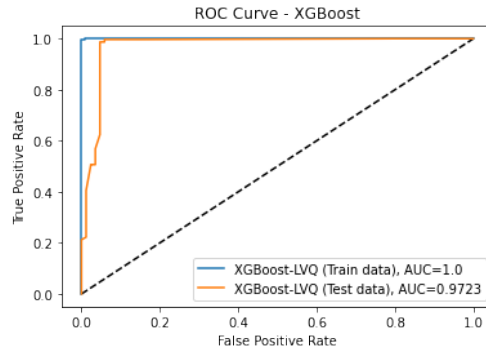


Figure 2.13: ROC curves for XGBoost-LVQ on training and test data

2.4.4.1 Feature Importance

Figures 2.14, 2.15, 2.16 and 2.17 present feature importance plots for bankruptcy prediction using XGBoost-Stepwise, XGBoost-Tabu, XGBoost-GA and XGBoost-LVQ classifiers. From the feature importance plots (Figs. 2.15, 2.16, and 2.17), pre-tax operating income/average assets is the most important feature that contributes to the prediction of bankruptcy. The features pre-tax operating income/average assets, impaired loans/equity, equity/net loans, ROAE and loan loss reserve/gross loan are the top 5 features that provided the highest forecasting accuracy using XGBoost-GA classifier while pre-tax operating income/average assets, equity/net loans, impaired loans/equity, capital funds/net loans and income net of distribution/average equity are the top 5 features for predicting bankruptcy using XGBoost-LVQ classifier. For the top 10 features identified by XGBoost-Ga, XGBoost-Tabu, and XGBoost-LVQ classifiers, there are common features (pre-tax operating income/average assets, impaired loans/equity, capital funds/net loans, capital funds/total assets, and non-performing loans/gross loans) that are shared by all three classifiers, as well as disjoint features. Pre-tax operating income divided by average assets, which is a key metric, is a measure of operational efficiency. The pre-tax operating income-to-average ratio is an important metric in determining the health status of a financial institution. This ratio indicates how effectively a financial institution is utilising its assets to generate earnings. The second most important variable is impaired loans to equity, a measure of asset quality. It measures the asset quality of a financial institution.

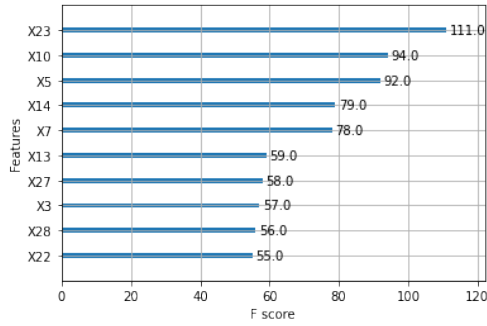


Figure 2.14: Feature importance using XGBoost-Stepwise

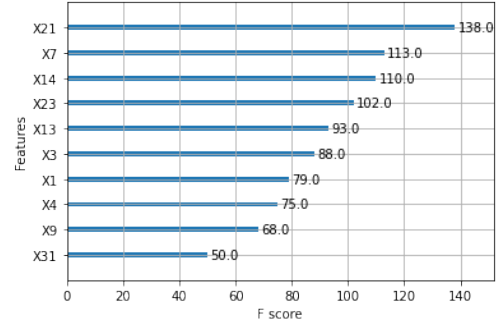


Figure 2.15: Feature importance using XGBoost-Tabu

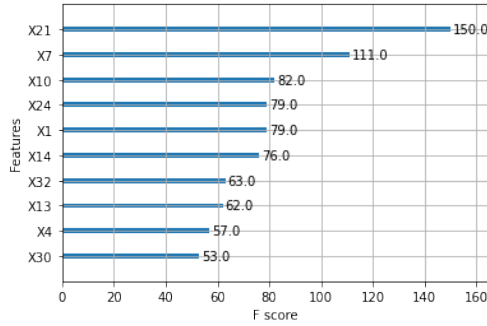


Figure 2.16: Feature importance using XGBoost-GA

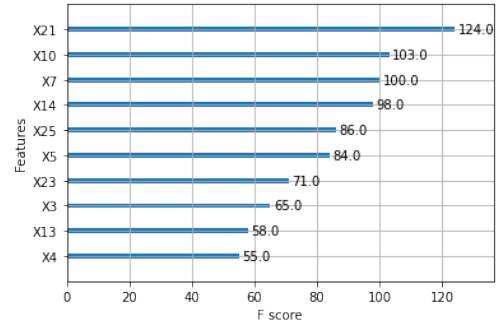


Figure 2.17: Feature importance using XGBoost-LVQ

2.4.4.2 Feature importance and SHAP Value

The impact of features in a tree model, such as XGBoost, is typically represented using a bar chart to illustrate global feature importance or partial dependence plots to show the effect of changing a feature. The SHAP summary plots replace typical bar charts of global feature importance, and the SHAP dependence plot provides an alternative to partial dependence plots. The standard feature importance bar charts give the relative importance of the features. Still, they are unable to represent the range and distribution of impacts the feature has on the model output. For the SHAP summary plot, features are sorted by their global implications, while dots representing the SHAP values are plotted horizontally. Each dot is coloured by the value of the feature from low (blue) to high (red).

From our results, XGBoost-GA, XGBoost-Tabu and XGBoost-LVQ achieved the highest prediction accuracy. Figures 2.19, 2.20, and 2.21 present the SHAP bar graphs for the three best models. Each of the models presents the first ten important features for bankruptcy prediction. As shown in the bar graphs of the three models, there is a set

of common as well as disjoint features; however, the overall ranking of the first ten features is similar. Features such as pre-tax operating income to average assets, impaired loans to equity, capital funds to net loans, capital funds to total assets, equity to total assets, non-performing loans to gross loans, and ROAE were identified as the most important features by all three models for bankruptcy prediction. It is worth noting that pre-tax operating income/average assets, as well as impaired loans to equity, were ranked as the first and second most important features by all three models.

We compared the top 10 important features between the existing method and the SHAP method. However, there were few changes in the ranking for the first 10 features; seven features in the top 10 matched for XGBoost-Tabu using the inbuilt feature importance and SHAP approach. According to Lundberg and Lee (2017), the feature importance obtained using the gain approach is inconsistent, making it challenging to compare feature attribution values reliably. The SHAP summary plot indicates that larger values of pre-tax operating income/average assets derive a lower risk of bankruptcy because red dots are plotted on the higher side while blue points are plotted at the lower side. The first 10 features extracted using the SHAP summary plot are consistent with expectations from risk analysts.

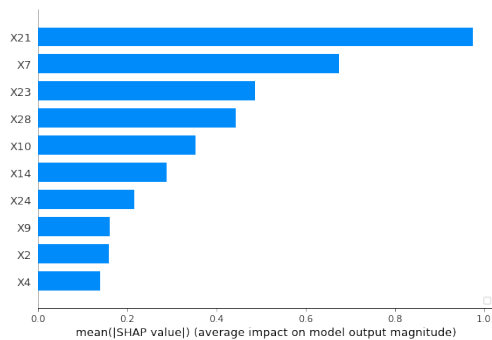


Figure 2.18: Mean absolute SHAP values of XGBoost - All features

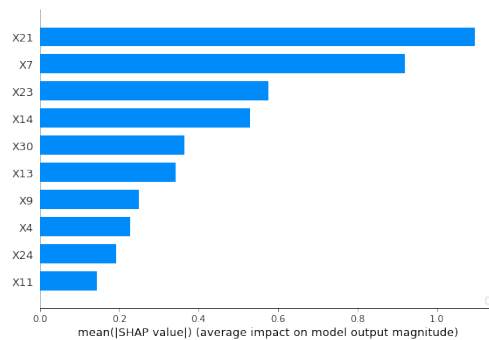


Figure 2.19: Mean absolute SHAP values of XGBoost - Tabu

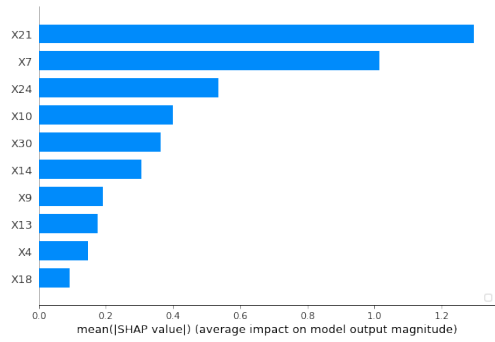


Figure 2.20: Mean absolute SHAP values of XGBoost - GA

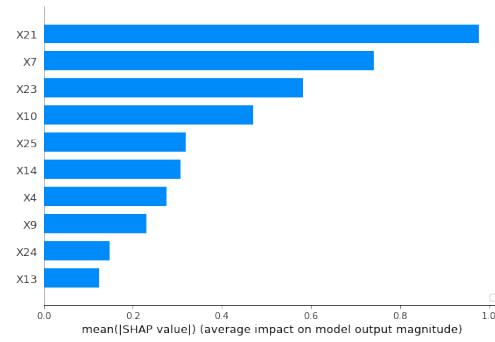


Figure 2.21: Mean absolute SHAP values of XGBoost - LVQ

Figures 2.22, 2.23, 2.24, and 2.25 present the SHAP summary plots for XGBoost-stepwise, XGBoost-Tabu, XGBoost-GA and XGBoost-LVQ, respectively. The top two essential features, according to the three models, are pre-tax operating income to average assets and impaired loans to equity. Higher values of pre-tax operating income/average assets lead to solvency of the banks, while lower values tend to lead to bankruptcy according to the SHAP graph. On the contrary, lower values of impaired loans to equity lead to solvency, while higher values lead to bankruptcy.

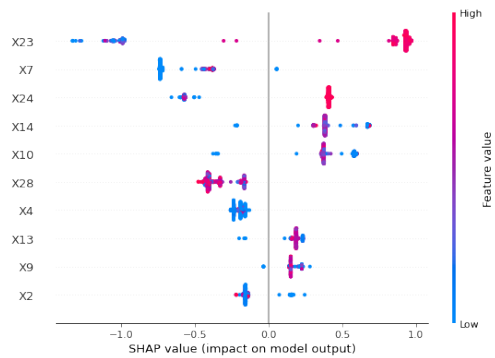


Figure 2.22: SHAP summary plot of XGBoost-Stepwise

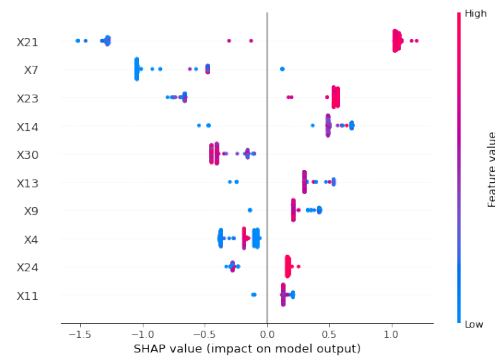


Figure 2.23: SHAP summary plot of XGBoost-Tabu

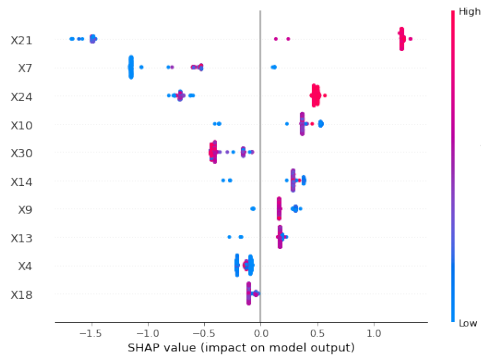


Figure 2.24: SHAP summary plot of XGBoost-GA

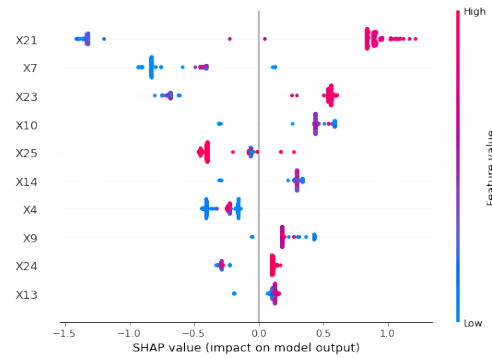


Figure 2.25: SHAP summary plot of XGBoost-LVQ

An analysis of the random sample bank from the test data is presented in a force plot (Figure 2.27) for XGBoost-Tabu, illustrating which features are associated with bankruptcy prediction. The graph illustrates the contribution of each attribute to the outcome's base value (excluding samples). It displays features that help move the model output above its base value, which is the average model output across the training dataset. Features that have a favourable impact on the prediction are highlighted in red, while those that have a negative impact are highlighted in blue. It demonstrates how several characteristics affect whether a bank will be solvent or fail. It displays the relationship between attributes and specific predictions. The features with the greatest impact are labelled after each group of features is sorted according to the strength of their influence. Whereas the base value (0.01262) is the mean forecast based on the test data, the output value (1.25) is the predicted value for the bank. According to Figure 2.27, pre-tax operating income/average assets, ROAA, capital funds/net loans, capital funds/total assets, equity/total assets, ROAE and equity/deposit and short term funding help to raise values over the basic value, whereas impaired loans/equity, net loans/total deposit and borrowing, cost/income and non-performing loans/gross loans lower values. Hence, whereas impaired loans/equity, net loans/total deposit and borrowing, cost/income and non-performing loans/gross loans have an adverse effect on the prediction, pre-tax operating income/average assets, ROAA, capital funds/net loans, capital funds/total assets, equity/total assets, ROAE and equity/deposit and short term funding have a good effect.



Figure 2.26: SHAP Force plot using XGBoost-Stepwise



Figure 2.27: SHAP Force plot using XGboost-Tabu



Figure 2.28: SHAP Force plot using XGboost-GA



Figure 2.29: SHAP Force plot using XGboost-LVQ

Table 2.19 presents the results of the McNemar test, which statistically compares the classification accuracy of different variable selection techniques for the XGBoost model on the hold-out data. Results show that there is no statistically significant difference in prediction performance among the XGBoost models using different variable selection techniques at a 5% significance level.

Table 2.19: McNemar test for pairwise performance comparison using XGBoost model

	SA	GA	Tabu	RFE	Stepwise	LVQ
ALL	NA	0.48	0.48	1.00	NA	0.48
SA		0.48	0.48	1.00	NA	0.48
GA			NA	1.00	0.48	NA
Tabu				1.00	0.48	NA
RFE					1.00	1.00
Stepwise						0.48

Note: SA-Simulated annealing, GA-genetic algorithm, RBE-recursive feature elimination, LVQ- learning vector quantization

2.4.5 Light Gradient Boosting Machine (LightGBM)

The hyperparameters of the LightGBM model were determined by using the Tree Parzen Estimator method with a 10-fold cross-validation based on the AUC score. The optimised hyperparameters are presented in Table 2.20.

Table 2.20: Optimised hyperparameters for LightGBM model

Name of hyperparameter	Value
learning rate	0.02
Number of estimators	200
Subsample	1
Maximum depth	7
Gamma	0
Minimum child weight	1
Colsample by tree	0.5
Alpha	10

Table 2.21 presents the outcomes of machine learning models for both training and testing data, utilising the LightGBM classifier with variables selected by various feature selection methods and all variables. The confusion matrix displayed in Table 2.22 presents the predictive performance of the different classifiers for both the training and

test sets. LightGBM using all variables, LightGBM-RFE and LightGBM-LVQ had the highest accuracy, precision, F1 score, MCC, kappa and AUC for the training data. The overall prediction accuracy is 99.96% (2371 out of 2372), with a precision of 99.92%, an F1 score of 0.9996, and an AUC score of 0.9996. The in-sample prediction for solvent banks is 99.92% (1185 out of 1186), and that for failed banks is 100% (1186 out of 1186), with a precision of 0.9992 and a kappa of 0.9992. The model produced a Type I error rate of 0.08% and a Type II error rate of 0.00%. LightGBM-Tabu and LightGBM-GA classifiers achieved an accuracy of 99.92% with a sensitivity of 0.9983. The in-sample prediction for failed banks is 100% (1186 out of 1186), and that for solvent banks is 99.83% (1184 out of 1186). LightGBM with all variables and LightGBM-RFE had a better out-of-sample accuracy of 98.47%, precision of 98.77%, F1 score of 0.9639 and an AUC of 0.9690. The out-of-sample forecast for failed banks using both LightGBM-RFE and LightGBM with all features is 94.12% (80 out of 85) and 99.67% (305 out of 306) for solvent banks. The out-of-sample prediction for LightGBM, using other feature selection techniques, yields performance metrics above 90% for all metrics. Table 2.21 presents the performance metrics of the different LightGBM models.

Figure 2.30 shows the ROC curve of each model for the test data. LightGBM-RFE and LightGBM, which utilise all features, demonstrate significant superiority over other models. While it is difficult to confirm a single preferred model from the curves, LightGBM-RFE and LightGBM using all features are the most promising candidates. Figures 2.32 and 2.31 display the ROC curves for LightGBM-RFE and LightGBM, using all features for both training and testing data.

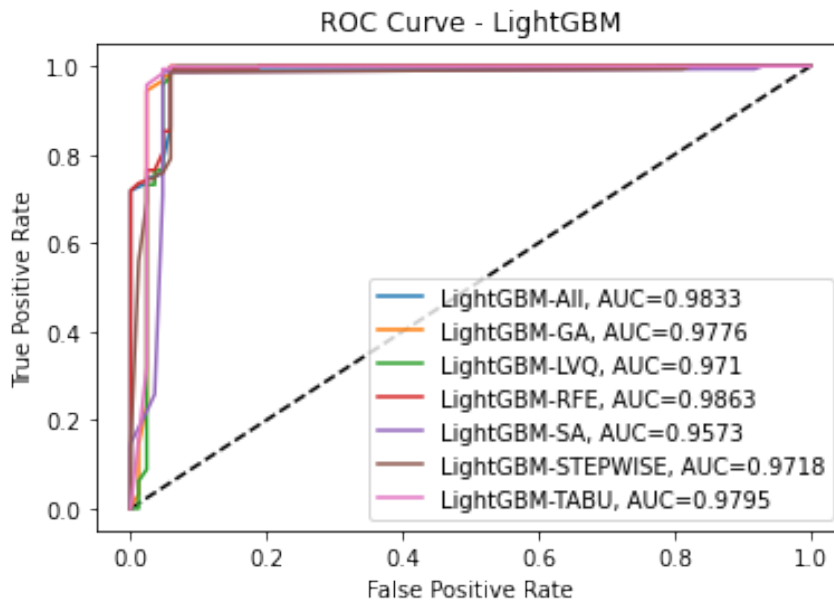


Figure 2.30: ROC curves for LightGBM models on test data

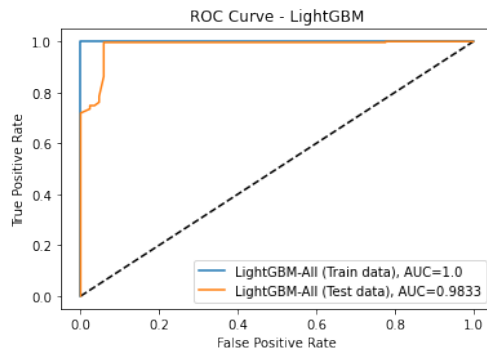


Figure 2.31: ROC curves for LightGBM using all features for training and test data

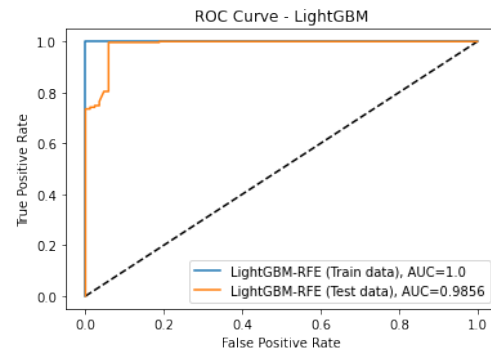


Figure 2.32: ROC curves for LightGBM-RFE on training and test data

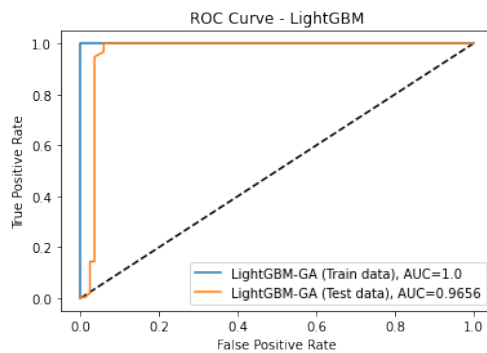


Figure 2.33: ROC curves for LightGBM-GA on training and test data

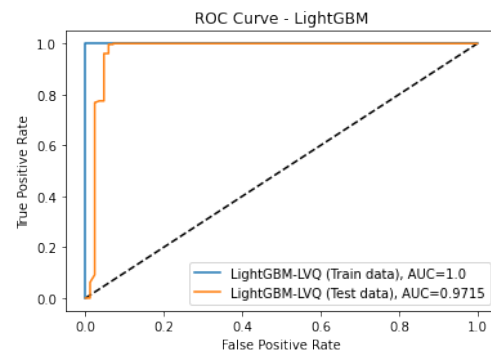


Figure 2.34: ROC curves for LightGBM-LVQ on training and test data

Table 2.21: Predictive performance of LightGBM models using different feature selection methods

	Accuracy	Sensitivity	Specificity	Precision	F1 score	MCC	Kappa	AUC
Training model								
LightGBM	99.96	1.0000	0.9992	0.9992	0.9996	0.9992	0.9992	0.9996
LightGBM -SA	99.7	0.9966	0.9975	0.9975	0.997	0.9941	0.9941	0.997
LightGBM -GA	99.92	1.0000	0.9983	0.9983	0.9992	0.9983	0.9983	0.9992
LightGBM -rfe	99.96	1.0000	0.9992	0.9992	0.9992	0.9992	0.9992	0.9996
LightGBM -tabu	99.92	1.0000	0.9983	0.9983	0.9992	0.9983	0.9983	0.9992
LightGBM -stepwise	99.87	1.0000	0.9975	0.9975	0.9987	0.9975	0.9975	0.9987
LightGBM-LVQ	99.96	1.0000	0.9992	0.9992	0.9996	0.9992	0.9992	0.9996
Testing model								
LightGBM	98.47	0.9412	0.9967	0.9877	0.9639	0.9546	0.9541	0.9833
LightGBM -SA	97.7	0.9412	0.9869	0.9524	0.9467	0.9321	0.9321	0.9573
LightGBM -GA	98.21	0.9412	0.9935	0.9756	0.9581	0.947	0.9467	0.9776
LightGBM -RFE	98.47	0.9412	0.9967	0.9877	0.9639	0.9546	0.9541	0.9863
LightGBM -Tabu	98.21	0.9412	0.9935	0.9756	0.9581	0.947	0.9467	0.9795
LightGBM -Stepwise	97.7	0.9412	0.9869	0.9524	0.9467	0.9321	0.9321	0.9718
LightGBM-LVQ	98.21	0.9412	0.9935	0.9756	0.9581	0.947	0.9467	0.9710

Note: SA-Simulated annealing, GA-genetic algorithm, RBE-recursive feature elimination, LVQ- learning vector quantization

Table 2.22: Performance of LightGBM classifiers on training and testing datasets

			Training data		Test data	
Features selection techniques			Predicted		Predicted	
			Bankrupt	Solvent	Bankrupt	Solvent
Actual	All variables	Bankrupt	1186	0	80	5
		Solvent	1	1185	1	305
	SA	Bankrupt	1182	4	80	5
		Solvent	3	1183	4	302
	GA	Bankrupt	1186	0	80	5
		Solvent	2	1184	2	304
	Tabu search	Bankrupt	1186	0	80	5
		Solvent	2	1184	2	304
	RFE	Bankrupt	1186	0	80	5
		Solvent	1	1185	1	305
	Stepwise	Bankrupt	1186	0	80	5
		Solvent	3	1183	4	302
	LVQ	Bankrupt	1186	0	80	5
		Solvent	1	1185	2	304

2.4.5.1 Feature importance

According to the results, LightGBM-RFE and LightGBM, which utilised all features, achieved the best predictive performance on the test data. Figures 2.35 and 2.36 display the feature importance plots for LightGBM, using all variables, and LightGBM-RFE, respectively, in predicting bankruptcy. From Figure 2.35, pre-tax operating income/average assets, ROAA, net loans/total assets, equity/net loans and loan loss reserve/gross loans are the top 5 variables that provided the highest forecasting accuracy. The top 5 important features for bankruptcy prediction using LightGBM-RFE are pre-tax operating income/average assets, income net of distribution/average equity, equity/net loans, loan loss reserve/gross loan and ROAA. From the selected features by LightGBM-RFE and LightGBM using all features, a set of common features and a set of disjoint features is identified for both models. Features such as pre-tax operating income/average assets, ROAA, equity/net loans, loan loss reserve/gross loan, non-operating items and taxes/average assets, unreserved impaired loans/equity and income net of distribution/average equity were among the top 10 most important features by both models. The overall ranking of features by both models, LightGBM-RFE and LightGBM, using all features is similar. The feature importance plots of the LightGBM method using GA (Fig. A.19), SA (Fig. A.20), Stepwise (Fig. A.21), Tabu (Fig. A.22), and LVQ (Fig. A.23) feature selection techniques are shown in Appendix A.

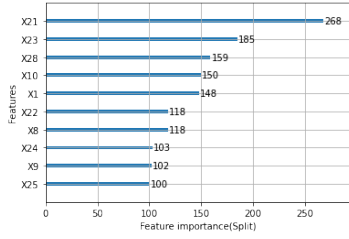


Figure 2.35: Feature Importance using all variables

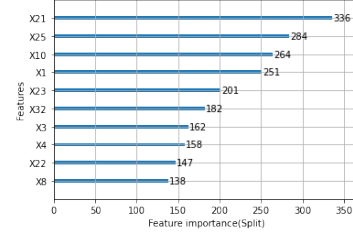


Figure 2.36: Feature Importance using RFE selected variables

Figure 2.37 is a horizontal binary tree. The first node in the tree indicates that pre-tax operating income divided by average assets is the optimal segmentation variable among the 33 features in the current subset of data, and the corresponding optimal segmentation point is 0.463. When pre-tax operating income/average assets is less than or equal to 0.463, it enters the second node as shown in Figure 2.37. The second node indicates that for the current subset, ROAA is the new optimal segmentation feature, with the corresponding optimal segmentation point at 0.429. When the pre-tax operating income divided by average assets is greater than 0.463, it enters the third node. The third node indicates that unreserved impaired loans/equity is the optimal segmentation feature for the current subset, with the corresponding optimal segmentation point being -0.047.

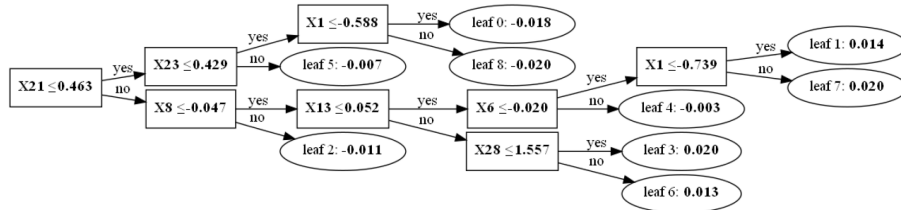


Figure 2.37: Tree diagram of LightGBM using all features

2.4.5.2 Feature importance and SHAP Value

To visually illustrate the importance of each feature for bankruptcy prediction, we utilised SHAP values to highlight the effect of each feature on model interpretability. For accuracy, the mean magnitude value of SHAP value $\sum_{j=1}^N |\phi_i^j|$ is used to characterise feature importance as stated in Equation 2.3. The first benefit of SHAP values is global interpretability- the SHAP values show how much each feature contributes to the dependent variable. The SHAP method generates rankings (Kvamme et al., 2018) of the features according to their level of contribution to predicting bankruptcy (Fitzpatrick and Mues, 2016). The higher the feature importance, the greater the contribution the

feature provides in predicting the model. The feature's level of contribution is calculated and presented in SHAP values. A positive SHAP value indicates a positive impact on the prediction, leading the model to predict solvency. In contrast, a negative SHAP value indicates a negative effect, leading the model to predict bankruptcy. The features are ordered by the degree to which they influence the model's prediction.

Another benefit is that local interpretability means that each observation receives its own set of SHAP values. This enables a deeper analysis of individual-level interpretation regarding the degree to which features affect bankruptcy prediction. The x-axis represents the average of the absolute SHAP value for each feature. The SHAP values technique was proposed in 2017 and is based on a technique used in game theory (Lundberg and Lee, 2017). Figure 2.38 presents the top 10 features sorted according to their importance with the highest mean SHAP values according to the SHAP method. According to the SHAP ranking, pre-tax operating income/average assets had the most influence in predicting bankruptcy. This was followed by the ROAA and net loans to total assets ratio.

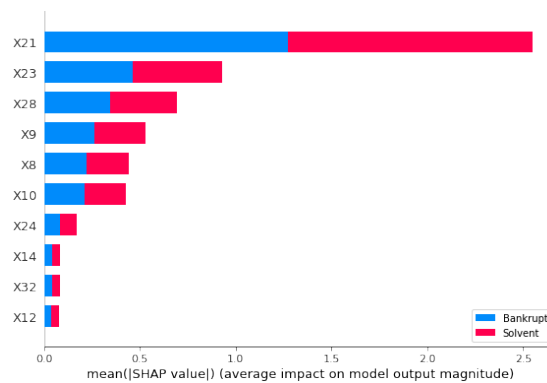


Figure 2.38: SHAP Bar Graph

Fig. 2.38 provides the relative importance of the features in the dataset, but does not reflect the distribution of the feature's impact on the output, or the relationship between the SHAP value of the feature and its exact impact. To better understand the attribution of importance based on SHAP values, Figure 2.39 (SHAP summary plot) shows the feature explanations. With the summary plot, we use dot charts to visualise the directional impact of the features. In this chart, the x-axis represents the SHAP value, and the y-axis displays all the features. Each point on the chart represents a single SHAP value for a prediction and a feature. The colour red indicates a higher value of a feature. Blue means a lower value of a feature. We can get a general sense of the impact of features' directionality based on the distribution of the red and blue dots. Figure 2.39 displays the top-most important features in the model according to SHAP analysis. The feature rankings (y-axis) indicate the importance of the features ranked by their importance. The SHAP value (x-axis) is an index that measures the influence

of a feature on the model. SHAP values distant from zero have a greater impact on the model's prediction.

The plot gives the most essential features for the model using the test dataset. The SHAP values for all features and all samples in the test data are aggregated and sorted. In each feature importance row, the attribution of all banks to the outcome is represented by dots of different colours, where red dots represent high-risk values and blue dots represent low-risk values. Thus, the dots with red colour refer to a larger feature value. The first feature is the most essential. The graph provides information on how each feature affects the model output. Pre-tax operating income/average assets had the most potent predictive power, followed by ROAA and net loans to total assets. Higher values of pre-tax operating income/average assets and ROAA, according to the graph, indicate solvency in a bank, while lower values indicate bankruptcy.

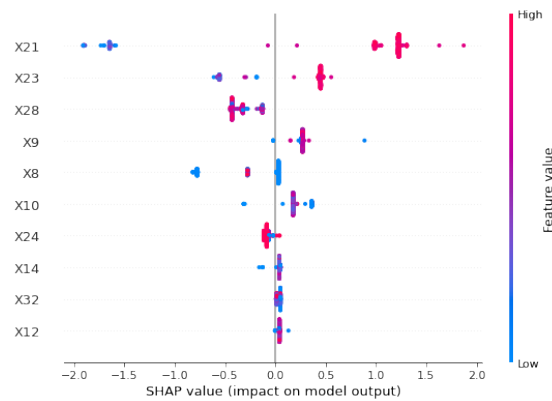


Figure 2.39: Aggregated SHAP values for top 10 features of LightGBM

The plot gives the most important features for the model using the test dataset. The SHAP values for all features and all samples in the test data are aggregated and sorted. The first feature is the most important. The graph provides information on how each feature affects the model output. Two important features are pre-tax operating income/average assets and ROAA. Low values of pre-tax operating income/average assets and ROAA, according to the graph, lead to a bank's bankruptcy, while high values lead to bankruptcy.

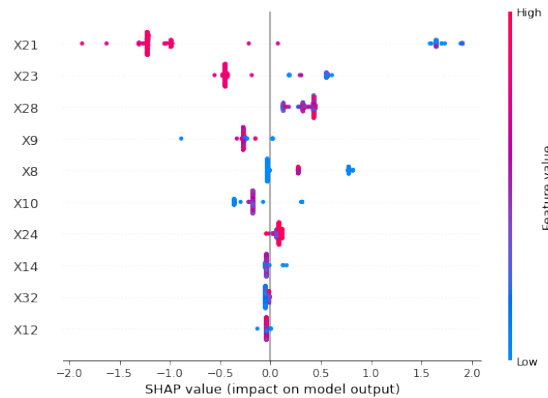


Figure 2.40: Aggregated SHAP Plot Graph

The inbuilt feature importance graph from LightGBM and feature importance using SHAP value identified a common set of features as well as disjoint features for bankruptcy prediction. However, the overall ranking of the features is similar. The most important features for LightGBM feature importance and highest SHAP value impact are pre-tax operating income/average assets, ROAA, and net loans to total assets. Both methods identified these three features as the most important, with the same ranking. Both LightGBM and SHAP analysis treat pre-tax operating income/average assets, ROAA, net loans to total assets, equity to net loans, equity to total assets, unreserved impaired loans to equity, and ROAE as critical features for bankruptcy prediction. Since positive SHAP values indicate that the probability of class prediction is increased, banks with higher values for pre-tax operating income/average assets, ROAA, and net loans to total assets are more likely to be solvent.

In the waterfall (Figures 2.41 and 2.42), the x-axis represents the target variable, and $f(x)$ is the predicted value of the model. The length of the bar gives the SHAP value for each feature. Figure 2.41 shows SHAP values for bankruptcy prediction using features X1 to X33. Figure 2.41 shows a prediction score of 1.711 for solvent banks, which is larger than the average prediction score of 0.002, indicating a likelihood of solvency. The top driving features are ordered from top to bottom according to their SHAP values, where the red bar represents a positive value and the blue bar represents a negative value. It is evident from the graph that high values of pre-tax operating income/average assets, ROAA, and equity as a percentage of total assets contribute most to a bank's ability to remain solvent. Pre-tax operating income/average assets has a SHAP value of 1.22, ROAA has a SHAP value of 0.45 and net loans/total assets has a SHAP value of -0.45. Pre-tax operating income/average assets contributed the most to the prediction, with ROAA coming in second. The sum of all SHAP values will be equal to $E[f(x)] - f(x)$. For Figure 2.42, a prediction score of -1.711 indicates a very low likelihood of solvency and high probability of bankruptcy, which is negatively impacted by a bank's pre-tax

operating income/average assets and ROAA values and high net loans/total asset values. Low values of pre-tax operating income/average assets, ROAA, and high values of net loans to total assets lead to bankruptcy.

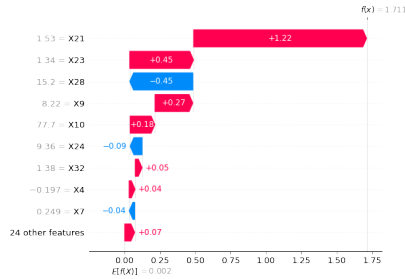


Figure 2.41: SHAP values of individual samples

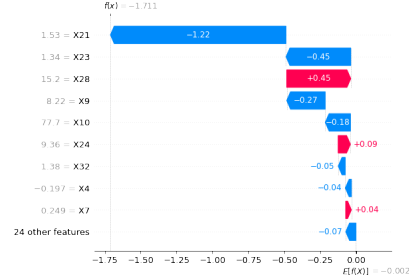


Figure 2.42: SHAP values of individual samples

An alternative way to visualize local explanations for tree models is a force plot. Figure 2.43 visualises a force plot for LightGBM using all features, which depicts a random sample bank of the test data to determine which features correspond to bankruptcy prediction. The graph (Figure 2.43) illustrates the relative contribution of features from the base value (i.e., no samples) to an outcome. It shows features that contribute to pushing the model output beyond the base value (the average model output over the training dataset we passed). Features that make the prediction higher (positive impact) are shown in red, while those that push the prediction lower (negative impact) are shown in blue. It illustrates how various features contribute to predicting whether a bank is solvent or bankrupt. It visualizes how features contribute to individual predictions. Each group of features is sorted by the magnitude of their impact, and the features with the most significant impact are labelled. The output value, 1.71, is the predicted value for the bank, while the base value (0.002042) is the mean prediction of the test data. From Figure 2.43, pre-tax operating income/average assets, ROAA, equity/total assets, and equity/net loans contribute to driving the values above the base value, while net loans/total assets, ROAE drive the value below the base value. Thus, pre-tax operating income/average assets, ROAA, equity/total assets and equity/net loans have a positive impact on the prediction. In contrast, net loans to total assets and ROAE have a negative impact on the prediction. The direction indicates the importance of each feature in driving it below or above the base value.

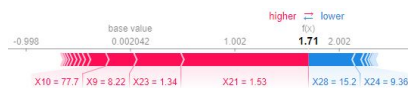


Figure 2.43: Force Plot for LightGBM with all features



Figure 2.44: Force Plot for LightGBM-RFE

Table 2.23 presents the results of the McNemar test, which statistically compares the classification accuracy of different variable selection techniques for the LightGBM model on the hold-out data. Results show that there is no statistically significant difference in prediction performance among the LightGBM models using the different variable selection techniques. The LightGBM model, which utilised different variables, performed well on the hold-out data.

Table 2.23: McNemar test for pairwise performance comparison using LightGBM model

	SA	GA	Tabu	RFE	Stepwise	LVQ
ALL	0.25	1.00	1.00	NA	0.25	1.00
SA		0.48	0.48	0.25	NA	0.48
GA			NA	1.00	0.48	NA
Tabu				1.00	0.48	NA
RFE					0.25	1.00
Stepwise						0.48

2.4.6 Online Sequential Extreme Learning Machine

2.4.6.1 OSELM-All variables

To investigate the effectiveness of the OSELM model, we experimented with different activation functions and feature selection methods. We compared the prediction performance of each OSELM model for both the training and validation sets. It is evident from Table 2.24 that the OSELM model with sigmoid activation function using all variables yields the best prediction outcome with an accuracy of 96.67%, precision of 93.93% and a kappa of 0.9334 for the training set. This is followed by the model, which utilises a hard limit activation function, achieving an accuracy of 96.50%, precision of 93.68%, and a kappa of 0.9300. However, the model with hardlimit activation function had the highest accuracy, sensitivity, kappa, F1 score and MCC for the test set. OSELM with hardlimit had an accuracy of 91.82%, a sensitivity of 81.93%, a kappa of 0.7574, and an F1 score of 0.8095.

The OSELM model with a sigmoid activation function, using all variables, accurately classifies 96.67% of the banks in the training dataset, with 79 misclassified banks, and 81.59% of the banks in the validation dataset, with 72 misclassified samples. The overall error rate for the training dataset is 3.33%, with a Type I error rate of 0.59% and a Type II error rate of 6.07%. However, the overall error rate for the test data set is 18.4%, with a Type I error rate of 4.90% and a Type II error rate of 67.06%. However, the model with hardlimit activation function had a better performance for the validation set using all measurement metrics than the other activation functions. This model achieved an accuracy of 91.82%, a kappa of 0.7574, and an F1 score of 0.8095. The error rate for the validation set of the model using hardlimit is 8.18%, with a Type I error rate of 4.90%

and a Type II error rate of 20.0%. The OSELM model with a sine function yielded the least accurate prediction outcome for the training dataset, with an accuracy of 85.83%, precision of 78.08%, and a kappa of 0.7167. The testing set yielded an accuracy of 51.41% and a kappa of 0.0036.

Table 2.24: Predictive performance of OSELM using all features

	Accuracy	Sensitivity	Specificity	Kappa	Precision	F1 score	MCC
Training model							
Rbf	0.9245	0.9275	0.9216	0.8491	0.922	0.9248	0.8491
Sig	0.9667	0.9393	0.9941	0.9334	0.9938	0.9658	0.9348
Hardlim	0.965	0.9368	0.9933	0.93	0.9929	0.964	0.9315
sin	0.8583	0.7808	0.9359	0.7167	0.9242	0.8464	0.7255
Testing model							
Rbf	0.6061	0.1294	0.7386	-0.1287	0.1209	0.125	-0.1289
Sig	0.8159	0.3294	0.951	0.3413	0.6512	0.4375	0.3697
Hardlim	0.9182	0.8	0.951	0.7574	0.8193	0.8095	0.7575
sin	0.5141	0.4824	0.5229	0.0036	0.2193	0.3015	0.0043

2.4.6.2 OSELM-Simulated Annealing

The OSELM model with sigmoid activation function, which used simulated annealing feature selection technique, had the best prediction outcome for the training set in comparison to other activation functions (radial basis function, hardlimit and sine) with an accuracy of 96.92%, precision of 99.82%, kappa of 0.9384 and F1 score value of 0.9683. The OSELM model, which utilised the radial basis function, performed slightly better in predicting solvent banks compared to the sigmoid activation function, although the sigmoid generally showed better performance. Additionally, the model with a sigmoid activation function achieved the highest accuracy, sensitivity, kappa, F1 score, and MCC for the test set. OSELM with hardlimit had a slightly better specificity and precision than the sigmoid activation function. The OSELM model, with a sigmoid activation function and a simulated annealing feature selection method, accurately classifies 96.92% of the banks in the training dataset, with 72 misclassifications, and achieves 92.33% accuracy in the test set, with 30 misclassified banks. This model had an overall error rate of 3.08% for the training dataset, with a Type I error rate of 0.17% and a Type II error rate of 5.99%. The overall error rate for the test data set is 3.67%, with a Type I error rate of 0.65% and a Type II error rate of 32.94%.

Table 2.25: Predictive performance of OSELM-SA

	Accuracy	Sensitivity	Specificity	Kappa	Precision	F1 score	MCC
Training model							
Rbf	0.9545	0.9427	0.9663	0.9089	0.9655	0.9539	0.9092
Sig	0.9692	0.9401	0.9983	0.9384	0.9982	0.9683	0.94
Hardlim	0.9654	0.9368	0.9941	0.9309	0.9937	0.9644	0.9324
sin	0.9182	0.8879	0.9486	0.8364	0.9452	0.9157	0.838
Testing model							
Rbf	0.3248	0.9529	0.1503	0.0495	0.2375	0.3803	0.1275
Sig	0.9233	0.6706	0.9935	0.7465	0.9661	0.7917	0.7652
Hardlim	0.3632	0.9882	0.1895	0.0867	0.253	0.4029	0.2049
sin	0.5243	0.5412	0.5196	0.0417	0.2383	0.3309	0.0501

2.4.6.3 OSELM-Genetic algorithm

We compared four OSELM models using different activation functions after the genetic algorithm technique was used to select the best combination of features for bankruptcy prediction. The results of the different OSELM models with different activation functions are presented in Table 2.26. The best OSELM model is the one that uses the sigmoid activation function for both the training set and the validation set. Based on the performance metrics used, the model with the sigmoid activation function yielded better results, except for sensitivity, despite achieving a high score. The accuracy, kappa, and F1 score for OSELM-sigmoid on the training data set were 96.59%, 0.9317, and 0.9647, respectively. In contrast, those of the validation data set were 91.82%, 0.7488, and 0.8000, respectively. The OSELM-sigmoid model has an overall error rate of 3.41%, with Type I and Type II error rates of 0.25% and 6.58%, respectively, for the training dataset. The overall error rate for the validation dataset is 8.18%, with a Type I error rate of 3.59% and a Type II error rate of 24.71%.

Table 2.26: Predictive performance of OSELM-GA

	Accuracy	Sensitivity	Specificity	Kappa	Precision	F1 score	MCC
Training model							
Rbf	0.9418	0.8853	0.9983	0.8836	0.9981	0.9383	0.8893
Sig	0.9659	0.9342	0.9975	0.9317	0.9973	0.9647	0.9336
Hardlim	0.965	0.9342	0.9958	0.93	0.9955	0.9639	0.9318
sin	0.8921	0.8533	0.9309	0.7841	0.925	0.8877	0.7865
Testing model							
Rbf	0.7826	0	1	0	NAN	0	0
Sig	0.9182	0.7529	0.9641	0.7488	0.8533	0.8	0.7511
Hardlim	0.8798	0.5059	0.9837	0.5808	0.8958	0.6466	0.6153
sin	0.4859	0.4706	0.4902	-0.0266	0.2041	0.2847	-0.0324

2.4.6.4 OSELM-Tabu Search

For the OSELM-tabu search method, the accuracy of the model using the sigmoid activation function is 97.09%, with a kappa value of 0.9418 for the training dataset. This model exhibits the best performance across most performance metrics. However, the model with the hardlimit activation function performed better for the test data set, although the improvement was very slight. The accuracy of the model using the hardlimit activation function is 59.59%, with a kappa of 0.1302 and an F1 Score of 0.3828. Generally, the OSELM model with tabu search feature selection performed poorly in predicting the test dataset for all activation functions. The OSELM-Tabu search feature selection technique with a sigmoid activation function misclassifies 69 banks in the training dataset and has an overall error rate of 2.91%, with Type I error rates of 0.08% and Type II error rates of 5.73%. The model with a hardlimit activation function had an overall error rate of 6.66%, with Type I and Type II error rates of 39.87% and 42.35%, respectively, for the test dataset. The model with hardlimit has a better prediction for the test dataset than the sigmoid activation function. Generally, the performance of the OSELM-Tabu search with all activation functions on the test data was poor.

Table 2.27: Predictive performance of OSELM-Tabu search

	Accuracy	Sensitivity	Specificity	Kappa	Precision	F1 score	MCC
Training model							
Rbf	0.9439	0.8879	1.0000	0.8879	1.0000	0.9406	0.8935
Sig	0.9709	0.9427	0.9992	0.9418	0.9991	0.9701	0.9433
Hardlim	0.9688	0.9444	0.9933	0.9376	0.9929	0.968	0.9387
sin	0.8917	0.8558	0.9275	0.7833	0.9219	0.8876	0.7853
Testing model							
Rbf	0.7826	0.0000	1.0000	0.0000	NAN	0.0000	0.0000
Sig	0.2251	0.4000	0.1765	-0.2285	0.1189	0.1833	-0.3942
Hardlim	0.5959	0.5765	0.6013	0.1302	0.2865	0.3828	0.1478
sin	0.4834	0.4588	0.4902	-0.0347	0.2000	0.2786	-0.0421

2.4.6.5 OSELM-Recursive Feature Elimination

The performance of the OSELM-RFE model across all activation functions for the training dataset ranges from 88.91% to 96.54%. The best-performing model for the training dataset is derived from the OSELM-RFE, which utilises the sigmoid activation function. The model achieves an accuracy of 96.54%, a kappa of 0.9309, and an F1 score of 0.9644. The OSELM-RFE search feature selection technique with a sigmoid activation function misclassifies 82 banks in the training dataset and has an overall error rate of 3.46%, with Type I error rates of 0.51% and Type II error rates of 6.41%. The OSELM-RFE with hardlimit activation function had the best accuracy for the test dataset compared to other activation functions, although it is relatively low. The accuracy is 43.48%, the

kappa is 0.1162, and the F1 score is 0.4138. The OSELM-RFE with hardlimit activation function misclassifies 221 of the test data set, resulting in a Type I error rate of 69.93% and a Type II error rate of 8.24%.

Table 2.28: Predictive performance of OSELM-RFE

	Accuracy	Sensitivity	Specificity	Kappa	Precision	F1 score	MCC
Training model							
Rbf	0.9359	0.9376	0.9342	0.8718	0.9345	0.936	0.8718
Sig	0.9654	0.9359	0.9949	0.9309	0.9946	0.9644	0.9325
Hardlim	0.9616	0.9359	0.9874	0.9233	0.9867	0.9606	0.9245
sin	0.8891	0.8651	0.9132	0.7782	0.9088	0.8864	0.7791
Testing model							
Rbf	0.2941	0.6941	0.1830	-0.063	0.1909	0.2995	-0.1245
Sig	0.3504	0.3412	0.3529	-0.1908	0.1278	0.1859	-0.2557
Hardlim	0.4348	0.9176	0.3007	0.1162	0.2671	0.4138	0.2071
sin	0.4936	0.5059	0.4902	-0.0026	0.2161	0.3028	-0.0032

2.4.6.6 OSELM-Stepwise regression

From Table 2.29, the OSELM-stepwise model with hardlimit activation function yields the best prediction outcome with an accuracy of 96.92%, precision of 93.93% and a kappa of 0.9384 for the training set. The OSELM-stepwise model with hardlimit activation function misclassifies 73 banks in the training dataset. The overall error rate for the training data set is 3.08%, with a Type I error rate of 0.59% and a Type II error rate of 5.56%. However, the model with a sigmoid activation function performed better on the test data set. This model achieved an accuracy of 69.05%, a kappa of 0.1094, and an F1 score of 0.3086. The error rate for the test data set is 5.10%, with a Type I error rate of 20.59% and a Type II error rate of 68.24%. The model misclassifies 121 banks in the test data set.

Table 2.29: Predictive performance of OSELM-Stepwise regression

	Accuracy	Sensitivity	Specificity	Kappa	Precision	F1 score	MCC
Training model							
Rbf	0.9338	0.8710	0.9966	0.8676	0.9961	0.9294	0.8746
Sig	0.9671	0.9376	0.9966	0.9342	0.9964	0.9661	0.9359
Hardlim	0.9692	0.9444	0.9941	0.9384	0.9938	0.9684	0.9396
sin	0.8853	0.8432	0.9275	0.7707	0.9208	0.8803	0.7734
Testing model							
Rbf	0.5882	0.1412	0.7124	-0.1376	0.1200	0.1297	-0.1384
Sig	0.6905	0.3176	0.7941	0.1094	0.3000	0.3086	0.1095
Hardlim	0.4834	0.9059	0.3660	0.1519	0.2841	0.4326	0.2432
sin	0.4501	0.4235	0.4575	-0.0795	0.1782	0.2509	-0.0982

2.4.6.7 OSELM-LVQ

For OSELM-LVQ, the model's accuracy using the sigmoid activation function is 96.75%, with a kappa value of 0.9351 for the training dataset. This model exhibits the best performance across all performance metrics. The OSELM-LVQ model with a sigmoid activation function misclassifies 77 banks in the training dataset, resulting in an overall error rate of 3.25%. The model also exhibits Type I and Type II error rates of 0.25% and 6.24%, respectively. The general performance of other activation functions, such as RBF, Hardlim, and sine, was good for the training data set. The accuracy of the OSELM-LVQ model with a sigmoid activation function is 90.54%, a kappa of 0.6700, and an F1 score of 0.7218. The model misclassifies 37 banks in the test data. The model with a sigmoid activation function had an overall error rate of 9.46%, with Type I error rates of 0.00% and Type II error rates of 43.53% for the test dataset. The other activation functions performed poorly on the test data. The hardlim activation function had the least accuracy on the test data.

Table 2.30: Predictive performance of OSELM-LVQ

	Accuracy	Sensitivity	Specificity	Kappa	Precision	F1 score	MCC
Training model							
Rbf	0.9545	0.9115	0.9975	0.9089	0.9972	0.9524	0.9123
Sig	0.9675	0.9376	0.9975	0.9351	0.9973	0.9665	0.9368
Hardlim	0.9629	0.9351	0.9907	0.9258	0.9902	0.9618	0.9272
sin	0.9376	0.9005	0.9747	0.8752	0.9727	0.9352	0.8776
Testing model							
Rbf	0.7826	0.0000	1.0000	0.0000	NAN	0.0000	0.0000
Sig	0.9054	0.5647	1.0000	0.67	1.0000	0.7218	0.7098
Hardlim	0.2916	0.7294	0.1699	-0.0508	0.1962	0.3092	-0.1054
sin	0.509	0.5647	0.4935	0.0387	0.2365	0.3333	0.048

Table 2.31 presents the results of the McNemar test, which statistically compares the classification accuracy of different variable selection techniques for the OSELM model on the hold-out data. As shown in Table 2.31, OSELM using all variables and hardlimit activation function outperforms OSELM-tabu with hardlimit activation function, OSELM-RFE with hardlimit activation function and OSELM-LVQ with sigmoid activation function at 5% statistical significance level. OSELM-SA with sigmoid activation function outperforms OSELM-All variables with hardlimit activation function, OSELM-GA with sigmoid activation function, OSELM-tabu with hardlimit activation function, OSELM-RFE with hardlimit activation function and OSELM-Stepwise with sigmoid activation function at 5% statistical significance level. Results suggest that OSELM-SA with sigmoid activation function only outperforms OSELM-LVQ with sigmoid activation function at a 10% statistical significance level. However, OSELM-ALL with hardlimit activation function does not significantly outperform OSELM-GA with sigmoid activation function, nor does it outperform OSELM-Stepwise with sigmoid activation function.

In addition, the results show that the prediction performance of OSELM-ALL with a hardlimit activation function, OSELM-GA with a sigmoid activation function, and OSELM-Stepwise with a sigmoid activation function do not differ significantly.

Table 2.31: McNemar test for pairwise performance comparison for OSELM

	SA	GA	Tabu	RFE	Stepwise	LVQ
ALL	11.02(0.00)	0.94(0.33)	49.80(0.00)	176.59(0.00)	0.30(0.58)	28.20(0.00)
SA		5.63(0.02)	80.01(0.00)	223.34(0.00)	8.11(0.00)	3.23(0.07)
GA			55.71(0.00)	201.97(0.00)	1.65(0.20)	19.31(0.00)
Tabu				56.92(0.00)	33.51(0.00)	89.13(0.00)
RFE					185.33(0.00)	234.32(0.00)
Stepwise						15.01(0.00)

2.4.7 All model comparison

Table 2.32 presents the results of the McNemar test comparing the prediction performance of the best models for XGBoost, LightGBM, SVM, and OSELM on the hold-out data. As shown in Table 2.32, LightGBM-RFE and LightGBM-All outperform OSELM-SA with sigmoid activation function at 5% significance level. XGBoost-GA, XGBoost-Tabu, and XGBoost-LVQ outperform OSELM-SA with a sigmoid activation function at a 5% significance level. However, XGBoost-GA, XGBoost-Tabu and XGBoost-LVQ do not significantly outperform LightGBM-RFE, LightGBM-All and SVM-Tabu using a polynomial activation function.

Table 2.32: McNemar test for pairwise performance comparison of best models

	OSELM-SA(sigmoid)	XGBoost-GA	LightGBM-All
SVM-Tabu (Polynomial)	164.01(0.00)	1.07(0.30)	1.79(0.18)
OSELM-SA(sigmoid)		173.19(0.00)	174.09(0.00)
XGBoost-GA			0.00(1.00)

2.5 Discussion

The purpose of this chapter was to investigate and compare the prediction accuracy of different machine learning models in bankruptcy prediction. The evaluation of the selected ML techniques was conducted using various performance metrics, including accuracy, precision, sensitivity, specificity, MCC, AUC, and kappa. The objective was to determine how these ML models can be exploited to predict bankruptcy in the US financial sector. Forecasting bank insolvency is of great interest to both policymakers and market participants. The chapter aimed to forecast bank failures by using publicly available information from the financial statements of US banking institutions from 2002 to 2019. We used the CAMEL features. All features were assessed four years

before the bankruptcy filing by the individual financial institutions. The study compared different ML models to predict bankruptcy utilising financial institutions in the USA. We also employed different feature selection methods within the ML framework to build and optimise the models using 10-fold cross-validation as well as feature importance analysis. In many ML tasks, including bankruptcy prediction, an imbalanced class arises that has the potential to affect the performance of the predictive model. Ignoring the imbalanced problem has the potential to produce biased models. The imbalanced nature of the data was addressed using the SMOTE technique.

The results suggest that there was a statistically significant difference in almost all financial ratios between failed and solvent banks. Our findings indicate that financial ratios had a notably increased impact on explaining bankruptcy, confirming what was expected based on previous studies. The results of the study are consistent with Climent, Momparler, and Carmona (2019), which also found that equity-to-net loans, liquid assets-to-total deposits and borrowing influence the health of financial institutions. Additionally, the study identified pre-tax operating income, average assets, return on average assets, and net loans to total assets as some of the most important features that drive bankruptcy, in contrast to what was found by Climent, Momparler, and Carmona (2019). Equity to net loans and loan loss provision to net interest revenue were identified as the most significant features for bankruptcy prediction. This outcome is consistent with Climent, Momparler, and Carmona (2019).

We implemented SVM using different kernel functions (linear, sigmoid, and polynomial) and feature selection techniques on both the training and test data. The best linear, sigmoid and polynomial kernel models have similar results for both training and test data. While linear (SVM-LVQ) and polynomial kernels (SVM-tabu) performed better than the sigmoid kernel (SVM-SA), the polynomial kernel achieved a slightly better result. The SVM-Tabu model with a polynomial kernel yielded the best performance, particularly on the test data (Accuracy = 0.9744, F1 score = 0.9419, kappa = 0.9255, AUC = 0.9686). The lowest-performing kernel-based model (SVM-SA) with a sigmoid function yielded an accuracy of 94.73% for the training data and 92.58% for the test data. SVM-tabu with polynomial kernel accurately classified 96.87% (1177 out of 1215) of the solvent banks and 99.22% (1148 out of 1157) of the failed banks in the training data. Also, it classifies 300 out of 304 solvent banks (98.68%) and 81 out of 87 failed banks (93.10%) in the test data.

Furthermore, the Type I and Type II errors are 3.13% and 0.78%, respectively, for the training data, and 1.32% and 0.52%, respectively, for the test data. SVM-Tabu (polynomial) outperforms SVM-LVQ (linear) and SVM-SA (sigmoid) by 2.30% and 4.86%, respectively, on the test data. According to the experimental results, we conclude that the prediction performance of SVM-tabu with a polynomial kernel function is more precise compared to that of the other models on the test data. When the results of the SVM

model are compared with those of other studies on bankruptcy prediction, our findings show similarity to the findings of (Min and Lee, 2005; Boyacioglu, Kara, and Baykan, 2009; Barboza, Kimura, and Altman, 2017; Erdogan, 2013), which achieved high accuracies for their models. However, the predictive ability of our model is better than these studies, which equally used SVM for bankruptcy prediction. Our results provide evidence in support of using SVM for bankruptcy prediction. This study explored the potential of online sequential extreme learning machine (OS-ELM), a proposed alternative to ELM and the standard feed-forward neural networks. The batch ELM requires all the training data to be available before training can begin. However, in real-world practical applications, learning must be an ongoing process, as the complete set of data is usually not available at once. When new data is to be introduced, batch learning must repeat the training process with both past data and the new data, which takes a considerable amount of time. The proposed OS-ELM can learn data one-by-one or in chunks (a block of data) with fixed or varying chunk sizes. Empirical results showed that OSELM, using all features and hardlimit activation function, as well as OSELM-SA with hardlimit activation function, were the most optimal models.

We also conducted empirical analysis using Extreme Gradient Boosting, which was developed from other boosting methods. We applied the XGBoost method to the features selected by the various feature selection techniques. We also included feature importance and partial dependence plots to identify which financial ratios were the primary driving force behind bankruptcy prediction. The features selected by stepwise regression had the highest accuracy for the training data. The research also shows that features selected using Tabu search with XGBoost show the highest accuracy of 98.21% for the out-of-sample data. Overall, features selected by Tabu search with XGboost offer the best prediction ability for both training and testing data. The feature importance of XGBoost model using the Tabu search feature selection technique indicated that the top 10 important financial ratios included pre-tax operating income/average assets, impaired loans/equity, capital funds/net loans, ROAA, capital funds/total assets, loan loss reserve/non-performing loans, loan loss reserve/gross loan, non-performing loans/gross loans, equity/total assets and liquid assets/total deposit and borrowing. Also the feature importance of XGBoost-Tabu using the SHAP importance feature selection suggested that the top 10 financial ratios for bankruptcy prediction included pre-tax operating income/average assets, impaired loans/equity, ROAA, capital funds/net loans, net loans/total deposit and borrowing, capital funds/total assets, equity/total assets, non-performing loans/gross loans, ROAE and equity/deposit and short term funding. The accuracy and AUC levels of the best machine learning model are generally higher than those reported in other studies (Jordan et al., 2010; Cleary and Hebb, 2016; Cox and Wang, 2014).

The study employs a state-of-the-art Machine Learning algorithm, specifically Light Gradient Boosting Machine (LightGBM), to classify financial institutions as either bankrupt

or non-bankrupt. The tree-based method was compared based on different features as inputs. The study assessed the performance of the tree-based LightGBM model for bankruptcy prediction using different feature selection techniques. We employed the Parzan technique to optimize the parameters of LightGBM. We also performed a comprehensive feature importance analysis using SHAP to identify useful features for bankruptcy prediction. The results showed that LightGBM, using all features and the RFE feature selection technique, outperformed LightGBM with other feature selection techniques employed in the study for both training and test data. Additionally, the study conducted a feature importance analysis of the input features to identify the most important financial ratios. More importantly, we adopted the SHAP to explore the correlation between features and the prediction of bankruptcy. The feature importance of the LightGBM model illustrates that pre-tax operating income/average assets, ROAA, net loans/total asset, equity/net loans, loan loss reserve/gross loan, non-operating items and taxes/average assets, unreserved impaired loans/equity, ROAE, equity/total assets and income net of distribution/average equity had significant influence on the final model. The feature importance and SHAP analysis illustrated that pre-tax operating income/average assets, ROAA, net loans/total asset, equity/total assets, non-operating items and taxes/average assets, unreserved impaired loans/equity, equity/net loans, ROAE, capital funds/net loans, liquid assets/deposit and short term funding and equity/liabilities are the top 10 features selected and had significant influence and significantly important for bankruptcy prediction. In credit risk analysis, policymakers, financial institutions, and market analysts can focus more on these features to make more informed bankruptcy decisions. By jointly considering accuracy, stability, and efficiency, the optimised LightGBM model is highly recommended for predicting bankruptcy with excellent performance. LightGBM contains two novel techniques: Gradient-based One-Side Sampling and Exclusive Feature Bundling, which deal with a large number of data instances and a large number of features, respectively. We find that pre-tax operating income/average assets, ROAA, equity/total assets, ROAE, and capital funds/net loans were consistently ranked among the most important features for both XGBoost and LightGBM SHAP importance feature selection techniques in determining the state of bankruptcy. The machine learning models used in the study performed better in both in-sample and out-of-sample predictions than traditional statistical and machine learning models. Other studies (Barboza, Kimura, and Altman, 2017; Chen et al., 2023) have also demonstrated a good prediction of bankruptcy. Our study demonstrated the superiority of machine learning models in predicting bankruptcy over other statistical and machine learning models in previous studies (Gogas, Papadimitriou, and Agrapetidou, 2018; Le and Viviani, 2018; Odom and Sharda, 1990; Boyacioglu, Kara, and Baykan, 2009; Climent, Momparler, and Carmona, 2019), confirming existing literature.

Pre-tax operating income to average assets is an indicator of a financial institution's

profitability, and it is by far the most influential financial ratio. It is a measure of a firm's asset productivity. It indicates the efficiency with which the financial institutions generate earnings from their assets before taxes and interest. A critical measure in bankruptcy prediction is the pre-tax operating income to average assets ratio, which offers insights into a company's operational efficiency, asset utilisation, and profitability. A high ratio typically indicates financial stability and a reduced risk of bankruptcy, whereas a low ratio may suggest financial distress and a greater likelihood of bankruptcy. A company's financial health and prospective risks can be more accurately evaluated by stakeholders by integrating this ratio into broader financial analysis and prediction models. The experimental results indicate that the optimised LightGBM outperformed the other machine learning algorithms, achieving the highest performance in terms of accuracy, AUC, precision, and F1-score. The results reveal that the optimised LightGBM algorithm is superior to other classifiers and outperforms other models such as XGBoost, SVM and OSELM. The experimental results are consistent with the theory and demonstrate that, with the aid of GOSS and EFB, LightGBM can significantly outperform XGBoost, SVM, and OSELM in terms of computational speed and memory consumption. The results also underscore the importance and value of adopting an efficient parameter optimisation strategy to enhance the predictive performance of ML models.

The study develops a machine learning framework that can serve as a universal model for predicting bankruptcy. While earlier research has made efforts to identify features crucial to bankruptcy prediction, it usually ignores the interpretation of these features in favour of the accuracy of the prediction techniques. The use of data visualization methods, such as SHAP, provides practitioners with the opportunity to identify the most crucial features and their contribution to bankruptcy. The information provided by the data visualisation using SHAP is beneficial for banks and other relevant stakeholders, enabling them to make informed decisions and provide more accurate bankruptcy predictions. Early prediction of bankruptcy enables central banks and governments to implement early relief measures for financial institutions in distress, thereby avoiding a potential cascading effect on the economy. Stakeholders can utilise the study's findings to develop policies that alleviate the economic impact of bankruptcy. The adoption of machine learning predictive models has far-reaching implications for central banks, financial institutions, investors, policy makers and individuals. The machine learning models could improve credit evaluation, thereby helping banks and other stakeholders to assess the probability of default of individuals. This could result in the reduction of non-performing loans and help create a stable financial sector. The study's findings could help banks and other financial institutions comply with Basel risk assessment regulations. Additionally, the conclusions of this study offer an alternative credit risk assessment that relies on data and leads to a shift away from human analysis-driven decisions. Central banks and other financial regulators

can utilise the study's findings to implement effective macro-prudential policies that pre-empt financial distress or mitigate the impact of a bank failure on the economy. In summary, the high predictive power of the models developed in this chapter should encourage bank managers to track the most critical features that were identified. Managers can avoid financial distress by taking timely action rather than waiting for intervention. Regulators may also find the model helpful in avoiding financial distress by taking timely actions and warning potentially distressed banks.

One of the limitations of the study is the lack of usage of hybrid methods for bankruptcy prediction. Future research should focus on utilising hybrid machine learning models. The use of hybrid methods has the potential to increase predictive power and reduce the level of overfitting. The hybrid models explore the advantages of different techniques. One possible limitation of the study is that it focuses on US data. Future research should also consider data from other economies. Also, the application of bankruptcy prediction should be extended to different industries as well. The use of environmental sustainable indicators should be considered in future research. The study relied only on quantitative data. Future research could explore qualitative data. The application of natural language processing to bankruptcy prediction is another area that warrants exploration in future research.

Chapter 3

Measuring Systemic Risk across banks in South Africa

3.1 Introduction

The 2007–2009 financial crisis prompted academics and financial regulators to re-evaluate how best to comprehend and control the banking industry (Callaghan et al., 2013). The interdependence of the financial system essentially caused the financial crisis that lasted from 2007 to 2009. Numerous scholars have noted that the severity of the most recent financial situation can be attributed, at least in part, to the fast growth of the structured finance industry and the increasing acceptance of securitisation in the US financial system (Brunnermeier, 2009; Longstaff, 2010). The U.S. housing bubble burst in 2007, and rising interest rates caused many homeowners to fall behind on their mortgage payments and face foreclosure, which set off the crisis. It contributed to a spread of financial distress from one financial institution to another. According to the Federal Deposit Insurance Corporation, 348 banks collapsed in the United States between January 2008 and March 2011, with 25 of those failures happening in 2008, 140 in 2009, and 157 in 2010. Numerous big financial institutions, such as Lehman Brothers, Merrill Lynch, Washington Mutual, AIG, Fannie Mae, Freddie Mac, and many more, failed and declared bankruptcy. The GFC brought attention to how crucial tail-risk interdependencies are in determining an institution's capacity to maintain system stability. To reduce risk spillovers and preserve the financial stability of both the market and individual market participants, it is essential to precisely measure the systemic impact of individual institutions as well as their susceptibility to extreme market movements. Benoit et al. (2017) offer an in-depth analysis of the literature on systemic risk, covering risk metrics and regulatory concerns. In the wake of the global financial crisis that occurred between 2007 and 2009, regulatory bodies and scholars have expressed concerns about the assessment of systemic risk. The crisis demonstrated how the collapse of major institutions may have detrimental effects on the financial sector as a whole

and eventually on the actual economy. According to Claessens et al. (2010), systemically important financial institutions' (SIFI) actions had a significant role in the current financial crisis. These SIFIs are institutions whose failure would disrupt the financial system and most likely have negative economic effects because of their size or interdependence. Basel III laws recommend that these kinds of institutions be subject to an extra capital fee. Policymakers and regulators have had to reevaluate the institutional structure for monitoring the stability of financial systems in light of the worldwide GFC that lasted from 2007 to 2009.

The crisis has amply illustrated that, despite the potential for individual risks to be predicted, financial shocks to just one firm can swiftly spread over a vast number of markets and institutions, so endangering the entire system. Since then, the macroprudential approach to bank regulation—which emphasises the overall health of the financial system—has taken centre stage in the policy agenda. In the past, the primary goal of financial regulation was to control the risk that individual banks posed by mandating that they maintain enough reserves to protect themselves from the inherent risk of their investments. A collapse in a few banks might spread to other banks and cause the whole financial system to become paralysed because the systemic risks resulting from the linkages between the institutions are disregarded. Therefore, it has been recognised that systemic risk, as opposed to just individual institutions, should receive greater attention to enhance financial stability through regulation. Financial regulators are considering new regulatory measures that acknowledge the interconnectedness of the banking system in light of the need to manage systemic risk. Since the global financial crisis, there has been a renewed discussion on how to regulate banks and other financial institutions to guarantee financial stability. Acknowledging that financial regulation has mostly been micro-focused on the risk of specific financial firms rather than the financial system as a whole has been crucial to this discussion. A systemic shock that impacts every institution in the system might result in systemic risk. Systemic risk has been the subject of economic literature that has addressed several facets of these phenomena. The objective of studies of Banulescu and Dumitrescu (2015), Girardi and Ergün (2013), and Tobias and Brunnermeier (2016) has been to measure the financial institutions' marginal contributions to the total risk of the domestic financial systems. By examining the spillovers from one country's unstable financial system to another, Alter and Beyer (2014), Eichengreen et al. (2012), and Kallestrup, Lando, and Murgoci (2016) incorporate a cross-border trade component into systemic risk analysis. Gray (2009) and Gómez-Puig and Sosvilla-Rivero (2014) discuss contagion about the issue of systemic risk in sovereign debt markets, which is examined by Reboredo and Ugolini (2015) and Acharya and Thakor (2016). As a result, the body of research has grown significantly, with academics putting out alternate systemic risk metrics that are based on market data. Tobias and Brunnermeier (2016) established the Conditional Value at Risk (CoVaR), a well-known market-based indicator of systemic risk.

The CoVaR measures risk spillovers across institutions and is defined as the Value at Risk (VaR) of institution i conditional on institution j being in distress. Tobias and Brunnermeier (2016) defines the conditioning financial distress event as the returns of institution j equal to the VaR of that particular institution. The majority of research studies look at the extent to which financially distressed institutions add to the overall systemic risk of the financial system. Nonetheless, a key factor in maintaining stability is the level of danger that failing financial systems pose to specific institutions. This occasionally examined perspective, known as systemic risk exposure, focuses on how susceptible financial institutions are to systemic shocks. In contrast to systemic risk contribution, systemic risk exposure is concerned with the increased risk that individual institutions face in the case of a market disruption. This work, therefore, adds to and expands upon previous research by employing an approach that can adapt to time-varying tail dependency. Sedunov (2016) asserts that institutions that are highly susceptible to crises but do not initially add much to systemic risk tend to amplify the financial system's distress. The majority of empirical research on measuring systemic risk has mainly examined developed financial markets in the US and Europe. It is not surprising that the existing methods are tested in the advanced economies, as the GFC started in the US and the European sovereign debt crisis began in Europe. In the African context, there are a handful of studies which address systemic risk, especially in South Africa. The effect of the GFC had a negative impact on economies in Africa, especially the South African economy. Even though financial institutions in South Africa were not at the centre of the GFC, the effect of the financial crisis had a negative impact on them and the South African economy in general. The potential for any disruptive crisis in the SA economy cannot be ignored. The effect of COVID-19 has been heavily felt in SA, leading to significant job losses and macroeconomic imbalances, which were marked by deterioration of the fiscal balance (Arndt et al., 2020). According to Chitiga-Mabugu et al. (2021), the South African economy has been on the decline with economic growth falling from 3% in 2010 to 1.5% in 2019. The negative impact of COVID-19 is reflected in the South African financial markets. According to Kells (2020), it was not surprising that the financial market experienced volatility during the pandemic. Banks, including those in the South African economy, were vulnerable during the economic downturn due to the likelihood of nonperforming loans. Given the crisis and its devastating effects, it is imperative to understand the level of riskiness in the financial sector, gauge the extent of damage, and identify the risk exposures in the economy. Cele and Mlitwa (2024) proposed the need to explore the impact of COVID-19 on the financial sector to evaluate the influence. Given the substantial presence of South African banks throughout the continent, it is critical to assess their risk profile as it will impact other financial institutions in neighbouring African nations. This could pose a cross-border systemic risk exposure of financial institutions in different African countries linked to South African banks.

This chapter assesses the possible distribution of systemic risk among South African financial institutions in light of the context above. The primary goal is to analyse the systemic risk potential of South African banks using current methods. Regulators will be able to assess which institutions may be potential candidates for rescue in the case of a crisis by understanding the risk exposure of the institutions. Furthermore, the extent of an organisation's exposure to systemic risk will enable researchers and policymakers to pinpoint the fundamental factors that contribute to their risk exposure.

The chapter makes three significant contributions. First, by offering a thorough examination of the systemic risk potential across South African banks, it closes substantial gaps in the literature. Additionally, it lists South Africa's systemically important financial institutions and investigates each one's risk contribution. The "CoVaR" approach is another tool used in this chapter to quantify each bank's systemic impact. The CoVaR quantifies an institution's contribution to systemic risk at different quantiles and evaluates the value-at-risk of the financial system conditional on an institution being in a distressed state.

This chapter assesses the possible distribution of systemic risk among South African financial institutions in light of the context above. The primary goal is to analyse the systemic risk potential of South African banks using current methods. Regulators will be able to assess which institutions may be potential candidates for rescue in the case of a crisis by understanding the risk exposure of the institutions. Furthermore, the extent of an organisation's exposure to systemic risk will enable researchers and policymakers to pinpoint the fundamental factors that contribute to their risk exposure.

The chapter makes three significant contributions. First, by offering a thorough examination of the systemic risk potential across South African banks, it closes substantial gaps in the literature. Additionally, it lists South Africa's systemically important financial institutions and investigates each one's risk contribution. The "CoVaR" approach is another tool used in this chapter to quantify each bank's systemic impact. The CoVaR quantifies an institution's contribution to systemic risk at different quantiles and evaluates the value-at-risk of the financial system conditional on an institution being in a distressed state.

We employ the data that is currently accessible for the seven biggest South African banks: ABSA, Capitec, First Rand, Investec, Nedbank, Standard Bank, and JSE. We employed the CoVaR and VaR metrics. Since our data is non-normal, we estimate our marginal models using GARCH, exponential GARCH, and GJR GARCH, utilising the normal, Student's *t*, skewed *t* distribution, and generalised error distribution to capture the skewness and excess kurtosis. We selected the normal distribution as a benchmark distribution even though the data was not normally distributed. Scholars have created new analytical techniques to investigate the impact of financial system interconnection

(Bisias et al., 2012). The specific goal of this thesis is to use CoVaR to identify systemically important banks. To help academics and policymakers assess the differences between various GARCH models, comparisons between them are made. We apply many model selection criteria to determine which of the GARCH models is the best. More specifically, the lowest Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) criteria will be used to verify the best-fitted model. To choose the optimal model, we further employed three back-testing techniques: conditional coverage, dynamic quantile testing, and unconditional coverage. Results based on the CoVaR measure indicated that ABSA is the most systemically important bank in SA for the period being considered in this thesis.

3.2 Literature Review

3.2.1 Introduction

This chapter provides a detailed summary of the existing literature on systemic risk and financial stability. Systemic risk and financial stability are interdisciplinary topics; hence, the literature is reviewed from different perspectives. In regular times, banking systems can help economic growth by funding investment opportunities and reducing volatility in the financial markets by facilitating risk sharing. However, during financial crises, the banking system itself can become a vehicle for amplifying and spreading financial shocks (Haldane, 2013). The 2007-2009 financial crisis showed the fragility of the banking system. The effects of the global financial crisis on the world economy highlighted the need to investigate the connectivity that exists in the international financial system. After the crisis, several studies were conducted to measure systemic risk or analyse contagion risk in sovereign debt markets. Since the global financial crisis, academics and regulators have developed multiple techniques to measure systemic risk. This is because “the size and complexity of the financial system imply a diversity of legal and institutional constraints, market practices, and participant characteristics as exogenous factors driving the system at any given time. Accordingly, there is a corresponding diversity of models and measures that emphasise the different aspects of systemic risk (Bisias et al., 2012).

3.2.2 Definition of systemic risk

The financial crisis has created a new focus on trying to define, measure, and regulate systemic risk. The existence of multiple possible definitions of systemic risk automatically implies that there will be more than one risk measure to capture the complex and adaptive nature of financial markets. Academics, practitioners and regulators are yet to come to a consensus on an appropriate definition for systemic risk. With such a widespread focus on systemic risk, there are multiple working definitions for the term.

According to Billio et al. (2012), there is no standard definition of systemic risk. Hansen (2013) points to the difficulty in identifying and quantifying systemic risk, unlike other risks in the financial sector. In the author's view, the plethora of existing measures has resulted from the lack of a proper definition and understanding of what systemic risk is about and as such, describes it as a "grab bag of scenarios" that provides the rationale for regulators to intervene in financial markets. Billio et al. (2012) describes systemic risk as any set of circumstances that threatens the stability of or public confidence in the financial system. Denkowska and Wanat (2020) defines systemic risk as a risk coming from the financial system itself due to a series of failures of financial institutions, which amplifies its exogenous version. Cummins and Weiss (2014) define systemic risk as "the risk that an event will trigger a loss of economic value or confidence in a substantial segment of the financial system that is serious enough to have a significant adverse effect on the real economy with a high probability". Acharya et al. (2017) describe systemic risk as "the propensity of a financial institution to be undercapitalised when the financial system as a whole is undercapitalised". As Summer (2003) notes, there appears to be consensus that systemic risk arises from the simultaneous failure of several financial institutions, despite the literature's wide range of definitions of the term. Furthermore, we can differentiate between systemic risk and a "normal" crisis due to the many definitions of the term. According to Eling and Pankoke (2016), a systemic risk denotes a financial services impairment that affects the entire economy, as opposed to a "normal" crisis that only affects the financial sector.

3.2.3 Review of systemic risk measures

The literature on systemic risk measures for financial institutions has been extensively explored, and numerous approaches have been proposed. For instance, the recent survey by Bisias et al. (2012) alone highlights over 30 measures of systemic risk. Each of these measures deals with different themes, employs different indicators and therefore has different policy implications, although their ultimate aim is to ensure a healthy global financial system. The existing measures can be grouped into broad categories.

3.2.3.1 Stress testing measures

Determination and prediction of the factors which affect the stability of the financial sector are crucial for its sustainability. Some of these factors include deterioration of the balance sheet due to non-performing loans, an increase in GDP, uncertainty, and a hike in interest rates, among others. When these factors are determined, precautions can be taken to avoid a financial crisis. These factors are usually determined through stress testing of the financial institutions. Stress testing is an essential tool for the assessment and classification of systemic risk. The use of macro stress tests to assess bank

solvency has developed rapidly over the past few years. This development was reinforced by the financial crisis, which resulted in substantial losses for banks and created general uncertainty about the banking sector's loss-bearing capacity. Stress testing is a basic method to determine whether the capital ratios of financial institutions are adequate. The global financial crisis and the prolonged post-crisis recovery have enhanced the relevance of effective methods for assessing the level of vulnerability of banking systems to systemic risks under the conditions of a changing macroeconomic situation (Zayernyuk et al., 2015).

Stress testing is a method used to measure the fragility of a portfolio or the whole financial system against various hypothetical scenarios (Başarır and Toraman, 2014). Stress-testing is considered the aggregate of multiple methods used by banks for assessing their vulnerability concerning exceptional, but probable events (Global Financial System, 2000). Stress tests usually evaluate the financial performance of financial institutions due to their financial solvency and liquidity (Vinals, 2012). In general, stress tests can be referred to as a risk measurement technique to measure the potential loss when an infrequent but probable risk occurs. The IMF defines stress testing as a variety of methods employed to evaluate the exposure of a portfolio to significant changes in the macroeconomic environment or to exceptional but feasible events (Blaschke et al., 2001). The stress-testing of banks is represented by a complex, multi-stage process of assessing the vulnerability of the banking system and its ability to withstand the impact of risks generated by several exceptional, but probable events (Zayernyuk et al., 2015).

Macroprudential stress test analyses the vulnerability of the whole financial system, while microprudential stress test examines the portfolio of financial institutions and evaluates the portfolios due to their risk management systems (Başarır and Toraman, 2014). One of the shortcomings of micro stress testing was that it was entirely portfolio-focused and ignored the health of the financial system or the build-up of systemic risk. However, macro stress test, which is usually used by central banks and supervisors, focuses on the health of the entire financial system rather than individual banks and is used as a crisis management tool (Taskinsoy, 2022). According to Bernanke (2013), macro tests add a forward-looking aspect to capital ratios, take care of tail risks, and ensure an adequate level of capital together with sufficient capital buffers at systemically important financial institutions to absorb losses. Borio, Drehmann, and Tsatsaronis (2014) argue that stress testing must have four common elements; (1) risk exposures that are pertinent to stress; (2) parameters of the scenario (exogenous) that send plausible shocks to risk exposures; (3) a model that simulates shocks at different levels to analyze and monitor propagation of various stress points through the system; (4) measurement of the outcome. Alfaro and Drehmann (2009) proposed a stress test that relies on macroeconomic factors such as GDP and observed that these macroeconomic factors weaken significantly before banking crises. Banks also use stress testing as a part

of their internal risk management tool to manage risks of their existing portfolio, gain data to make a regular workflow plan (Vinals, 2012). The 10-by-10-by-10 approach by Duffie (2013) monitors systemic risk by analysing the exposure of systemically important financial institutions to a list of stressful scenarios. Under the different scenarios, the financial institutions will report their gains and losses. According to Blaschke et al. (2001), the selection of stress tests depends on several factors, including the type of risk, the type of shock to apply, the scenario being considered, and the assets to be shocked over some time, as well as the magnitude of the shock. Stress tests, a tool for risk management, develop severe but possible scenarios about the market, interest, exchange rate, credit and liquidity risks and try to determine how strong the banking sector is against these scenarios (Başarır and Toraman, 2014). The stress-testing of banks is represented by a complex multiple-stage process of assessing the vulnerability of the banking system and also its ability to stand against the impact of risks generated by several exceptional, but probable events. According to Rebonato (2010), stress-testing is not necessarily an accurate instrument that can be used with scientific scrupulousness but instead demands application of econometric methods, expert evaluation and a series of sober-minded assumptions. The key advantage of stress-testing is that a credit organisation relying on the results of stress-testing can foresee potential risks, assess them, and allocate funds in advance for loss recovery (Zayernyuk et al., 2015).

3.2.3.2 Forward looking risk measures

Another category of systemic risk measures is a set of forward-looking risk measures, which might help focus regulatory scrutiny on emergent risk factors and other exposures before they begin to appear in traditional financial statements. Gray and Jobst (2010) use market implied expected losses to construct their systemic risk measure, known as the contingent claims analysis (CCA). This approach not only measures the individual contributions of financial institutions to overall contingent liabilities in the event of systemic distress, but also measures the potential risk transfer to the government. CCA, which is a generalisation of the option pricing theory (OPT) by Black and Scholes (1973), as well as Merton (1974), is forward-looking by construction, providing a consistent framework based on current market conditions rather than on historical experience (Gray and Jobst, 2010). CCA is a framework that combines market-based and balance sheet information to obtain a comprehensive set of company financial risk indicators (Saldías, 2013). According to Board et al. (2009), CCA is a tool to enhance systemic risk analysis and to identify systemically important financial institutions and help establish a regulatory framework that can cope with risk arising from systemic linkages. The CCA approach produces time-varying point estimates of risk indicators that can be periodically updated, becoming more comprehensive than alternative measurement approaches to systemic risk (Jobst and Gray, 2013). Based on the Merton (1974) approach to credit risk, CCA has three principles: (1) the economic value of

liabilities is derived from assets; (2) liabilities have different priorities, and (3) assets follow a stochastic process (Gray, Echeverría, and Luna, 2006; Gray and Jobst, 2010).

3.2.3.3 Measures of illiquidity and insolvency

The third strand focuses on illiquidity and insolvency of financial institutions or the financial system. Following the 2007-2008 GFC, the notion that systemic risk is primarily the risk of cascading bank defaults has evolved into the belief that it encompasses both cascading bank defaults and funding liquidity shocks (Hurd et al., 2014). Both shocks impede the operations of the remaining undefaulted banks. Bank stress testing has mainly focused on solvency, which evaluates the exposure of bank portfolios to a macro-stress scenario and compares this exposure with the bank's capital to assess capital adequacy (Schuermann, 2014). According to Gai and Kapadia (2010), insolvency of a bank refers to a bank whose net-worth is non-positive and has the potential to disrupt the asset side of the balance sheet of each creditor. According to Jovanović, Todorovic, and Grbic (2017), liquidity crisis, solvency crisis, profitability crisis and lack of competitiveness threaten the functioning of financial institutions and the economy as a whole. Although all these factors are closely related and essential for the survival of the financial institution, the inability to discharge liabilities regularly (liquidity crisis) is a sign of financial crisis. The crisis is exacerbated whenever financial institutions are unable to pay liabilities, including interest and repayment of financial debts, leading to an insolvency crisis. In a study conducted by Chung and Tam (1993) to predict insolvency, they found gross profit to total assets, retained profit to total assets, working capital/total assets, market value of equity/total liabilities, quick liquidity ratio, interest costs, liabilities/total assets, and return on assets as indicators for insolvency in New Zealand.

In practice, the determination of insolvency takes into consideration numerous factors, including financial statements, adequacy of capital and amount of existing debt; and this exercise is complex (Cieri, Ganske, and Lennox, 1998). The current models of systemic risk treat funding illiquidity and insolvency as two distinct and separate phenomena (Hurd et al., 2014). While the Basel III framework emphasises the need for a unified stress testing approach, the assessment of solvency and liquidity risk has remained largely fragmented (Cont, Kotlicki, and Valderrama, 2020). The need for interactions between liquidity risk and solvency in the context of financial stability has been emphasised in many theoretical and empirical studies (Schmitz, Sigmund, and Valderrama, 2019; Cetina, Lelyveld, and Anand, 2015; Cornett et al., 2011; Imbierowicz and Rauch, 2014; Pierret, 2015).

3.2.3.4 Cross sectional measures

The recent global financial crisis has led bank supervisors and regulators to rethink the rationale of banking regulation. There have been attempts to measure systemic risk using cross-sectional methods that estimate the co-dependence of financial institutions and the financial system. Some of the approaches rely on public market data such as stock returns, option prices or CDS spreads. Huang, Zhou, and Zhu (2009) proposed the distress insurance premium, which is an insurance premium to cover distressed losses in a banking system. Within the framework, the systemic importance of each bank can be adequately defined as its marginal contribution to the hypothetical distress insurance premium (DIP) of the whole banking system. The DIP studies the time variation and cross-section of the systemic risk contributions of large complex financial institutions (LCFIs). The marginal contribution of each bank under this framework adds to the aggregate systemic risk. Other approaches, such as the CoVaR measure proposed by Adrian and Brunnermeier (2008) and marginal expected shortfall (MES) proposed by (Acharya et al., 2017), use market-based indicators on systemic risk contributions. Adrian and Brunnermeier (2008) propose to measure systemic risk via an increase in the conditional Value-at-Risk (CoVaR) of the financial system. This measure, denominated ΔCoVaR , can quantify the risk contribution of individually systemically important institutions, which are so interconnected and significant that they can cause risk spillover effects on others. The Co-Risk of Chan-Lau et al. (2009) measures the tail co-dependence between credit default swaps of various financial institutions. Allen, Bali, and Tang (2012) propose a systemic risk index called CATFIN, which associates systemic risk with the VaR of the financial system. Brownlees and Engle (2017) introduces the SRISK to measure the expected capital shortfall of a firm conditional on a severe market decline. This metric is a function of the size of the firm, its degree of leverage, and its expected equity loss conditional on a market decline, which is denominated as Long Run Marginal Expected Shortfall (LRMES). Acharya et al. (2017) introduce the marginal expected shortfall (MES) and systemic expected shortfall (SES). The MES is defined as the expected decrease of an institution's net equity return conditional on a market decline. The SES is a linear combination of MES and leverage.

3.3 Methodology

3.3.1 Introduction

This chapter discussed the empirical methodology that was employed to measure systemic risk. The first section discusses the data sources and descriptions. It discloses the banks included in the study and highlights the process of selecting these banks. The second section discusses the GARCH models used in the study. In addition, the distributions used for the study were discussed. Model selection criteria were also discussed.

The chapter ends with measures of systemic risk.

3.3.2 Data Source and description

This section focuses on the time series data for each of the six financial institutions and the JSE for the risk analysis. We investigate systemic risk in the South African financial sector using data from six major banks, namely Firststrand, Standard, Absa, Capitec, Nedbank, and Investec. These are the largest banks in South Africa with a market capitalisation of 2 billion USD. These banks are among the top 10 financial institutions in South Africa by market capitalisation. As a proxy for the financial system, we used the Johannesburg Stock Exchange financials index. We collected 5553 daily equity prices of the banks and the JSE with a sample period from February 19 2002, to May 31 2023. Even though other banks had data before 2002, the availability of Capitec data determined the start date for the period of analysis. The data covers the recent GFC of 2007/2008 and the global health pandemic caused by COVID-19. The data was extracted from Thomson Reuters DataStream and is denominated in Rands. We computed the daily log-returns for each bank and the banking system using the formula.

$$r_t = \log \left(\frac{P_t}{P_{t-1}} \right) \quad (3.1)$$

where P_t and P_{t-1} are the equity prices at day t and $t - 1$, respectively, and $r[t]$ is the return. We divided the data into in-sample (19/02/2002 to 27/02/2019) and out-of-sample (28/02/2019 to 31/05/2023). The in-sample data was used to estimate the model, and the out-of-sample data was used to forecast. The estimation sample consisted of 4442 observations, while the forecasting data had a total of 1110 observations.

3.3.3 Univariate Volatility GARCH Models

3.3.3.1 Standard GARCH

Although the ARCH model has been widely used in financial modelling for its ability to capture volatility clustering, limitations such as the inability to capture long-range dependence and non-linear relationships between variables have limited the ARCH model in the modelling of practical financial data. In practice, the application of the ARCH model often requires a larger order, which will not only increase the number of parameters to be estimated but also lead to multicollinearity (Asteriou and Hall, 2011). To address the existing problems in the ARCH model, Bollerslev extended it by adding an autoregressive term to get the GARCH model. It has been widely used for modelling volatility in financial time series data. The GARCH model allows for the estimation of time-varying volatility that is dependent on past observations and captures the volatility clustering phenomenon, which is commonly observed in financial data. Bollerslev

(1986) added another term to the original ARCH model to allow the σ_t to not only depend on the value of past squared observations ε_{t-i}^2 but also on the value of past variance σ_{t-1}^2 . The simplest form of GARCH(1,1) can be shown as

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (3.2)$$

The generalized form of 3.2 as proposed by Bollerslev (1986) is GARCH(p,q) and can be shown as

$$\sigma_t^2 = \omega + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \quad (3.3)$$

GARCH(p,q) model has the parameter constraints of $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ for $i = 1, \dots, q$ and $j = 1, \dots, p$ so that the conditional variance is strictly positive. The stationary condition for GARCH(1,1) process is $\alpha_1 + \beta_1 < 1$. The term $\sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2$ contains past values of the process, ε_t , which mean the process depends on its past. The term $\sum_{j=1}^p \beta_j \sigma_{t-j}^2$ contains past volatilities and allows the conditional standard deviation to exhibit periods with more persistent volatility compared to an ARCH model. The σ_t^2 denotes the conditional variance, ε_t^2 denotes the residuals from the mean process.

The sum $\alpha_i + \beta_j$ measures the persistence of a shock to the conditional variance in an equation. The main feature of this model is that it captures volatility clustering in the data through the persistent parameter $\alpha + \beta$ with the stated restrictions to ensure a uniquely stationary process and positivity of the conditional variance. However, if the persistence parameter equals one, the GARCH model converges to the integrated GARCH model where the long-term volatility bears an explosive process. In GARCH, the impact of positive and negative news on the conditional variance is symmetrical. The model restricts all coefficients to be greater than zero. It is thus unable to explain the negative correlation between returns and volatility.

The original GARCH model has some limitations. It assumes that the conditional variance is linearly dependent on past squared error terms and that the errors are independently and identically distributed with a normal distribution. These assumptions may not hold in real-world financial data, leading to model misspecification and inaccurate predictions. Several extensions have been proposed to the GARCH model due to the violation of the non-negativity condition and inability of the GARCH model to account for leverage effects, despite its ability to account for volatility clustering and leptokurtosis. One key restriction of the GARCH model is that it enforces a symmetric response of volatility to positive and negative shocks. To address these limitations, various extensions to the GARCH model, such as GJR-GARCH, EGARCH, GARCH-M, IGARCH, TGARCH, have been proposed. The GJR-GARCH model introduced by Glosten, Jagannathan, and Runkle (1993) includes an additional parameter to capture the asymmetric effect of positive and negative shocks on volatility. The EGARCH model proposed by

Nelson (1991) allows for the modelling of the leverage effect, which is the tendency for negative shocks to have a larger impact on volatility than positive shocks. These extended models have been shown to provide more accurate predictions of volatility in financial time series data than the original GARCH model. Some of these extensions considered in the study include the EGARCH and GJR-GARCH. These two popular asymmetric models are considered in this thesis.

3.3.3.2 Exponential GARCH

The exponential GARCH (EGARCH) model, which allows for asymmetric and leverage effects, was proposed by Nelson (1991). EGARCH was the first model to incorporate the asymmetric volatility (Nelson, 1991). The EGARCH(p,q) incorporates the asymmetric impact of positive and negative shocks on volatility, whereby the latter is believed to produce greater levels of volatility, despite having the same magnitude. Empirical studies have shown that the EGARCH provides a more accurate result compared to the standard GARCH (Alberg, Shalit, and Yosef, 2008). This model is specified in logarithmic form, which suggests that parameters are unrestricted and are thereby allowed to take negative values while ensuring a positive conditional variance. In addition, the conditional variance is written as a function of past standardised innovations, rather than past innovations. Formally, the volatility dynamics of an EGARCH (1,1) can be written as

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] \quad (3.4)$$

where β represents the persistence parameter, α and γ represent the size and the sign (leverage) effect, respectively. The above specification exhibits an asymmetric effect when $\gamma \neq 0$. More specifically, if the leverage parameter (γ) is negative, this suggests that negative news affects volatility more than positive news. Conversely, if returns and volatility are positively correlated, γ will be positive, indicating positive shocks will have a higher impact on volatility than negative shocks, which is irregularly the case. If $\gamma \neq 0$ and significant, $(\alpha + \gamma)$ is the effect on volatility of a previous positive return, whereas $\gamma - \alpha$ is the corresponding effect when the previous return has been negative.

The model captures the asymmetric volatility through the variable gamma (γ). The sign of gamma determines the size of the asymmetric volatility and whether the asymmetric volatility is positive or negative. If $\gamma = 0$, there is symmetry, i.e. no asymmetric volatility. If $\gamma < 0$, negative shocks will increase volatility more than positive shocks. If $\gamma > 0$, positive shocks increase volatility more than negative shocks. Given that the model uses the log of the variance (σ_t^2), it means that if the parameters are negative, the

variance will still be positive. The model is not subject to the non-negativity constraints (Nelson, 1991). ω is the intercept for the variance (constant), β is the coefficient for the logged GARCH term, $\ln(\sigma_{t-1}^2)$ is the logged GARCH term, γ is the scale of the asymmetric volatility/leverage, $\gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}}$ is the last period's shock which is standardized,

$\left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right]$ is the parameter that takes into account the absolute value of last periods volatility shock. It replaces the ARCH term.

3.3.3.3 GJR GARCH

The GJR-GARCH model Glosten, Jagannathan, and Runkle (1993) is similar to EGARCH (1,1) in incorporating the asymmetric impact of positive and negative shocks. The GJR-GARCH model is an extension of GARCH with an additional term to account for asymmetry. The volatility equation of a GJR-GARCH is given by

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \gamma I_{t-1} \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (3.5)$$

where $I_{t-1} = 1$ if $\varepsilon_{t-1} < 0$ and $I_{t-1} = 0$ if $\varepsilon_{t-1} \geq 0$. A defining feature of this model is that a positive shock will increase volatility by α_t , whereas a negative shock will increase volatility by $\alpha_t + \gamma_t$ at t . However, in contrast to the EGARCH model, the leverage effect exists when $\gamma > 0$, indicating that past bad news has a stronger impact on current volatility than past good news. If $\gamma < 0$, then past positive returns increase current volatility more than past negative returns. The persistence in this model relies on α , β and γ_k with k representing the average value of standardised errors. parametric restrictions are similar to the standard GARCH whereby $\omega \geq 0$, $\alpha \geq 0$ and $\beta \geq 0$. The GJR-GARCH model of (Glosten, Jagannathan, and Runkle, 1993) models positive and negative shocks on the conditional variance asymmetrically via the use of the indicator function I .

3.3.3.4 Persistence

The persistence in volatility, which is exhibited by financial time series data, can be determined using the coefficients of GARCH models. The persistence of a GARCH is computed as the sum of the GARCH and ARCH coefficients ($\beta + \alpha$). The volatility ($\beta + \alpha$) in most financial time series models is very close to one (Ahmed et al., 2018). If $\alpha + \beta < 1$, the model ensures positive conditional variance as well as stationarity. If $\alpha + \beta = 1$, the model is an exponential decay model and the half-life becomes infinite, suggesting that the model is strictly stationary. If $\alpha + \beta > 1$, the GARCH model

is considered to be non-stationary, suggesting that the volatility detonates towards infinity (Ahmed et al., 2018). It also indicates that the conditional variance is unstable, unpredictable and non-stationary (Kuhe, 2018).

3.3.3.5 Half life

Half-life volatility measures the mean-reverting speed of price or returns. The half-life volatility is given as $Half - life = \frac{\ln(0.5)}{\ln(\alpha + \beta)}$. The value of $\alpha + \beta$ influences the mean reverting speed (Ahmed et al., 2018). This suggests that when $\alpha + \beta$ is closer to one, the volatility shocks of the half-life will be longer.

3.3.4 Conditional distribution

The GARCH model supports a range of univariate distributions, including normal, Student's t, skewed Student's t, and generalised error.

3.3.4.1 Normal distribution

The normal distribution is a spherical distribution entirely described by its first two moments: the mean and variance. The random variable X is said to be normally distributed with mean μ and variance σ^2 , with density given by

$$f(x) = \frac{e^{\frac{-0.5(x - \mu)^2}{\sigma^2}}}{\sigma\sqrt{2\pi}} \quad (3.6)$$

The residuals, ε , standardized by σ yield the standard normal density given by

$$f\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma}f(z) = \frac{1}{\sigma} \left(\frac{e^{-0.5z^2}}{\sqrt{2\pi}} \right) \quad (3.7)$$

The normal distribution has zero skewness and zero excess kurtosis.

3.3.4.2 Student distribution

The student model is an alternative to the normal distribution for fitting the standardised innovations. It is described by a shape parameter ν . For standardisation, the three-parameter representation is given as

$$f(x) = \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\sqrt{\beta\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(x - \alpha)^2}{\beta\nu}\right)^{-\frac{(\nu + 1)}{2}} \quad (3.8)$$

where α , β , and ν are the location, scale and shape parameters, respectively, and Γ is the Gamma function. It is a unimodal and symmetric distribution where the location parameter α is the mean of the distribution while the variance is $var(x) = \frac{\beta\nu}{(\nu-2)}$. The standardised student's distribution is given by

$$f\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}f(z) = \frac{1}{\sigma} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{(\nu-2)}\pi\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{Z^2}{(\nu-2)}\right)^{-\left(\frac{\nu+1}{2}\right)} \quad (3.9)$$

The student distribution has zero skewness and excess kurtosis equal to $\frac{6}{(\nu-4)}$ for $\nu > 4$.

3.3.4.3 The generalised error distribution

The generalised error distribution (GED) is a three-parameter distribution that belongs to the exponential family with conditional density given by

$$f(x) = \frac{\kappa e^{-0.5\left|\frac{x-\alpha}{\beta}\right|^\kappa}}{2^{1+\kappa^{-1}}\beta\Gamma(\kappa^{-1})} \quad (3.10)$$

with α , β and κ representing the location, scale and shape parameters. Since the distribution is symmetric and unimodal, the location parameter is also the mode, median and mean of the distribution. All odd moments beyond the mean are zero by symmetry. The variance and kurtosis are given by

$$Var(x) = \beta^2 2^{\frac{2}{\kappa}} \frac{\Gamma(3\kappa^{-1})}{\Gamma(\kappa^{-1})}$$

$$Ku(x) = \frac{\Gamma(5\kappa^{-1})\Gamma(\kappa^{-1})}{\Gamma(3\kappa^{-1})\Gamma(3\kappa^{-1})}$$

3.3.5 Model selection criteria

Model selection has become an essential focus in recent years in statistical analysis, econometrics, machine learning, and big data analytics (Burnham and Anderson, 2002). A wide variety of model selection methods have been proposed and used over the past decades. Selecting an optimal model can be difficult in statistical analysis. As a result, different model selection criteria are used to choose the best model from possible alternatives. After fitting different GARCH models to the returns using other distributions, it is essential to sort the models according to the value of the information criterion. Thus, it is crucial to select the most appropriate model from a set of candidate models. Information criteria refer to model selection methods that are based on likelihood functions (Ding, Tarokh, and Yang, 2018). The Akaike's Information Criterion (AIC),

Bayesian Information Criterion (BIC), SIC and HQIC are among the common statistical tools that have been used to measure model performance and select the best model from a set of potential models.

3.3.5.1 Akaike Information Criterion (AIC)

The AIC was introduced by Akaike (1973) and is calculated by obtaining the maximum value of the likelihood function for the model and the penalty term due to the number of estimated parameters in the model. The AIC balances the goodness of fit of the model against its complexity by incorporating a penalty for the number of parameters. This criterion implies that adding more parameters to the model improves the goodness of the fit, but also increases the penalty imposed by adding more parameters. The Akaike information criterion (AIC) assesses the goodness-of-fit of the model without overfitting. In selecting between two models, the model with the least AIC is considered the most appropriate. The AIC is given as:

$$AIC = 2k - 2\ln(\hat{L}) \quad (3.11)$$

where k is the number of parameters and $\hat{L}$ is the maximum likelihood of the model.

3.3.5.2 Bayesian Information Criterion (BIC)

The Bayesian information criterion (BIC), also known as the Schwarz information criterion, was introduced by Schwarz (1978) for model selection. It provides a balance between the goodness of fit of the model and its complexity using a different penalty term. It is based on the likelihood function and introduces a stronger penalty term for the number of parameters. The model with the least BIC value is considered the best. Generally, the BIC is defined as:

$$BIC = k\ln(n) - 2\ln(\hat{L}) \quad (3.12)$$

where n is the number of observations, k is the number of parameters in the model.

3.3.5.3 Shibata Information Criterion (SIC)

The Shibata Information Criterion (SIC) is a model selection technique which is closely related to AIC but adjusted for finite sample sizes (Shibata, 1976). The SIC chooses the model that best balances goodness of fit and model complexity. Shibata (1976) demonstrated that the AIC tends to choose complex models. The Shibata Information Criterion (SIC) was developed to penalise overfitting. The SIC is given as

$$SIC = 2k - 2\ln(\hat{L}) + \frac{2k(k+1)}{n-k-1} \quad (3.13)$$

where k is the number of parameters, n is the sample size.

3.3.5.4 Hannan-Quinn Information Criterion (HQIC)

The Hannan-Quinn information criterion (HQIC) was designed to have the slowest growing penalty term for complexity compared to the AIC and the BIC (Hannan and Quinn, 1979). The HQIC balances the goodness of fit of a statistical model against the model's complexity. The HQIC penalises models for having more parameters. The HQIC was originally defined as:

$$HQIC = -2\ln(\hat{L}) + 2k\ln(\ln(n)) \quad (3.14)$$

3.3.6 Measures of Systemic Risk

3.3.6.1 CoVaR

Value-at-Risk (VaR) measures the worst expected loss over a specific time interval at a given confidence level. VaR is defined as the percentage of asset value that a bank might lose with q probability over a pre-set horizon T . The value of VaR by definition is negative in general. The confidence level (quantile) q and the period T are the two parameters in VaR. We considered 1%, 2.5% and 5% quantiles and daily return in this thesis. Conditional value-at-Risk (CoVaR) is defined as the VaR of the financial system on condition that a financial institution is in financial distress (Tobias and Brunnermeier, 2016). CoVaR measures the degree to which a tail event in a financial institution spills over and causes a tail event in another institution or financial system.

3.3.6.2 Δ CoVaR and systemic risk contribution

We used the Δ CoVaR to measure the systemic risk contribution of each financial institution. Δ is defined as the difference between the $\alpha\%$ -CoVaR of the financial system conditional on an institution being in financial distress, and the $\alpha\%$ -CoVaR of the financial system conditional on an institution being at the median state. Thus Δ CoVaR measures how much each financial institution adds to the risk of the entire system when that institution moves from regular operation to a state of distress. According to Tobias and Brunnermeier (2016), Δ CoVaR is the difference between CoVaR conditional on an institution being in distress and CoVaR conditional on an institution being in a normal state. Δ CoVaR is used to rank financial institutions in terms of their individual threat or risk contribution to the whole financial system. Thus, it measures the contribution of each institution to systemic risk in the entire system. Also, Δ CoVaR measures the exposure of individual financial institutions to systemic risk. According to Tobias and Brunnermeier (2016), in addition to calculating the risk of financial institutions in isolation using their VaR, it is equally important to include the systemic risk contribution

of the financial institution to the whole financial system. The marginal contribution of each financial institution is denoted by

$$\Delta CoVaR_q^{j|i} = CoVaR^{j|X^i=VaR_q^i} - CoVaR^{j|X^i=median^i} \quad (3.15)$$

Equation 3.15 provides the individual risk contribution to the risk of the entire financial system. In this thesis, we measure the risk contribution of each financial institution to potential systemic risk in the South African banking system and identify banks that are deemed systemically important.

3.4 Econometric Analysis and Results

3.4.1 Exploratory data analysis

Financial time series such as stock prices, returns, and other financial series are known to exhibit stylised facts which are critical for model specification, estimation, and forecasting (Abdalla, 2012). Table 3.1 shows the descriptive statistics of the equity price levels of the banks and the financial system. All the series have very high standard deviations. Capitec Bank has the lowest minimum and maximum prices. Standard Bank and Absa Bank have negative skewness, while the rest of the institutions exhibit positive skewness.

Table 3.1: Descriptive statistics of the daily stock prices

	Minimum	Maximum	Mean	Std. dev	Skewness	Kurtosis
JSE	1586.28	15578.12	7304.377	3746.984	0.2476	-1.1457
FNB	3003.3	46829.1	19846.71	12736.84	0.3206	-1.3856
Standard	853.1	7547.6	3713.884	1531.279	-0.0120	-0.5946
Absa	266.3	2200.8	1321.417	460.7443	-0.7179	-0.2408
Capitec	44.4	139613.1	29522.71	36016.2	1.2035	0.2775
Nedbank	1022.7	6073.8	3102.054	1209.402	0.1996	-1.0891
Investec	1033.5	13762.1	4893.932	2397.196	0.9036	1.0765

Figure 3.1 depicts the general trend characterised by the daily stock prices of the banks and JSE for the specified period in the study. The plot provides a general overview of the characteristics of each equity price. It can be seen from Figure 3.1 that the prices are non-stationary. The JSE prices demonstrated an upward trend from February 2002 until January 2016, when there was a sudden drop in the prices. It started the upward trend immediately with a few fluctuations. From March 2020, the stock price for the JSE dropped below the average price and fluctuated around the mean until late 2020, when it started the upward trend again. Between February 2002 and January 2013, the JSE stock prices, although showing an upward trend, were below the average stock price.

The FNB stock prices exhibited an upward trend from February 2002 until February 2018, followed by a drop and fluctuation until March 2020, where it dropped significantly. There were price fluctuations from March 2020 until late 2020, and the price trend began to rise again. Although a general upward trend existed between February 2002 and January 2014, the FNB stock prices were below the average price. The stock prices for Standard Bank demonstrated a particular momentum and trended upward until March 2015, after which it exhibited a downward movement till January 2016. It again displayed an upward trend until March 2020, after which it dropped below the average stock price. From late 2020, the stock price demonstrated a general upward trend with fluctuations. The Absa stock prices showed a general upward trend from February 2002 until December 2006, after which it dropped and fluctuated below the mean. During the GFC period, the stock prices showed a downward movement. However, from June 2009, it exhibited an upward trend until December 2015, dropping significantly and fluctuating until December 2017. There was a significant drop in the stock price in March 2020, below the average price, and it fluctuated until January 2021, when it showed an upward trend again. The Capitec bank stock prices fluctuated relatively below average from January 2002 to January 2010. After that, the stock prices began to display an upward trend. From March 2020, the stock price dropped significantly and fluctuated until late 2020 before going upward.

The Nedbank stock prices showed a downward trend from February 2002 until around August 2004, after which it exhibited an upward trend. The stock prices showed a general upward and downward trend, with values below the average until January 2012. During the GFC, the stock prices for Nedbank experienced a downward trend and only picked up after the crisis. There was a downward trend from January 2019, with a sharper downward trend in March 2020, where it fell below the average stock price. The stock prices remained relatively stable until October 2020, when they began trending upwards. From Figure 3.1, Investec stock prices showed a general upward trend until they peaked in March 2007, after which they exhibited a downward trend until March 2009. After the GFC period, the stock prices began to trend upward until 2016. Between January 2016 and March 2020, the stock prices showed fluctuations. There was a significant drop at the start of COVID (March 2020), and it remained stable until October 2020 and trended upward. During the GFC, all financial institutions except Capitec Bank experienced a decline in their stock prices. There was a general decline in the stock prices of all banks and the JSE during the COVID-19 era. It can be observed from Figure 3.1 that all the financial prices have stochastic trends. Thus, the stock prices for all the financial institutions and the JSE are non-stationary.

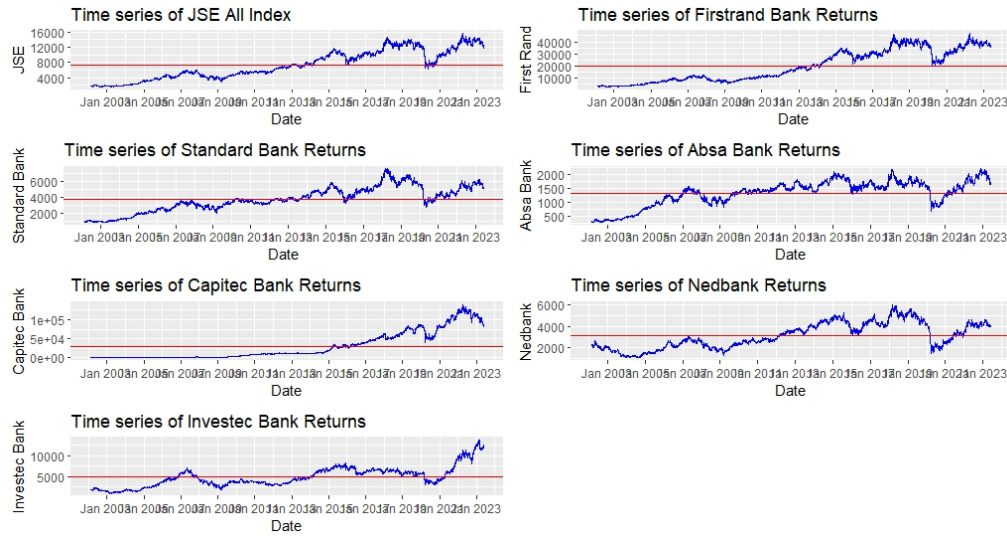


Figure 3.1: Daily stock prices of each series

Figure 3.2 shows the quantile-quantile (QQ) plot of the distributions of the daily returns of each bank from 2002 to 2023 with reference lines and confidence bands. The QQ plot is a scatter plot that compares the quantiles from a theoretical and empirical distribution. Thus, it compares the distribution of the data with a theoretical distribution. For all the banks and the JSE, the QQ plots show that the distributions of their daily returns have fatter tails than a normal distribution. The returns from the banks and the JSE thus have extreme values, as expected from financial data.

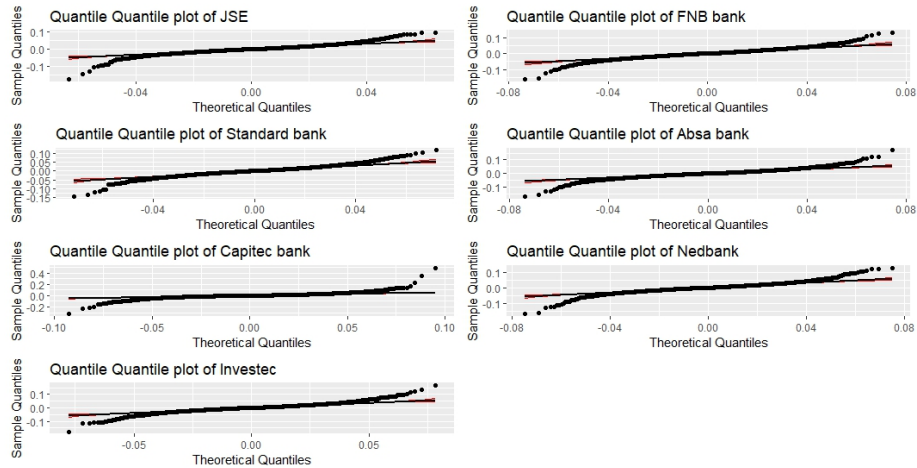


Figure 3.2: Quantile Quantile plots of daily stock returns

Figure 3.3 shows the distribution of the returns of the banks and JSE using histograms and density plots. Although it is difficult to judge the distribution of the stock returns

using the histogram, the density plots show significant deviations from normal distribution. The tails of all the banks and JSE are heavier than those of a normal distribution. Thus, the density plots of the returns of all banks and JSE substantially differ from a normal distribution.

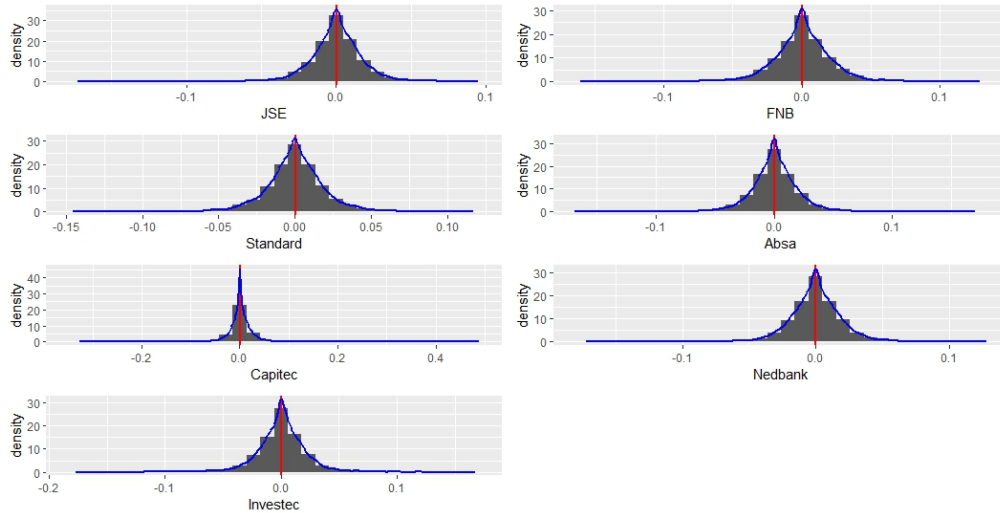


Figure 3.3: Histogram and density plot of daily stock returns

Table 3.2 reports descriptive statistics for the returns of all series considered. The average returns and the corresponding standard deviation are similar across all the markets. The differences between the maximum and minimum price returns show that the price return range for Capitec Bank is greater than all banks and the JSE index too. For all seven return series, their means are similar and very close to zero, which supports the market efficiency theory that expected returns are equal to zero. In line with usual stylised facts of financial time series, returns appear to be skewed and leptokurtic, indicating that they are not normally distributed. Negative values for skewness were more pronounced for the JSE Index than for Firststrand, Standard Bank, Absa, Nedbank and Investec, with only Capitec Bank showing positive skewness. All returns showed high values for the kurtosis statistic, consistent with fat tails in the returns distribution. The kurtosis values, which are all greater than three, indicate the returns exhibit fat tails.

In addition, the Jarque-Bera (JB) statistic confirms the departure from normality for all return series at the 5% level of significance. Based on the result of the Jarque-Bera test, we assumed that the underlying distributions of the stock returns in the GARCH model were Student t, skewed Student t, and generalised error distribution to account for the asymmetry and heavy tails in the seven financial time series. The results of the Dickey and Fuller (1979) and Phillips and Perron (1988) non-stationarity tests and the Kwiatkowski et al. (1992) stationarity test confirm that all the return series were stationary. Ljung-Box test for autocorrelation at up to the twelfth lag suggests the presence of

significant serial correlation for the weekly returns of all the banks at the 0.05 level of significance. This motivates specifying models for the conditional mean and variance to capture this predictability. The Engle's ARCH LM test provides strong evidence of serial dependence and heteroscedasticity for all returns. The ARCH-Lagrange multiplier (ARCH-LM) statistic indicated that ARCH effects could be found in all return series.

One important feature of financial returns is volatility clustering. Volatility clustering describes the tendency of significant changes in asset returns to follow large changes and small changes to follow small changes. It is evident in Figure 2 that there is volatility clustering in the data. Figure 2 suggests that periods of high price returns are usually followed by periods of high returns, while low price returns are likely to be followed by low price returns. The volatilities, as shown in Figure 3.4, are time varying. This volatility clustering suggests that the ARCH and GARCH models can estimate the residual. One significant point to note is that volatility occurs in clusters. The results of the ARCH effect test are presented in Table 3.2. The model is significant at 5%, thereby rejecting the null hypothesis suggesting that there is ARCH effects. A GARCH model can be applied to capture the changing variance.

Table 3.2: Summary statistics of daily stock returns

Statistic	JSE	FNB	Standard	ABSA	Capitec	Nedbank	Investec
Minimum	-0.1723	-0.1606	-0.1455	-0.1689	-0.3274	-0.1717	-0.1773
Maximum	0.0946	0.1291	0.1170	0.1697	0.4890	0.1282	0.1676
Mean	0.0003	0.0004	0.0003	0.0003	0.0012	0.0001	0.0003
Std. Dev	0.0170	0.0197	0.0191	0.0198	0.0251	0.0200	0.0209
Skewness	-0.3098	-0.1742	-0.0709	-0.1016	1.1525	-0.1670	-0.1068
Kurtosis	6.3342	4.2099	3.8505	5.9227	45.3562	5.8418	5.1823
ACF	0.021	0.003	-0.012	0.002	-0.093	0.039	-0.001
JB	9370.3*	4128.1*	3434.4*	8124.4*	477124*	7920.3*	6223.3*
ADF	-18.919*	-19.364*	-18.838*	-18.779*	-18.281*	-18.938*	-18.745*
PP	-73.292*	-75.055*	-76.189*	-74.912*	-82.497*	-71.587*	-74.789*
KPSS	0.1065	0.0970	0.1732	0.2700	0.4636	0.0472	0.0579
Q(12)	49.701 (0.000)	54.537 (0.000)	57.522 (0.000)	36.902 (0.000)	63.36 (0.000)	37.673 (0.000)	30.322 (0.002)
ARCH	455.11 (0.000)	318.89 (0.000)	187.94 (0.000)	528.78 (0.000)	55.227 (0.000)	464.42 (0.000)	374.42 (0.000)

Figure 3.4 displays the returns for all six banks and also for the JSE index. All the return series are volatile, indicating behaviour that is common in financial time series data. The data document the presence of large volatility clusters followed by periods of low volatility. These facts are consistent with the presence of different regimes of bull and bear market conditions, usually associated with high and low dependence.

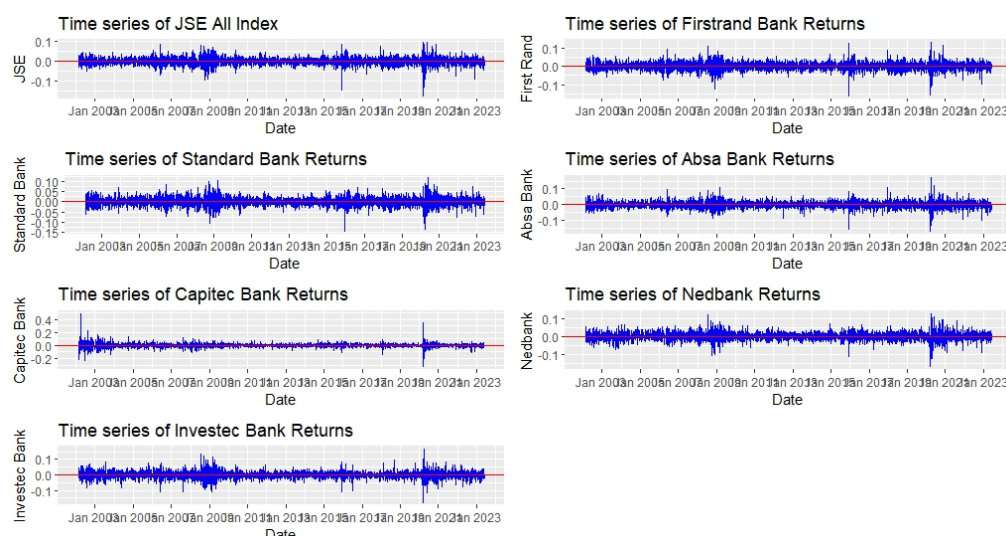


Figure 3.4: Daily stock returns

Table 3.3 displays the pairwise correlation coefficient estimations of stock returns for all banks used in the study and the Index from February 19, 2002, to May 31, 2023. The correlation coefficient ranges from 0.2330 to 0.9124 for Pearson correlation, 0.1472 to 0.7278 for Kendall Tau correlation and 0.2112 to 0.8944 for Spearman Rank correlation. The dependence (bivariate) between the banks and the index is positive and high, as demonstrated by the correlation coefficients for Pearson, Kendall Tau and Spearman rank. All the correlation coefficients are statistically significant at the 5% level of significance. Generally, the Pearson correlation is higher than Kendall's tau and Spearman's rank measure, possibly due to the high kurtosis as reported earlier and possible outliers. The general decrease in the correlation coefficient for Kendall's tau and Spearman's rank could be attributed to the loss of information and precision when converting the continuous returns into ranked variables. Kendall's tau and Spearman's rank measures are both non-parametric approaches (less sensitive to outliers) and consider the order of the items and not necessarily the numerical values.

Table 3.3: Kendall Correlation between daily stock returns

Kendall's Tau							
JSE	1.0000						
FNB	0.7278	1.0000					
Standard	0.7203	0.5268	1.0000				
Absa	0.6227	0.4732	0.5031	1.0000			
Capitec	0.2633	0.2161	0.2056	0.2084	1.0000		
Nedbank	0.5989	0.4645	0.4851	0.4626	0.2018	1.0000	
Investec	0.3566	0.3125	0.3145	0.2950	0.1472	0.3157	1.0000

3.4.2 Results of Marginal models

The log returns of the daily stock price of each of the financial institutions and JSE were modelled with different GARCH models (SGARCH, GJRGARCH, EGARCH) and different innovations (Student t distribution, skewed student t distribution and generalised error distribution) using data from 18 February 2002 to 31 May 2023. As the returns are not normally distributed, we used the student-t, skewed student t and generalised error distribution.

3.4.2.1 Model selection using Information criteria

The following information criteria, such as Akaike Information criteria (AIC), Bayes Information criteria (BIC), Shibata Information criteria (SIC) and Hannan-Quinn Information criteria (HQIC), were used to select the optimal model. Table 3.5, 3.6, and 3.4 display the AIC, BIC, SIC, HQIC and log-likelihood values for each of the financial institutions for different competing GARCH models using student's t, skewed student's t and generalised error distributions. Based on the various information criteria for model selection, ARMA (1,2)-GJRGARCH(1,1) is the best estimated model for JSE returns using the generalised error distribution. For the daily FNB stock returns, the optimal model using the information criteria is ARMA(2,1)-EGARCH(1,1) using the generalised error distribution. Also, the optimal models for the daily stock returns for Standard Bank, Absa, Capitec, Nedbank and Investec Banks are ARMA(1,2)-GJRGARCH-GED, ARMA(1,2)-EGARCH-GED, ARMA(1,1)-EGARCH-sstd, ARMA(1,2)-GJRGARCH-GED, and ARMA(2,2)-EGARCH-GED, respectively.

Table 3.4: GARCH models using information criteria

Models	Information criteria	student t distribution	skewed student t distribution	Generalised error distribution
Investec Bank				
ARMA(2,2) - SGARCH(1,1)	AIC	-5.1734	-5.1736	-5.1808
	BIC	-5.1626	-5.1616	-5.1701
	SIC	-5.1734	-5.1736	-5.1808
	HQIC	-5.1696	-5.1694	-5.1771
ARMA(2,2) - EGARCH(1,1)	AIC	-5.1791	-5.1793	-5.1868
	BIC	-5.1671	-5.1662	-5.1749
	SIC	-5.1791	-5.1793	-5.1868
	HQIC	-5.1749	-5.1747	-5.1826
ARMA(2,2) - GJRGARCH(1,1)	AIC	-5.1762	-5.1762	-5.1838
	BIC	-5.1643	-5.1631	-5.1719
	SIC	-5.1762	-5.1763	-5.1838
	HQIC	-5.1720	-5.1717	-5.1796

Table 3.5: GARCH models using information criteria

Models	Information criteria	Student t distribution	Skewed student t distribution	Generalised error distribution
JSE				
ARMA(1,2)- SGARCH(1,1)	AIC	-5.5763	-5.5759	-5.5785
	BIC	-5.5667	-5.5652	-5.5690
	SIC	-5.5763	-5.5759	-5.5785
	HQIC	-5.5730	-5.5722	-5.5752
ARMA(1,2)- EGARCH(1,1)	AIC	-5.5789	-5.5786	-5.5814
	BIC	-5.5681	-5.5667	-5.5707
	SIC	-5.5789	-5.5786	-5.5814
	HQIC	-5.5751	-5.5745	-5.5777
ARMA(1,2)- GJRGARCH(1,1)	AIC	-5.5794	-5.5791	-5.5818
	BIC	-5.5687	-5.5671	-5.5711
	SIC	-5.5794	-5.5791	-5.5818
	HQIC	-5.5756	-5.5749	-5.5781
FNB				
ARMA(2,1)- SGARCH(1,1)	AIC	-5.2196	-5.2194	-5.2217
	BIC	-5.2101	-5.2086	-5.2122
	SIC	-5.2196	-5.2194	-5.2217
	HQIC	-5.2163	-5.2156	-5.2184
ARMA(2,1)- EGARCH(1,1)	AIC	-5.2223	-5.2221	-5.2248
	BIC	-5.2115	-5.2101	-5.2141
	SIC	-5.2223	-5.2221	-5.2249
	HQIC	-5.2185	-5.2179	-5.2211
ARMA(2,1)- GJRGARCH(1,1)	AIC	-5.2222	-5.2219	-5.2245
	BIC	-5.2115	-5.2100	-5.2138
	SIC	-5.2222	-5.2219	-5.2245
	HQIC	-5.2184	-5.2178	-5.2207
Standard bank				
ARMA(1,2)- SGARCH(1,1)	AIC	-5.2971	-5.2967	-5.3011
	BIC	-5.2875	-5.2860	-5.2916
	SIC	-5.2971	-5.2967	-5.3011
	HQIC	-5.2937	-5.2930	-5.2978
ARMA(1,2)- EGARCH(1,1)	AIC	-5.2979	-5.2976	-5.3018
	BIC	-5.2872	-5.2856	-5.2910
	SIC	-5.2979	-5.2976	-5.3018
	HQIC	-5.2942	-5.2934	-5.2980
ARMA(1,2)- GJRGARCH(1,1)	AIC	-5.2987	-5.2983	-5.3029
	BIC	-5.2879	-5.2864	-5.2922
	SIC	-5.2987	-5.2983	-5.3029
	HQIC	-5.2949	-5.2941	-5.2992

Table 3.6: GARCH models using information criteria

Models	Information criteria	student t distribution	skewed student t distribution	Generalised error distribution
Absa				
ARMA(1,2)- SGARCH(1,1)	AIC	-5.2556	-5.2554	-5.2603
	BIC	-5.2461	-5.2446	-5.2507
	SIC	-5.2556	-5.2554	-5.2603
	HQIC	-5.2523	-5.2516	-5.2570
ARMA(1,2)- EGARCH(1,1)	AIC	-5.2601	-5.2598	-5.2648
	BIC	-5.2494	-5.2479	-5.2541
	SIC	-5.2601	-5.2598	-5.2648
	HQIC	-5.2564	-5.2557	-5.2611
ARMA(1,2)- GJRGARCH(1,1)	AIC	-5.2572	-5.2569	-5.2619
	BIC	-5.2465	-5.2450	-5.2512
	SIC	-5.2572	-5.2569	-5.2619
	HQIC	-5.2534	-5.2527	-5.2582
Capitec				
ARMA(1,1)- SGARCH(1,1)	AIC	-5.1718	-5.1723	
	BIC	-5.1634	-5.1627	
	SIC	-5.1718	-5.1723	
	HQIC	-5.1688	-5.1690	
ARMA(1,1)- EGARCH(1,1)	AIC	-5.1745	-5.1749	
	BIC	-5.1650	-5.1642	
	SIC	-5.1745	-5.1749	
	HQIC	-5.1712	-5.1712	
ARMA(1,1)- GJRGARCH(1,1)	AIC	-5.1737	-5.1740	
	BIC	-5.1641	-5.1633	
	SIC	-5.1737	-5.1740	
	HQIC	-5.1703	-5.1703	
Nedbank				
ARMA(1,2)- SGARCH(1,1)	AIC	-5.2376	-5.2373	-5.2436
	BIC	-5.2281	-5.2266	-5.2341
	SIC	-5.2376	-5.2373	-5.2436
	HQIC	-5.2343	-5.2336	-5.2403
ARMA(1,2)- EGARCH(1,1)	AIC	-5.2389	-5.2385	-5.2447
	BIC	-5.2282	-5.2266	-5.2339
	SIC	-5.2389	-5.2386	-5.2447
	HQIC	-5.2352	-5.2344	-5.2409
ARMA(1,2)- GJRGARCH(1,1)	AIC	-5.2390	-5.2387	-5.2448
	BIC	-5.2283	-5.2267	-5.2341
	SIC	-5.2390	-5.2387	-5.2448
	HQIC	-5.2353	-5.2345	-5.2411

Table 3.7 reports the parameter estimates and model diagnostics of the GARCH models for each of the financial institutions and the JSE. Overall, all the GARCH models perform well. The constant parameter, omega, is statistically significant for FNB, Absa, Capitec, Nedbank, and Investec at a 5% level of significance, and at a 10% level of significance for the JSE. However, it is not statistically significant for the GARCH model for the standard bank and therefore not useful. In all the GARCH models, both α and β are statistically significant at the 5% level of significance and different from zero; therefore, it is reasonable to suggest time-varying volatility of the residuals. It is expected that the sign of α in EGARCH models must be negative and statistically significant. The large coefficients of β_1 indicate that shocks to conditional variance persist for a long time and suggest that volatility is persistent. The sum of the two parameters ($\alpha_1 + \beta_1$) is less than one for all the models, indicating that the models do not result in explosive volatility predictions. Since the sum of the parameters is close to one, it suggests that the volatility dies down slowly, thus it reverts to the mean slowly. The sum of α and β ranges from 0.9308 for FNB to the high of 0.9650 for Standard Bank, indicating high persistence in the conditional volatility. Likewise, the GARCH (1,1) coefficients are positive, confirming the non-negativity condition of the model and validating the existence of asymmetry. The gamma coefficients for all the GARCH models are greater than zero and statistically significant at 5%, indicating that there are substantial leverage effects in each of the daily returns for all financial institutions. For JSE, Standard Bank and Nedbank daily stock returns, negative shocks tend to increase volatility more than positive shocks of the same magnitude. When bad news hits the market and returns are negative, volatility rises strongly. However, for FNB, Absa, Capitec and Investec daily returns, the volatility brought by positive shocks is greater than that of negative shocks of the same magnitude. This is an indication that for the asymmetric models (GJR-GARCH and EGARCH), volatility reacts differently to bad news than to good news. We can confirm that there is no serial correlation in all GARCH models' residuals based on the Weighted Ljung-Box test on standardised residuals. The null hypothesis of no serial correlation is not rejected. The Adjusted Pearson Goodness of fit test suggests that the conditional error terms do not follow a normal distribution for all the models under consideration for each financial data set.

Table 3.7: Estimated parameters of GARCH (1,1) model results

Parameter	JSE	FNB	Standard	ABSA	Capitec	Nedbank	Investec
mu	0.0003 (0.018)	0.0004 (0.002)	0.0003 (0.062)	0.0002 (0.036)	0.0013 (0.000)	0.0001 (0.453)	0.0004 (0.000)
ar1	0.7156 (0.0000)	0.7655 (0.0000)	0.7385 (0.0000)	0.6961 (0.000)	-0.6057 (0.000)	0.7235 (0.000)	-0.1702 (0.000)
ar2		-0.0249 (0.045)					0.8317 (0.000)
ma1	-0.7162 (0.000)	-0.7943 (0.000)	-0.7652 (0.000)	-0.7109 (0.000)	0.5842 (0.000)	-0.7364 (0.00)	0.1555 (0.000)
ma2	-0.0420 (0.00)		-0.0297 (0.005)	-0.0272 (0.002)		-0.0237 (0.107)	-0.8468 (0.000)
omega	0.0000 (0.0642)	-0.1545 (0.000)	0.0000 (0.428)	-0.1983 (0.000)	-0.1128 (0.000)	0.0000 (0.000)	-0.1493 (0.000)
alpha1	0.0464 (0.000)	-0.0500 (0.000)	0.0380 (0.000)	-0.0442 (0.000)	-0.0349 (0.001)	0.0416 (0.000)	-0.0464 (0.000)
beta1	0.9065 (0.000)	0.9808 (0.000)	0.9270 (0.000)	0.9753 (0.000)	0.9848 (0.000)	0.9219 (0.000)	0.9813 (0.000)
gamma1	0.0575 (0.000)	0.1232 (0.000)	0.0380 (0.002)	0.1531 (0.000)	0.1851 (0.000)	0.0337 (0.002)	0.1473 (0.000)
shape	1.3593 (0.000)	1.3687 (0.000)	1.3283 (0.000)	1.2410 (0.000)	2.9627 (0.000)	1.2778 (0.000)	1.2220 (0.000)
skew					1.0303 (0.000)		
LogL	2190.257	2036.557	2082.214	2033.407	1921.193	2025.294	1958.544
AIC	-3.9678	-3.6886	-3.7715	-3.6828	-3.4790	-3.6681	-3.5469
BIC	-3.9405	-3.6613	-3.7442	-3.6556	-3.4517	-3.6408	-3.5196
Shibata	-3.9678	-3.6886	-3.7716	-3.6829	-3.4791	-3.6682	-3.5469
HQIC	-3.9575	-3.6783	-3.7612	-3.6725	-3.4687	-3.6578	-3.5365
Weighted Ljung-Box Test on standardized residuals							
Lag [1]	2.390 (0.122)	1.125 (0.289)	0.3661 (0.545)	0.318 (0.573)	0.0172 (0.896)	2.096 (0.148)	0.704 (0.401)
Lag [2]	2.627 (0.999)	3.212 (0.989)	1.640 (1.000)	1.864 (1.000)	2.693 (0.667)	3.254 (0.987)	2.552 (1.000)
Lag [5]	3.785 (0.983)	5.552 (0.825)	4.026 (0.973)	4.239 (0.962)	5.968 (0.262)	6.238 (0.708)	6.566 (0.942)
Weighted ARCH LM Test							
Lag[3]	2.797 (0.094)	1.848 (0.174)	1.132 (0.288)	0.033 (0.856)	1.031 (0.310)	1.021 (0.312)	0.054 (0.816)
Lag[5]	2.847 (0.313)	1.863 (0.502)	2.034 (0.464)	0.685 (0.828)	1.777 (0.523)	1.786 (0.521)	0.210 (0.963)
Lag[7]	3.600 (0.408)	2.365 (0.640)	4.041 (0.341)	0.843 (0.938)	2.381 (0.637)	4.288 (0.307)	0.479 (0.980)

Table 3.8 shows the persistence and half-life volatility of GARCH models for the daily returns of the various financial institutions. Evidence from persistence shows that all the daily stock returns of each financial institution and JSE can be modelled and predicted using GARCH since all the persistence values are less than one. The ARMA(1,1)-EGARCH-SSTD, the optimal model for Capitec bank daily returns, exhibited the highest persistence of 0.9848, while ARMA(1,2)-EGARCH-GED for Absa daily returns had the least persistence value of 0.9753. Generally, the persistence values range from 0.9753 to 0.9848, indicating high persistence for each of the models. The half-life volatility of the models ranges from 10 to 19 days. The ARMA(1,2)-GJRGARCH(1,1)-GED model for the Standard Bank daily returns has 19 days of mean-reverting, while the ARMA(2,1)-EGARCH(1,1)-GED model for the FNB daily returns has mean-reverting days of 10. The half-life of the models for the daily returns for JSE, FNB, Standard Bank, Absa, Capitec, Nedbank and Investec for mean reverting takes about 14, 10, 19, 10, 13, 19 and 10 days, respectively. This should boost the confidence of stakeholders and suggest that any drop in the price of the stocks can be regained in the future. Overall, we find that the marginal models fit the univariate time series well. Specifically, our results indicate that the EGARCH adequately models the first four conditional moments of the daily stock returns of FNB, Absa, Capitec and Investec. In contrast, the GJRGARCH models the JSE, Standard Bank and Nedbank daily stock returns. Furthermore, our results confirm that the skewed student t assumption for the returns' conditional distribution is adequate to model the excess of kurtosis and skewness present in the data for Capitec Bank. The generalised error distribution is sufficient to model the third and fourth moments of JSE, FNB, Standard Bank, Absa, Nedbank, and Investec Bank returns.

Table 3.8: Persistence and half-life volatility of GARCH models

Institution	Model	Persistence	Half-life volatility
JSE	ARMA(1,2)-GJRGARCH(1,1)-GED	0.9816	14.3618
FNB	ARMA(2,1)-EGARCH(1,1)-GED	0.9808	9.6595
Standard bank	ARMA(1,2)-GJRGARCH(1,1)-GED	0.9840	19.4341
Absa	ARMA(1,2)-EGARCH(1,1)-GED	0.9753	9.7079
Capitec	ARMA(1,1)-EGARCH(1,1)-SSTD	0.9848	13.4662
Nedbank	ARMA(1,2)-GJRGARCH(1,1)-GED	0.9803	18.6390
Investec	ARMA(2,2)-EGARCH(1,1)-GED	0.9813	10.2953

Figures 3.5, 3.6, 3.7, 3.8, 3.9, 3.10, and 3.11 show the model diagnostics of each of the GARCH models utilised in the study. The QQ-plot at the bottom left of each graph indicates that the residuals of each of the GARCH models follow a normal distribution. The residual kernel graph for each of the models also confirms the normality of the residuals. The news impact curves suggest that good news and bad news have different impacts on volatility. The news impact for each of the models is centred at $\varepsilon_{t-1} = 0$. The news impact curve for each of the models has a steeper slope on the negative side than on the positive side. Negative return shocks cause higher volatility than positive return shocks of the same magnitude.

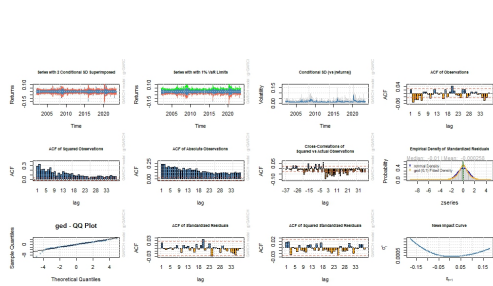


Figure 3.5: GARCH model diagnostics for JSE

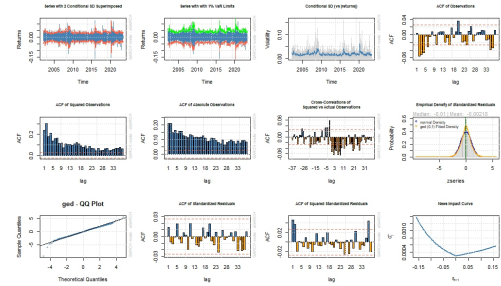


Figure 3.6: GARCH model diagnostics for FNB

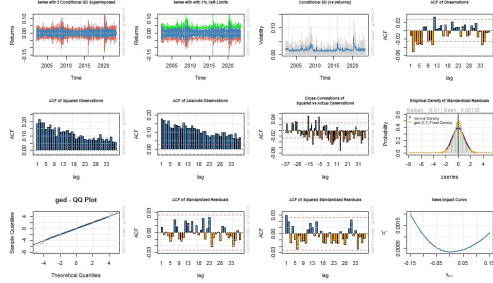


Figure 3.7: GARCH model diagnostics for Standard bank

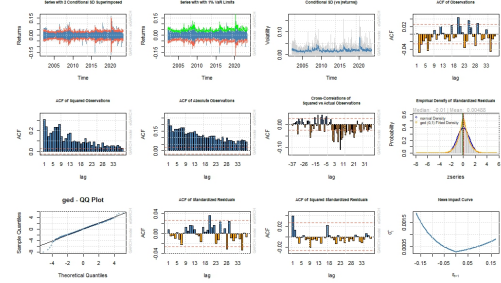


Figure 3.8: GARCH model diagnostics for Absa bank

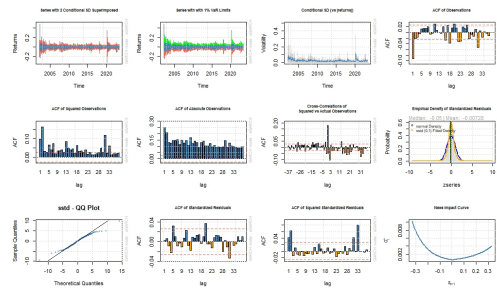


Figure 3.9: GARCH model diagnostics for Capitec bank

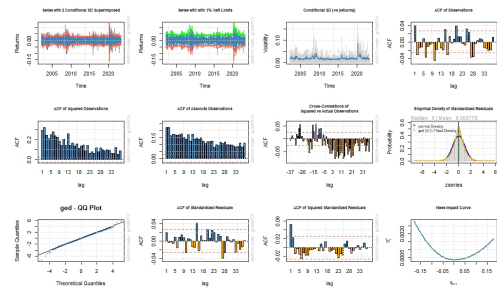


Figure 3.10: GARCH model diagnostics for Nedbank

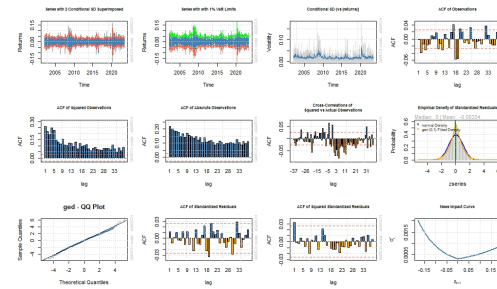


Figure 3.11: GARCH model diagnostics for Investec

Figure 3.12 display the QQ plots for each of the models used in the study. The QQ plots indicate that the data follow the theoretical distributions of each model.

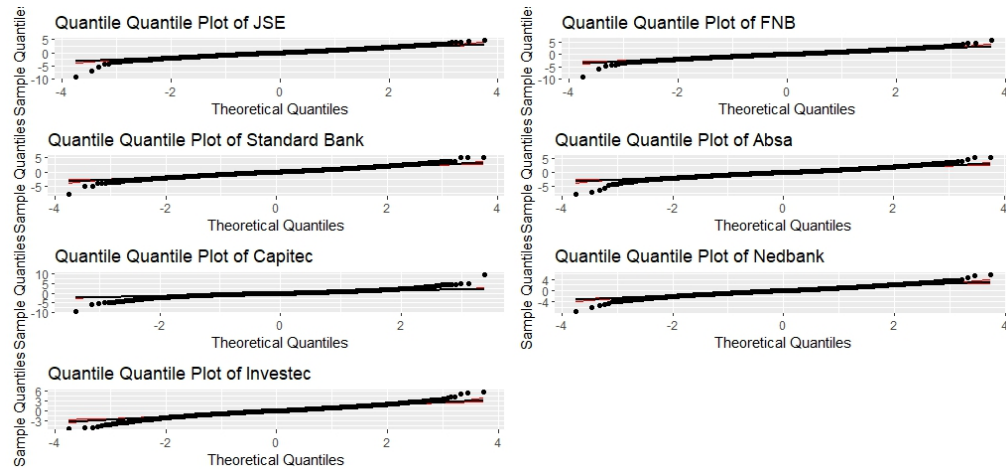


Figure 3.12: Quantile quantile plot of standardised residuals

3.4.2.2 Backtesting of GARCH Models

We backtest our optimal GARCH models for the unconditional and conditional coverage properties under both 1% and 5% quantile levels. We utilise Kupiec's POF-test to statistically test the model's accuracy in estimating the proportion of exceptions (unconditional coverage). Table 3.9 shows the GARCH model's performance for the stock returns for each financial institution and JSE on a sample of 1000 out-of-sample data for one-day-ahead VaR estimation. Table 3.9 shows the expected number of exceptions (violations) and the number of observed violations for each time series and each GARCH model using a 1%, 2.5%, and 5% VaR confidence level. None of the models considerably exceeds the threshold. Each of the models for the different time series shows considerable consistency. The models have the precision to estimate risk and accurately estimate extreme tail losses. The unconditional coverage and independence properties must both be present for a VaR measure to be considered accurate. These types of evaluations that jointly evaluate the qualities give the chance to identify VaR measures that are lacking in one way or another (Campbell, 2005). The joint test of Christoffersen's Interval Forecast was applied to test for both the frequency of exceptions and independence of exceptions. The results of the joint test are presented in Table 3.9. Results of the joint test reject the accuracy of the GARCH model at the 5% confidence level for Absa returns. However, the joint test result confirms the accuracy of the rest of the GARCH models for each stock's returns at all confidence levels.

Table 3.9: Kupiec's POF and Joint test of Christoffersen backtesting results

Institution	Confidence level	Test statistic	Critical value	P-value	Test result
Kupiec					
JSE	1%	0.831	6.635	0.362	Accept
	2.5%	0.625	5.024	0.429	Accept
	5%	2.826	3.841	0.093	Accept
FNB	1%	2.189	6.635	0.139	Accept
	2.5%	1.374	5.024	0.241	Accept
	5%	0.329	3.841	0.566	Accept
Standard	1%	1.437	6.635	0.231	Accept
	2.5%	0.041	5.024	0.841	Accept
	5%	1.284	3.841	0.257	Accept
Absa	1%	1.437	6.635	0.231	Accept
	2.5%	0.356	5.024	0.551	Accept
	5%	5.524	3.841	0.019	Rejected
Capitec	1%	0.105	6.635	0.746	Accept
	2.5%	0.169	5.024	0.681	Accept
	5%	2.388	3.841	0.122	Accept
Nedbank	1%	2.189	6.635	0.139	Accept
	2.5%	0.356	5.024	0.551	Accept
	5%	0.000	3.841	1.000	Accept
Investec	1%	7.827	6.635	0.005	Accept
	2.5%	1.849	5.024	0.174	Accept
	5%	1.984	3.841	0.159	Accept
Joint					
JSE	1%	2.833	9.210	0.243	Accept
	2.5%	0.655	7.378	0.721	Accept
	5%	2.833	5.991	0.243	Accept
FNB	1%	2.647	9.210	0.266	Accept
	2.5%	1.375	7.378	0.503	Accept
	5%	0.735	5.991	0.693	Accept
Standard	1%	3.183	9.210	0.204	Accept
	2.5%	0.183	7.378	0.912	Accept
	5%	2.074	5.991	0.335	Accept
Absa	1%	3.183	9.210	0.204	Accept
	2.5%	1.779	7.378	0.411	Accept
	5%	6.928	5.991	0.031	Reject
Capitec	1%	0.268	9.210	0.875	Accept
	2.5%	1.253	7.378	0.535	Accept
	5%	2.841	5.991	0.242	Accept
Nedbank	1%	7.823	9.210	0.02	Accept
	2.5%	1.779	7.378	0.411	Accept
	5%	0.855	5.991	0.652	Accept
Investec	1%	8.496	9.210	0.014	Accept
	2.5%	1.850	7.378	0.397	Accept
	5%	2.105	5.991	0.349	Accept

Table 3.10: Backtesting results at 1%, 2.5% and 5%

Institution	Alpha	Expected Exceed	Actual VaR Exceed	Actual	MSE	MAE	DAC
JSE	1%	10	13	1.3	0.00045	0.01394	0.50200
	2.5%	25	29	2.9	0.00045	0.01394	0.50200
	5%	50	62	6.2	0.00045	0.01394	0.50200
FNB	1%	10	15	1.5	0.00050	0.01531	0.49800
	2.5%	25	31	3.1	0.00050	0.01531	0.49800
	5%	50	54	5.4	0.00050	0.01531	0.49800
Standard bank	1%	10	14	1.4	0.00052	0.01539	0.49500
	2.5%	25	26	2.6	0.00052	0.01539	0.49500
	5%	50	58	5.8	0.00052	0.01539	0.49300
Absa	1%	10	14	1.4	0.00064	0.01710	0.47600
	2.5%	25	28	2.8	0.00064	0.01710	0.47600
	5%	50	67	6.7	0.00064	0.01710	0.47600
Capitec bank	1%	10	9	0.9	0.00074	0.01539	0.50000
	2.5%	25	23	2.3	0.00074	0.01539	0.50000
	5%	50	61	6.1	0.00074	0.01539	0.50000
Nedbank	1%	10	15	1.5	0.00072	0.01753	0.48600
	2.5%	25	28	2.8	0.00072	0.01753	0.48600
	5%	50	50	5.0	0.00072	0.01753	0.48600
Investec	1%	10	20	2.0	0.00061	0.01688	0.47500
	2.5%	25	32	3.2	0.00061	0.01688	0.47500
	5%	50	60	6.0	0.00061	0.01688	0.47500

Figures 3.13, 3.14, 3.15, 3.16, 3.17, 3.18, and 3.19 depict the backtesting effect of the optimal GARCH models at 99% confidence intervals for the returns of all financial institutions. The figures exhibit the one-day-ahead VaR computed using the last 1000 days and compared to the returns of the same out-of-sample data. The black lines denote the daily forecasted VaR estimates by the GARCH models, while the red points illustrate returns lower than the VaR or the exceedances. The red points in Figure 3.13 illustrate exactly 13 exceedances as reported in Table 3.10. From Figure 3.13, the exceedances are not so far from the GARCH model.

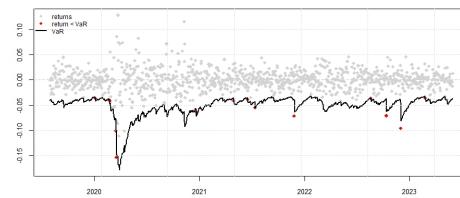
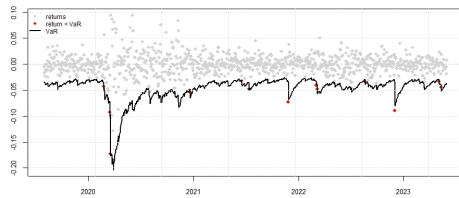


Figure 3.13: VaR backtesting plot for JSE **Figure 3.14:** VaR backtesting plot for FNB

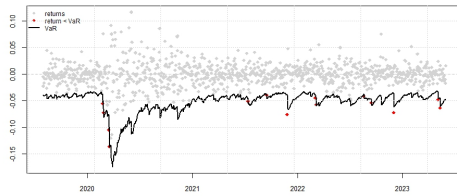


Figure 3.15: VaR backtesting plot for Standard bank

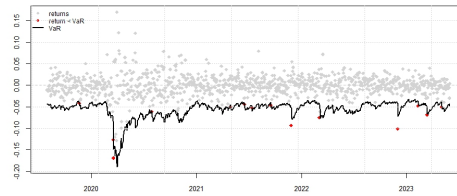


Figure 3.16: VaR backtesting plot for ABSA

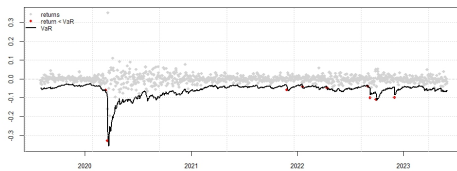


Figure 3.17: VaR backtesting plot for Capitec bank

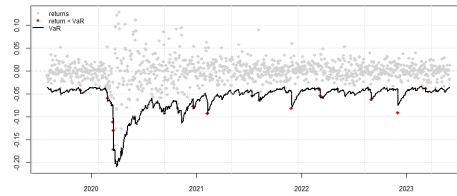


Figure 3.18: VaR backtesting plot for Ned-bank

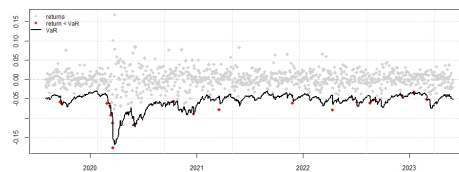


Figure 3.19: VaR backtesting plot for Investec

3.4.3 Systemic Risk Contribution

To measure the marginal contributions of individual institutions to systemic risk in the South African financial system, we compute the ΔCoVaR . The ΔCoVaR estimates the systemic contribution of each bank within the region and makes it possible to rank these banks in order of importance. It is essential to determine which financial institutions are systemically important for practical regulatory work. ΔCoVaR values are negative, but the interpretation is based on absolute values, as we are only interested in the additional VaR imposed on the banking system by bank i when it moves from normality to distress. Thus, the bank with the largest absolute ΔCoVaR contributes most to systemic risk. The negative signs are omitted to follow the conventional way of reporting value-at-risk numbers, and we express the ΔCoVaR values in percentages. Table 3.11 shows the 5% VaR and Table 3.12 shows the 1% VaR for all financial institutions. From Table 3.11, Capitec and Investec show the highest value at risk measures equal to -7.5433% and -5.3421%, respectively, while Absa had the least value at risk. This suggests that Capitec might lose 7.5433% at maximum with a 95% degree of confidence. From Table 3.12, Capitec had the most significant value at risk, followed by Investec. There is a noticeable difference between the VaR of Capitec and the rest of the financial institutions at both 5% and 1% risk levels, showing a higher potential loss in Capitec equity returns compared to other financial institutions used in the study.

Table 3.11 and Table 3.12 display the 5% CoVaR and 1% CoVaR for each of the financial institutions and the corresponding ranking. The result suggests that all the financial institutions exhibited similar systemic risk trends at the 5% risk level. Thus, the 5% CoVaR for all financial institutions is very close, indicating the spillover effect is identical for different financial institutions. From Tables 3.11 and 3.12, it is evident that financial institutions with higher VaR do not have higher values of CoVaR. For instance, Capitec has the highest 5% and 1% VaR, but Nedbank contributes more risk to the financial system. This indicates that individual risk has little to do with the spillover effect. The fact that Nedbank has the highest CoVaR is not sufficient to conclude that Nedbank contributes most to the systemic risk, as CoVaR alone is unable to convey that information. There is a need to compute ΔCoVaR . Table 3.11 reports the ΔCoVaR estimated at 5% significance level, while Table 3.12 provides the ΔCoVaR estimated at 1% and ranks the individual financial institutions according to their systemic risk contribution. These contributions are measured by the average of the institutions' daily ΔCoVaR . The systemic risk contributions of the banks have been ranked along with their corresponding VaR rankings. Large ΔCoVaR values imply high systemic risk contribution and vice versa. The 5% ΔCoVaR values span from -4.1745 for Absa Bank, through -4.3098 for Nedbank. The variation in the numbers indicates the unique contribution by each institution to the riskiness of the financial system. Also, the 1% ΔCoVaR values range from -7.6666 for Absa to -8.5029 for Nedbank. Our results reveal that Nedbank is, on average, the most systemically important bank in South Africa for the chosen data

period with an average ΔCoVaR of -4.3098% and -8.5029% at 5% and 1% risk levels, respectively. Practically, 4.3098 basis points is added to the VaR of the financial system when Nedbank moves from a normal state into a distress state using the 5% risk level. Standard Bank has the second highest contribution to the risk of the South African system with a ΔCoVaR of -4.3098% and -8.5026% at 5% and 1% risk levels respectively. The least contributor to systemic risk is Absa Bank for both risk levels. This indicates that the activities of Absa Bank make the financial system less risky compared to the other selected banks in the study. The results do not suggest any discernible pattern between the ΔCoVaR and an institution's VaR. For instance, Nedbank, the highest systemic risk contributor, has a VaR value ranked as number 3. In contrast, Capitec, with the highest VaR, ranks 4th in terms of systemic risk contribution for both risk levels.

Table 3.11: Systemic Risk Contributions at 5%

Institution	5% VaR	Rank	5% CoVaR	Rank	ΔCoVaR	Rank
FNB	4.8723	4	4.3395	3	4.3073	3
Standard	4.7067	5	4.3416	2	4.3098	2
Absa	4.6884	6	4.2023	5	4.1745	6
Capitec	7.5433	1	4.2711	4	4.2481	4
Nedbank	4.9049	3	4.3417	1	4.3098	1
Investec	5.3421	2	4.2022	6	4.1746	5

Table 3.12: Systemic Risk Contributions at 1%

Institution	1% VaR	Rank	1% CoVaR	Rank	ΔCoVaR	Rank
FNB	9.1965	5	8.5332	3	8.5011	3
Standard	9.2405	4	8.5345	2	8.5026	2
Absa	8.7878	6	7.6944	6	7.6666	6
Capitec	15.7428	1	7.8974	4	7.8744	4
Nedbank	10.1054	3	8.5348	1	8.5029	1
Investec	10.1610	2	7.6951	5	7.6675	5

3.5 Conclusion and policy implication

This study investigates the volatility of the stock returns of FNB, Standard Bank, Absa, Capitec, Nedbank and Investec Bank using three variants of GARCH models, namely sGARCH, GJRGARCH and EGARCH. The research evaluates the risk in the South African stock market using these selected GARCH models and different innovations. To ensure the models are accurate, we backtested the GARCH models using Kupiec's POF and Christoffersen's Independence test. We obtained and analysed the daily stock returns for each selected financial institution from Thomson Reuters. The VaR is computed employing GARCH models and compares losses between the upper and lower VaR thresholds using the confidence interval at 95%, 97.5%, and 99%. We also computed the VaR, CoVaR, and ΔCoVaR using the optimal GARCH models. The study used student t, skewed student t and generalised error innovations with AIC, BIC, SIC and HQIC to select the best model.

It was found that the most optimal model for JSE, FNB Standard Bank, Absa, Capitec, Nedbank and Investec returns are GJRGARCH(1,1), EGARCH(1,1), GJRGARCH-GED, EGARCH-GED, EGARCH-sstd, GJRGARCH-GED, and EGARCH-GED, respectively. All the returns were skewed and leptokurtic, indicating that they are not normal. The study also investigated the effects of different distributional assumptions for the innovation process of the GARCH models for both model fitting and forecasting. Overall, it was found that non-normal distributional assumptions gave better results for model fitting as well as forecasting. This is because non-normal distributions can account for features such as excess kurtosis and asymmetry of the high-frequency financial returns. The asymmetrical GARCH models outperform the symmetrical GARCH model in predicting volatility in all returns. The Jarque-Bera test confirms the departure of the daily returns from normality. All the return series are stationary as indicated from the results of Dickey and Fuller (1979), Phillips and Perron (1988) and Kwiatkowski et al. (1992). All the return series exhibited serial dependence and heteroscedasticity.

The GARCH models are acceptable in risk projection for JSE, FNB, Standard Bank, Absa, Capitec, Nedbank, and Investec at 95%, 97.5%, and 99% confidence intervals, according to the Kupiec Probability of Failure Test and Christoffersen's Independence test, except for Absa at 95%. Thus, the GARCH model for Absa returns was rejected by both the Kupiec test and the Christofferson test at a 95% confidence level. Finally, both the unconditional (Kupiec) and conditional (Christoffersen) coverage tests for the correct number of exceedances for all the return series revealed a non-rejection of the models at a 1%, 2.5% and 5% level of significance except for Absa at 5%, which is similar to results obtained for the percentages of violation rates. This confirms that unconditional (Kupiec) and conditional (Christoffersen) coverage tests for the correct number of exceedances are reliable. The asymmetric GARCH models have been demonstrated

to be better at risk forecasting and capturing the leverage effect. Results for VaR estimates at 95% and 99% indicate that Capitec Bank has the highest potential for loss. In contrast, Absa Bank is the safest index with a loss probability at both 5% and 1% risk levels. Evidence from the VaR analysis of the selected models revealed that the risk of VaR losses was higher at a 99% confidence level than at a 95% confidence level. At the individual financial institution level, the empirical evidence suggests that Capitec Bank is the most risky and should be of concern to investors. The empirical evidence based on the CoVaR results suggests that Nedbank has the highest level of systemic risk at both 5% and 1% risk levels. The results from ΔCoVaR indicate that the most systemically important financial institution in South Africa is Nedbank at both 1% and 5% risk levels. The next risky banks include Standard Bank, FNB, Capitec, Investec and Absa.

The findings of the study offer financial institutions, creditors, investors and other relevant stakeholders a perspective on systemic risk within the South African economy. It requires constant supervision by the central bank to identify the systemically important institutions and formulate policies that will guide and prevent financial crises. In addition, the findings of the study will inform financial institutions on the need to have sufficient capital reserves to minimise the effects of financial crisis and other shocks in other countries or even within the same economy. Also, the study will assist regulators in imposing stricter supervision on systemically important banks.

Chapter 4

Time-Varying Risk Analysis in South African Equity Markets

4.1 Introduction

The recent financial crisis, such as the GFC of 2007-2008, has highlighted the need for predicting extreme events in the financial sector. These financial crises have negative effects on the entire economy, leading to severe downturns. According to McNeil, Frey, and Embrechts (2015), the ability of econometric models to accurately predict extreme events depends on the model's ability to capture non-linear and asymmetric dependencies of the financial returns. Models based on linear relationships are unable to capture the tail behaviour of the joint probability distribution. The financial crisis has demonstrated that there are non-linear dependencies in the financial markets or structural breaks (Wen et al., 2021; Wen et al., 2020). The risks in the financial markets are dynamic and time-varying. GARCH, which has been used in modelling conditional heteroscedasticity and volatility clustering, require distributional assumptions that may not address the non-linear dependence structure or adapt to structural regime changes. Also, there is a greater chance of model misspecification when using GARCH since the marginal and dependence structures are modelled at the same time. The copula approach developed by Reboredo and Ugolini (2015) addressed the limitations of the GARCH model by allowing for separate modelling of the marginal and the dependence structure. The copula approach also captures the extreme tail non-linear interdependence between the financial markets (Ji, Bouri, and Roubaud, 2018). However, the time-varying copula only exploits the first and second order moments of the observations. Given these developments and limitations, the study used the generalised autoregressive score (GAS) model, also known as score-driven models of Creal, Koopman, and Lucas (2013), to estimate time-varying risk analysis for the top six banks in South Africa. The GAS model goes beyond just the first and second moments and captures the full density structure (Creal, Koopman, and Lucas, 2013). It also uses scores to update the time-varying parameters, which enables new information to be

used to enhance the performance of the model. The advantages of the GAS model are well documented in the literature. Makatjane (2024) used the Markov-switching GAS-copula on the Botswana stock market returns and found that GAS models were the best to forecast VaR and ES. Subramoney, Chinhamu, and Chifurira (2021) also demonstrated that combining GAS with heavy-tailed distributions improves VaR estimation for cryptocurrencies which exhibit extreme tail behaviour.

Risk management requires accurate estimation of downside risk measures such as Value at Risk (VaR) and Expected Shortfall (ES). Over the past three decades, tail risk measures, such as Value-at-Risk (VaR) and Expected Shortfall (ES), have been playing a prominent role as the standard risk metrics in global banking and insurance regulatory frameworks, such as Basel III/IV (Banking Supervision, 2016). Value-at-Risk (VaR) and Expected Shortfall (ES), two tail risk indicators, have been widely used as standard risk metrics in global banking, capital requirements, and insurance regulatory frameworks during the past three decades, including Basel III (Banking Supervision, 2016). Since J.P. Morgan originally launched it in 1994, VaR has become a well-known market risk indicator and has been extensively embraced by financial organisations and examined by scholars. VaR has become a widely accepted metric for financial market risk in insurance and financial institutions, acting as a standard risk management instrument for both regulatory and internal risk management. VaR is, by definition, the highest likely loss at a given probability level over a target period. The VaR is the greatest expected loss of a portfolio, as estimated by a predetermined period, a confidence interval, and a given distribution of losses and gains (Jorion, 1997). VaR is defined as a tail quantile of the probability distribution of a financial return. Consequently, VaR is a conditional quantile that is located in the lower tail of the distribution of the return on a portfolio. It entails the determination of the maximum loss that a specific portfolio can sustain at a specified probability level. Following the financial crisis of 2007–2008, there was discussion on the suitability of Value at-Risk as a risk indicator for establishing regulatory capital requirements. VaR, though widely utilised under Basel II and Basel III regulatory frameworks, has been criticised for its inability to capture the magnitude of losses beyond the quantile threshold. It received criticism for failing to account for tail risk, or significant losses brought on by unexpected and extreme financial market events. VaR has been criticised for being utilised as a risk metric, despite being widely employed in the financial sector. However, as Artzner et al. (1999) point out, although being easy to quantify, it has several disadvantages. VaR has some unfavourable characteristics despite its significance, popularity and reputation as the market risk measure. VaR does not offer any information regarding prospective returns that exceed the quantile, despite its ease of application. When risk is assessed, for instance, it fails to meet the sub-additive property, which might have significant repercussions. The measure for a portfolio may be less than the aggregate of the measures of the components of the portfolio, as it does not consider the magnitude of the potential loss beyond the Var and

is not a sub-additive risk measure (Du and Escanciano, 2017; Gordy and Juneja, 2010). Furthermore, if VaR serves as the primary risk metric, the absence of sub-additivity may result in inadequate investment selections. VaR is however not a perfect risk measure. Artzner et al. (1999) established that VaR is not a coherent risk metric since it does not satisfy the axioms of subadditivity and convexity. Because VaR does not take into consideration the return distribution below the predefined confidence level, it is unable to incorporate tail risk. According to Artzner et al. (1999), a risk measure must adhere to four axioms to be deemed coherent: translation invariance, positive homogeneity, subadditivity, and monotonicity. According to Danielsson et al. (2005), financial institutions may be under-hedged if sub-additivity is broken because they fail to assess the true degree of risk.

In response to these concerns, the Expected Shortfall (ES) has been suggested as a risk measure for future regulatory frameworks (Du and Escanciano, 2017; Embrechts et al., 2014). ES is an alternative risk metric that Artzner et al. (1999) offers to address the inadequacies of the VaR. Expected Shortfall addresses this limitation by incorporating the severity of extreme losses, making it a coherent risk measure. Several changes have been implemented to the financial market regulation and banking supervision as a result of the 2007-2008 GFC. One significant change is the Third Basel Accord Banking Supervision (2016), which sets a new emphasis on ES as a risk measure, complementing and in some cases substituting the recognised VaR measure. The ES, a coherent risk measure Acerbi (2002) and Acerbi and Tasche (2002), has attracted significant attention as an alternative risk measure in recent years. The Basel Committee on Banking Supervision now recommends it as a risk measure (Banking Supervision, 2016). The Basel Committee replaced Value-at-Risk with Expected Shortfall as the model for market risk and regulatory capital needs in 2016, releasing new minimum capital rules for market risk. ES is defined as the expected value of potential loss exceeding the VaR level or the average return conditional on the return being below VaR. ES refers to the expected loss given that the loss exceeds VaR. ES is the expected return on an asset if the returns fall below a specific quantile of its distribution, particularly its VaR. The conditional expectation of exceedances beyond the VaR is referred to as ES. According to Artzner et al. (1999), ES has the appealing property of subadditivity, which means that the measure for a portfolio cannot be greater than the sum of the measures for the individual portions of the portfolio, in contrast to the VaR. In comparison to VaR, Artzner et al. (1999) demonstrate that ES satisfies the sub-additive characteristic and is a cohesive risk measure. VaR is not a risk measure that can account for the more extreme events, hence replacing VaR with ES is motivated in part by the fact that ES offers more potent theoretical properties (Artzner et al., 1999) and also takes portfolio tail risk into account.

The definitions of these two risk measures make this inadequacy clear:

$$\begin{aligned} VaR_{t+1,q} &= F_{t+1}^{-1}(q) \quad q \in (0,1) \\ ES_{t+1,q} &= \mathbb{E}[Y_{t+1} | Y_{t+1} \leq VaR_{t+1,q}, \mathcal{F}_t] \\ Y_{t+1} | \mathcal{F}_t &\sim F_{t+1} \end{aligned} \tag{4.1}$$

That is, the q – *quantile* of a particular return, Y_{t+1} , is known as q – *VaR* (VaR at a given confidence level). The average value of the same return series, q – *ES* (ES at a given confidence level), is determined by assuming that the return is lower than its q – *VaR* value. Consequently, in contrast to q – *VaR*, which represents a single point on the return's distribution curve, q – *ES* accounts for every point below the distribution as well as the single point in the distribution (tail risk). The formulation does, however, make it clear that the ES is dependent upon the VaR estimate. Although ES is theoretically appealing as a risk measure, it is not elicitable, which means that there is no scoring function for which the expectation is minimised by the actual ES (Gneiting, 2011). Despite being promoted as a superior risk indicator, ES is susceptible to tails and, in contrast to VaR, may result in higher periodic capital charges (Chang et al., 2019). But more importantly, the lack of elicibility of the ES poses concerns (Fissler and Ziegel, 2016). In addition, the reality that ES is not elicitable implies that the accurate ES forecast is not the exclusive minimizer of the expectation of any loss function. The ES is not elicitable because the level set for ES is not convex (Acerbi and Szekely, 2014). This has created problems for the practical estimation and evaluation of ES (Rockafellar, Royset, and Miranda, 2014; Lazar and Zhang, 2019). This dearth of elicibility poses a challenge for the estimation and backtesting of ES. A risk measure is elicitable if there exists a loss function that enables the risk measure to be the solution to minimising the expected loss. Thus, a risk measure is elicitable if the true forecast of the measure is the unique minimiser of the expectation of at least one scoring function. Fissler and Ziegel (2016) demonstrated that VaR and ES are jointly elicitable, facilitating the estimation of joint models for these two risk measures.

Recently, Fissler and Ziegel (2016) addressed the deficiencies of ES by introducing several joint loss functions for VaR and ES that are jointly elicitable. Fissler and Ziegel (2016) propose a set of scoring functions that the true VaR and ES minimise. These scoring functions enable the joint estimation and evaluation of VaR and ES. The primary justification for combining forecasts is that, if competing forecasts exist, a combination can provide a pragmatic integration of the information that is present in the individual predictions. This framework enables the development of joint models that overcome the limitations of separate estimation and provides more reliable forecasts for risk management. The direct modelling of the VaR and ES has the advantage of circumventing a distributional assumption. This is appealing because there is no consensus on the distribution of daily returns, and the distribution could vary over time.

VaR is elicitable but not coherent in the sense of Artzner et al. (1999). ES which was proposed as a replacement for VaR is a coherent risk measure but not elicitable. Thus, both risk measures suffer practical and coherent risk weaknesses (Fissler and Ziegel, 2016; Nolde and Ziegel, 2017). While VaR models are elicitable, ES lacks this property. However, at the higher level, the joint (VaR, ES) is elicitable. Fissler and Ziegel (2016) proved that Expected Shortfall is elicitable if combined with Value at Risk. It is demonstrated by Fissler and Ziegel (2016) that ES can only be elicited in conjunction with VaR and cannot be done so individually. This indicates that no loss function exists that, when minimised, would produce an accurate estimate of ES. Nevertheless, they demonstrate that VaR and ES are jointly elicitable and offer a loss function family that, when jointly minimised with regard to both risk measures, yields an estimate of ES (and VaR) that is consistent. The claim that ES lacks elicibility and cannot be backtested has been debated (Gneiting, 2011).

The financial market exhibits a dynamic behaviour over time with fluctuating risk levels as a result of policy changes and macroeconomic conditions. While VaR and ES are valuable risk measures, they cannot capture extreme losses and the dynamic behaviour of the financial market over time. In addition, VaR and ES models assume risk is static or changes predictably. However, in reality, risk varies with time and spreads rapidly. Although ES improves on the VaR, it is still not able to accurately predict extreme events. Thus, VaR and ES fail to capture time-varying risk. The use of joint VaR and ES bridges the limitations of both VaR and ES and explores alternative risk management frameworks that address the time-varying limitations of VaR and ES. Despite its theoretical advantages, the application of the Fissler–Ziegel (FZ) joint model to real-world financial markets, particularly in the context of emerging economies such as South Africa, remains underexplored. Given the increasing complexity of the financial markets, volatility in asset prices and regulatory shift towards ES under Basel III, it is important for studies to assess the effectiveness of the joint VaR-ES model and evaluate whether it improves model accuracy and robustness compared to VaR and ES. The study addresses the gap by modelling risk through joint VaR and ES using the GAS model. The study quantifies time-varying tail risks in the SA market in agreement with the Basel III regulatory framework.

The complex and nonlinear nature of the financial market requires an efficient and robust modelling framework that accommodates time-varying volatility and tail dependence. GAS model provides a flexible framework to capture the time-varying, complex and nonlinear characteristics of the financial markets. In this chapter, we examined the predictive power of several GAS models for forecasting joint VaR and ES measures in the South African equity market. The time-varying copula introduced by Patton (2006) to capture VaR only exploits the first and second moments of the observations. The study fills this gap by using the generalised autoregressive score (GAS) provided by

Creal, Koopman, and Lucas (2013) to estimate the joint VaR and ES. This approach captures all the necessary information because it uses the full density structure. The GAS model is suitable for different distributions which enabling the modeling of heavy tails. Also, the study applied the the Fissler and Ziegel joint framework for time-varying risk analysis, with the aim of providing regulators, investors, and financial institutions with a more reliable tool for downside risk measurement and capital adequacy assessment. The study evaluated the influence of different symmetric and asymmetric distributions on the predictive abilities of the GAS model. The study employed different symmetric and asymmetric distributions in the univariate GAS framework to estimate the joint VaR and ES model. This enables the model to capture fat-tails, skewness, time-varying scale and shape parameters. We used the normal, skewed normal, student-t, skewed student-t, Asymmetric t-Student with two decay parameters, Asymmetric t-Student with a left-tail decay parameter and asymmetric Laplace distribution. To the best of our knowledge, this is the first study that uses different innovations for the GAS models in a South African context.

The literature on forecast combination in the VaR and ES context is surprisingly limited for the top six banks in South Africa, although forecast combination strategies have demonstrated beneficial results in various disciplines (Atiya, 2020; Thomson et al., 2019). Although research has been conducted (Junior and Alagidede, 2020; Owusu Junior, Alagidede, and Tiwari, 2021) on the combination of VaR and ES, few studies have examined the risk among the top six South African banks despite their systemic importance to the South African economy and the African economy generally. There has been recent empirical work that has advanced risk measurement in the South African financial sector (Manguzvane and Ngobese, 2023). Accurate measurement of downside risk using joint VaR and ES is critical to bank risk management as well as regulatory assessment. Despite much empirical research, there remains a gap in joint modelling and backtesting of VaR and ES for South Africa's largest banks. There is limited empirical application of elicitable joint VaR-ES frameworks for the top South African banks. There is a paucity of studies that apply joint VaR-ES to individually assess tail risks for the top six South African banks. We are unaware of any studies that have investigated the integration of VaR and ES forecasts for the South African market especially using the top six banks, despite the increasing interest in forecasting ES in recent years.

This study implements a joint-VaR-ES estimation for the top six banks and conducts backtesting of the risk models using joint scoring functions. Addressing this gap in the South African financial sector provides regulators, financial institutions and other stakeholders with risk measures that align with elicitable risk frameworks. To the best of our knowledge, this is the first study investigating the predictive power of combining methods for VaR and ES in an emerging market like the South African economy. In this study, we focused our empirical analysis on the South African stock market as this

economy is a typical emerging market with a rapid economic growth and extensive integration into the global financial markets. The South African financial system, which comprises both local and foreign banks, plays an important role in the regional and international economy. The South African banks are important players in the African economy with significant presence in the continent's financial sector. Given the importance of these financial institutions to the African economy, a spillover effect has far reaching implication beyond the South African borders. This crucial role of the SA financial sector makes it essential to assess the risk levels of the financial institutions and evaluate the stability of the system. The financial industry in SA is very well developed and linked to other economies across the world. Being an emerging market, there is a large degree of global financial integration, making it susceptible to contagion. The SA economy is highly volatile with rising inflation, which has the potential to increase financial risk exposure. The Johannesburg Stock Exchange (JSE) is exposed to other external crisis thereby requiring robust risk management. Due to the exposure and interconnectedness of the South African financial system, it is essential to develop risk models that assess the risk in the system.

The chapter makes a number of contributions. This study proposes the use of Fissler and Ziegel (2016) joint VaR and ES loss functions for estimating joint VaR and ES prediction. Forecasting combinations from an ensemble of heterogeneous risk models has the potential to enhance the forecasting accuracy of VaR and ES in the South African market. Secondly, we explored a wide range of alternative risk models to forecast both VaR and ES with recent data, including periods of the COVID-19 pandemic. Our risk models include the class of observation-driven time series models, which is the generalized autoregressive score (GAS) model of (Creal, Koopman, and Lucas, 2013). Also, the study implemented the joint VaR and ES using the time varying GAS model to capture tail risk behaviour and conditional distributions in bank's returns. The study investigated the performance of each backtesting method to evaluate risk models with different conditional distributions. The findings of the study will assist regulators, financial institutions and investors to manage risk by using models that have been evaluated.

4.2 Generalised Autoregressive Score Models

The time-varying parameters that govern how the returns distribution appears must be precisely and accurately estimated to manage financial risk. citetcox1981statistical separated models driven by observations from models with dynamic parameters. The generalised autoregressive conditional heteroscedasticity (GARCH) of Bollerslev (1986), the autoregressive conditional heteroscedasticity (ARCH) model of Engle (1982), and the dynamic conditional scoring model (DCS) put forth by Creal, Koopman, and Lucas (2012) and Harvey (2013) are examples of observation-driven models whose parameter evolution is dependent on the function of the observed values. In contrast

to parameter-driven time series models (Cox et al., 1981), the Generalised Autoregressive Score (GAS) models, also known as dynamic conditional score models, were first presented by Harvey (2013) and Creal, Koopman, and Lucas (2013) as a class of observation-driven time series models. Regarding volatility and VaR modelling, the GAS models have proven to be a competitive substitute for the GARCH models. GAS emerged as a result of the challenges associated with managing time-varying characteristics found in financial time series data. According to Ardia, Boudt, and Catania (2018), the univariate GAS model is a powerful tool for predicting asset distributions that are fat-tailed, skewed, and exhibiting volatility clustering.

The scaled score of the likelihood function serves as the model's driving mechanism, setting the GAS technique apart from other observation-driven methods. One characteristic that sets GAS models apart is their use of a score (or loss) function as the driver of time-variation in the parameters of non-linear models (Ardia, Boudt, and Catania, 2016). According to Creal, Koopman, and Lucas (2013), the model uses the score of the conditional density as a driving mechanism for time varying over time. The technique by which the parameters are updated over time using the scaled score log-likelihood is the crucial component of those models. This specification enables a better understanding of both past and future pricing behaviour as well as the time-varying volatility of the returns. The GAS specification provides a framework for introducing time-varying parameters in a range of nonlinear models. With the benefit of utilising the entire density of the returns rather than just the first and second order moments, it includes well-known observation-driven models like the GARCH models, autoregressive conditional density models (ACD), which Engle and Russell (1998) introduced, and Poisson models for count data (Creal, Koopman, and Lucas, 2012). GAS models may model all kinds of time series data and offer a driving mechanism for time-varying parameters by employing the score function of the prediction model density at time t rather than merely upper moments and means. According to Laporta, Merlo, and Petrella (2018), the GAS thereby takes advantage of the entire density of the returns as opposed to the first and second order moments. The DCS model's parameters are determined by the standardised score, which guarantees robust estimation and good adaptation to non-normal data (Lucas and Zhang, 2016).

The following specification of the basic GAS is based on Creal, Koopman, and Lucas (2013). Let $N \times 1$ vector y_t represent the dependent variable, f_t represent the time-varying parameter, x_t is a vector of covariates, and θ a vector of static parameters. Let the filtration $\{f_t, F_t\}$ be given as

$$\mathcal{F}_t = \{Y^{t-1}, F^{t-1}, X^t\}, \text{ for } t = 1, \dots, n \quad (4.2)$$

where $Y^t = \{y_1, \dots, y_t\}$, $F^t = \{f_0, f_1, \dots, f_t\}$ and $X^t = \{x_1, \dots, x_t\}$. It is assumed that y_t is generated by the observation density and the parameter vector θ driven by the score of

the conditional distribution given as

$$y_t \sim p(y_t | f_t, \mathcal{F}_t; \theta) \quad (4.3)$$

The time-varying parameter f_t is updated using an autoregressive equation given as

$$f_{t+1} = \omega + \sum_{i=1}^p A_i S_{t-i+1} + \sum_{j=1}^q B_j f_{t-j+1}, \quad i = 1, \dots, p, \quad j = 1, \dots, q \quad (4.4)$$

where ω is a vector of constants, A_i and B_j are matrices of requisite dimensions, and $s_t = s_t(y_t, f_t, \mathcal{F}_t; \theta)$ a function of historical data. It follows that the three unknown coefficients are all functions of θ , (i.e. $\omega = \omega(\theta)$, $A_i = A_i(\theta)$, $B_i = B_i(\theta)$) evaluated at $r_t = \ln P_t - \ln P_{t-1}$, (log-returns in conventional notations). With a realised y_t , f_t is updated to f_{t+1} in 4.4 via

$$s_t = S_t \cdot \nabla_t, \quad \nabla_t = \frac{\partial \ln p(y_t | f_t, \mathcal{F}_t; \theta)}{\partial f_t} \quad S_t = S(t, f_t, \mathcal{F}_t; \theta) \quad (4.5)$$

where $S(\cdot)$ is a matrix. The scaling matrix that adjusts the shape of the score specified is given as

$$S_t = I_{t|t-1}^{-1}, \quad I_{t|t-1} = E_{t-1}[\nabla_t \nabla_t'] \quad (4.6)$$

where E_{t-1} denotes an expectation with respect to $p(y_t | f_t, \mathcal{F}_t; \theta)$. The equations 4.3, 4.4, 4.5 define the GAS model with orders p and q .

4.2.1 Selected distributions

We follow the empirical evidence that financial returns are skewed and heavy-tailed with-varying variances or volatility clustering (Cajueiro and Tabak, 2005; McNeil and Frey, 2000; McNeil, Frey, and Embrechts, 2015) and use distributions that are flexible to incorporate these stylised facts in forecasting the (VaR, ES) for EMEs equities in a univariate GAS framework; skewed-Gaussian (SNORM), student-t distribution (STD), skewed-student-t distribution (SSTD) (Fernández and Steel, 1998); asymmetric student-t with two tail decay parameters (AST), asymmetric student-t with one tail decay parameter (AST1) (Zhu and Galbraith, 2010; Zhu and Galbraith, 2011); and asymmetric Laplace distribution (ALD) (Kotz, Kozubowski, and Podgorski, 2012).

4.2.1.1 STD, SSTD, SNORM

The normal distribution is not a heavy-tailed distribution and, therefore, not robust to extreme outliers. The normal distribution is symmetric with respect to its mean. This implies the mean and median of a normal distribution coincide. However, in skewed (asymmetric) distributions, the mean and the median can be different. The use of normal distribution to model and analyse financial time series data that does not exhibit

symmetry is inappropriate. The skew normal distribution proposed by Azzalini (1985) is an extension of the normal distribution that allows for the presence of skewness. The skew normal distribution is a continuous probability distribution that generalises the normal distribution to allow for non-zero skewness (Benavoli, Azzimonti, and Piga, 2020). It has been used to model skewed data in many applied fields such as finance, medicine, and economics. The shape parameter is used to model the skewness of the data. The probability density function (PDF) of the univariate skew-normal distribution is given as:

$$f(x; \mu, \sigma, \lambda) = \frac{2}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left(\lambda \left(\frac{x - \mu}{\sigma}\right)\right) \quad (4.7)$$

where $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, $\lambda \in \mathbb{R}$ is the shape parameter. Also, ϕ and Φ are the PDF and cumulative distribution function (CDF) of a standard univariate normal distribution, respectively.

The Student-t distribution is commonly used in finance and risk management to model returns for which the tails of the normal distribution are almost invariably found to be too thin. Bollerslev (1987) incorporated asymmetries and thicker tails into conditional variance models through the use of thick-tailed distributions for the standardized innovations. Subsequently, the use of the skewed versions of the Student-t for the conditional distribution has been explored. Hansen (1994) was the first to consider a skewed Student's t distribution to model skewness in conditional distributions of financial returns. Fernández and Steel (1998) introduced a skewness parameter to any univariate probability density function (pdf) which is unimodal and symmetric, resulting in a skewed version of the Student-t equivalent to that of Hansen (1994). Since then, several skew extensions of the Student-t distribution have been proposed for financial applications (Fernández and Steel, 1998; Theodossiou, 1998; Bauwens and Laurent, 2005; Jones and Faddy, 2003; Azzalini and Capitanio, 2003).

The PDF of the Student-t distribution is given as

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\beta\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(x - \alpha)^2}{\beta\nu}\right)^{-\left(\frac{\nu+1}{2}\right)} \quad (4.8)$$

where α , β and ν are the location, scale and shape parameters, respectively, Γ is the Gamma function.

To account for the stylized facts, we considered the STD, SSTD and SNORM distributions which is distributed on past observations as follows

$$r_t | \mathcal{F}_t \sim SSTD(r_t, \mu, \sigma_t, \zeta, \nu) \quad (4.9)$$

where $\mathcal{F}_t$ is the autoregressive conditional distribution of r_t , $\mu \in \mathbb{R}$ denotes the location parameter, $\sigma_t > 0$ denotes time-varying scale (volatility), ζ and $\nu > 2$ denote the skewness and shape parameters, respectively. Following Bauwens and Laurent (2005), Ardia, Boudt, and Catania (2016) reparametrise 4.9 such that $E_{t-1}[r_t] = \mu$ and $Var_{t-1}[r_t] = \sigma_t^2$. In special cases when $\zeta = 1$, the SSTD becomes STD and SNORM when $\zeta = \infty$.

4.2.1.2 AST and AST1

The skewed versions of the Student-t distributions use two parameters which control skewness and thickness of the two tails together. Thus, in general, the skewness parameter mainly controls the asymmetry of the central part of a distribution. Therefore, a class of generalised asymmetric Student-t (AST) distributions with one skewness parameter and two tail parameters offers the potential to improve the ability to fit and forecast data in the tail regions. Zhu and Galbraith (2010) propose the generalised asymmetric Student-t (AST) distribution in which two parameters control the thickness of the two tails and a third allows skewness to change independently of the tail parameters. The AST distribution has skewness and two-tailed parameters that check the left (right) tail behaviour of returns. The probability density function of AST is given as

$$AST(r_t; \theta) = \begin{cases} \frac{1}{\sigma} \left[1 + \frac{1}{\vartheta_1} \left(\frac{r_t - \mu}{2\alpha\sigma K(\vartheta_1)} \right)^2 \right]^{-(\vartheta_1+1)/2} & r_t \leq \mu, \\ \frac{1}{\sigma} \left[1 + \frac{1}{\vartheta_2} \left(\frac{r_t - \mu}{2(1-\alpha)\sigma K(\vartheta_2)} \right)^2 \right]^{-(\vartheta_2+1)/2} & r_t > \mu \end{cases} \quad (4.10)$$

where $\theta = (\alpha, \vartheta_1, \vartheta_2, \mu, \sigma)^T$ with skewness, left tail, right tail, location and scale parameters, respectively. Further $K(\vartheta) = (\Gamma(\vartheta + 1)/2) / [\sqrt{\pi}\vartheta\Gamma(\vartheta/2)]$, for $\{\vartheta_1 + \vartheta_2\} \in \vartheta$ where $\Gamma(\cdot)$ is the Gamma function.

When one of the tail parameters goes to infinity (say $\vartheta_2 \rightarrow \infty$), the AST behaves as a Student-t on the left side and as a Gaussian on the right side, implying one heavy tail and one exponential tail. With these two tail parameters, the AST can accommodate empirical distributions of daily returns of financial assets that are often skewed and have one heavy tail and one relatively thin tail. From the AST probability density function, the effect of the shape parameter on the distribution is observed.

4.2.1.3 Asymmetric Laplace distributions ALD

According to Poon and Granger (2003), there is empirical evidence indicating that daily returns are leptokurtic or fat-tailed, with a minor negative skew, and cannot be sufficiently described by mean and variance alone. The classical Laplace distribution is one of the most widely utilised symmetric distributions for modelling leptokurtic data. A family of distributions known as the asymmetric Laplace distributions was proposed because the distribution above is inappropriate for simulating data having an asymmetric empirical distribution. Kotz, Kozubowski, and Podgorski (2012) developed the asymmetric Laplace distribution (ALD), which illustrates flexibility in the face of high tails and skewness inherent in financial data. The Asymmetric Laplace distributions (ALD) were initially used to model financial data by Hinkley and Revankar (1977). As suggested by Lu, Huang, and Gerlach (2010), who used frequentist methods and an exponential smoothing procedure, the advantage of the ALD is that, in addition to the flexible excess kurtosis and skewness properties, dynamics in these higher moments can be allowed via a time-varying shape parameter. ALD is a skewed family of distributions which is the most suitable skewed generalization of the classical Laplace law (Kotz, Kozubowski, and Podgorski, 2012). The most suitable skewed simplification of classical Laplace law, which robustly measures fat-tail, skewness, and leptokurtic characteristics of financial returns, is ALD (Kotz, Kozubowski, and Podgorski, 2012; Yu and Zhang, 2005) with a generalised probability density function specified as

$$f(r_t; \theta, \sigma, \kappa) = \frac{\kappa(1-\kappa)}{\sigma} \exp\left(-\frac{(r_t - \theta)}{\sigma} [\kappa - I(r_t \leq \theta)]\right) \quad (4.11)$$

where $0 < \kappa < 1$, $\sigma > 0$, $-\infty < \theta < \infty$ are skewness, scale, and location parameters, respectively and $I(\cdot)$ is the indication function. We employ Kotz, Kozubowski, and Podgorski (2012), Blasques, Koopman, et al. (2014) maximum likelihood estimator jointly with one-step ahead forecasts to estimate the parameters. ALD is skewed to left when $\kappa > \frac{1}{2}$, and skewed to right when $\kappa < \frac{1}{2}$.

4.2.2 Backtesting methods

Value-at-Risk has become one of the most popular risk measurement techniques in finance and among practitioners. The ability of models to predict VaRs accurately is of utmost importance since VaR models are helpful only if they predict future risks accurately. Despite the wide use and common acceptance of VaR as a risk management tool, the method has frequently been criticised for being incapable of producing reliable risk estimates. After estimating VaR, we need to evaluate the quality of the VaR forecasts for the different models. To verify that the results acquired from VaR calculations are consistent and reliable, the models should always be backtested with appropriate statistical methods. Actual profits and losses of a portfolio are compared

to the corresponding VaR estimates. When a VaR model is estimated, it is essential to check its reliability and accuracy. The statistical procedure for examining the appropriate estimation of VaR is called backtesting. Backtesting is a statistical procedure where actual profits and losses are systematically compared to corresponding VaR estimates. Backtesting aims to estimate whether the amount of losses predicted by VaR is correct. The backtesting technique relies on quantitative tests which scrutinise the model's performance in terms of accuracy and efficiency concerning a defined criterion. The performance of the VaR model is measured by applying several different tests, including the dynamic quantile (DQ) test, the actual over expected (AE) exceedance ratio, the conditional coverage (CC) test, and the unconditional coverage (UC) test. The appropriate estimation of VaR, referred to as backtesting, implements both unconditional and conditional coverage tests for the correct number of exceedances. The unconditional coverage test checks if the frequency exceptions, during the selected time interval, are synchronised with the chosen confidence interval. The most commonly used test in this group is the Kupiec test. The conditional coverage test examines if the exceedances over VaR levels are serially independent, and the most widely known test for this group is the Christoffersen independence test. Since VaR is still used for regulatory backtesting and for setting the capital multiplier, we first consider individual backtests for VaR. The predictive ability of each model was evaluated using the Actual over Expected (AE) exceedance ratio, the Unconditional Coverage (UC) test of Kupiec et al. (1995), the Conditional Coverage (CC) test of Christoffersen (1998) and the Dynamic Quantile (DQ) test of Engle and Manganelli (2004), which can be interpreted as an overall goodness-of-fit test for the estimated VaR. The AE ratio tracks the actual number of times that the returns have exceeded the estimated VaR over the expected VaR violations. The closer the ratio gets to one, the better the model estimates VaR. Basel III changes the regulatory market risk measure from VaR at the 99% confidence level to ES at the 97.5% confidence level. When Basel 3 was introduced, though, a major criticism against using ES for regulation was the lack of available backtesting due to the risk measure not being elicitable (Gneiting, 2011). However, Fissler, Ziegel, and Gneiting (2015) showed that ES is jointly elicitable with VaR. Academics have developed a recent literature on joint VaR and ES backtesting.

4.2.2.1 Unconditional coverage

The Kupiec's test, which was developed by Kupiec et al. (1995), is a VaR test that is based on failure rates. It is also referred to as the proportion of failures (POF) test or unconditional coverage test. It is a backtest which is based on the number of VaR exceptions, i.e. the number of holding periods when portfolio losses exceed VaR forecasts. More precisely, it tests the null hypothesis that the number of exceptions, which is assumed to follow a binomial distribution, is consistent with the confidence level. It tests the difference between the observed and the expected number of VaR violations. The

POF test aims to determine whether the actual failure rate differs significantly from the failure rate suggested by the confidence level.

The null hypothesis for the test is given as

$$\begin{aligned} H_0 : \pi &= \hat{\pi} = \frac{T_1}{T} \\ H_1 : \pi &\neq \hat{\pi} = \frac{T_1}{T} \end{aligned} \quad (4.12)$$

where T is the number of observations and x is the number of exceptions.

Kupiec et al. (1995) proposed a likelihood ratio (LR) test of unconditional coverage (LR_{UC}) to evaluate the accuracy of the model. Since the VaR is based on a confidence level α , when we observe actual exceedances (T_1) out of T observations (with $T = T_0 + T_1$ and T_0 the number of no VaR violations), we observe an empirical frequency of $\hat{\pi} = T_1/T$ proportion of excessive losses. The Kupiec's test checks whether $\hat{\pi}$ is statistically significantly different from α . It is possible to show that the test of the null hypothesis $H_0 : \hat{\pi} = \alpha$ is given by the following likelihood ratio test statistic. The test examines whether the failure rate equals the expected value. The LR test statistic is given by

$$LR = -2\ln \left[\frac{(1-\alpha)^{T-T_1} \alpha^{T_1}}{\left[1 - \frac{T_1}{T}\right]^{T-T_1} \left(\frac{T_1}{T}\right)^{T_1}} \right] \sim \chi_1^2 \quad (4.13)$$

If the VaR model is correctly specified, the failure rate ($\hat{\pi} = T_1/T$) should be equal to the prespecified VaR level (α). The null hypothesis $H_0 : \pi = \alpha$ is evaluated through the likelihood ratio test. Violation means that if $VaR_t > r_t$, violation occurs; the opposite case indicates the non-violation. Under the null hypothesis ($H_0 : \pi = \hat{\pi}$), the LR statistic follows a chi-square distribution with one degree of freedom. Thus, the model is rejected if the violation number is too high or too low. If the value of LR is smaller than the critical value, the null hypothesis is accepted, indicating that the estimation of VaR is reliable. The drawback of this test is that it may fail to reject a model having a violation clustering.

4.2.2.2 Conditional coverage

The LR_{UC} test fails to detect if violations are not randomly distributed. The Christoffersen test extends the Kupiec likelihood ratio test and accounts for serial independence violations. Thus, the test indicates clustering of exceedances. Christoffersen (1998) proposed a LR test of conditional coverage LR_{cc} to remove the lack of Kupiec et al. (1995) test. The LR_{cc} test investigates both equality of failure rate and expected one, and also

independently distributed violations. Christoffersen (1998) developed a likelihood ratio statistic to test the joint assumption of unconditional coverage and independence of VaR exceedances. The main advantage of this test is that it can reject a VaR model that forecasts either too many or too few clustered exceedances. In contrast, its main limit is the need for (at least) several hundred observations to be accurate.

Christoffersen's independence test depicts whether an outcome of an estimate (violation or not) is dependent or independent of the previous estimate. The test first categorises the outcomes via an indicator, $I_t = (0, 1)$, where 0 refers to no violation and 1 refers to a violation at time t . There are four possible categories, $I_{t,t+1} = (00, 01, 10, 11)$ and represented by n_{ij} as shown in Table 4.1.

Table 4.1: Christoffersen's Markov Chain representation

	$I_{t-1} = 0$	$I_{t-1} = 1$	Total
$I_t = 0$	n_{00}	n_{10}	$n_{00} + n_{10}$
$I_t = 1$	n_{01}	n_{11}	$n_{01} + n_{11}$
Total	$n_{00} + n_{01}$	$n_{10} + n_{11}$	N

The violation probabilities π_i for $i = 0, 1$ are expressed as $\pi_0 = \frac{n_{01}}{n_{00} + n_{01}}$, $\pi_1 = \frac{n_{11}}{n_{10} + n_{11}}$, $\pi = \frac{n_{11} + n_{01}}{n_{00} + n_{01} + n_{10} + n_{11}}$. The n_{ij} is the number of days when condition j occurred assuming that condition i occurred on the previous day. $n(00)$ is no VaR violation at time I and on $I - 1$ day, $n(01)$ is no VaR violation at time I but there is VaR violation on $I - 1$ day, $n(10)$ is VaR violation on time I but no VaR violation at time $I - 1$ day, $n(11)$ is VaR violation at time I followed at other VaR violation at time $I - 1$ day.

The log-likelihood ratio becomes

$$LR_{IND} = -2 \ln \left[\ln \frac{(1 - \pi)^{n_{00} + n_{10}} \pi^{n_{01} + n_{11}}}{(1 - \pi_0)^{n_{00}} \pi_0^{n_{01}} (1 - \pi_1)^{n_{10}} \pi_1^{n_{11}}} \right] \quad (4.14)$$

The LR_{IND} function follows χ_1^2 distribution. This test only tests the independence of two consecutive days. Therefore, it ignores the violations above or below two days. To address the problem of the independence test, a composite independence test is proposed. The composite or joint test is calculated by combining Christoffersen's independence test with Kupiec's unconditional coverage test. The simplified format for the composite or joint test statistic is defined as

$$LR_{CC} = LR_{UC} + LR_{ind}$$

The LR_{cc} statistic follows χ_2^2 distribution with two degrees of freedom. The composite test LR_{cc} can detect a violation from POF and IND tests combined. It comes with a decreased ability to detect a violation of only one of the two tests.

The LR_{cc} test statistic under the null hypothesis shows that the failures are independent and equal to the expected one. It is given by

$$LR_{CC} = -2\ln\left[\frac{(1-\alpha)^{n_0}\alpha^{n_1}}{(1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}}\right] \quad (4.15)$$

where n_{ij} is the number of observations with value i followed by j for $i, j = 0, 1$ and $\pi_{ij} = n_{ij} / \sum_j n_{ij}$ are the probabilities for $i, j = 1$. It denotes that the violation occurred, otherwise indicates the opposite case. The LR_{cc} statistic follows a chi-square distribution with two degrees of freedom.

4.2.2.3 Dynamic Quantile Test

The dynamic quantile test proposed by Engle and Manganelli (2004) overcomes some drawbacks of the conditional coverage test of Christoffersen (1998) using a linear regression model that links current violations to past violations. The DQ test of Engle and Manganelli (2004) is based on a linear regression where the dependent variable is the sequence of violations and the independent variables are past violations or any other variables in the information set. The test examines whether the violations are uncorrelated with any variable that belongs to the information set when the VaR is calculated. The dynamic quantile test proposed by Engle and Manganelli (2004) overcomes some drawbacks of the conditional coverage test of Christoffersen (1998) using a linear regression model that links current violations to past violations. The DQ test of Engle and Manganelli (2004) is based on a linear regression where the dependent variable is the sequence of violations and the independent variables are past violations or any other variables in the information set. The test examines whether the violations are uncorrelated with any variable that belongs to the information set when the VaR is calculated.

We define the auxiliary variable $Hit_t(\alpha) = I_t(\alpha) - \alpha$, so that $Hit_t(\alpha) = I_t(\alpha) - \alpha$ if $r_t < VaR_{t|t-1}(\alpha)$ and $Hit_t(\alpha) = -\alpha$ otherwise, where $I_t(\alpha)$ is equal to 1 if $r_t < VaR_{t|t-1}(\alpha)$ and equal to 0, otherwise. The null hypothesis of this test is that the sequence of hits (Hit_t) is uncorrelated with any variable that belongs to the information set Ω_{t-1} available when the VaR was calculated and it has a mean value of zero, which implies, in particular, that the hits are not autocorrelated. The dynamic quantile test is a Wald test of the null hypothesis that all slopes in the regression model

$$Hit_t(\alpha) = \delta_0 + \sum_{i=1}^p \delta_i Hit_{t-i} + \sum_{j=1}^q \delta_{p+j} X_j + \epsilon_t \quad (4.16)$$

are zero, where X_j are explanatory variables contained in Ω_{t-1} . The test statistic has an asymptotic χ^2_{p+q+1} distribution.

4.3 Empirical Results

We used six different distributional assumptions (norm, snorm, std, sstd, ast, ast1, ald) of the univariate Generalised Autoregressive Score (GAS) model to make one-step ahead (VaR and ES) rolling forecasts using the out-of-sample data at the 1% and 2.5% levels of significance. We fixed the length of the out-of-sample period to 1110 and re-estimated the model parameters at the weekly frequency (thus every five new observations).

4.3.1 Univariate GAS FZL functions

Given that seven models are used, and not all seven models can be used as the optimal models of choice for forecasting, we conducted further tests to determine the predictive ability of each model. We applied the multivariate version of Diebold and Mariano (2002) and Mariano and Preve (2012) (MDP) to test the null hypothesis of equal predictive ability (EPA) of the seven competing forecasting models used in this study. The MDP test of EPA result is shown in Table 4.2. Although all seven models are robust and adequate tail risk measures, results suggest their predictive capabilities differ. The results of the MDP test reject the null hypothesis of EPA for all banks and JSE at both 1% and 2.5% levels of significance. This result confirms the robustness of the MDP test in discriminating between different forecast models. Since EPA does not hold for all models, it becomes necessary to rank the models in order of predictive ability.

Table 4.2: Multivariate Diebold-Mariano test of equal predictive accuracy

Financial Institution	α	W	P-value
JSE	1%	44.668	0.000
	2.5%	71.005	0.000
FNB	1%	40.752	0.000
	2.5%	81.316	0.000
Standard	1%	49.424	0.000
	2.5%	82.725	0.000
Absa	1%	66.404	0.000
	2.5%	89.213	0.000
Capitec	1%	52.823	0.000
	2.5%	98.687	0.000
Nedbank	1%	44.063	0.000
	2.5%	63.869	0.000
Investec	1%	61.84	0.000
	2.5%	94.152	0.000

4.3.2 MAE and RMSE ranking of univariate GAS (VaR, ES) model forecasts per distributional innovation

Following the results of the MDP analysis, it is evident that the seven distributions do not have equal EPA. It is, therefore, essential to rank the seven distributions in order of predictive ability since the FZL has a consistent scoring function that permits comparative tests on competing models. Following the work of Blazsek and Hernández (2018), we computed the MAE and RMSE for each of the seven distributions across the different banks at 1% and 2.5% risk thresholds to rank the models. The result of the MAE and RMSE for both 1% and 2.5% for all banks and JSE is displayed in Table 4.3. The model with the least MAE and RMSE is the most optimal. It is observed from Table 4.3 that the optimal forecast model across all financial institutions for both 1% and 2.5% risk levels is the model with the Ald distribution as it shows the least MAE and RMSE. The result is consistent with a study by Taylor (2019), which found the ALD the most optimal to jointly (VaR, ES) estimate and forecast GAS models.

4.3.3 Backtesting and model ranking of joint (VaR, ES) models

Tables B.1, B.2, B.3, and B.4 report 1-day joint (VaR, ES) backtesting results for the seven distributions considered in the study at the 1% and 2.5% levels of significance for all banks and JSE. The best-performing model for the joint VaR and ES at the 1% risk level based on the maximum absolute deviations (AD), UC, CC, DQ, QL for the JSE stock returns is GAS-Ald with the least maximum absolute deviation at 62.67723. In contrast, the best-performing model based on the average absolute deviations is GAS-Ast and GAS-Ast1 with the least absolute deviations of 21.8807 each. The worst-performing model, with the largest maximum absolute deviations and highest average absolute deviations at 321.2407 and 58.0783, respectively, is GAS-snorm. Although the result might seem to offer support for the GAS-Ast or GAS-Ast1 with the least average absolute deviation, it is important to note that these models lead to higher values for the maximum absolute deviation, MAE, RMSE and quantile loss values in comparison to GAS-Ald. At the 2.5% risk level, GAS-norm and GAS-ald are the best models. The GAS-norm model had the best UC and CC values, and slightly overestimated the model (AE=1.01). The GAS-ald model has the least QL, DQ and ADmax values but slightly underestimates the model (AE=0.43). Using a multiple of results, the GAS-ald model is preferred since it gives the least QL, MAE and RMSE values. Using similar criteria, the best-performing models for FNB, Standard Bank, Absa, Capitec, Nedbank, and Investec for both 1% and 2.5% are provided in Table 4.4.

For each backtesting result across all banks and the JSE, the model with the Ald distribution has the least loss value for 1% and 2.5% risk levels. This result is consistent with the models selected by the MAE and RMSE for each financial institution. We

evaluated the market risk models using different statistical backtesting techniques, including the Unconditional test of Kupiec et al. (1995) (UC), the Conditional Coverage test of Christoffersen (1998) (CC), and the Dynamic Quantile Test of Engle and Manganelli (2004) (DQ). In addition, we performed the joint VaR and ES model comparison using the Quantile Loss (QL) of each model. Each test has a null hypothesis of a correctly specified forecast model for the joint VaR and ES at the 1% and 2.5% levels. The UC, CC, and DQ tests present test statistics with associated p-values, while the QL presents a single value. The most optimal model has the least QL. We backtested the FZL forecast of the optimal distributional innovation using the UC, CC, DQ, and QL at the selected risk levels. All the backtesting models except DQ show evidence that the ALD using the 1% and 2.5% levels are correctly specified. CC, UC, and DQ test statistics, which show p-values close to unity, are preferred. The QL test statistics, which provide smaller Quantile losses, are selected. Table 2 shows the results of the backtesting process. As seen in Table 4.4, the distributions can robustly estimate the FZL forecast for financial stock returns as they are not rejected by the UC and CC tests as inadequate and incorrectly specified at the different levels. However, the DQ test did not accept them as correctly specified. The result of the DQ test confirms the study of Braione and Scholtes (2016), which claims that the DQ is robust and generally accounts for a temporal interdependence among a series of violations. The quantile loss ratio at 1% to the quantile loss at 2.5% compares the performance of the joint VaR and ES model forecasts in line with Basel III. If the quantile loss ratio is less than one, the model at 1% is deemed to have outperformed the model at 2.5% and vice versa (Ardia, Boudt, and Catania, 2016). It is evident from the results that 1% FZL models perform better than the 2.5% FZL models, varying between 6.45% and 25.60% for all financial institutions except Capitec (-6.17%). For example, the 1% FZL model for JSE outperforms the 2.5% FZL by 9.61% in terms of average QL. Thus, the 1% FZL model has a lower average QL than the 2.5% for all the financial institutions except Capitec.

4.3.4 Characteristic (VaR, ES) estimate for 1% univariate GAS models

To compare the capital requirement, tail risk, and riskiness of the financial institutions, we use the characteristic FZL (CFZL) values (single values of FZL) estimated from the unconditional location parameters from the OOS FZL forecast of the best-performing models. The characteristic FZL (CFZL) values (single values of FZL) were obtained from the unconditional location parameters from the OOS FZL forecast of the ALD, to compare the risk exposure and capital required for loss absorption of the financial institutions. For the seven financial institutions, the characteristic (VaR, ES) forecast values are estimated from the unconditional location parameters from the respective distribution for 1% and 2.5% out-of-sample (VaR, ES) forecasts. As tail risk estimates (and jointly sharing features with VaR and ES) and as loss functions, smaller values

are preferred to larger ones. As displayed in Table 4.5 in the Rank column, the (VaR, ES) values are ranked according to their magnitudes in an ascending order. The least negative forecast value is ranked 1, and the most negative is ranked 7. Table 4.5 exhibits the maximum losses expected for each financial institution for the forecasted period. The results show that the joint VaR and ES estimates are all around 3%.

To put it in perspective, the (VaR, ES) forecast values in Table 4.5 indicate the riskiness, tail risk, and capital requirement inherent in the respective equities of the financial institutions. Besides the mathematical and theoretical strength of the (VaR, ES) gained by drawing information from two tail measures (VaR and ES), it is intuitively necessary for empirical risk modelling because it can provide a better risk measure than either separately. This same statement was put forward by Patton, Ziegel, and Chen (2019). Table 4.5 shows that the capital required to absorb losses at the 1% risk threshold is least for Absa, followed by Investec, Nedbank and largest for JSE. This implies that Absa has the least risk profile among the financial institutions, while JSE is the riskiest financial institution at the $\alpha = 0.01$ level. At a 2.5% risk threshold, Absa once again was ranked first and requires the least capital to absorb losses, followed by Investec, and the largest for JSE. It is important to note that FNB was ranked fourth as the least risky institution at the 1% risk level but was ranked third as the least risky institution at the 2.5% risk level. Nedbank, which was ranked third at the 1% risk level, was ranked fourth at the 2.5% risk level. Results show that Absa, Investec, Standard Bank, and Capitec maintained their first, second, fifth, and sixth positions respectively, at both 1% and 2.5% risk thresholds, while Nedbank and FNB alternate third and fourth positions. Our results suggest that at the 1% and 2.5% risk levels, Absa is the safest financial institution, followed by Investec. Following the results of the study, investors may adopt the rankings to enhance their trading and investment strategies for optimal utility while minimising the risk to avoid potential loss. Banks, central banks and regulators may equally adopt the rankings in the study to assist with their risk assessment and financial stability oversight responsibility to deliver economic growth from the financial markets of the banks. This could also assist financial supervisors in designing policies that will help monitor underlying risks of each financial institution to either prevent bank collapses or mitigate the risks. The thesis confirms the study of Patton, Ziegel, and Chen (2019) which contends that FZL values drawn from both VaR and ES provide a better empirical risk modelling than either VaR or ES separately. This approach offers coherent risk measure estimates which banks, central banks and regulators can use for effective risk management.

Table 4.3: MAE and RMSE ranking of univariate GAS(VaR, ES) model forecasts per distributional innovations

Institution	Innovation	$\alpha=1\%$			$\alpha=2.5\%$		
		MAE	RMSE	Rank	MAE	RMSE	Rank
JSE	norm	3.6549	7.056	4	3.5921	4.6638	3
	snorm	4.7855	15.5155	7	4.1894	8.1677	7
	std	3.6013	6.2805	3	3.5977	4.4779	4
	sstd	3.5087	5.572	2	3.5289	4.2273	2
	ast	3.7022	7.319	5	3.6818	4.9606	5
	ast1	3.7022	7.319	5	3.6818	4.9606	5
	ald	3.1629	3.9876	1	3.3549	3.7411	1
FNB	norm	3.5476	6.4786	2	3.5125	4.4519	3
	snorm	4.0473	10.7511	5	3.7364	6.0955	5
	std	3.6361	7.3210	4	3.5858	4.8505	4
	sstd	3.5783	7.1934	3	3.5117	4.7657	2
	ast	4.0677	14.1662	6	3.7440	7.6178	6
	ast1	4.0677	14.1662	6	3.7440	7.6178	6
	ald	3.0730	4.3441	1	3.2635	3.7927	1
Standard	norm	3.4965	5.9497	4	3.4840	4.2643	4
	snorm	4.2727	11.1673	7	3.8964	6.3426	5
	std	3.4576	5.6409	3	3.4672	4.1992	2
	sstd	3.4569	5.0610	2	3.4751	4.0722	3
	ast	3.6953	7.5971	5	3.5829	5.0165	6
	ast1	3.6953	7.5971	5	3.5829	5.0165	6
	ald	3.0085	3.5024	1	3.2200	3.4877	1
Absa	norm	3.6214	7.3122	4	3.5264	4.7464	4
	snorm	4.6855	14.7600	5	4.0677	7.8317	7
	std	3.5410	6.9174	2	3.4931	4.6533	2
	sstd	3.5711	6.6641	3	3.5098	4.6207	3
	ast	4.0407	10.3283	6	3.7260	6.1266	5
	ast1	4.0407	10.3283	6	3.7260	6.1266	5
	ald	2.9928	3.9827	1	3.2044	3.6457	1
Capitec	norm	3.8156	11.0998	3	3.6289	6.1978	4
	snorm	3.9064	15.8985	5	3.5570	8.0760	2
	std	3.8304	9.5830	4	3.7038	5.7602	5
	sstd	3.6761	8.3146	2	3.5890	5.2398	3
	ast	4.6091	11.4133	6	4.1381	6.7945	6
	ast1	4.6091	11.4133	6	4.1381	6.7945	6
	ald	3.0930	6.0985	1	3.2585	4.3861	1
Nedbank	norm	3.5271	6.5791	3	3.4776	4.4648	2
	snorm	5.0508	16.5842	7	4.2365	8.6408	7
	std	3.5216	6.5219	2	3.5221	4.5421	3
	sstd	3.5822	6.8833	4	3.5229	4.6856	4
	ast	3.7508	8.3876	5	3.5757	5.2534	5
	ast1	3.7508	8.3876	5	3.5757	5.2534	5
	ald	3.0654	4.3323	1	3.2695	3.8135	1
Investec	norm	3.5813	6.2866	4	3.4977	4.4128	4
	snorm	4.2083	11.2582	7	3.8152	6.3459	7
	std	3.7093	7.4671	5	3.5893	4.9433	5
	sstd	3.7501	8.0238	6	3.5952	5.1629	6
	ast	3.4078	8.1828	2	3.2864	4.9230	2
	ast1	3.4078	8.1828	2	3.2864	4.9230	2
	ald	3.0344	4.4006	1	3.2368	3.8413	1

Table 4.4: Backtesting results of selected univariate GAS (VaR, ES) models

Institution	Distribution	Alpha	UC	CC	DQ	QL	QL ratio
JSE	ald	1%	0.43 (0.51)	0.58 (0.75)	589.04 (0.000)	0.2106	0.9039
	ald	2.5%	11.61 (0.00)	11.87 (0.00)	358.73 (0.00)	0.2330	
FNB	ald	1%	0.97 (0.33)	1.08 (0.58)	526.55 (0.00)	0.2422	0.9355
	ald	2.5%	4.95 (0.03)	5.47 (0.06)	408.73 (0.00)	0.2589	
Standard	ald	1%	1.76 (0.18)	1.85 (0.40)	561.97 (0.00)	0.1636	0.7440
	ald	2.5%	1.84 (0.16)	2.65 (0.27)	525.33 (0.00)	0.2199	
Absa	ald	1%	0.43 (0.51)	0.58 (0.75)	551.31 (0.00)	0.2152	0.8084
	ald	2.5%	0.54 (0.46)	0.91 (0.63)	557.66 (0.00)	0.2662	
Capitec	ald	1%	0.97 (0.33)	1.08 (0.58)	322.79 (0.00)	0.3010	1.0617
	ald	2.5%	8.52 (0.00)	8.87 (0.01)	250.09 (0.00)	0.2835	
Nedbank	ald	1%	0.43 (0.51)	0.58 (0.75)	580.75 (0.00)	0.2514	0.8459
	ald	2.5%	0.54 (0.46)	0.91 (0.63)	550.50 (0.00)	0.2972	
Investec	ald	1%	0.31 (0.58)	0.62 (0.73)	750.54 (0.00)	0.2913	0.8742
	ald	2.5%	0.06 (0.81)	0.13 (0.94)	692.77 (0.00)	0.3332	

Table 4.5: Characteristic FZL forecast values for selected distributional innovations

1%				2.5%			
Financial institution	Innovation	CFZL	Rank	Financial institution	Innovation	CFZL	Rank
ABSA	ald	-3.1098	1	ABSA	ald	-3.3819	1
Investec	ald	-3.1540	2	Investec	ald	-3.4114	2
Nedbank	ald	-3.1733	3	FNB	ald	-3.4311	3
FNB	ald	-3.1786	4	Nedbank	ald	-3.4317	4
Standard bank	ald	-3.1908	5	Standard bank	ald	-3.4493	5
Capitec	ald	-3.2293	6	Capitec	ald	-3.4650	6
JSE	ald	-3.3439	7	JSE	ald	-3.5942	7

Table 4.6: Estimated information criteria and Log-Likelihood

Model	GAS-norm	GAS-snorm	GAS-std	GAS-sstd	GAS-ast	GAS-astl	GAS-ald
JSE							
AIC	-24794.38	-24117.73	-24854.45	-24460.06	-23738.45	-23740.45	-24831.85
BIC	-24768.79	-24085.73	-24822.46	-24421.67	-23693.66	-23702.06	-24799.86
np	4	5	5	6	7	6	5
LogL	12401.19	12063.86	12432.23	12236.03	11876.23	11876.23	12420.93
FNB							
AIC	-23024.78	-22532.17	-23095.67	-22772.14	-21985.18	-21987.18	-23087.6
BIC	-22999.18	-22500.18	-23063.68	-22733.75	-21940.39	-21948.79	-23055.6
np	4	5	5	6	7	6	5
LogL	11516.39	11271.09	11552.84	11392.07	10999.59	10999.59	11548.8
Standard							
AIC	-23504.42	-22970.10	-23511.96	-23148.14	-22342.84	-22344.84	-23539.31
BIC	-23478.83	-22938.11	-23479.97	-23109.75	-22298.04	-22306.45	-23507.31
np	4	5	5	6	7	6	5
LogL	11756.21	11490.05	11760.98	11580.07	11178.42	11178.42	11774.65
Absa							
AIC	-23355.44	-22886.03	-23457.25	-23072.01	-22078.98	-22080.98	-23638.83
BIC	-23329.85	-22854.04	-23425.26	-23033.61	-22034.18	-22042.59	-23606.84
np	4	5	5	6	7	6	5
LogL	11681.72	11448.02	11733.63	11542.00	11046.49	11046.49	11824.41
Capitec							
AIC	-21654.65	-20272.85	-22123.75	-21656.85	-20411.05	-20413.06	-22932.98
BIC	-21629.05	-20240.86	-22091.76	-21618.46	-20366.26	-20374.66	-22900.98
np	4	5	5	6	7	6	5
LogL	10831.32	10141.43	11066.88	10834.43	10212.52	10212.53	11471.49

Table 4.7: Estimated information criteria and Log-Likelihood

Model	GAS-norm	GAS-snorm	GAS-std	GAS-sstd	GAS-ast	GAS-astl	GAS-ald
Nedbank							
AIC	-23356.10	-22971.64	-23505.70	-23148.97	-22085.80	-22087.80	-23485.09
BIC	-23330.51	-22939.65	-23473.71	-23110.58	-22041.01	-22049.41	-23453.10
np	4	5	5	6	7	6	5
LogL	11682.05	11490.82	11757.85	11580.49	11049.90	11049.90	11747.55
Investec							
AIC	-22948.25	-22154.52	-22986.48	-22610.45	-21906.30	-21908.36	-23150.87
BIC	-22922.66	-22122.53	-22954.48	-22572.05	-21861.51	-21869.96	-23118.88
np	4	5	5	6	7	6	5
LogL	11478.13	11082.26	11498.24	11311.22	10960.15	10960.18	11580.44

Table 4.8: Backtest results of univariate GAS (VAR, ES) models

Institution	Distribution	UC	CC	AE	DQ	QL
JSE	norm	1.92 (0.17)	2.39 (0.30)	1.44	696.68 (0.000)	0.5006
	snorm	22.18 (0.00)	37.80 (0.00)	2.70	1351.94 (0.00)	1.5862
	std	3.65 (0.06)	4.24 (0.12)	1.62	773.88 (0.00)	0.4553
	sstd	4.68 (0.03)	5.34 (0.07)	1.71	857.31 (0.00)	0.4106
	ast	28.55 (0.00)	30.57 (0.00)	2.97	1187 (0.00)	0.6745
	ast1	28.55 (0.00)	30.57 (0.00)	2.97	1187 (0.00)	0.6745
	ald	0.43 (0.51)	0.58 (0.75)	0.81	589.04 (0.00)	0.2106
FNB	norm	2.72 (0.10)	3.25 (0.20)	1.53	865.59 (0.00)	0.5072
	snorm	8.41 (0.00)	11.61 (0.00)	1.98	1101.99 (0.00)	1.0123
	std	4.68 (0.03)	5.34 (0.07)	1.71	883.39 (0.00)	0.5902
	sstd	3.65 (0.06)	4.24 (0.12)	1.62	878.08 (0.00)	0.5796
	ast	20.19 (0.00)	24.33 (0.00)	2.61	831.57 (0.00)	1.1587
	ast1	20.19 (0.00)	24.33 (0.00)	2.61	831.57 (0.00)	1.1587
	ald	0.97 (0.33)	1.08 (0.58)	0.72	526.55 (0.00)	0.2422
Standard	norm	3.65 (0.06)	4.75 (0.09)	1.62	907.48 (0.00)	0.4608
	snorm	22.18 (0.00)	25.99 (0.00)	2.70	1529.58 (0.00)	1.2143
	std	2.72 (0.10)	3.25 (0.20)	1.53	898.60 (0.00)	0.4276
	sstd	7.07 (0.01)	7.88 (0.02)	1.89	1446.40 (0.00)	0.4689
	ast	22.18 (0.00)	23.85 (0.00)	2.70	1475.37 (0.00)	0.7865
	ast1	22.18 (0.00)	23.85 (0.00)	2.70	1475.37 (0.00)	0.7865
	ald	1.76 (0.18)	1.85 (0.40)	0.63	561.97 (0.00)	0.1636

4.4 Conclusion

This chapter utilises the time-varying joint (VaR, ES) risk model to assess the daily returns of selected financial institutions in the SA equity market. The joint (VaR, ES) risk model provides a measure of tail risk that facilitates comparative backtesting of competing models through ranking using a scoring function. Models from financial institutions are comparable to the standardised capital requirement approach by regulatory bodies using the elicibility feature of joint (VaR, ES) models and the FZL scoring function. According to Nolde and Ziegel (2017), the FZL function provides grounds for comparing different models when assessing new forecasting models with existing ones. We applied six different asymmetrical distributions (STD, SNORM, SSTD, AST, AST1, and ALD) to the univariate GAS framework to capture both heavy-tailed and volatility clustering that characterise financial returns. The forecasting abilities of the different distributions were tested using the MDM test to check the equal predictive accuracy (EPA) of the models. The results indicate that although all the models were adequate for tail risk forecasting, their predictive abilities were different. The models were ranked using the MAE and RMSE to select the optimal model and backtest the FZL forecasts.

The results suggest that the GAS model with heavy-tailed distributions is appropriate for modelling financial returns. However, the best model across all financial institutions for both the 1% and 2.5% risk levels is the model with the Ald distribution. We compared the 1% models to 2.5% models by computing QL ratios as required by Basel III. The QL ratios indicate that, apart from Capitec, the 1% level risk models for the joint (VaR, ES) performed better than those of 2.5% for all returns of all financial institutions. This suggests that the specified confidence level influences the joint (VaR, ES) model. We considered a 1-day joint (VaR, ES) backtesting for the distributions at 1% and 2.5% levels of significance utilising UC, CC, DQ and QL tests. The best-performing model for the joint VaR and ES at 1% and 2.5% risk levels for the returns of all financial institutions is Gas-Ald. Results suggest that Absa was the least risky financial institution and required the least capital to absorb its losses at the 1% and 2.5% risk thresholds. At the same time, the JSE remains the most risky financial institution. The study provides evidence to support Patton, Ziegel, and Chen (2019) that the FZL values drawn from both VaR and ES provide a better empirical risk modelling than either VaR or ES separately. The application of the FZL function for joint VaR and ES to financial returns provides coherent risk measure estimation for enhanced investment and trading strategies for risk management portfolio optimisation.

Table 4.9: Backtest results of univariate GAS (VAR, ES) models

Institution	Distribution	UC	CC	AE	DQ	QL
Absa	norm	9.84 (0.00)	10.30 (0.01)	2.07	1054 (0.00)	0.6395
	snorm	26.36 (0.00)	36.24 (0.00)	2.88	1491.71 (0.00)	1.6374
	std	9.84 (0.00)	10.30 (0.01)	2.07	1054 (0.00)	0.5855
	sstd	14.66 (0.00)	15.91 (0.00)	2.34	298.55 (0.00)	0.6332
	ast	33.11 (0.00)	33.12 (0.00)	3.15	1631.03 (0.00)	1.1688
	ast1	33.11 (0.00)	33.12 (0.00)	3.15	1631.03 (0.00)	1.1688
	ald	0.43 (0.51)	0.58 (0.75)	0.81	551.31 (0.00)	0.2152
Capitec	norm	24.24 (0.00)	97.72 (0.00)	2.79	1632.84 (0.00)	0.8037
	snorm	1.25 (0.26)	7.26 (0.03)	1.35	548.52 (0.00)	1.0771
	std	3.65 (0.06)	4.24 (0.12)	1.62	718.72 (0.00)	0.7540
	sstd	7.07 (0.01)	7.74 (0.02)	1.89	878.93 (0.00)	0.6889
	ast	42.93 (0.00)	44.52 (0.00)	3.51	2294.45 (0.00)	1.5135
	ast1	42.93 (0.00)	44.52 (0.00)	3.51	2294.45 (0.00)	1.5135
	ald	0.97 (0.33)	1.08 (0.58)	0.72	322.79 (0.00)	0.3010
Nedbank	norm	7.07 (0.01)	10.58 (0.01)	1.89	938.76 (0.00)	0.5355
	snorm	34.48 (0.00)	46.88 (0.00)	3.24	1741.69 (0.00)	2.0050
	std	8.41 (0.00)	11.61 (0.00)	1.98	962.97 (0.00)	0.5436
	sstd	9.84 (0.00)	10.82 (0.00)	2.07	1131.83 (0.00)	0.6315
	ast	14.66 (0.00)	14.88 (0.00)	2.34	1286.57 (0.00)	0.8652
	ast1	14.66 (0.00)	14.88 (0.00)	2.34	1286.57 (0.00)	0.8652
	ald	0.43 (0.51)	0.58 (0.75)	0.81	580.75 (0.00)	0.2514
Investec	norm	11.37 (0.00)	11.73 (0.00)	2.16	1528.02 (0.00)	0.6516
	snorm	28.55 (0.00)	29.43 (0.00)	2.97	1539.94 (0.00)	1.2447
	std	16.43 (0.00)	16.59 (0.00)	2.43	1446.00 (0.00)	0.7721
	sstd	18.28 (0.00)	19.73 (0.00)	2.52	1538.65 (0.00)	0.8665
	ast	8.41 (0.00)	8.97 (0.00)	1.98	908.12 (0.00)	0.6861
	ast1	8.41 (0.00)	8.97 (0.00)	1.98	908.12 (0.00)	0.6861
	ald	0.31 (0.58)	0.62 (0.73)	1.17	750.54 (0.00)	0.2913

Chapter 5

Elicitability and risk model comparison of financial institutions

5.1 Introduction

The Global financial crisis of 2007-2009 highlighted the role of tail-risk interdependencies among institutions in determining the stability of the financial system. Being able to accurately quantify the systemic impact of individual institutions and assess their vulnerability to extreme market movements is key to mitigating risk spillovers and maintaining the financial stability of individual market players and the market itself. A recent comprehensive survey of the literature on systemic risk, including measures of risk and regulatory issues, is provided by (Benoit et al., 2017). Systemic risk within the banking system has received considerable attention from academics and regulators in the aftermath of the pandemic. The concepts of systemic risk and financial contagion hold significant importance within the domain of global economics. Systemic risk occurs when the distress of one financial institution spreads to others, potentially causing a domino effect that could lead to broader economic impacts (Bisias et al., 2012). Systemic risk is a term used to describe the possibility of a disturbance occurring within the financial system, which has the potential to propagate throughout various institutions, markets, and economies, resulting in substantial adverse outcomes. The potential ramification of systemic risk emphasize the need for proactive risk management measures. The field of risk management has garnered significant attention over an extended period, with heightened interest following the rapid onset of the 2008 financial crisis. After the global financial crisis in 2008-2009, the concept of systemic risk has gained significant attention from regulators and researchers in the field (Geng, Liu, and Zhao, 2024).

Since financial markets have become increasingly integrated and complex, these markets have undergone several unprecedented crises such as the Global Financial crisis, and later, the financial crisis induced by the Covid-19 pandemic. Hence, risk management techniques have become increasingly important to enable portfolio managers to anticipate and manage such crises. The Value at Risk (VaR) model is one of the most widely adopted risk management tools in the banking industry since the Basel Accord III of the Bank for International Settlements indicated using VaR for risk assessment. The VaR refers to the maximum expected loss of a portfolio given a distribution of losses and gains, a predefined period, and a given confidence interval (Jorion, 1997). When measuring VaR, it is essential to choose a volatility model and a conditional distribution that best fits the returns. Studies have shown that the generalized autoregressive conditional heteroskedasticity (GARCH) family models can effectively describe the volatility clustering and heteroskedasticity phenomenon. These models include the GARCH model by Bollerslev (1986). An extension of the GARCH model, specifically an asymmetric GARCH model, has been developed to describe the leverage effect of volatility. These include the threshold ARCH (TARCH) model (Zakoian, 1994), the exponential generalized autoregressive conditional heteroskedasticity (EGARCH) model (Nelson, 1991), the Glosten Jagannathan Runkle GARCH (GJRARCH) model (Glosten, Jagannathan, and Runkle, 1993), the asymmetric power autoregressive conditional heteroskedasticity (APARCH) model (Ding, Granger, and Engle, 1993), and the component generalised autoregressive conditional heteroskedasticity (CGARCH) model (Engle and Lee, 1993). Since financial time series data exhibit thick tails and skewness, the normal distribution cannot describe these stylised facts. Therefore, other distributions such as Student's t , and generalized error distributions, skewed generalized error distribution (GED) of Fernández and Steel (1998) have been used to model the thickness of tails. As the accuracy of VaR measurement determines the accurate assessment of the tail risk, it is critical to evaluate the quality of each model (Escanciano and Pei, 2012). Given the many models available for risk analysis, it becomes tedious to identify a method that produces the best model using a particular criterion. To reduce the uncertainty when the comparative methods fail to identify the best model, an enhanced model selection technique becomes necessary. This issue is further exacerbated when there are multiple competing models. It is not realistic to anticipate that a single model can outperform other competing models, either because the various specifications are statistically similar or because the data does not provide sufficient information to distinguish the models (Hansen and Lunde, 2005; Hansen, Lunde, and Nason, 2011).

Several studies have used the AIC, BIC, HQIC and SIC for risk model ranking and selection. Other studies used the root mean squared error (RMSE), mean absolute error (MAE) and log-likelihood (Blazsek and Hernández, 2018; Troster et al., 2019). However, these techniques do not rank every model simultaneously as a set. Secondly,

there is no uniform scoring function in these methods to support model ranking. Also, these methods do not provide information about the statistically significant differences among the models. The forecasted VaR are usually backtested using different techniques such as the actual over expected exceedance ratio, the mean and the maximum absolute deviations of violating returns of McAleer and da Veiga (2008), the unconditional coverage test of Kupiec et al. (1995), the conditional and unconditional coverage (UC) test of Christoffersen (1998) and the dynamic quantile (DQ) test of Engle and Manganelli (2004). Also, Hansen (2005) proposed the SPA test, which can be used to compare the VaR forecasting accuracy of different models. However, the SPA test requires the selection of a benchmark model. To obtain a better understanding of each model's performance, the model confidence set (MCS) approach proposed by Hansen, Lunde, and Nason (2011) overcomes the shortcomings of the SPA test. It is usually used to identify the optimal set of models. Also, despite the relevance of these backtesting measures, they fail to discriminate alternative models based on the predictive accuracy of VaR. To overcome the drawbacks of the backtesting methods mentioned above, we implement the model confidence set (MCS) method developed by Hansen, Lunde, and Nason (2011). The MCS technique helps to choose a model that best fits the data from a list of possible candidate models. Model confidence set (MCS) test (Hansen, Lunde, and Nason, 2011) is utilized to test whether there is a significant difference between different models in terms of bringing accurate forecasts.

In summary, Basel III's goal of reducing regulatory arbitrage is essentially supported by tail risk modelling using the (VaR, ES) and FZL. This feature makes it possible to use a consistent scoring method to rank comparable risk models. The study seeks to examine risk selection and ranking of equities of sampled financial institutions in SA using the (VaR, ES) and FZL function. In an attempt to rank the models, we used the MCS procedure proposed by Hansen, Lunde, and Nason (2011) to rank and select Superior Set Models (SSMs) of the (VaR, ES) model forecasts. The null hypothesis of equal predictive ability (EPA) (Diebold and Mariano, 2002) is used to arrive at the SSM. We use the 97.5% and 99% confidence levels in selecting SSMs in the MCS procedure. The volatile and complex nature of the financial market requires robust methods that can compare and select optimal risk forecasting models, especially models that estimate downside risk measures such as VaR, ES and joint VaR and ES. The MCS procedure of Hansen, Lunde, and Nason (2011) provides a solution by identifying a set of superior models from a candidate number of models. This approach affords banks, supervisors and other stakeholders the confidence to trust a set of models rather than commit to a single model.

This chapter makes important contributions to the literature on tail risk analysis in the South African market. Firstly, there have been few studies on the risk measurement of the South African market. This chapter constructs GARCH models based on the normal, Student's t , generalised error, skewed normal, skewed Student's t , skewed

generalised error distributions, and uses the rolling window approach to predict the out-of-sample VaR of the South African stock returns. Secondly, to the best of our knowledge, this is the first study to model tail risks in the South African market equities using the joint (VaR, ES) model based on FZL functions conforming to the current Basel III regulatory framework. The use of risk modelling and selection techniques that are consistent with Basel regulatory frameworks has the potential to bolster confidence in the investor community. Thirdly, unlike other authors who used the LR_{UC} , LR_{CC} , DQ, and SPA tests to compare the VaR forecasting performance of different models, this chapter uses the MCS approach to compare and analyze the VaR forecast accuracy. Finally, this chapter comprehensively investigates the impact of different model re-estimation strategies, different quantile levels, and different forecasting periods on the VaR forecasting performance of GARCH models. It selects a relatively optimal risk measurement model for investors and regulatory authorities to manage risks.

5.2 Model Confidence Set Procedure

The model confidence set (MCS) technique evaluates and analyses the prediction abilities of various models. The null hypothesis that all chosen models have the same predictive capacity for user-defined loss functions/measurement error is tested by this sequential testing process, which is based on the hypothesis of equal predictive ability (EPA). Through the use of the MCS technique, empiricists can create a set of models known as the superior set of models (SSM), in which the EPA null hypothesis is accepted with a given level of confidence. Using an elimination rule that is consistent with the statistical test, the worst-performing model is eliminated at each stage until the hypothesis of EPA is not rejected for all models included in the SSM.

The MCS procedure considers the quantile as a loss function. The VaR estimate of the model i at level τ is represented by $\widehat{VaR}_{i,t}^\tau$, where the loss function related to this model is defined as

$$\ell_{i,t} = \ell(r_t, \widehat{VaR}_{i,t}^\tau) \quad (5.1)$$

The asymmetric loss function deals with several large negative returns that surpass the VaR estimate that is represented as:

$$\ell(r_t, \widehat{VaR}_{i,t}^\tau) = \begin{cases} (\tau - 1)(r_t - \widehat{VaR}_{i,t}^\tau) & \text{if } r_t < \widehat{VaR}_{i,t}^\tau, \\ \tau(r_t - \widehat{VaR}_{i,t}^\tau) & \text{if } r_t \geq \widehat{VaR}_{i,t}^\tau \end{cases} \quad (5.2)$$

The average loss differences and loss differences were taken into account when developing the EPA test. The loss differences between models i and j are represented by $d_{ij,\tau}$

while $d_{i,t}$ represents the average (loss) differences between model i and other competing models that exist in the initial M set of models, with respect to time t :

$$d_{ij,t} = \ell_{i,t} - \ell_{j,t} \text{ for all } i, j \in M \quad (5.3)$$

$$d_{i.,t} = \frac{1}{m-1} \sum_{j \in M} d_{ij,t} \text{ for all } i, j \in M \quad (5.4)$$

Intuitively, i is preferred as alternative to j when $d_{ij,t} < 0$. The null and alternative hypotheses for the test statistics (EPA) are described as follows

$$H_0 : E[d_{i.}] = 0 \text{ for all } i \in M \quad (5.5)$$

$$H_0 : E[d_{i.}] \neq 0 \text{ for some } i \in M \quad (5.6)$$

The t-statistic is given as

$$t_{i.} = \frac{\bar{d}_i}{\sqrt{\widehat{VaR}(\bar{d}_{i.})}} \quad (5.7)$$

where $\bar{d}_{i.} = \frac{1}{m-1} \sum_{j \in M} d_{ij}$

5.3 Evaluation of VaR prediction performance using GARCH

This section investigates the impact of different forecasting periods, different quantile levels and different model re-estimation strategies on the out-of-sample VaR prediction performance of GARCH models for each of the returns of the financial institutions. Following the practice of Hansen, Lunde, and Nason (2011), we used 0.2 as the significance level of the MCS approach. We divided the data into estimation and forecasting samples and used the rolling window approach to produce a one-day-ahead VaR forecast. We estimated standard GARCH, EGARCH, and GJR-GARCH using the normal, Student's t, skewed Student's t, and generalised error distributions.

5.3.1 Evaluation of VaR prediction performance under different quantiles

We set the quantile levels to 1%, 2.5%, and 5%, and re-estimated the model every 30 days. We used 1000 observations to calculate the out-of-forecast VaR. The MCS results and the final composition of the SSM under different quantiles are shown in Tables 5.1,

5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10 and 5.11 for each bank and JSE for each bank and JSE. At the 1% and 2.5% quantile levels, the GJRGARCH, standard GARCH, and EGARCH models survived with different distributions for the JSE returns. The VaR forecasting ability of the GARCH, EGARCH and GJRGARCH models is excellent, as all of them survived. However, at the 5% quantile level, all the models except GARCH-std survived.

Generally, as the quantile level increases, only one model was excluded, indicating that the gap in forecasting ability between the GARCH models with different distributions for the JSE returns is not too noticeable. The number of surviving models is largest under the normal, skewed student distribution and generalised error distribution, regardless of the quantile level. At each quantile level, the loss values of each model based on different distributions are different. For each of the models, the loss value of GARCH, EGARCH and GJRGARCH with the generalised error distribution shows better risk forecasting ability than other distributions for the JSE returns. The loss value of GJRGARCH is the smallest based on the generalised error distribution, and the VaR forecasting accuracy is higher than for other distributions and models. For the JSE stock returns, only the GJRGARCH-ged model has p-values equalling 1.0000 for both test statistics and has the least loss value under different quantile levels. For FNB stock returns, all the models with different distributions survived at the 1%, 2.5%, and 5% quantile levels. This indicates that the GARCH, EGARCH and GJRGARCH models have good VaR forecasting abilities. At 1% and 5% quantile levels, the GJRGARCH-ged model has p-values equalling 1.0000 for both test statistics and the least loss value. However, at the 2.5% quantile level, EGARCH-ged and GJRGARCH-ged both had a p-value of 1.0000 for both test statistics, indicating these two models are the best competing models. Between these two competing models, GJRGARCH-ged has the least loss value and therefore shows a relatively optimal VaR forecasting ability compared to EGARCH-ged. All the models with different distributions were retained for quantile levels 1% and 2.5% for Standard bank stock returns. However, at 5%, the GARCH model has the least number of survivals, as two of them were eliminated. The EGARCH and GJRGARCH had the highest number of survivals at the 5% quantile level. For all quantile levels, the GJRGARCH-ged model has the least loss value and p-values of 1.0000 for both test statistics.

All models for Absa stock returns at the different quantile levels and different distributions survived. At the 1% quantile level, EGARCH-ged has a p-value of 1.0000 for both test statistics and the least loss value. Even though EGARCH-ged and GJRGARCH-ged have a p-value of 1.0000 for both test statistics, EGARCH-ged has the least loss value and has the most optimal VaR forecasting ability than GJRGARCH-ged. At the 5% quantile level, GJRGARCH-ged is the most optimal model. At the different quantile levels for Capitec stock returns, all the models with different distributions survived. At

1%, EGARCH-norm and EGARCH-sstd have a p-value of 1.0000 for both test statistics. However, EGARCH-sstd has the lowest loss value. The only optimal model at the 2.5% quantile level is the EGARCH-sstd model with the least loss value and p-value of 1.0000 for both test statistics. The most optimal model for the 5% quantile level is GJRGARCH-std. All the models survived for each of the quantile levels for the Nedbank stock returns. The quantile levels have no significant impact on the VaR forecasting accuracy for the different models. However, only GJRGARCH-ged has the least loss value and is considered the most optimised model for 1%, 2.5%, and 5% quantile levels for the returns of Nedbank. At the 1% quantile levels, two models (sGARCH-norm, EGARCH-norm) were eliminated for the Investec stock returns. However, all models survived at the 2.5% and 5% quantile levels. For the 1% quantile level, the GJRGARCH-ged is the most optimal model with the least loss value. For both 2.5% and 5% quantile levels, EGARCH-ged is the most optimal model. It is evident from Table 5.12 that the quantile level has no impact on the VaR forecasting accuracy of the type of GARCH models generally for all banks and JSE.

Table 5.1: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
JSE (0.01)							
GARCH-norm	11	0.7273	0.8716	12	1.0833	0.8646	7.3525×10^{-6}
GARCH-std	7	-0.2562	1.0000	7	0.7750	0.9656	7.2007×10^{-6}
GARCH-sstd	6	-0.3283	1.0000	3	0.5093	0.9964	7.1925×10^{-6}
GARCH-ged	3	-0.6469	1.0000	2	0.3821	0.9998	7.1633×10^{-6}
EGARCH-norm	12	1.0084	0.7100	11	1.0730	0.8684	7.4139×10^{-6}
EGARCH-std	9	0.5577	0.9408	10	1.0301	0.8864	7.2662×10^{-6}
EGARCH-sstd	8	0.3128	0.9922	5	0.7034	0.9798	7.2459×10^{-6}
EGARCH-ged	5	-0.3904	1.0000	4	0.6493	0.9846	7.1943×10^{-6}
GJRGARCH-norm	10	0.6369	0.9126	9	0.9015	0.9342	7.2765×10^{-6}
GJRGARCH-std	4	-0.6135	1.0000	8	0.8271	0.9558	7.1544×10^{-6}
GJRGARCH-sstd	2	-0.6609	1.0000	6	0.7681	0.9668	7.1456×10^{-6}
GJRGARCH-ged	1	-0.7170	1.0000	1	-0.3821	1.0000	7.0862×10^{-6}
JSE (0.025)							
GARCH-norm	10	0.5102	0.9512	9	0.9217	0.9612	1.3615×10^{-6}
GARCH-std	8	0.3305	0.9880	8	0.9161	0.9614	1.3582×10^{-6}
GARCH-sstd	7	0.2777	0.9944	7	0.8823	0.9684	1.3578×10^{-6}
GARCH-ged	2	-0.7339	1.0000	3	0.2509	1.0000	1.3479×10^{-6}
EGARCH-norm	12	0.8311	0.7982	11	1.2502	0.8398	1.3711×10^{-6}
EGARCH-std	11	0.5469	0.9386	12	1.3543	0.7808	1.3651×10^{-6}
EGARCH-sstd	9	0.4002	0.9768	10	0.9652	0.9498	1.3619×10^{-6}
EGARCH-ged	1	-1.2696	1.0000	2	0.0208	1.0000	1.3444×10^{-6}
GJRGARCH-norm	4	-0.4305	1.0000	4	0.5084	0.9986	1.3505×10^{-6}
GJRGARCH-std	6	-0.3258	1.0000	6	0.7292	0.9902	1.3515×10^{-6}
GJRGARCH-sstd	5	-0.3664	1.0000	5	0.7072	0.9920	1.3509×10^{-6}
GJRGARCH-ged	3	-0.5462	1.0000	1	-0.0208	1.0000	1.3440×10^{-6}

Table 5.2: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
JSE (0.05)							
GARCH-norm	8	0.7239	0.7210	11	1.5696	0.4432	2.2468×10^{-5}
GARCH-sstd	11	1.4792	0.2250	10	1.3219	0.6452	2.2607×10^{-5}
GARCH-ged	5	-0.2188	1.0000	8	1.1457	0.7590	2.2346×10^{-5}
EGARCH-norm	7	0.6094	0.7852	4	1.0234	0.8360	2.2568×10^{-5}
EGARCH-std	10	0.8134	0.6644	9	1.2305	0.7112	2.2725×10^{-5}
EGARCH-sstd	9	0.7760	0.6874	5	1.0816	0.8012	2.2698×10^{-5}
EGARCH-ged	6	0.4300	0.8750	3	0.8146	0.9258	2.2481×10^{-5}
GJRGARCH-norm	2	-1.0037	1.0000	2	0.7383	0.9472	2.2018×10^{-5}
GJRGARCH-std	4	-0.8718	1.0000	7	1.1259	0.7752	2.2136×10^{-5}
GJRGARCH-sstd	3	-0.9056	1.0000	6	1.0912	0.7958	2.2125×10^{-5}
GJRGARCH-ged	1	-1.0100	1.0000	1	-0.7383	1.0000	2.1986×10^{-5}
FNB (0.01)							
GARCH-norm	10	0.310	0.7436	9	1.0321	0.8786	7.7575×10^{-6}
GARCH-std	9	0.5178	0.9456	6	0.8692	0.9360	7.6030×10^{-6}
GARCH-sstd	8	0.4485	0.9668	7	0.9109	0.9242	7.5937×10^{-6}
GARCH-ged	7	-0.2288	1.0000	5	0.7829	0.9592	7.5349×10^{-6}
EGARCH-norm	11	1.3127	0.4546	12	1.5311	0.5726	7.7523×10^{-6}
EGARCH-std	6	-0.5637	1.0000	10	1.1724	0.8052	7.5182×10^{-6}
EGARCH-sstd	1	-1.0441	1.0000	4	0.7491	0.9646	7.4835×10^{-6}
EGARCH-ged	2	-1.0271	1.0000	2	0.4888	0.9960	7.4418×10^{-6}
GJRGARCH-norm	12	1.3315	0.4416	11	1.4500	0.6262	7.6611×10^{-6}
GJRGARCH-std	4	-0.9644	1.0000	8	0.9337	0.9176	7.4508×10^{-6}
GJRGARCH-sstd	3	-1.0161	1.0000	3	0.7420	0.9660	7.4285×10^{-6}
GJRGARCH-ged	5	-0.9405	1.0000	1	-0.4888	1.0000	7.3825×10^{-6}

Table 5.3: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
FNB (0.025)							
GARCH-norm	12	0.7434	0.8514	9	1.0381	0.8776	1.4772×10^{-5}
GARCH-std	10	0.6322	0.9054	12	1.1432	0.8248	1.4729×10^{-5}
GARCH-sstd	9	0.4821	0.9586	10	1.0882	0.8526	1.4706×10^{-5}
GARCH-ged	5	-0.4241	1.0000	3	0.3904	0.9994	1.4602×10^{-5}
EGARCH-norm	11	0.6841	0.8836	11	1.1277	0.8330	1.4734×10^{-5}
EGARCH-std	8	0.1524	0.9998	8	0.9893	0.8952	1.4659×10^{-5}
EGARCH-sstd	7	-0.1686	1.0000	7	0.8895	0.9314	1.4629×10^{-5}
EGARCH-ged	2	-0.6861	1.0000	2	0.2086	1.0000	1.4553×10^{-5}
GJRGARCH-norm	6	-0.3850	1.0000	4	0.6173	0.9886	1.4615×10^{-5}
GJRGARCH-std	4	-0.4551	1.0000	5	0.7331	0.9722	1.4608×10^{-5}
GJRGARCH-sstd	3	-0.4741	1.0000	6	0.8174	0.9520	1.4602×10^{-5}
GJRGARCH-ged	1	-0.7043	1.0000	1	-0.2086	1.0000	1.4523×10^{-5}
FNB (0.05)							
GARCH-norm	7	0.4884	0.8902	5	0.7834	0.9482	2.4180×10^{-5}
GARCH-std	12	0.9223	0.6680	11	1.2558	0.7418	2.4287×10^{-5}
GARCH-sstd	11	0.9003	0.6834	12	1.3828	0.6530	2.4272×10^{-5}
GARCH-ged	5	-0.2817	1.0000	4	0.7684	0.9514	2.4091×10^{-5}
EGARCH-norm	8	0.5404	0.8686	6	0.7947	0.9454	2.4242×10^{-5}
EGARCH-std	10	0.6601	0.8158	8	0.8804	0.9170	2.4278×10^{-5}
EGARCH-sstd	9	0.6084	0.8398	7	0.8731	0.9204	2.4248×10^{-5}
EGARCH-ged	6	-0.1371	1.0000	3	0.6310	0.9796	2.4103×10^{-5}
GJRGARCH-norm	1	-0.8265	1.0000	2	0.4167	0.9968	2.3910×10^{-5}
GJRGARCH-std	4	-0.7095	1.0000	9	0.9944	0.8790	2.3974×10^{-5}
GJRGARCH-sstd	3	-0.7276	1.0000	10	1.1151	0.8238	2.3961×10^{-5}
GJRGARCH-ged	2	-0.7575	1.0000	1	-0.4167	1.0000	2.3883×10^{-5}

Table 5.4: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Standard bank (0.01)							
GARCH-norm	11	1.0613	0.5802	12	1.2546	0.7154	8.0447×10^{-6}
GARCH-std	4	-1.0273	1.0000	6	1.0427	0.8360	7.7315×10^{-6}
GARCH-sstd	2	-1.0822	1.0000	4	0.9135	0.8886	7.7379×10^{-6}
GARCH-ged	1	-1.1412	1.0000	2	0.6594	0.9652	7.6374×10^{-6}
EGARCH-norm	9	1.0383	0.6010	8	1.1080	0.8074	8.3069×10^{-6}
EGARCH-std	12	1.1066	0.5418	11	1.2280	0.7306	8.0459×10^{-6}
EGARCH-sstd	10	1.0406	0.5990	9	1.1791	0.7616	8.0420×10^{-6}
EGARCH-ged	8	1.0301	0.6092	10	1.1860	0.7570	7.9483×10^{-6}
GJRGARCH-norm	7	-0.2391	1.0000	7	1.0815	0.8210	7.7985×10^{-6}
GJRGARCH-std	6	-0.9928	1.0000	5	1.0240	0.8434	7.5843×10^{-6}
GJRGARCH-sstd	3	-1.0436	1.0000	3	0.8666	0.9050	7.5806×10^{-6}
GJRGARCH-ged	5	-1.0122	1.0000	1	-0.6594	1.0000	7.5074×10^{-6}
Standard bank (0.025)							
GARCH-norm	9	0.6675	0.8102	11	1.1189	0.7714	1.5040×10^{-5}
GARCH-std	6	-0.3224	1.0000	6	0.7184	0.9468	1.4954×10^{-5}
GARCH-sstd	7	-0.0298	1.0000	12	1.1520	0.7506	1.4975×10^{-5}
GARCH-ged	3	-0.7369	1.0000	2	0.5391	0.9850	1.4861×10^{-5}
EGARCH-norm	10	0.7262	0.7724	7	0.7504	0.9382	1.5251×10^{-5}
EGARCH-std	11	0.7575	0.7526	8	0.8142	0.9186	1.5203×10^{-5}
EGARCH-sstd	12	0.7636	0.7484	9	0.8228	0.9152	1.5211×10^{-5}
EGARCH-ged	8	0.5936	0.8462	5	0.7090	0.9500	1.5069×10^{-5}
GJRGARCH-norm	1	-0.7531	1.0000	3	0.5508	0.9836	1.4831×10^{-5}
GJRGARCH-std	2	-0.7483	1.0000	4	0.5605	0.9822	1.4802×10^{-5}
GJRGARCH-sstd	4	-0.7356	1.0000	10	0.8587	0.9022	1.4813×10^{-5}
GJRGARCH-ged	5	-0.6852	1.0000	1	-0.5391	1.0000	1.4718×10^{-5}

Table 5.5: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Standard bank (0.05)							
GARCH-norm	10	1.1325	0.4108	10	2.2482	0.1044	2.4573×10^{-5}
GARCH-ged	6	0.0286	1.0000	3	0.9905	0.8180	2.4447×10^{-5}
EGARCH-norm	7	0.3966	0.8860	6	1.3522	0.5642	2.4541×10^{-5}
EGARCH-std	8	0.4453	0.8626	5	1.0437	0.7822	2.4576×10^{-5}
EGARCH-sstd	9	0.4825	0.8420	8	1.7747	0.2776	2.4590×10^{-5}
EGARCH-ged	5	-0.0249	1.0000	2	0.4766	0.9886	2.4439×10^{-5}
GJRGARCH-norm	2	-0.5507	1.0000	7	1.5412	0.4294	2.4303×10^{-5}
GJRGARCH-std	3	-0.4651	1.0000	4	1.0272	0.7958	2.4366×10^{-5}
GJRGARCH-sstd	4	-0.3449	1.0000	9	2.1946	0.1160	2.4388×10^{-5}
GJRGARCH-ged	1	-0.7627	1.0000	1	-0.4766	1.0000	2.4209×10^{-5}
Absa (0.01)							
GARCH-norm	10	1.1939	0.5222	5	1.4460	0.5858	9.1935×10^{-6}
GARCH-std	11	1.2327	0.4890	11	1.8524	0.3110	8.8911×10^{-6}
GARCH-sstd	12	1.3128	0.4284	10	1.7962	0.3434	8.9125×10^{-6}
GARCH-ged	7	0.0072	1.0000	12	1.8702	0.3014	8.7978×10^{-6}
EGARCH-norm	8	0.7280	0.8500	3	1.1597	0.7904	9.0310×10^{-6}
EGARCH-std	3	-1.8478	1.0000	7	1.5736	0.4926	8.6287×10^{-6}
EGARCH-sstd	1	-1.9876	1.0000	4	1.3108	0.6886	8.6455×10^{-6}
EGARCH-ged	2	-1.8493	1.0000	1	-0.3961	1.0000	8.5128×10^{-6}
GJRGARCH-norm	9	1.1712	0.5410	8	1.6356	0.4464	8.9501×10^{-6}
GJRGARCH-std	5	-0.5648	1.0000	6	1.5206	0.5322	8.7016×10^{-6}
GJRGARCH-sstd	6	-0.5317	1.0000	9	1.7670	0.3610	8.7234×10^{-6}
GJRGARCH-ged	3	-0.8557	1.0000	2	0.3961	0.9962	8.5780×10^{-6}

Table 5.6: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Absa (0.025)							
GARCH-norm	10	1.1276	0.6346	10	1.4802	0.6004	1.6903×10^{-5}
GARCH-std	11	1.4898	0.3500	11	1.4977	0.5870	1.6716×10^{-5}
GARCH-sstd	12	1.5039	0.3402	12	1.6358	0.4862	1.6753×10^{-5}
GARCH-ged	5	-0.6516	1.0000	5	0.9767	0.8954	1.6465×10^{-5}
EGARCH-norm	9	0.8007	0.8342	7	1.0463	0.8726	1.6960×10^{-5}
EGARCH-std	2	-1.0074	1.0000	3	0.8374	0.9324	1.6461×10^{-5}
EGARCH-sstd	8	-0.3333	1.0000	4	0.9330	0.9080	1.6523×10^{-5}
EGARCH-ged	1	-1.3330	1.0000	1	-0.1193	1.0000	1.6288×10^{-5}
GJRGARCH-norm	6	-0.5785	1.0000	8	1.0711	0.8640	1.6503×10^{-5}
GJRGARCH-std	4	-0.6524	1.0000	6	0.9977	0.8908	1.6441×10^{-5}
GJRGARCH-sstd	7	-0.5702	1.0000	9	1.1490	0.8296	1.6471×10^{-5}
GJRGARCH-ged	3	-0.8094	1.0000	2	0.1193	1.0000	1.6311×10^{-5}
Absa (0.05)							
GARCH-norm	10	1.0786	0.5606	6	1.2006	0.7484	2.7883×10^{-5}
GARCH-std	11	1.5172	0.2398	10	1.5499	0.4782	2.7890×10^{-5}
GARCH-sstd	12	1.5885	0.2130	11	1.5513	0.4762	2.7960×10^{-5}
GARCH-ged	5	-0.6842	1.0000	4	1.0940	0.8060	2.7529×10^{-5}
EGARCH-norm	7	0.6314	0.8376	5	1.1057	0.8014	2.7886×10^{-5}
EGARCH-std	8	0.7509	0.7696	9	1.5028	0.5118	2.7824×10^{-5}
EGARCH-sstd	9	0.8774	0.6974	8	1.4994	0.5132	2.7898×10^{-5}
EGARCH-ged	4	-0.7554	1.0000	2	0.7961	0.9268	2.7476×10^{-5}
GJRGARCH-norm	2	-1.1053	1.0000	3	1.0877	0.8092	2.7294×10^{-5}
GJRGARCH-std	3	-0.7689	1.0000	7	1.4558	0.5482	2.7370×10^{-5}
GJRGARCH-sstd	6	-0.6061	1.0000	12	2.6223	0.0694	2.7453×10^{-5}
GJRGARCH-ged	1	-1.1270	1.0000	1	-0.7961	1.0000	2.6963×10^{-5}

Table 5.7: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Capitec (0.01)							
GARCH-norm	9	0.6357	0.7602	8	1.9806	0.2058	1.0439×10^{-5}
GARCH-std	7	0.4194	0.9094	6	1.4595	0.5032	1.0248×10^{-5}
GARCH-sstd	4	-0.2173	1.0000	3	0.9302	0.8600	1.0177×10^{-5}
EGARCH-norm	2	-0.5333	1.0000	2	0.1482	1.0000	1.00602×10^{-5}
EGARCH-std	3	-0.4845	1.0000	4	0.9545	0.8468	1.00603×10^{-5}
EGARCH-sstd	1	-0.9109	1.0000	1	-0.1482	1.0000	9.9949×10^{-6}
GJRGARCH-norm	8	0.4511	0.8880	9	2.2423	0.1306	1.0323×10^{-5}
GJRGARCH-std	6	0.3658	0.9350	7	1.7276	0.3296	1.0257×10^{-5}
GJRGARCH-sstd	5	-0.0383	1.0000	5	1.1823	0.7076	1.0190×10^{-5}
Capitec (0.025)							
GARCH-norm	5	0.2080	0.9988	4	0.9558	0.9256	1.7067×10^{-5}
GARCH-std	6	0.3520	0.9916	8	1.3538	0.6998	1.7060×10^{-5}
GARCH-sstd	2	-1.1995	1.0000	3	0.8900	0.9456	1.6970×10^{-5}
EGARCH-norm	7	0.7467	0.8902	5	1.1047	0.8546	1.7309×10^{-5}
EGARCH-std	3	-0.8373	1.0000	2	0.3264	0.9994	1.6836×10^{-5}
EGARCH-sstd	1	-1.4066	1.0000	1	-0.3264	1.0000	1.6800×10^{-5}
GJRGARCH-norm	8	0.7547	0.8860	7	1.2301	0.7812	1.7129×10^{-5}
GJRGARCH-std	9	0.8559	0.8324	9	1.6174	0.5014	1.7145×10^{-5}
GJRGARCH-sstd	4	-0.0197	1.0000	6	1.1491	0.8290	1.7038×10^{-5}

Table 5.8: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Capitec (0.05)							
GARCH-norm	6	0.0557	1.0000	3	0.5792	0.9908	2.6377×10^{-5}
GARCH-std	1	-0.8656	1.0000	2	0.1628	1.0000	2.6217×10^{-5}
GARCH-sstd	4	-0.1916	1.0000	8	1.1717	0.8186	2.6331×10^{-5}
EGARCH-norm	8	0.5199	0.9432	6	0.6818	0.9818	2.6528×10^{-5}
EGARCH-std	7	0.3671	0.9864	5	0.6245	0.9878	2.6445×10^{-5}
EGARCH-sstd	9	0.7204	0.8494	9	1.2194	0.7880	2.6583×10^{-5}
GJRGARCH-norm	5	-0.0318	1.0000	4	0.6076	0.9894	2.6357×10^{-5}
GJRGARCH-std	2	-0.7827	1.0000	1	-0.1628	1.0000	2.6193×10^{-5}
GJRGARCH-sstd	3	-0.5291	1.0000	7	0.7225	0.9772	2.6267×10^{-5}
Nedbank (0.01)							
GARCH-norm	11	0.9431	0.7316	6	1.0483	0.8270	9.5237×10^{-6}
GARCH-std	9	0.5810	0.8964	11	1.1406	0.7746	9.1823×10^{-6}
GARCH-sstd	10	0.7046	0.8510	9	1.1346	0.7778	9.2109×10^{-6}
GARCH-ged	2	-0.9994	1.0000	3	0.7627	0.9242	8.9629×10^{-6}
EGARCH-norm	12	1.0837	0.6078	12	1.2072	0.7262	9.7352×10^{-6}
EGARCH-std	7	0.3724	0.9574	5	1.0406	0.8308	9.1455×10^{-6}
EGARCH-sstd	8	0.4739	0.9284	7	1.0628	0.8198	9.1654×10^{-6}
EGARCH-ged	1	-1.2221	1.0000	2	0.7505	0.9276	8.9874×10^{-6}
GJRGARCH-norm	6	0.1380	0.9990	4	0.9655	0.8622	9.1170×10^{-6}
GJRGARCH-std	4	-0.8253	1.0000	8	1.1323	0.7792	8.7657×10^{-6}
GJRGARCH-sstd	5	-0.8011	1.0000	10	1.1357	0.7776	8.7956×10^{-6}
GJRGARCH-ged	3	-0.9141	1.0000	1	-0.7505	1.0000	8.6033×10^{-6}

Table 5.9: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Nedbank (0.025)							
GARCH-norm	7	0.1927	0.9932	3	0.7186	0.9412	1.7299×10^{-5}
GARCH-std	9	0.6760	0.8348	9	0.9964	0.8458	1.7375×10^{-5}
GARCH-sstd	8	0.6541	0.8458	8	0.9766	0.8554	1.7382×10^{-5}
GARCH-ged	5	-0.6401	1.0000	4	0.7908	0.9186	1.7165×10^{-5}
EGARCH-norm	12	1.0115	0.6610	12	1.1622	0.7608	1.7726×10^{-5}
EGARCH-std	10	0.8721	0.7482	10	1.0082	0.8388	1.7453×10^{-5}
EGARCH-sstd	11	0.8936	0.7364	11	1.0147	0.8356	1.7473×10^{-5}
EGARCH-ged	6	-0.0692	1.0000	7	0.8646	0.8960	1.7241×10^{-5}
GJRGARCH-norm	1	-1.0961	1.0000	2	0.5249	0.9800	1.6996×10^{-5}
GJRGARCH-std	4	-0.7685	1.0000	6	0.8539	0.8980	1.7041×10^{-6}
GJRGARCH-sstd	3	-0.7836	1.0000	5	0.8252	0.9048	1.7050×10^{-6}
GJRGARCH-ged	2	-0.8269	1.0000	1	-0.5249	1.0000	1.6813×10^{-6}
Nedbank (0.05)							
GARCH-norm	6	0.1198	0.9988	4	0.7352	0.9408	2.7626×10^{-5}
GARCH-std	7	0.5341	0.8880	7	1.0182	0.8416	2.7735×10^{-5}
GARCH-sstd	9	0.5832	0.8662	10	1.0552	0.8270	2.7755×10^{-5}
GARCH-ged	5	-0.7943	1.0000	3	0.7210	0.9434	2.7515×10^{-5}
EGARCH-norm	12	1.0574	0.6042	12	1.5232	0.5048	2.7912×10^{-5}
EGARCH-std	10	0.8530	0.7376	6	0.9220	0.8816	2.7841×10^{-5}
EGARCH-sstd	11	0.8576	0.7354	5	0.9211	0.8816	2.7852×10^{-5}
EGARCH-ged	8	0.5809	0.8674	9	1.0344	0.8362	2.7712×10^{-5}
GJRGARCH-norm	1	-0.9427	1.0000	2	0.5779	0.9768	2.7291×10^{-5}
GJRGARCH-std	3	-0.8201	1.0000	11	1.0592	0.8254	2.7411×10^{-5}
GJRGARCH-sstd	4	-0.8158	1.0000	8	1.0193	0.8410	2.7425×10^{-5}
GJRGARCH-ged	2	-0.8418	1.0000	1	-0.5779	1.0000	2.7193×10^{-5}

Table 5.10: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Investec (0.01)							
GARCH-std	9	0.7453	0.8494	6	1.2021	0.7924	9.5169×10^{-6}
GARCH-sstd	8	0.5549	0.9404	4	1.0479	0.8734	9.4498×10^{-6}
GARCH-ged	5	-0.1213	1.0000	3	0.8818	0.9348	9.2975×10^{-6}
EGARCH-std	7	0.1367	1.0000	7	1.2576	0.7578	9.3246×10^{-6}
EGARCH-sstd	3	-0.8463	1.0000	5	1.1111	0.8434	9.2431×10^{-6}
EGARCH-ged	1	-1.4182	1.0000	2	0.1520	1.0000	9.0247×10^{-6}
GJRGARCH-norm	10	1.5822	0.2786	10	2.1742	0.1920	9.7109×10^{-6}
GJRGARCH-std	6	0.0847	1.0000	9	1.7724	0.3938	9.3184×10^{-6}
GJRGARCH-sstd	4	-0.7816	1.0000	8	1.4720	0.6098	9.2383×10^{-6}
GJRGARCH-ged	2	-1.1202	1.0000	1	-0.1520	1.0000	9.0051×10^{-6}
Investec (0.025)							
GARCH-norm	11	0.8361	0.8660	7	1.1444	0.8898	1.7902×10^{-5}
GARCH-std	10	0.7722	0.8938	6	1.0939	0.9104	1.7962×10^{-5}
GARCH-sstd	12	0.8866	0.8424	9	1.2590	0.8384	1.7955×10^{-5}
GARCH-ged	9	0.6528	0.9400	8	1.2365	0.8492	1.7836×10^{-5}
EGARCH-norm	7	0.1147	1.0000	12	1.4901	0.7046	1.7784×10^{-5}
EGARCH-std	5	-0.4437	1.0000	4	0.8399	0.9732	1.7722×10^{-5}
EGARCH-sstd	2	-0.9693	1.0000	3	0.8340	0.9742	1.7687×10^{-5}
EGARCH-ged	1	-1.2999	1.0000	1	-0.6502	1.0000	1.7515×10^{-5}
GJRGARCH-norm	4	-0.4549	1.0000	5	0.9416	0.9520	1.7670×10^{-5}
GJRGARCH-std	8	0.2708	0.9996	11	1.3713	0.7802	1.7781×10^{-5}
GJRGARCH-sstd	6	0.0553	1.0000	10	1.3493	0.7904	1.7768×10^{-5}
GJRGARCH-ged	3	-0.7953	1.0000	2	0.6502	0.9914	1.7588×10^{-5}

Table 5.11: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
				Investec (0.05)			
GARCH-norm	12	1.1628	0.6116	9	1.6035	0.6522	2.8400×10^{-5}
GARCH-std	9	0.9999	0.7350	6	1.5105	0.7138	2.8516×10^{-5}
GARCH-sstd	11	1.1346	0.6336	10	1.6158	0.6430	2.8502×10^{-5}
GARCH-ged	10	1.1299	0.6370	8	1.5342	0.6954	2.8369×10^{-5}
EGARCH-norm	3	-0.9237	1.0000	12	1.8003	0.5134	2.8036×10^{-5}
EGARCH-std	5	-0.5170	1.0000	3	0.9308	0.9702	2.8154×10^{-5}
EGARCH-sstd	2	-1.0479	1.0000	2	0.7879	0.9890	2.8101×10^{-5}
EGARCH-ged	1	-1.6565	1.0000	1	-0.7879	1.0000	2.7919×10^{-5}
GJRGARCH-norm	6	-0.3979	1.0000	4	1.2096	0.8896	2.8127×10^{-5}
GJRGARCH-std	8	0.6251	0.9384	11	1.6781	0.6010	2.8278×10^{-5}
GJRGARCH-sstd	7	0.3589	0.9928	5	1.4966	0.7220	2.8258×10^{-5}
GJRGARCH-ged	4	-0.5532	1.0000	7	1.5279	0.6992	2.8127×10^{-5}

Table 5.12: Summary of SSM composition for different quantile levels

	1%	2.5%	5%		1%	2.5%	5%
Models	Distributions						
	JSE				JSE		
GARCH	4	4	3	norm	3	3	3
EGARCH	4	4	4	std	3	3	2
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
	FNB				FNB		
GARCH	4	4	4	norm	3	3	3
EGARCH	4	4	4	std	3	3	2
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
	Standard bank				Standard bank		
GARCH	4	4	2	norm	3	3	3
EGARCH	4	4	4	std	3	3	2
GJRGARCH	4	4	4	sstd	3	3	2
				ged	3	3	3
	Absa				Absa		
GARCH	4	4	4	norm	3	3	3
EGARCH	4	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
	Capitec				Capitec		
GARCH	3	3	3	norm	3	3	3
EGARCH	3	3	3	std	3	3	3
GJRGARCH	3	3	3	sstd	3	3	3
				ged			
	Nedbank				Nedbank		
GARCH	4	4	4	norm	3	3	3
EGARCH	4	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
	Investec bank				Investec bank		
GARCH	3	4	4	norm	1	3	3
EGARCH	3	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3

5.3.2 Evaluation of VaR prediction performance under different forecasting periods

We estimated the performance of the GARCH models under different forecasting periods. The number of out-of-forecast periods includes 1000, 1200 and 1400 observations. The models were re-estimated every 30 days at 1% quantile level. The MCS approach results and summaries of the final SSM compositions under different forecasting periods are shown in Tables 5.13, 5.14, 5.15, 5.16, 5.17, 5.18, 5.19, 5.20, 5.21, 5.22 and 5.23. Regardless of the forecasting period for the JSE stock returns, all the models survived, indicating that the out-of-sample forecasting ability of the models is good. As the forecasting period increases, the difference in forecasting accuracy between the models becomes smaller. For each of the forecasting periods, the out-of-sample forecasting accuracy of the models based on all distributions is relatively better. Thus, all the models exhibit comparable EPAs. From the MCS approach results under different forecasting periods in Table 5.13, when H is set to 1000, only the GJRGARCH-ged model has p values equaling 1.0000 for both test statistics, and the loss value is the smallest for this model. Thus, the GJRGARCH-ged model shows relatively optimal VaR forecasting ability. As the forecasting period increased to 1200 and 1400, the optimal model remained the GJRGARCH-ged model.

All models survived for FNB stock returns irrespective of the forecasting period and distribution. From the results shown in Tables 5.14 and 5.15, the GJRGARCH-ged model has p-values equal to 1.0000 for both tests and has the lowest loss value for each of the forecasting periods considered in the study. Thus, the GJRGARCH-ged model is regarded as the optimal model for VaR forecasting for FNB data. All models survived for standard bank stock returns regardless of the forecasting period and distribution. The GJRGARCH-ged model has p-values equalling 1.0000 for both tests and has the lowest loss value for each of the forecasting periods. Thus, the GJRGARCH-ged model is considered the most optimal model for VaR forecasting for FNB data. For the Absa stock returns, all the models survived for each of the forecasting periods considered. As the forecasting period increased from 1000 to 1200 and 1400, in addition to the EGARCH-ged model, the p-values of the test statistics became 1.0000 for the GJRGARCH-ged model. However, if the forecasting ability of the model is considered from the loss value perspective, the EGARCH-ged model remains the optimal model. All the models survived for each of the forecasting periods for the Capitec daily stock returns. The optimal model for each of the forecasting periods is the EGARCH-sstd. Irrespective of the forecasting period used for the Nedbank returns, the GJRGARCH-ged model remains the optimal model. For Investec stock returns, the optimal model for the out-of-sample VaR forecasting is GJRGARCH-ged for each of the forecasting periods being considered in the study.

Table 5.13: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
JSE (H=1000)							
GARCH-norm	11	0.7273	0.8716	12	1.0833	0.8646	7.3525×10^{-6}
GARCH-std	7	-0.2562	1.0000	7	0.7750	0.9656	7.2007×10^{-6}
GARCH-sstd	6	-0.3283	1.0000	3	0.5093	0.9964	7.1925×10^{-6}
GARCH-ged	3	-0.6469	1.0000	2	0.3821	0.9998	7.1633×10^{-6}
EGARCH-norm	12	1.0084	0.7100	11	1.0730	0.8684	7.4139×10^{-6}
EGARCH-std	9	0.5577	0.9408	10	1.0301	0.8864	7.2662×10^{-6}
EGARCH-sstd	8	0.3128	0.9922	5	0.7034	0.9798	7.2459×10^{-6}
EGARCH-ged	5	-0.3904	1.0000	4	0.6493	0.9846	7.1943×10^{-6}
GJRGARCH-norm	10	0.6369	0.9126	9	0.9015	0.9342	7.2765×10^{-6}
GJRGARCH-std	4	-0.6135	1.0000	8	0.8271	0.9558	7.1544×10^{-6}
GJRGARCH-sstd	2	-0.6609	1.0000	6	0.7681	0.9668	7.1456×10^{-6}
GJRGARCH-ged	1	-0.7170	1.0000	1	-0.3821	1.0000	7.0862×10^{-6}
JSE (H=1200)							
GARCH-norm	9	0.4327	0.9690	8	0.7026	0.9704	6.6953×10^{-6}
GARCH-std	6	-0.1990	1.0000	4	0.5747	0.9874	6.6089×10^{-6}
GARCH-sstd	5	-0.2506	1.0000	2	0.3648	0.9988	6.6039×10^{-6}
GARCH-ged	4	-0.3992	1.0000	3	0.3759	0.9984	6.5932×10^{-6}
EGARCH-norm	12	0.9118	0.7614	12	0.9917	0.8850	6.7680×10^{-6}
EGARCH-std	11	0.7371	0.8548	11	0.8202	0.9454	6.6698×10^{-6}
EGARCH-sstd	10	0.5263	0.9422	9	0.7293	0.9636	6.6552×10^{-6}
EGARCH-ged	7	0.0091	1.0000	10	0.7460	0.9604	6.6261×10^{-6}
GJRGARCH-norm	8	0.1322	1.0000	5	0.6030	0.9836	6.6345×10^{-6}
GJRGARCH-std	3	-0.5996	1.0000	7	0.6248	0.9808	6.5653×10^{-6}
GJRGARCH-sstd	2	-0.6136	1.0000	6	0.6073	0.9836	6.5600×10^{-6}
GJRGARCH-ged	1	-0.6234	1.0000	1	-0.3648	1.0000	6.5257×10^{-6}

Table 5.14: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
JSE (H=1400)							
GARCH-norm	11	0.7189	0.8568	12	1.1064	0.8352	6.5203×10^{-6}
GARCH-std	6	-0.1838	1.0000	4	0.4340	0.9974	6.4009×10^{-6}
GARCH-sstd	5	-0.2328	1.0000	3	0.4156	0.9976	6.3963×10^{-6}
GARCH-ged	4	-0.4164	1.0000	2	0.4081	0.9980	6.3841×10^{-6}
EGARCH-norm	12	1.0526	0.6458	11	1.0737	0.8532	6.5536×10^{-6}
EGARCH-std	9	0.4288	0.9728	9	0.7887	0.9528	6.4399×10^{-6}
EGARCH-sstd	8	0.2005	0.9976	7	0.7590	0.9606	6.4262×10^{-6}
EGARCH-ged	7	-0.1450	1.0000	8	0.7764	0.9562	6.4070×10^{-6}
GJRGARCH-norm	10	0.5941	0.9196	10	1.0213	0.8748	6.4568×10^{-6}
GJRGARCH-std	1	-0.7710	1.0000	5	0.6231	0.9848	6.3480×10^{-6}
GJRGARCH-sstd	2	-0.7475	1.0000	6	0.6443	0.9810	6.3423×10^{-6}
GJRGARCH-ged	3	-0.7333	1.0000	1	-0.4081	1.0000	6.3162×10^{-6}
FNB (H=1000)							
GARCH-norm	10	0.9268	0.7458	9	1.0237	0.8832	7.7596×10^{-6}
GARCH-std	9	0.5226	0.9440	6	0.8756	0.9344	7.6063×10^{-6}
GARCH-sstd	8	0.4536	0.9654	7	0.9148	0.9232	7.5971×10^{-6}
GARCH-ged	7	-0.2400	1.0000	5	0.7796	0.9596	7.5369×10^{-6}
EGARCH-norm	11	1.3132	0.4560	12	1.5340	0.5712	7.7563×10^{-6}
EGARCH-std	6	-0.5750	1.0000	10	1.1361	0.8268	7.5202×10^{-6}
EGARCH-sstd	1	-1.0388	1.0000	4	0.7368	0.9662	7.4866×10^{-6}
EGARCH-ged	2	-1.0186	1.0000	2	0.4936	0.9956	7.4457×10^{-6}
GJRGARCH-norm	12	1.3194	0.4502	11	1.4397	0.6318	7.6624×10^{-6}
GJRGARCH-std	4	-0.9622	1.0000	8	0.9428	0.9148	7.4535×10^{-6}
GJRGARCH-sstd	3	-1.0127	1.0000	3	0.7345	0.9670	7.4305×10^{-6}
GJRGARCH-ged	5	-0.9371	1.0000	1	-0.4936	1.0000	7.3855×10^{-6}

Table 5.15: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
FNB (H=1200)							
GARCH-norm	10	0.8446	0.7872	9	0.9549	0.8846	7.1944×10^{-6}
GARCH-std	9	0.4219	0.9680	6	0.8420	0.9274	7.0600×10^{-6}
GARCH-sstd	8	0.3497	0.9828	8	0.8607	0.9202	7.0509×10^{-6}
GARCH-ged	7	-0.1211	1.0000	7	0.8438	0.9268	7.0173×10^{-6}
EGARCH-norm	12	1.2856	0.4712	12	1.4003	0.6366	7.2020×10^{-6}
EGARCH-std	6	-0.3337	1.0000	10	0.9597	0.8830	7.0081×10^{-6}
EGARCH-sstd	5	-0.7579	1.0000	4	0.7711	0.9464	6.9789×10^{-6}
EGARCH-ged	4	-0.8214	1.0000	3	0.7069	0.9640	6.9529×10^{-6}
GJRGARCH-norm	11	1.2022	0.5388	11	1.3902	0.6432	7.1154×10^{-6}
GJRGARCH-std	1	-1.0611	1.0000	5	0.8033	0.9398	6.9271×10^{-6}
GJRGARCH-sstd	2	-1.0446	1.0000	2	0.6769	0.9702	6.9056×10^{-6}
GJRGARCH-ged	3	-0.9670	1.0000	1	-0.6769	1.0000	6.8803×10^{-6}
FNB (H=1400)							
GARCH-norm	10	1.1117	0.5966	10	1.2749	0.7270	7.0521×10^{-6}
GARCH-std	9	0.5679	0.9248	7	1.0196	0.8708	6.8954×10^{-6}
GARCH-sstd	8	0.5677	0.9248	9	1.0847	0.8378	6.8894×10^{-6}
GARCH-ged	7	0.0955	1.0000	8	1.0706	0.8432	6.8531×10^{-6}
EGARCH-norm	12	1.4093	0.3756	11	1.6318	0.4848	7.0098×10^{-6}
EGARCH-std	6	-0.7686	1.0000	6	0.9219	0.9080	6.8111×10^{-6}
EGARCH-sstd	4	-0.9816	1.0000	5	0.9041	0.9142	6.7884×10^{-6}
EGARCH-ged	5	-0.9059	1.0000	4	0.9012	0.9146	6.7730×10^{-6}
GJRGARCH-norm	11	1.3117	0.4448	12	1.6888	0.4476	6.9437×10^{-6}
GJRGARCH-std	1	-1.3768	1.0000	3	0.7960	0.9514	6.7395×10^{-6}
GJRGARCH-sstd	2	-1.2683	1.0000	2	0.7724	0.9570	6.7219×10^{-6}
GJRGARCH-ged	3	-1.1763	1.0000	1	-0.7724	1.0000	6.6963×10^{-6}

Table 5.16: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Standard bank (H=1000)							
GARCH-norm	11	1.0613	0.5802	12	1.2546	0.7154	8.0447×10^{-6}
GARCH-std	4	-1.0273	1.0000	6	1.0427	0.8360	7.7315×10^{-6}
GARCH-sstd	2	-1.0822	1.0000	4	0.9135	0.8886	7.7379×10^{-6}
GARCH-ged	1	-1.1412	1.0000	2	0.6594	0.9652	7.6374×10^{-6}
EGARCH-norm	9	1.0383	0.6010	8	1.1080	0.8074	8.3069×10^{-6}
EGARCH-std	12	1.1066	0.5418	11	1.2280	0.7306	8.0459×10^{-6}
EGARCH-sstd	10	1.0406	0.5990	9	1.1791	0.7616	8.0420×10^{-6}
EGARCH-ged	8	1.0301	0.6092	10	1.1860	0.7570	7.9483×10^{-6}
GJRGARCH-norm	7	-0.2391	1.0000	7	1.0815	0.8210	7.7985×10^{-6}
GJRGARCH-std	6	-0.9928	1.0000	5	1.0240	0.8434	7.5843×10^{-6}
GJRGARCH-sstd	3	-1.0436	1.0000	3	0.8666	0.9050	7.5806×10^{-6}
GJRGARCH-ged	5	-1.0122	1.0000	1	-0.6594	1.0000	7.5074×10^{-6}
Standard (H=1200)							
GARCH-norm	8	0.7558	0.8068	8	0.9858	0.8800	7.3387×10^{-6}
GARCH-std	6	-0.9097	1.0000	7	0.9083	0.9066	7.1296×10^{-6}
GARCH-sstd	3	-0.9976	1.0000	5	0.7655	0.9498	7.1281×10^{-6}
GARCH-ged	1	-1.0360	1.0000	2	0.6201	0.9788	7.0646×10^{-6}
EGARCH-norm	9	0.9793	0.6670	9	1.0534	0.8548	7.5533×10^{-6}
EGARCH-std	11	1.3151	0.3862	11	1.3779	0.6624	7.3859×10^{-6}
EGARCH-sstd	10	1.2485	0.4390	10	1.3355	0.6908	7.3797×10^{-6}
EGARCH-ged	12	1.4146	0.3158	12	1.3856	0.6564	7.3120×10^{-6}
GJRGARCH-norm	7	-0.6236	1.0000	4	0.7509	0.9532	7.1327×10^{-6}
GJRGARCH-std	4	-0.9716	1.0000	6	0.8574	0.9218	7.0134×10^{-6}
GJRGARCH-sstd	2	-1.0242	1.0000	3	0.6946	0.9678	7.0072×10^{-6}
GJRGARCH-ged	5	-0.9568	1.0000	1	-0.6201	1.0000	6.9635×10^{-6}

Table 5.17: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Standard bank (H=1400)							
GARCH-norm	8	0.9578	0.6814	9	1.1767	0.7916	7.2578×10^{-6}
GARCH-std	5	-0.8983	1.0000	6	0.9373	0.8946	7.0251×10^{-6}
GARCH-sstd	6	-0.8562	1.0000	4	0.8813	0.9146	7.0273×10^{-6}
GARCH-ged	1	-1.1148	1.0000	2	0.7171	0.9594	6.9694×10^{-6}
EGARCH-norm	9	1.0782	0.5902	8	1.1518	0.8024	7.4224×10^{-6}
EGARCH-std	10	1.2616	0.4344	11	1.3421	0.6836	7.2438×10^{-6}
EGARCH-sstd	11	1.2665	0.4320	12	1.3593	0.6708	7.2411×10^{-6}
EGARCH-ged	12	1.2988	0.4048	10	1.3011	0.7124	7.1786×10^{-6}
GJRGARCH-norm	7	-0.4422	1.0000	7	0.9621	0.8856	7.0366×10^{-6}
GJRGARCH-std	3	-1.0778	1.0000	5	0.8927	0.9096	6.8928×10^{-6}
GJRGARCH-sstd	2	-1.0945	1.0000	3	0.8402	0.9258	6.8905×10^{-6}
GJRGARCH-ged	4	-1.0471	1.0000	1	-0.7171	1.0000	6.8439×10^{-6}
Absa (H=1000)							
GARCH-norm	10	1.1942	0.5212	5	1.4459	0.5846	9.1924×10^{-6}
GARCH-std	11	1.2321	0.4900	11	1.8546	0.3100	8.8906×10^{-6}
GARCH-sstd	12	1.3123	0.4292	10	1.7937	0.3454	8.9117×10^{-6}
GARCH-ged	7	0.0090	1.0000	12	1.8724	0.3002	8.7976×10^{-6}
EGARCH-norm	8	0.7256	0.8524	3	1.1575	0.7924	9.0307×10^{-6}
EGARCH-std	3	-1.8448	1.0000	7	1.5792	0.4880	8.6290×10^{-6}
EGARCH-sstd	1	-1.9814	1.0000	4	1.3160	0.6826	8.6461×10^{-6}
EGARCH-ged	2	-1.8503	1.0000	1	-0.3979	1.0000	8.5123×10^{-6}
GJRGARCH-norm	9	1.1675	0.5452	8	1.6310	0.4504	8.9491×10^{-6}
GJRGARCH-std	5	-0.5620	1.0000	6	1.5309	0.5226	8.7021×10^{-6}
GJRGARCH-sstd	6	-0.5298	1.0000	9	1.7705	0.3600	8.7234×10^{-6}
GJRGARCH-ged	4	-0.8552	1.0000	2	0.3979	0.9960	8.5778×10^{-6}

Table 5.18: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Absa (H=1200)							
GARCH-norm	10	1.1330	0.5736	7	1.3619	0.6848	8.3997×10^{-6}
GARCH-std	12	1.5181	0.3014	11	1.9316	0.2982	8.1931×10^{-6}
GARCH-sstd	11	1.3967	0.3820	10	1.7885	0.3826	8.1946×10^{-6}
GARCH-ged	7	0.3278	0.9852	12	2.0078	0.2566	8.1241×10^{-6}
EGARCH-norm	8	0.8120	0.8048	4	1.2087	0.7834	8.2948×10^{-6}
EGARCH-std	3	-1.6872	1.0000	9	1.4985	0.5892	7.9908×10^{-6}
EGARCH-sstd	2	-1.7509	1.0000	6	1.3247	0.7096	7.9981×10^{-6}
EGARCH-ged	1	-1.7509	1.0000	1	-0.1004	1.0000	7.9038×10^{-6}
GJRGARCH-norm	9	0.8159	0.8018	8	1.3628	0.6834	8.1919×10^{-6}
GJRGARCH-std	6	-0.7130	1.0000	5	1.2385	0.7666	8.0082×10^{-6}
GJRGARCH-sstd	5	-0.7681	1.0000	3	1.1890	0.7948	8.0126×10^{-6}
GJRGARCH-ged	4	-0.9206	1.0000	2	0.1004	1.0000	7.9174×10^{-6}
Absa (H=1400)							
GARCH-norm	10	1.0370	0.6348	9	1.2386	0.7758	7.9110×10^{-6}
GARCH-std	12	1.4619	0.3178	12	1.7599	0.3978	7.7488×10^{-6}
GARCH-sstd	11	1.2282	0.4836	10	1.5788	0.5272	7.7420×10^{-6}
GARCH-ged	8	0.6180	0.8950	11	1.6396	0.4800	7.6944×10^{-6}
EGARCH-norm	9	0.6499	0.8836	6	0.9897	0.8970	7.7942×10^{-6}
EGARCH-std	3	-1.3620	1.0000	8	1.1069	0.8464	7.5728×10^{-6}
EGARCH-sstd	1	-1.7089	1.0000	3	0.7641	0.9596	7.5600×10^{-6}
EGARCH-ged	2	-1.3804	1.0000	1	-0.0156	1.0000	7.5133×10^{-6}
GJRGARCH-norm	7	0.3375	0.9820	5	0.9219	0.9174	7.6919×10^{-6}
GJRGARCH-std	6	-0.6362	1.0000	7	1.0118	0.8878	7.5841×10^{-6}
GJRGARCH-sstd	5	-0.7356	1.0000	4	0.8710	0.9330	7.5800×10^{-6}
GJRGARCH-ged	4	-0.7901	1.0000	2	0.0156	1.0000	7.5150×10^{-6}

Table 5.19: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Capitec (H=1000)							
GARCH-norm	9	0.6357	0.7602	8	1.9806	0.2058	1.0439×10^{-5}
GARCH-std	7	0.4194	0.9094	6	1.4595	0.5032	1.0248×10^{-5}
GARCH-sstd	4	-0.2173	1.0000	3	0.9302	0.8600	1.0177×10^{-5}
EGARCH-norm	2	-0.5333	1.0000	2	0.1482	1.0000	1.00602×10^{-5}
EGARCH-std	3	-0.4845	1.0000	4	0.9545	0.8468	1.00603×10^{-5}
EGARCH-sstd	1	-0.9109	1.0000	1	-0.1482	1.0000	9.9949×10^{-6}
GJRGARCH-norm	8	0.4511	0.8880	9	2.2423	0.1306	1.0323×10^{-5}
GJRGARCH-std	6	0.3658	0.9350	7	1.7276	0.3296	1.0257×10^{-5}
GJRGARCH-sstd	5	-0.0383	1.0000	5	1.1823	0.7076	1.0190×10^{-6}
Capitec (H=1200)							
GARCH-norm	7	0.4116	0.9374	8	1.6452	0.3986	9.3246×10^{-6}
GARCH-std	9	0.9423	0.6202	6	1.4462	0.5400	9.3024×10^{-6}
GARCH-sstd	6	0.2815	0.9814	3	0.7949	0.9294	9.2244×10^{-6}
EGARCH-norm	1	-0.8101	1.0000	2	0.0142	1.0000	9.0414×10^{-6}
EGARCH-std	3	-0.3938	1.0000	5	1.1568	0.7528	9.1028×10^{-6}
EGARCH-sstd	2	-0.7938	1.0000	1	-0.0142	1.0000	9.0364×10^{-6}
GJRGARCH-norm	5	0.2400	0.9902	9	1.8015	0.3062	9.2587×10^{-6}
GJRGARCH-std	8	0.6644	0.8012	7	1.5266	0.4744	9.2988×10^{-6}
GJRGARCH-sstd	4	0.2241	0.9930	4	0.9564	0.8630	9.2251×10^{-6}

Table 5.20: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Capitec (H=1400)							
GARCH-norm	9	0.6171	0.8200	8	1.7394	0.3688	9.5421×10^{-6}
GARCH-std	7	0.3681	0.9436	6	1.4211	0.5808	9.3934×10^{-6}
GARCH-sstd	4	-0.1324	1.0000	4	0.9194	0.8908	9.3413×10^{-6}
EGARCH-norm	3	-0.3663	1.0000	2	0.2672	0.9994	9.2706×10^{-6}
EGARCH-std	2	-0.5276	1.0000	3	0.8522	0.9172	9.2244×10^{-6}
EGARCH-sstd	1	-0.9622	1.0000	1	-0.2672	1.0000	9.1718×10^{-6}
GJRGARCH-norm	8	0.4913	0.8856	9	1.7858	0.3412	9.4732×10^{-6}
GJRGARCH-std	6	0.2999	0.9676	7	1.5053	0.5228	9.3963×10^{-6}
GJRGARCH-sstd	5	-0.0707	1.0000	5	1.0831	0.8064	9.3434×10^{-6}
Nedbank (H=1000)							
GARCH-norm	11	0.9431	0.7316	6	1.0483	0.8270	9.5237×10^{-6}
GARCH-std	9	0.5810	0.8964	11	1.1406	0.7746	9.1823×10^{-6}
GARCH-sstd	10	0.7046	0.8510	9	1.1346	0.7778	9.2109×10^{-6}
GARCH-ged	2	-0.9994	1.0000	3	0.7627	0.9242	8.9629×10^{-6}
EGARCH-norm	12	1.0837	0.6078	12	1.2072	0.7262	9.7352×10^{-6}
EGARCH-std	7	0.3724	0.9574	5	1.0406	0.8308	9.1455×10^{-6}
EGARCH-sstd	8	0.4739	0.9284	7	1.0628	0.8198	9.1654×10^{-6}
EGARCH-ged	1	-1.2221	1.0000	2	0.7505	0.9276	8.9874×10^{-6}
GJRGARCH-norm	6	0.1380	0.9990	4	0.9655	0.8622	9.1170×10^{-6}
GJRGARCH-std	4	-0.8253	1.0000	8	1.1323	0.7792	8.7657×10^{-6}
GJRGARCH-sstd	5	-0.8011	1.0000	10	1.1357	0.7776	8.7956×10^{-6}
GJRGARCH-ged	3	-0.9141	1.0000	1	-0.7505	1.0000	8.6033×10^{-6}

Table 5.21: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Nedbank (H=1200)							
GARCH-norm	11	0.9061	0.7614	6	0.9909	0.8614	8.6135×10^{-6}
GARCH-std	9	0.6053	0.8978	11	1.1686	0.7798	8.3360×10^{-6}
GARCH-sstd	10	0.6612	0.8784	9	1.1014	0.8136	8.3509×10^{-6}
GARCH-ged	3	-0.8858	1.0000	3	0.7319	0.9324	8.1708×10^{-6}
EGARCH-norm	12	1.0658	0.6490	12	1.1749	0.7754	8.7631×10^{-6}
EGARCH-std	7	0.1606	0.9958	7	1.0272	0.8496	8.2777×10^{-6}
EGARCH-sstd	8	0.2629	0.9854	5	0.9824	0.8652	8.2891×10^{-6}
EGARCH-ged	1	-1.1350	1.0000	2	0.7138	0.9370	8.1507×10^{-6}
GJRGARCH-norm	6	0.1100	0.9992	4	0.9066	0.8888	8.2777×10^{-6}
GJRGARCH-std	5	-0.7905	1.0000	10	1.1506	0.7902	8.0184×10^{-6}
GJRGARCH-sstd	4	-0.8022	1.0000	8	1.0624	0.8316	8.0268×10^{-6}
GJRGARCH-ged	2	-0.8871	1.0000	1	-0.7138	1.0000	7.9060×10^{-6}
Nedbank (H=1400)							
GARCH-norm	11	1.0827	0.6270	10	1.2296	0.7396	8.2425×10^{-6}
GARCH-std	8	0.4134	0.9550	11	1.2669	0.7158	7.9062×10^{-6}
GARCH-sstd	9	0.5565	0.9160	9	1.2153	0.7480	7.9261×10^{-6}
GARCH-ged	2	-1.0143	1.0000	2	0.7643	0.9298	7.7781×10^{-6}
EGARCH-norm	12	1.1574	0.5506	12	1.3245	0.6738	8.3640×10^{-6}
EGARCH-std	6	0.0913	0.9998	5	0.9503	0.8750	7.8751×10^{-6}
EGARCH-sstd	7	0.1902	0.9936	4	0.9039	0.8926	7.8856×10^{-6}
EGARCH-ged	1	-1.1299	1.0000	3	0.7847	0.9258	7.7798×10^{-6}
GJRGARCH-norm	10	0.7293	0.8514	8	1.1055	0.8150	7.9444×10^{-6}
GJRGARCH-std	5	-0.9025	1.0000	7	1.0852	0.8254	7.5945×10^{-6}
GJRGARCH-sstd	4	-0.9148	1.0000	6	1.0088	0.8590	7.6029×10^{-6}
GJRGARCH-ged	3	-0.9490	1.0000	1	-0.7643	1.0000	7.5130×10^{-6}

Table 5.22: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Investec (H=1000)							
GARCH-std	9	0.7453	0.8494	6	1.2021	0.7924	9.5169×10^{-6}
GARCH-sstd	8	0.5549	0.9404	4	1.0479	0.8734	9.4498×10^{-6}
GARCH-ged	5	-0.1213	1.0000	3	0.8818	0.9348	9.2975×10^{-6}
EGARCH-std	7	0.1367	1.0000	7	1.2576	0.7578	9.3246×10^{-6}
EGARCH-sstd	3	-0.8463	1.0000	5	1.1111	0.8434	9.2431×10^{-6}
EGARCH-ged	1	-1.4182	1.0000	2	0.1520	1.0000	9.0247×10^{-6}
GJRGARCH-norm	10	1.5822	0.2786	10	2.1742	0.1920	9.7109×10^{-6}
GJRGARCH-std	6	0.0847	1.0000	9	1.7724	0.3938	9.3184×10^{-6}
GJRGARCH-sstd	4	-0.7816	1.0000	8	1.4720	0.6098	9.2383×10^{-6}
GJRGARCH-ged	2	-1.1202	1.0000	1	-0.1520	1.0000	9.0051×10^{-6}
Investec (H=1200)							
GARCH-norm	11	1.6719	0.2678	10	1.9266	0.3460	9.1845×10^{-6}
GARCH-std	9	0.5670	0.9526	7	1.2966	0.7882	8.8021×10^{-6}
GARCH-sstd	8	0.2144	0.9986	5	0.9873	0.9242	8.7199×10^{-6}
GARCH-ged	7	-0.4320	1.0000	4	0.9188	0.9430	8.6205×10^{-6}
EGARCH-norm	12	1.7679	0.2200	12	2.1387	0.2340	9.0393×10^{-6}
EGARCH-std	5	-1.2194	1.0000	9	1.9199	0.3502	8.5878×10^{-6}
EGARCH-sstd	1	-1.9684	1.0000	3	0.6329	0.9884	8.4596×10^{-6}
EGARCH-ged	2	-1.5911	1.0000	2	0.3639	0.9994	8.3752×10^{-6}
GJRGARCH-norm	10	0.9874	0.7476	11	1.9913	0.3086	8.8602×10^{-6}
GJRGARCH-std	6	-0.7075	1.0000	8	1.7645	0.4452	8.6099×10^{-6}
GJRGARCH-sstd	3	-1.2809	1.0000	6	1.0401	0.9062	8.4877×10^{-6}
GJRGARCH-ged	4	-1.2788	1.0000	1	-0.3639	1.0000	8.3295×10^{-6}

Table 5.23: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Investec (H=1400)							
GARCH-norm	11	1.6378	0.2826	9	1.9440	0.3220	8.5145×10^{-6}
GARCH-std	9	0.4891	0.9642	7	1.4555	0.6550	8.1827×10^{-6}
GARCH-sstd	8	0.0516	1.0000	6	0.8974	0.9402	8.1020×10^{-6}
GARCH-ged	6	-0.6775	1.0000	5	0.7793	0.9664	8.0245×10^{-6}
EGARCH-norm	12	1.7679	0.2208	12	2.0739	0.2576	8.4064×10^{-6}
EGARCH-std	7	-0.5614	1.0000	11	1.9902	0.2964	8.0575×10^{-6}
EGARCH-sstd	1	-1.8710	1.0000	3	0.3066	0.9998	7.9083×10^{-6}
EGARCH-ged	2	-1.5568	1.0000	2	0.2521	1.0000	7.8848×10^{-6}
GJRGARCH-norm	10	1.1695	0.6032	10	1.9457	0.3210	8.2732×10^{-6}
GJRGARCH-std	5	-0.9611	1.0000	8	1.6501	0.5236	8.0157×10^{-6}
GJRGARCH-sstd	3	-1.3842	1.0000	4	0.4204	0.9986	7.8949×10^{-6}
GJRGARCH-ged	4	-1.1418	1.0000	1	-0.2521	1.0000	7.8559×10^{-6}

Table 5.24: Summary of SSM composition for different forecasting periods

	1000	1200	1400		1000	1200	1400
Models	Distributions						
		JSE				JSE	
GARCH	4	4	4	norm	3	3	3
EGARCH	4	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
		FNB				FNB	
GARCH	4	4	4	norm	3	3	3
EGARCH	4	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
		Standard bank				Standard bank	
GARCH	4	4	4	norm	3	3	3
EGARCH	4	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
		Absa				Absa	
GARCH	4	4	4	norm	3	3	3
EGARCH	4	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
		Capitec				Capitec	
GARCH	3	3	3	norm	3	3	3
EGARCH	3	3	3	std	3	3	3
GJRGARCH	3	3	3	sstd	3	3	3
				ged			
		Nedbank				Nedbank	
GARCH	4	4	4	norm	3	3	3
EGARCH	4	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
		Investec bank				Investec bank	
GARCH	3	4	4	norm	1	3	3
EGARCH	3	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3

5.3.3 Evaluation of VaR prediction performance under different model re-estimation strategies

We estimated the performance of the models under different model re-estimation strategies. The different re-estimation strategies include 30, 60, and 90 days with a forecast length of 1000 at a 1% quantile level. Tables 5.25, 5.26, 5.27, 5.28, 5.29, 5.30, 5.31, 5.32, 5.33, 5.34 and 5.35 report the MCS approach results and Table 5.36 provide a summary of final SSM compositions under different model re-estimation strategies. All the models survived regardless of the re-estimation strategy adopted for JSE, FNB, Standard Bank, Absa, Capitec and Nedbank stock returns. The GJRARCH-ged is the optimal model for FNB, Standard, and Nedbank stock returns, while EGARCH-ged is the optimal model for Absa, and EGARCH-sstd is the optimal model for Capitec stock returns. Two models (GARCH-norm, EGARCH-norm) were eliminated among the set of GARCH models for the Investec stock returns, with the most optimal model being GJRARCH-ged. From the perspective of the GARCH models, almost all models based on the different distributions survived. This shows that GARCH, EGARCH and GJRARCH models have excellent out-of-sample VaR predictive ability. In addition, when the model is re-estimated every 60 and 90 days, the standard GARCH, EGARCH and GJRARCH models based on almost all distributions survived, and the difference between the loss values of the models was not large. Thus, these models can be considered equivalent in terms of their VaR forecast abilities. According to the MCS approach results in the Table 5.36, we further discuss the forecast ability of the GARCH family models in SSM. Regardless of which model re-estimation strategy is chosen, the number and types of models that survive are the same except for Investec, and the VaR loss values of the GARCH family models are not significantly different. Therefore, the model re-estimation strategy has little impact on the forecast ability of the GARCH family models. It is observed that the p-values of the test statistics of GJRARCH-ged, EGARCH-ged and EGARCH-sstd models are all 1.0000, and the loss value is the least for this model. We can conclude that the out-of-sample VaR prediction performance of the GJRARCH-ged model is relatively optimal for JSE, FNB, Standard Bank, Nedbank and Investec. In contrast, the optimal model for Absa is EGARCH-ged and that for Capitec is EGARCH-ssts. When the model was re-estimated every 60 days, the optimal models had a p-value of 1.0000 for both test statistics.

Table 5.25: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
JSE (Refit.every=30)							
GARCH-norm	11	0.7273	0.8716	12	1.0833	0.8646	7.3525×10^{-6}
GARCH-std	7	-0.2562	1.0000	7	0.7750	0.9656	7.2007×10^{-6}
GARCH-sstd	6	-0.3283	1.0000	3	0.5093	0.9964	7.1925×10^{-6}
GARCH-ged	3	-0.6469	1.0000	2	0.3821	0.9998	7.1633×10^{-6}
EGARCH-norm	12	1.0084	0.7100	11	1.0730	0.8684	7.4139×10^{-6}
EGARCH-std	9	0.5577	0.9408	10	1.0301	0.8864	7.2662×10^{-6}
EGARCH-sstd	8	0.3128	0.9922	5	0.7034	0.9798	7.2459×10^{-6}
EGARCH-ged	5	-0.3904	1.0000	4	0.6493	0.9846	7.1943×10^{-6}
GJRGARCH-norm	10	0.6369	0.9126	9	0.9015	0.9342	7.2765×10^{-6}
GJRGARCH-std	4	-0.6135	1.0000	8	0.8271	0.9558	7.1544×10^{-6}
GJRGARCH-sstd	2	-0.6609	1.0000	6	0.7681	0.9668	7.1456×10^{-6}
GJRGARCH-ged	1	-0.7170	1.0000	1	-0.3821	1.0000	7.0862×10^{-6}
JSE (Refit.every=60)							
GARCH-norm	11	0.7349	0.8692	12	1.1048	0.8572	7.3594×10^{-6}
GARCH-std	7	-0.3167	1.0000	7	0.7895	0.9636	7.2041×10^{-6}
GARCH-sstd	5	-0.3876	1.0000	3	0.4820	0.9978	7.1958×10^{-6}
GARCH-ged	2	-0.6777	1.0000	2	0.3639	1.0000	7.1684×10^{-6}
EGARCH-norm	12	1.0054	0.7146	10	1.0836	0.8676	7.4229×10^{-6}
EGARCH-std	9	0.5574	0.9420	11	1.0841	0.8676	7.2749×10^{-6}
EGARCH-sstd	8	0.3073	0.9932	5	0.6764	0.9838	7.2542×10^{-6}
EGARCH-ged	6	-0.3562	1.0000	4	0.6478	0.9854	7.2060×10^{-6}
GJRGARCH-norm	10	0.6664	0.8996	9	0.9078	0.9340	7.2872×10^{-6}
GJRGARCH-std	4	-0.6020	1.0000	8	0.8241	0.9574	7.1646×10^{-6}
GJRGARCH-sstd	3	-0.6499	1.0000	6	0.7617	0.9692	7.1570×10^{-6}
GJRGARCH-ged	1	-0.7099	1.0000	1	-0.3639	1.0000	7.0967×10^{-6}

Table 5.26: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
JSE (Refit.every=90)							
GARCH-norm	11	0.7201	0.8748	12	1.0690	0.8770	7.3613×10^{-6}
GARCH-std	7	-0.3591	1.0000	5	0.6436	0.9860	7.2104×10^{-6}
GARCH-sstd	6	-0.3866	1.0000	3	0.5795	0.9914	7.2064×10^{-6}
GARCH-ged	1	-0.7289	1.0000	2	0.3395	0.9998	7.1723×10^{-6}
EGARCH-norm	12	0.9867	0.7304	11	1.0604	0.8808	7.4261×10^{-6}
EGARCH-std	9	0.5300	0.9512	9	0.8966	0.9402	7.2814×10^{-6}
EGARCH-sstd	8	0.3450	0.9884	6	0.7368	0.9744	7.2659×10^{-6}
EGARCH-ged	5	-0.4084	1.0000	4	0.6209	0.9878	7.2128×10^{-6}
GJRGARCH-norm	10	0.6634	0.9026	10	0.9046	0.9388	7.2962×10^{-6}
GJRGARCH-std	4	-0.5613	1.0000	8	0.8593	0.9510	7.1804×10^{-6}
GJRGARCH-sstd	3	-0.5707	1.0000	7	0.8564	0.9522	7.1779×10^{-6}
GJRGARCH-ged	2	-0.7078	1.0000	1	-0.3395	1.0000	7.1078×10^{-6}
FNB (Refit.every=30)							
GARCH-norm	10	0.310	0.7436	9	1.0321	0.8786	7.7575×10^{-6}
GARCH-std	9	0.5178	0.9456	6	0.8692	0.9360	7.6030×10^{-6}
GARCH-sstd	8	0.4485	0.9668	7	0.9109	0.9242	7.5937×10^{-6}
GARCH-ged	7	-0.2288	1.0000	5	0.7829	0.9592	7.5349×10^{-6}
EGARCH-norm	11	1.3127	0.4546	12	1.5311	0.5726	7.7523×10^{-6}
EGARCH-std	6	-0.5637	1.0000	10	1.1724	0.8052	7.5182×10^{-6}
EGARCH-sstd	1	-1.0441	1.0000	4	0.7491	0.9646	7.4835×10^{-6}
EGARCH-ged	2	-1.0271	1.0000	2	0.4888	0.9960	7.4418×10^{-6}
GJRGARCH-norm	12	1.3315	0.4416	11	1.4500	0.6262	7.6611×10^{-6}
GJRGARCH-std	4	-0.9644	1.0000	8	0.9337	0.9176	7.4508×10^{-6}
GJRGARCH-sstd	3	-1.0161	1.0000	3	0.7420	0.9660	7.4285×10^{-6}
GJRGARCH-ged	5	-0.9405	1.0000	1	-0.4888	1.0000	7.3825×10^{-6}

Table 5.27: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
FNB (Refit.every=60)							
GARCH-norm	10	0.9343	0.7450	9	1.0393	0.8788	7.7594×10^{-6}
GARCH-std	9	0.5100	0.9494	6	0.8668	0.9394	7.6023×10^{-6}
GARCH-sstd	8	0.4490	0.9672	8	0.9018	0.9306	7.5946×10^{-6}
GARCH-ged	7	-0.2730	1.0000	4	0.7761	0.9620	7.5324×10^{-6}
EGARCH-norm	11	1.3088	0.4614	12	1.5370	0.5740	7.7524×10^{-6}
EGARCH-std	6	-0.5946	1.0000	10	1.1165	0.8430	7.5166×10^{-6}
EGARCH-sstd	2	-1.0386	1.0000	5	0.7840	0.9596	7.4854×10^{-6}
EGARCH-ged	1	-1.0436	1.0000	2	0.4561	0.9978	7.4368×10^{-6}
GJRGARCH-norm	12	1.3439	0.4338	11	1.4344	0.6440	7.6633×10^{-6}
GJRGARCH-std	4	-0.9479	1.0000	7	0.8934	0.9328	7.4535×10^{-6}
GJRGARCH-sstd	3	-1.0002	1.0000	3	0.7657	0.9644	7.4335×10^{-6}
GJRGARCH-ged	5	-0.9327	1.0000	1	-0.4561	1.0000	7.3819×10^{-6}
FNB (Refit.every=90)							
GARCH-norm	10	0.9283	0.7502	9	1.0400	0.8812	7.7564×10^{-6}
GARCH-std	9	0.4779	0.9616	7	0.8896	0.9364	7.6023×10^{-6}
GARCH-sstd	8	0.4185	0.9750	8	0.9292	0.9242	7.5953×10^{-6}
GARCH-ged	7	-0.3595	1.0000	3	0.7366	0.9702	7.5313×10^{-6}
EGARCH-norm	11	1.3124	0.4600	12	1.5499	0.5688	7.7595×10^{-6}
EGARCH-std	6	-0.5911	1.0000	10	1.0482	0.8792	7.5219×10^{-6}
EGARCH-sstd	2	-1.0042	1.0000	5	0.7868	0.9612	7.4932×10^{-6}
EGARCH-ged	1	-1.0410	1.0000	2	0.4370	0.9986	7.4466×10^{-6}
GJRGARCH-norm	12	1.3290	0.4492	11	1.4382	0.6430	7.6656×10^{-6}
GJRGARCH-std	4	-0.9396	1.0000	6	0.8326	0.9512	7.4613×10^{-6}
GJRGARCH-sstd	3	-0.9768	1.0000	4	0.7602	0.9658	7.4436×10^{-6}
GJRGARCH-ged	5	-0.9161	1.0000	1	-0.4370	1.0000	7.3938×10^{-6}

Table 5.28: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Standard (Refit.every=30)							
GARCH-norm	11	1.0613	0.5802	12	1.2546	0.7154	8.0447×10^{-6}
GARCH-std	4	-1.0273	1.0000	6	1.0427	0.8360	7.7315×10^{-6}
GARCH-sstd	2	-1.0822	1.0000	4	0.9135	0.8886	7.7379×10^{-6}
GARCH-ged	1	-1.1412	1.0000	2	0.6594	0.9652	7.6374×10^{-6}
EGARCH-norm	9	1.0383	0.6010	8	1.1080	0.8074	8.3069×10^{-6}
EGARCH-std	12	1.1066	0.5418	11	1.2280	0.7306	8.0459×10^{-6}
EGARCH-sstd	10	1.0406	0.5990	9	1.1791	0.7616	8.0420×10^{-6}
EGARCH-ged	8	1.0301	0.6092	10	1.1860	0.7570	7.9483×10^{-6}
GJRGARCH-norm	7	-0.2391	1.0000	7	1.0815	0.8210	7.7985×10^{-6}
GJRGARCH-std	6	-0.9928	1.0000	5	1.0240	0.8434	7.5843×10^{-6}
GJRGARCH-sstd	3	-1.0436	1.0000	3	0.8666	0.9050	7.5806×10^{-6}
GJRGARCH-ged	5	-1.0122	1.0000	1	-0.6594	1.0000	7.5074×10^{-6}
Standard (Refit.every=60)							
GARCH-norm	11	1.0766	0.5662	12	1.2611	0.6906	8.0758×10^{-6}
GARCH-std	3	-1.0603	1.0000	6	1.0268	0.8228	7.7560×10^{-6}
GARCH-sstd	2	-1.0904	1.0000	4	0.9120	0.8740	7.7681×10^{-6}
GARCH-ged	1	-1.1357	1.0000	2	0.6954	0.9452	7.6694×10^{-6}
EGARCH-norm	9	1.0398	0.5962	7	1.1060	0.7900	8.3545×10^{-6}
EGARCH-std	12	1.0942	0.5490	11	1.2138	0.7220	8.0741×10^{-6}
EGARCH-sstd	10	1.0405	0.5960	9	1.1729	0.7478	8.0806×10^{-6}
EGARCH-ged	8	1.0345	0.6006	10	1.1858	0.7414	7.9885×10^{-6}
GJRGARCH-norm	7	-0.1681	1.0000	8	1.1140	0.7872	7.8401×10^{-6}
GJRGARCH-std	6	-1.0049	1.0000	5	1.0084	0.8296	7.5991×10^{-6}
GJRGARCH-sstd	4	-1.0486	1.0000	3	0.8693	0.8872	7.6017×10^{-6}
GJRGARCH-ged	5	-1.0184	1.0000	1	-0.6954	1.0000	7.5284×10^{-6}

Table 5.29: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Standard (Refit.every=90)							
GARCH-norm	11	1.0921	0.5592	12	1.2661	0.6920	8.0758×10^{-6}
GARCH-std	3	-1.0847	1.0000	6	1.0439	0.8252	7.7609×10^{-6}
GARCH-sstd	2	-1.1317	1.0000	4	0.9410	0.8710	7.7695×10^{-6}
GARCH-ged	1	-1.1520	1.0000	2	0.6893	0.9482	7.6745×10^{-6}
EGARCH-norm	9	1.0449	0.5994	8	1.1076	0.7956	8.3598×10^{-6}
EGARCH-std	12	1.0938	0.5582	11	1.2081	0.7342	8.0856×10^{-6}
EGARCH-sstd	10	1.0513	0.5936	9	1.1763	0.7540	8.0908×10^{-6}
EGARCH-ged	8	1.0268	0.6136	10	1.1817	0.7512	8.0002×10^{-6}
GJRGARCH-norm	7	-0.1529	1.0000	7	1.1039	0.7976	7.8503×10^{-6}
GJRGARCH-std	6	-1.0123	1.0000	5	1.0216	0.8378	7.6106×10^{-6}
GJRGARCH-sstd	4	-1.0447	1.0000	3	0.9089	0.8814	7.6128×10^{-6}
GJRGARCH-ged	5	-1.0286	1.0000	1	-0.6893	1.0000	7.5411×10^{-6}
Absa (Refit.every=30)							
GARCH-norm	10	1.1939	0.5222	5	1.4460	0.5858	9.1935×10^{-6}
GARCH-std	11	1.2327	0.4890	11	1.8524	0.3110	8.8911×10^{-6}
GARCH-sstd	12	1.3128	0.4284	10	1.7962	0.3434	8.9125×10^{-6}
GARCH-ged	7	0.0072	1.0000	12	1.8702	0.3014	8.7978×10^{-6}
EGARCH-norm	8	0.7280	0.8500	3	1.1597	0.7904	9.0310×10^{-6}
EGARCH-std	3	-1.8478	1.0000	7	1.5736	0.4926	8.6287×10^{-6}
EGARCH-sstd	1	-1.9876	1.0000	4	1.3108	0.6886	8.6455×10^{-6}
EGARCH-ged	2	-1.8493	1.0000	1	-0.3961	1.0000	8.5128×10^{-6}
GJRGARCH-norm	9	1.1712	0.5410	8	1.6356	0.4464	8.9501×10^{-6}
GJRGARCH-std	5	-0.5648	1.0000	6	1.5206	0.5322	8.7016×10^{-6}
GJRGARCH-sstd	6	-0.5317	1.0000	9	1.7670	0.3610	8.7234×10^{-6}
GJRGARCH-ged	3	-0.8557	1.0000	2	0.3961	0.9962	8.5780×10^{-6}

Table 5.30: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Absa (Refit.every=60)							
GARCH-norm	10	1.2072	0.5098	5	1.4829	0.5542	9.2031×10^{-6}
GARCH-std	9	1.1501	0.5568	11	1.9410	0.2702	8.8921×10^{-6}
GARCH-sstd	12	1.3460	0.4082	10	1.8630	0.3062	8.9171×10^{-6}
GARCH-ged	7	-0.1417	1.0000	12	1.9471	0.2672	8.7991×10^{-6}
EGARCH-norm	8	0.7545	0.8300	3	1.1682	0.7800	9.0773×10^{-6}
EGARCH-std	2	-1.9076	1.0000	7	1.5828	0.4778	8.6338×10^{-6}
EGARCH-sstd	1	-2.0290	1.0000	4	1.3054	0.6870	8.6558×10^{-6}
EGARCH-ged	3	-1.8896	1.0000	1	-0.4828	1.0000	8.5196×10^{-6}
GJRGARCH-norm	11	1.2197	0.5004	8	1.6756	0.4156	8.9859×10^{-6}
GJRGARCH-std	5	-0.5548	1.0000	6	1.5576	0.4986	8.7175×10^{-6}
GJRGARCH-sstd	6	-0.5055	1.0000	9	1.8442	0.3162	8.7434×10^{-6}
GJRGARCH-ged	4	-0.8430	1.0000	2	0.4828	0.9888	8.5986×10^{-6}
Absa (Refit.every=90)							
GARCH-norm	10	1.1886	0.5212	5	1.4615	0.5694	9.1969×10^{-6}
GARCH-std	9	1.0823	0.6064	12	1.9563	0.2596	8.8908×10^{-6}
GARCH-sstd	12	1.3617	0.3996	10	1.8716	0.3018	8.9172×10^{-6}
GARCH-ged	7	-0.1998	1.0000	11	1.9499	0.2630	8.7986×10^{-6}
EGARCH-norm	8	0.7917	0.8066	3	1.1877	0.7654	9.1031×10^{-6}
EGARCH-std	2	-1.9190	1.0000	7	1.5885	0.4740	8.6400×10^{-6}
EGARCH-sstd	1	-2.0129	1.0000	4	1.3430	0.6560	8.6633×10^{-6}
EGARCH-ged	3	-1.8949	1.0000	1	-0.4934	1.0000	8.5272×10^{-6}
GJRGARCH-norm	11	1.2100	0.5050	8	1.6616	0.4208	8.9895×10^{-6}
GJRGARCH-std	5	-0.5635	1.0000	6	1.5390	0.5116	8.7229×10^{-6}
GJRGARCH-sstd	6	-0.5028	1.0000	9	1.8220	0.3298	8.7494×10^{-6}
GJRGARCH-ged	4	-0.8506	1.0000	2	0.4934	0.9892	8.6066×10^{-6}

Table 5.31: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Capitec (Refit.every=30)							
GARCH-norm	9	0.6357	0.7602	8	1.9806	0.2058	1.0439×10^{-5}
GARCH-std	7	0.4194	0.9094	6	1.4595	0.5032	1.0248×10^{-5}
GARCH-sstd	4	-0.2173	1.0000	3	0.9302	0.8600	1.0177×10^{-5}
EGARCH-norm	2	-0.5333	1.0000	2	0.1482	1.0000	1.00602×10^{-5}
EGARCH-std	3	-0.4845	1.0000	4	0.9545	0.8468	1.00603×10^{-5}
EGARCH-sstd	1	-0.9109	1.0000	1	-0.1482	1.0000	9.9949×10^{-6}
GJRGARCH-norm	8	0.4511	0.8880	9	2.2423	0.1306	1.0323×10^{-5}
GJRGARCH-std	6	0.3658	0.9350	7	1.7276	0.3296	1.0257×10^{-5}
GJRGARCH-sstd	5	-0.0383	1.0000	5	1.1823	0.7076	1.0190×10^{-5}
capitec (Refit.every=60)							
GARCH-norm	9	0.6221	0.8036	8	1.8437	0.2628	1.0448×10^{-5}
GARCH-std	8	0.5438	0.8564	6	1.4558	0.5082	1.0279×10^{-5}
GARCH-sstd	5	0.3185	0.9672	4	0.9880	0.8316	1.0236×10^{-5}
EGARCH-norm	2	-0.8087	1.0000	2	0.0694	1.0000	1.0042×10^{-5}
EGARCH-std	3	-0.4920	1.0000	3	0.9002	0.8764	1.0078×10^{-5}
EGARCH-sstd	1	-0.9170	1.0000	1	-0.0694	1.0000	1.0014×10^{-5}
GJRGARCH-norm	7	0.4614	0.9060	9	2.3365	0.1040	1.0341×10^{-5}
GJRGARCH-std	6	0.4496	0.9114	7	1.7915	0.2892	1.0287×10^{-5}
GJRGARCH-sstd	4	0.0006	1.0000	5	1.1821	0.7080	1.0216×10^{-5}

Table 5.32: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Capitec (Refit.every=90)							
GARCH-norm	9	0.6066	0.7834	9	1.9782	0.2226	1.0427×10^{-5}
GARCH-std	8	0.4485	0.8898	6	1.4668	0.5224	1.0261×10^{-5}
GARCH-sstd	5	0.0169	1.0000	4	0.9854	0.8516	1.0207×10^{-5}
EGARCH-norm	3	-0.2364	1.0000	2	0.2797	0.9998	1.0138×10^{-5}
EGARCH-std	2	-0.4887	1.0000	3	0.9089	0.8866	1.0066×10^{-5}
EGARCH-sstd	1	-0.9054	1.0000	1	-0.2797	1.0000	1.0001×10^{-5}
GJRGARCH-norm	7	0.3519	0.9442	7	1.4863	0.5106	1.0301×10^{-5}
GJRGARCH-std	6	0.3010	0.9620	8	1.7187	0.3546	1.0258×10^{-5}
GJRGARCH-sstd	4	-0.1090	1.0000	5	1.1571	0.7474	1.0193×10^{-5}
Nedbank (Refit.every=30)							
GARCH-norm	11	0.9431	0.7316	6	1.0483	0.8270	9.5237×10^{-6}
GARCH-std	9	0.5810	0.8964	11	1.1406	0.7746	9.1823×10^{-6}
GARCH-sstd	10	0.7046	0.8510	9	1.1346	0.7778	9.2109×10^{-6}
GARCH-ged	2	-0.9994	1.0000	3	0.7627	0.9242	8.9629×10^{-6}
EGARCH-norm	12	1.0837	0.6078	12	1.2072	0.7262	9.7352×10^{-6}
EGARCH-std	7	0.3724	0.9574	5	1.0406	0.8308	9.1455×10^{-6}
EGARCH-sstd	8	0.4739	0.9284	7	1.0628	0.8198	9.1654×10^{-6}
EGARCH-ged	1	-1.2221	1.0000	2	0.7505	0.9276	8.9874×10^{-6}
GJRGARCH-norm	6	0.1380	0.9990	4	0.9655	0.8622	9.1170×10^{-6}
GJRGARCH-std	4	-0.8253	1.0000	8	1.1323	0.7792	8.7657×10^{-6}
GJRGARCH-sstd	5	-0.8011	1.0000	10	1.1357	0.7776	8.7956×10^{-6}
GJRGARCH-ged	3	-0.9141	1.0000	1	-0.7505	1.0000	8.6033×10^{-6}

Table 5.33: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Nedbank (Refit.every=60)							
GARCH-norm	11	0.9421	0.7304	6	1.0546	0.8236	9.5300×10^{-6}
GARCH-std	9	0.5795	0.8968	11	1.1567	0.7588	9.1865×10^{-6}
GARCH-sstd	10	0.7111	0.8486	10	1.1475	0.7660	9.2175×10^{-6}
GARCH-ged	2	-1.0036	1.0000	3	0.7551	0.9272	8.9650×10^{-6}
EGARCH-norm	12	1.0844	0.6082	12	1.2170	0.7172	9.7425×10^{-6}
EGARCH-std	7	0.3139	0.9724	5	1.0412	0.8296	9.1415×10^{-6}
EGARCH-sstd	8	0.4466	0.9354	7	1.0595	0.8224	9.1664×10^{-6}
EGARCH-ged	1	-1.2633	1.0000	2	0.7375	0.9298	8.9643×10^{-6}
GJRGARCH-norm	6	0.1967	0.9948	4	0.9737	0.8574	9.1299×10^{-6}
GJRGARCH-std	4	-0.8201	1.0000	8	1.1425	0.7688	8.7707×10^{-6}
GJRGARCH-sstd	5	-0.7906	1.0000	9	1.1436	0.7686	8.8075×10^{-6}
GJRGARCH-ged	3	-0.9121	1.0000	1	-0.7375	1.0000	8.6086×10^{-6}
Nedbank (Refit.every=90)							
GARCH-norm	11	0.9536	0.7228	7	1.0847	0.8072	9.5640×10^{-6}
GARCH-std	9	0.6487	0.8754	11	1.1612	0.7600	9.2311×10^{-6}
GARCH-sstd	10	0.7724	0.8276	10	1.1502	0.7692	9.2667×10^{-6}
GARCH-ged	3	-0.9050	1.0000	3	0.7767	0.9248	9.0178×10^{-6}
EGARCH-norm	12	1.0809	0.6116	12	1.2339	0.7048	9.7494×10^{-6}
EGARCH-std	6	0.1355	0.9990	5	1.0518	0.8252	9.1473×10^{-6}
EGARCH-sstd	8	0.3510	0.9652	6	1.0742	0.8128	9.1801×10^{-6}
EGARCH-ged	1	-1.3717	1.0000	2	0.7107	0.9360	8.9733×10^{-6}
GJRGARCH-norm	7	0.2538	0.9848	4	1.0118	0.8408	9.1672×10^{-6}
GJRGARCH-std	4	-0.8154	1.0000	9	1.1424	0.7734	8.8068×10^{-6}
GJRGARCH-sstd	5	-0.7822	1.0000	8	1.1398	0.7748	8.8484×10^{-6}
GJRGARCH-ged	2	-0.9121	1.0000	1	-0.7107	1.0000	8.6323×10^{-6}

Table 5.34: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Investec (Refit.every=30)							
GARCH-std	9	0.7453	0.8494	6	1.2021	0.7924	9.5169×10^{-6}
GARCH-sstd	8	0.5549	0.9404	4	1.0479	0.8734	9.4498×10^{-6}
GARCH-ged	5	-0.1213	1.0000	3	0.8818	0.9348	9.2975×10^{-6}
EGARCH-std	7	0.1367	1.0000	7	1.2576	0.7578	9.3246×10^{-6}
EGARCH-sstd	3	-0.8463	1.0000	5	1.1111	0.8434	9.2431×10^{-6}
EGARCH-ged	1	-1.4182	1.0000	2	0.1520	1.0000	9.0247×10^{-6}
GJRGARCH-norm	10	1.5822	0.2786	10	2.1742	0.1920	9.7109×10^{-6}
GJRGARCH-std	6	0.0847	1.0000	9	1.7724	0.3938	9.3184×10^{-6}
GJRGARCH-sstd	4	-0.7816	1.0000	8	1.4720	0.6098	9.2383×10^{-6}
GJRGARCH-ged	2	-1.1202	1.0000	1	-0.1520	1.0000	9.0051×10^{-6}
Investec (Refit.every=60)							
GARCH-norm							$\times 10^{-6}$
GARCH-std	8	0.6707	0.9072	6	1.1362	0.8620	9.4326×10^{-6}
GARCH-sstd	9	0.6709	0.9072	7	1.2098	0.8216	9.4248×10^{-6}
GARCH-ged	5	-0.2903	1.0000	3	0.7844	0.9716	9.2319×10^{-6}
EGARCH-std	7	0.3142	0.9950	8	1.3881	0.7036	9.2917×10^{-6}
EGARCH-sstd	4	-0.8057	1.0000	5	0.9894	0.9194	9.1984×10^{-6}
EGARCH-ged	1	-1.3310	1.0000	2	0.5878	0.9940	9.0557×10^{-6}
GJRGARCH-norm	10	1.6668	0.2556	10	2.1059	0.2372	9.6766×10^{-6}
GJRGARCH-std	6	-0.1100	1.0000	9	1.8417	0.3814	9.2572×10^{-6}
GJRGARCH-sstd	2	-1.2022	1.0000	4	0.9179	0.9398	9.1232×10^{-6}
GJRGARCH-ged	3	-1.1172	1.0000	1	-0.5878	1.0000	8.9567×10^{-6}

Table 5.35: Comparison of Value-at-Risk model using the MCS approach

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Investec (Refit.every=90)							
GARCH-std	9	0.8186	0.8378	7	1.2351	0.7932	9.4836×10^{-6}
GARCH-sstd	8	0.6179	0.9232	6	1.0805	0.8664	9.4127×10^{-6}
GARCH-ged	5	-0.1856	1.0000	3	0.8778	0.9348	9.2519×10^{-6}
EGARCH-std	7	0.4939	0.9616	8	1.6651	0.4930	9.3164×10^{-6}
EGARCH-sstd	4	-0.9281	1.0000	5	0.9662	0.9098	9.1988×10^{-6}
EGARCH-ged	2	-1.3048	1.0000	2	0.6015	0.9852	9.0431×10^{-6}
GJRGARCH-norm	10	1.6058	0.2948	10	2.1138	0.2272	9.6666×10^{-6}
GJRGARCH-std	6	0.0659	1.0000	9	2.0861	0.2420	9.2756×10^{-6}
GJRGARCH-sstd	1	-1.4292	1.0000	4	0.8783	0.9348	9.1150×10^{-6}
GJRGARCH-ged	3	-1.1896	1.0000	1	-0.6015	1.0000	8.9551×10^{-6}

Table 5.36: Summary of SSM composition for different model re-estimation strategies

	30	60	90		30	60	90
Models	Distributions						
	JSE				JSE		
GARCH	4	4	3	norm	3	3	3
EGARCH	4	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
	FNB				FNB		
GARCH	4	4	4	norm	3	3	3
EGARCH	4	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
	Standard bank				Standard bank		
GARCH	4	4	2	norm	3	3	3
EGARCH	4	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
	Absa				Absa		
GARCH	4	4	4	norm	3	3	3
EGARCH	4	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
	Capitec				Capitec		
GARCH	3	3	3	norm	3	3	3
EGARCH	3	3	3	std	3	3	3
GJRGARCH	3	3	3	sstd	3	3	3
				ged			
	Nedbank				Nedbank		
GARCH	4	4	4	norm	3	3	3
EGARCH	4	4	4	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3
	Investec bank				Investec bank		
GARCH	3	3	3	norm	1	1	1
EGARCH	3	3	3	std	3	3	3
GJRGARCH	4	4	4	sstd	3	3	3
				ged	3	3	3

5.3.4 GAS Empirical Results

To examine the performance of the GAS models using the different innovations used to forecast VaR, the complete data of daily returns is divided into two samples: an out-sample period from April 22, 2005, to April 15, 2015, for a total of 2567 observations, and a forecast or validation period, containing the remaining 1000 observations. A fixed rolling window approach is also implemented to produce 1-day ahead forecasts of VaRs at quantile levels 1% and 2.5%. Following the practice of Hansen, Lunde, and Nason (2011), the significance level of the MCS approach was at 0.2 and evaluated using 5000 bootstrap replications. This implies that if the corresponding p-value of the statistics $T_{R,M}$ or $T_{max,M}$ is less than 0.2, then these models will be removed from the SSM. We assessed the performance of the models using the asymmetric quantile loss function of González-Rivera, Lee, and Mishra (2004), which penalizes observations that are far away from the selected empirical quantile. The normal, skewed normal, Student's t, skewed Student's t, asymmetric Student-t, asymmetric student-t and asymmetric Laplace distribution are abbreviated as norm, snorm, std, sstd, ast, ast1, and ald in the Tables. Naturally, the higher the number of eliminated models, the higher the heterogeneity of the competing forecasts. On the contrary, if the final set of models contains a significant portion of the starting set, then the competing models are statistically equivalent in terms of their forecast ability of future VaR levels.

Tables 5.37, 5.38 and 5.39 present the SSM from Hansen, Lunde, and Nason (2011)'s MCS for both 1% and 2.5% VaR forecasts. In these tables, the MCS procedure of Hansen, Lunde, and Nason (2011) is applied to obtain the set of models with superior predictive ability in terms of the supplied VaR forecasts. After that, the selected models are ranked according to their $T_{R,M}$ and $T_{max,M}$ statistics. The p-values of the $T_{R,M}$ and $T_{max,M}$ statistics are reported in the third and sixth columns of Tables 5.37, 5.38 and 5.39, respectively. For the final column, it is, however, reserved for the expected loss given a VaR violation (i.e., the difference between realized and evaluated VaR measures). In this study, the SSM set is used to find the true model that provides more precise VaR forecasts. Regarding the rank of the accepted models for the JSE stock returns, the GAS-Ald is the best model to predict the 1%-VaR of the JSE daily returns. These results are related to the $T_{max,M}$ statistic and the expected loss. Whereas $T_{R,M}$ statistic considers the GAS-sstd followed by the GAS-std as the best models to predict 1%-VaR of JSE returns.

As Hansen, Lunde, and Nason (2011), if the final SSM contains a significant portion of the starting M^0 set, then the competing models are statistically equivalent in terms of their forecast ability of future VaR levels. Based on the Tmax and M statistic, we show that GARCH-Ald consistently outperformed all other competing models to forecast 1% VaR of daily returns for JSE, FNB, Standard Bank, Absa, Capitec, Nedbank, and Investec. Also, the $T_{R,M}$ statistic showed that GAS-Ald performed better than other

competing models for all daily returns except JSE (GAS-sstd) and Investec (GAS-norm) at 1% VaR forecast. The GAS-Ald model is the best to predict the 2.5% VaR of the JSE daily returns using the $T_{max, M}$ statistic and the expected loss. However, GAS-norm is the best model, followed by GAS-Ald, to predict 2.5% VaR when using the $T_{R, M}$ statistic. Based on the $T_{max, M}$ statistic, we show that GAS-Ald consistently outperformed all other competing models to forecast 2.5% VaR of daily returns for JSE, Standard Bank, Absa, Capitec, Nedbank, and Investec, except FNB, with GAS-norm. Also, the $T_{R, M}$ statistic showed that GAS-Ald performed better than other competing models for all daily returns except JSE (GAS-norm), FNB (GAS-norm), Nedbank (GAS-norm) and Investec (GAS-norm) at 2.5% VaR forecast.

For 1% VaR forecasts, we find JSE(0), FNB(2), Standard Bank (2), Absa (2), Capitec (2), Nedbank (0) and Investec (0) models are eliminated (number of eliminated models in bracket) out of 6. The results suggest that FNB, Standard Bank, Capitec, and Absa exhibited higher levels of heterogeneity than JSE, Nedbank, and Investec. The returns for JSE, Nedbank and Investec exhibited higher levels of homogeneity as no model was eliminated. Generally, the returns for all the banks exhibited fair levels of homogeneity. For the 5% VaR forecast models, the same can be said of the returns of each bank except Capitec, which eliminated three models. The homogeneity in the superior set of models (SSM) implies the models within the set exhibit similar predictive performance. This suggests the different models are robust to variations in the data and provide consistent results, indicating that any conclusions drawn are not dependent on any specific chosen model. The homogeneous models have similar predictive capabilities, which enhances the stability of the predictions by the different models and reduces the risk of significant variations when using other models from the SSM. The homogeneity in the models also indicates that each model captures the underlying data structure similarly and does not overfit the model. These results imply that the volatility of returns for each bank in South Africa may not be affected by some stylised facts, such as the leverage effect, or by complex nonlinear conditional volatility dynamics. However, it exhibits excess kurtosis and heavy-tailed distributions.

Table 5.37: SSM of univariate GAS (VaR, ES) model forecasts

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
JSE (0.01)							
norm	4	-2.0220	1.0000	4	0.7256	0.9758	-2.7080
snorm	7	1.2115	0.4488	5	1.4065	0.6296	-1.6462
std	2	-2.2068	1.0000	2	0.5732	0.9922	-2.7457
sstd	1	-2.2805	1.0000	3	0.6808	0.9812	-2.7424
Ast	6	0.4909	0.8980	7	2.4318	0.1056	-2.4013
Ast1	5	0.4909	0.8980	6	2.4318	0.6296	-2.4013
Ald	3	-2.1165	1.0000	1	-0.5732	1.0000	-2.7979
P-value	0.4488						
JSE (0.025)							
norm	1	-3.2028	1.0000	2	0.8433	0.9512	-2.9948
snorm	5	1.0103	0.5872	4	1.4163	0.6380	-2.4854
std	3	-3.1303	1.0000	3	0.9939	0.9012	-2.9984
sstd	4	-2.4234	1.0000	5	2.3093	0.1432	-2.9435
Ast	6	1.6873	0.2018	6	4.4547	0.0010	-2.6634
Ast1	7	1.6873	0.2018	7	4.4547	0.0010	-2.6634
Ald	2	-3.1495	1.0000	1	-0.8433	1.0000	-3.0410
P-value	0.2018						
FNB (0.01)							
norm	2	-1.3739	1.0000	2	0.5718	0.9722	-2.5851
snorm	5	0.9621	0.5684	3	1.1020	0.7494	-2.0640
std	3	-0.4070	1.0000	4	1.2656	0.6268	-2.5059
sstd	4	-0.1476	1.0000	5	1.5343	0.4334	-2.4685
Ald	1	-1.4059	1.0000	1	-0.5718	1.0000	-2.6418
P-value	0.5684						
Eliminated							
Ast							
Ast1							
FNB (0.025)							
norm	1	-1.7431	1.0000	1	-0.1020	1.0000	-2.8948
snorm	5	0.9187	0.5868	3	1.1151	0.7318	-2.6164
std	3	-0.4148	1.0000	4	1.6236	0.3772	-2.8311
sstd	4	0.3532	0.9320	5	1.9220	0.2316	-2.7842
Ald	2	-1.4019	1.0000	2	0.1020	1.0000	-2.8903
P-value	0.5868						
Eliminated							
Ast							
Ast1							
Standard (0.01)							
norm	3	-1.1602	1.000	3	1.0212	0.7814	-2.6278
snorm	5	1.3176	0.292	4	1.4046	0.5024	-1.8819
std	2	-1.2306	1.000	2	0.9028	0.8440	-2.6558
sstd	4	-0.7577	1.000	5	1.8264	0.2504	-2.5707
Ald	1	-1.5784	1.000	1	-0.9028	1.0000	-2.7363
P-value	0.292						
Eliminated							
Ast							
Ast1							

Table 5.38: SSM of univariate GAS (VaR, ES) model forecasts

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Standard (0.025)							
norm	2	-1.4530	1.0000	2	0.7132	0.9276	-2.8948
snorm	5	1.2948	0.3008	4	1.4150	0.5086	-2.5178
std	3	-1.1404	1.0000	3	1.1074	0.7344	-2.8851
sstd	4	-0.2776	1.0000	5	2.5664	0.0566	-2.8238
Ald	1	-1.7344	1.0000	1	-0.7132	1.0000	-2.9268
P-value	0.3008						
Eliminated							
Ast							
Ast1							
Absa (0.01)							
norm	2	-1.0671	1.0000	3	1.7063	0.3184	-2.3904
snorm	5	1.2458	0.3232	2	1.4009	0.5038	-1.4395
std	3	-1.0054	1.0000	4	1.8918	0.2228	-2.4185
sstd	4	-0.6493	1.0000	5	2.4463	0.0716	-2.3518
Ald	1	-1.8222	1.0000	1	-1.4009	1.0000	-2.6148
P-value	0.3232						
Eliminated							
Ast							
Ast1							
Absa (0.025)							
norm	2	-1.5411	1.0000	2	1.2691	0.6036	-2.7346
snorm	5	1.2095	0.3576	3	1.3836	0.5176	-2.2827
std	3	-0.9255	1.0000	4	2.1420	0.1474	-2.7152
sstd	4	-0.1446	1.0000	5	3.1557	0.0140	-2.6507
Ald	1	-1.8738	1.0000	1	-1.2691	1.0000	-2.8128
P-value	0.3576						
Eliminated							
Ast							
Ast1							
Capitec (0.01)							
norm	4	0.0726	0.9996	3	1.4103	0.4756	-2.2533
snorm	5	1.2569	0.3852	4	1.4215	0.4698	-1.7879
std	3	-0.8369	1.0000	2	1.1257	0.6950	-2.3698
sstd	2	-0.9266	1.0000	5	1.4553	0.4434	-2.3452
Ald	1	-1.5836	1.0000	1	-1.1257	1.0000	-2.5417
P-value	0.3852						
Eliminated							
Ast							
Ast1							
Capitec (0.025)							
norm	4	0.9261	0.6118	3	1.0846	0.6756	-2.7364
std	2	-0.5605	1.0000	2	0.4918	0.9628	-2.7997
sstd	3	0.7806	0.7110	4	1.2819	0.5478	-2.7551
Ald	1	-1.0469	1.0000	1	-0.4918	1.0000	-2.8332
P-value	0.6118						
Eliminated							
snorm							
Ast							
Ast1							

Table 5.39: SSM of univariate GAS (VaR, ES) model forecasts

Models	$Rank_{R,M}$	t_{ij}	$p_{R,M}$	$Rank_{max,M}$	t_i	$p_{max,M}$	Loss
Nedbank (0.01)							
norm	2	-2.6066	1.0000	3	0.9787	0.9018	-2.5065
snorm	7	1.4250	0.3172	4	1.6975	0.4244	-1.0623
std	3	-2.4348	1.0000	2	0.9636	0.9104	-2.4946
sstd	4	-1.6702	1.0000	5	1.9385	0.2778	-2.3668
Ast	6	0.5133	0.8858	7	2.8144	0.0412	-2.0619
Ast1	5	0.5133	0.0058	6	2.8144	0.2778	-2.0619
Ald	1	-2.6757	1.0000	1	-0.9636	1.0000	-2.6153
P-value	0.3172						
Nedbank (0.025)							
norm	1	-3.3122	1.0000	2	0.2291	1.0000	-2.8043
snorm	7	1.3738	0.3688	4	1.7049	0.4266	-2.1049
std	3	-2.9328	1.0000	3	0.2301	1.0000	-2.8015
sstd	4	-1.6360	1.0000	5	2.1065	0.2092	-2.7046
Ast	6	0.8499	0.7058	7	3.0273	0.0252	-2.5164
Ast1	5	0.8499	0.7058	6	3.0273	0.2092	-2.5164
Ald	2	-3.0582	1.0000	1	-0.2291	1.0000	-2.8163
P-value	0.3688						
Investec (0.01)							
norm	1	-3.1674	1.0000	3	1.7023	0.4212	-2.3240
snorm	6	0.9141	0.6816	2	1.4748	0.5858	-1.7533
std	3	-0.9331	1.0000	4	2.6680	0.0644	-2.2087
sstd	7	0.9688	0.6478	7	3.1789	0.0208	-2.0578
Ast	5	0.5185	0.9090	6	3.1241	0.0244	-2.0766
Ast1	4	0.5184	0.9090	5	2.8681	0.0644	-2.0766
Ald	2	-3.0101	1.0000	1	-1.4748	1.0000	-2.5011
P-value	0.6478						
Investec (0.025)							
norm	1	-4.3361	1.0000	2	0.3056	0.9998	-2.6914
snorm	4	0.7767	0.7808	3	1.3881	0.6426	-2.3984
std	3	-1.4653	1.0000	4	2.3406	0.1338	-2.6056
sstd	5	0.8363	0.7462	7	2.9802	0.0306	-2.5221
Ast	7	0.9118	0.6980	6	2.8945	0.0380	-2.4956
Ast1	6	0.9118	0.6980	5	2.8945	0.1338	-2.4956
Ald	2	-3.0580	1.0000	1	-0.3056	1.0000	-2.7050
P-value	0.698						

5.4 Conclusion and policy implication

The study investigates the volatility of financial returns of the JSE, FNB, Standard Bank, Absa, Capitec, Nedbank and Investec for the period from 19/02/2002 to 31/05/2023. The chapter focused on the predictive and forecast ability of various GARCH and GAS models for daily financial returns under different error distributions. Particularly, the chapter investigates the forecasting ability of three different generalised autoregressive conditional heteroscedasticity (GARCH) models: standard GARCH, EGARCH, and GJR-GARCH, based on four different distributions (normal, student-t, skewed student-t, and GED). We constructed different sets of GARCH models for joint (VaR, ES) to forecast and examine the performance of all selected equity returns based on seven different distributions (norm, snorm, std, sstd, Ast, Ast1, Ald). We compare alternative model specifications in terms of their VaR forecasting performances. Specifically, we compared VaR forecasts delivered by SGARCH, EGARCH and GJR-GARCH with different innovations and the joint (VaR, ES) forecasts from the GARCH models. We also constructed different sets of GAS models for (VaR, ES) to forecast and examine the performance of all selected financial institution equity returns. The model comparison is performed using the MCS procedure recently proposed by Hansen, Lunde, and Nason (2011). We utilise 12 GARCH models of different distributional assumptions to predict the downside tail risk of the daily returns of financial institutions in SA, and use the MCS approach to evaluate the VaR prediction performance of the models. The MCS technique is beneficial when several different models are available and it is not apparent which one performs better. The MCS sequence of tests delivers the Superior Set of Models having Equal Predictive Ability. The ability of the MCS procedure has been empirically investigated on an out-of-sample analysis of the 1%, 2.5%, and 5% VaR forecasts based on daily returns of six financial institutions (FNB, Standard Bank, Absa, Capitec, Nedbank, Investec) and the JSE. We construct the SSM for each equity using different symmetric and asymmetric distributional innovations through the GAS technique to fit (VaR, ES) models and score them by the FZL function. The model comparisons were performed using the R package MCS.

We investigated the impact of different model re-estimation strategies, different forecasting periods and different quantile levels on out-of-sample VaR prediction performance of the selected GARCH models. The quantile level did not generally affect the VaR forecasting performance of the GARCH family models for all the returns except for the returns of Capitec and Investec. At the 1% quantile, the GJR-GARCH using the generalised error distribution was the best model for JSE, FNB, Standard Bank, Nedbank, and Investec. The best model for Absa returns was EGARCH-ged, while that for Capitec bank was EGARCH-sstd. At the 2.5% quantile, there was no change in the most suitable model for each of the financial returns, except for Investec, which had EGARCH-ged as the preferred model. At the 5% quantile, Capitec returns had

GJRARCH-std as the preferred model while Investec had EGARCH-ged as the most optimal model. The rest of the equity returns had the same chosen models under 1% and 2.5% for the 5% quantile. The different forecasting periods did not affect the VaR prediction performance of the GARCH family models. Also, the other model re-estimation strategies had little impact on the forecasting accuracy of the GARCH family models. Empirical results show that three equities (JSE, Nedbank, Investec) contain all seven (7) initial models in the SSM between the two percentiles. The rest of the equities contain (5) initial models in the SSM between the two percentiles, except Capitec, which had four models at the 2.5% percentile. Capitec equity returns exhibited the most heterogeneity as per the SSM, percentile dependent, and time-varying, suggesting that it may be more challenging to model the tail risk of Capitec returns than the rest of the returns.

Empirical results from the joint (VaR, ES) models suggest that generally, the risk models exhibited homogeneity, their tail risks are time-invariant and percentile-independent. The homogeneity in the SSM models indicates that not one model vastly outperforms others within the set, reducing any burden on risk managers in selecting a single best model. It also increases the confidence that any of the models within the set will provide reliable and robust results since their performance is comparable. The risk associated with relying on a single best model is reduced, providing a hedge against underfitting or overfitting the model. Also, the homogeneity in the different sets of models suggests that the models will perform similarly with unseen data and give a good balance between bias and variability in the data. Although homogeneity of models within the SSM has significant benefits, it poses challenges in identifying the best model. Since modelling volatility is linked to portfolio optimization, hedging decisions and derivative pricing models, our findings can be helpful to investors and policymakers while developing the risk management strategies. Specifically, investors could forecast future daily financial returns volatility more precisely. In addition, the study will help investors to make proper asset-allocation and hedging decisions by understanding the effect of extreme events on the financial return VaR measure. The result of this study will help policymakers to formulate effective hedging strategies that could avoid the adverse impact of extreme events. The study contributes to the need to accurately forecast and select the best tail risk measure for risk management. Also, the study fits within the Basel III framework for comparative backtesting to reduce regulatory arbitrage.

Chapter 6

Summary, Conclusions and Recommendations

6.1 Introduction

This last chapter comprises the summary, conclusions, and recommendations from the study. The chapter begins by providing an overview of the main findings, followed by limitations of the study, recommendations for both investment and policy that may be implied from this study, and areas for further research.

6.2 Summary of findings and policy implications

6.2.1 Machine learning approach in bankruptcy prediction

Bankruptcy prediction has become an essential phenomenon for financial institutions, governments, investors and other stakeholders over the last two or three decades to hedge against potential economic damage. Experiences from the 2007-2008 GFC suggest that advanced prediction models are required to help prevent bank failures. The purpose of the study was to investigate and compare the prediction accuracy of machine learning models and also to use these models to predict bankruptcy. The study compared the machine learning models to predict bankruptcy using data from US banks. This thesis aims to contribute to the already extensive literature on bankruptcy prediction. Although several machine learning models have been used in bankruptcy prediction, the literature utilising LightGBM, XGBoost, OSELM and SVM on a large amount of data from financial institutions in the US is relatively narrow. This thesis presents novel methods combined with feature selection techniques to predict bankruptcy. Feature selection is an optimization problem that aims to improve the performance of models and reduce their computational time by eliminating redundant features. The feature selection techniques selected the most relevant financial ratios to predict bankruptcy and also identify the importance of each feature in the prediction process.

The research develops new efficient and effective algorithms to select a smaller feature subset and tune SVM, XGBoost and LightGBM hyperparameters simultaneously. The study also addressed the class imbalance problem by using the SMOTE technique.

The study implemented SVM using linear, sigmoid and polynomial kernel functions to predict bankruptcy. To obtain a model with better explanatory power and stability, we used the grid-search technique with a 10-fold cross-validation to find optimal parameters of the SVM model. Particularly, we utilized a grid-search technique using 10-fold cross-validation to choose optimal values of the upper bound C and the kernel parameter g that are most important in SVM model selection. According to the results, the best model for the linear, sigmoid and polynomial kernels has similar results for the training and test data. Empirical findings show that although the prediction performance of the SVM using different feature selection methods and different kernel functions is satisfactory, the SVM-Tabu had the best prediction performance using the linear kernel. The SVM-Tabu search with a polynomial kernel function outperformed the rest of the models based on the experimental results. Thus, we conclude that the SVM-Tabu search using the polynomial kernel function can be a valuable model for bank regulators, financial supervisors, central banks and other stakeholders in identifying distressed banks.

Also, the study used the online sequential extreme learning machine (OS-ELM) to predict bankruptcy. This approach, unlike the batch ELM, can learn data one-by-one or chunk-by-chunk with varying or fixed chunk size. Results suggest that OSELM using all features and hardlimit activation function, as well as OSELM-SA with hardlimit activation function, performed better than the rest of the OSELM models. Also, the study used the online sequential extreme learning machine (OS-ELM) to predict bankruptcy. This approach unlike the batch ELM can learn data one-by-one or chunk-by-chunk with varying or fixed chunk size. Results suggest that OSELM using all features and hardlimit activation function as well as OSELM-SA with hardlimit activation function performed better than the rest of the OSELM models. This thesis applies a state-of-the-art Machine Learning algorithm, which was developed from boosting methods, specifically XGBoost, to classify financial institutions as either bankrupt or non-bankrupt. In general, XGBoost with the Tabu search feature provides the most accurate prediction capabilities for both training and testing data. Furthermore, the SHAP importance feature selection using XGBoost-Tabu indicated that the top 10 financial ratios for bankruptcy prediction include pre-tax operating income to average assets, impaired loans to equity, return on average assets (ROAA), capital funds to net loans, net loans to total deposit and borrowing, capital funds to total assets, equity to total assets, non-performing loans to gross loans, return on average equity (ROAE), and equity to deposit and short term funding. In summary, the high predictive power of the model should encourage bank managers to track the variables that were identified to cause the failure of financial institutions. Managers can avoid financial distress by taking

timely action rather than waiting for government intervention. Regulators may also find the model useful to identify and warn potentially distressed banks.

The study used the novel LightGBM model with Gradient-based One-Side Sampling and Exclusive Feature Bundling to deal with a large number of data instances and a large number of features, respectively, to classify financial institutions as either bankrupt or solvent. We employed the Parzan technique to optimize the parameters of the LightGBM. We also performed a comprehensive feature importance analysis using SHAP to identify useful features for bankruptcy prediction. Additionally, we implemented a thorough feature importance analysis with SHAP to identify essential features for bankruptcy prediction. The results indicated that LightGBM using all features and the RFE feature selection technique outperformed LightGBM with other feature selection techniques for both training and test data. Additionally, the study assessed the feature importance of the input features to determine the most important financial ratios. More importantly, we implemented the SHAP method to investigate the correlation between bankruptcy prediction and the features. The feature importance and SHAP analysis revealed that the top 10 features with significant impact on bankruptcy prediction include pre-tax operating income to average assets, return on average assets (ROAA), net loans to total assets, equity to total assets, non-operating items and taxes to average assets, unreserved impaired loans to equity, equity to net loans, return on average equity (ROAE), capital funds to net loans, liquid assets to deposit and short term funding, and equity to liabilities. In credit risk analysis, policy-makers, financial institutions and market analysts can concentrate more on these selected features to make informed bankruptcy decisions. By jointly considering the accuracy, stability and efficiency, the optimized LightGBM model is highly recommended to predict bankruptcy with excellent performance.

6.2.2 Measuring Systemic Risk across banks in South Africa

Chapter 3 applies existing systemic risk measures to analyse the risk of individual banks and their systemic risk potential to the entire financial system in South Africa. We computed the VaR, CoVaR, and ΔCoVaR for each financial institution to determine the inherent risk and their risk contribution to the financial system. The results suggest that the most systemically important financial institution is Nedbank at both 1% and 5% risk levels. However, Capitec bank is the most risky financial institution. This result shows the need for extensive supervision of Nedbank as it poses a greater threat to the financial system than other banks. There is a need for financial regulators, the central bank and other shareholders to closely monitor Nedbank, Standard Bank and FNB.

6.2.3 Time-varying risk analysis in South African Equity market

The chapter performs an empirical investigation of time-varying tail risk and market risk analysis on financial returns from six financial institutions and the JSE using GAS models to estimate and forecast the joint VaR and ES. The joint FZL function draws from both VaR and ES, making it conducive for comparative backtesting of competing forecasting models. A consistent scoring function was used to measure tail risk. We accounted for skewness and the leptokurtic nature of the financial return series by using norm, SNORM, STD, SSTD, AST, AST1, and ALD distributions. The forecasting ability of the different models was tested using the MDM test, and the results suggest that the predictive capabilities of the models differ at both 1% and 2.5%. The distributions were ranked and selected using the MAE and RMSE. Results indicate that the optimal forecast model across all financial institutions is the model with the Ald distribution for both risk levels. We evaluated the market risk models using different statistical backtesting techniques, including the UC, CC, DQ and QL. Results indicate that all the backtesting models except DQ show evidence that the ALD using the 1% and 2.5% levels are correctly specified. It was observed that the models using 1% risk levels outperformed those using 2.5%. The characteristic FZL value indicates that the risk exposure of Absa is the least compared to other financial institutions and requires the least capital to absorb any loss. The JSE has the highest risk profile. We provide evidence in support of the study of Patton et al. (2019), which argues that FZL values drawn from a joint VaR and ES provide a better empirical risk modelling than either VaR or ES separately.

6.2.4 Elicitability and risk model comparison of financial institutions

Chapter 5 investigates the forecasting abilities of standard GARCH, GJR-GARCH, and EGARCH, based on normal, Student t, skewed Student t and GED. We implemented a variety of GAS models for the joint (VaR, ES) to forecast and compare the performance of the models using seven distinct distributions (norm, snorm, std, sstd, Ast, Ast1, Ald). The model comparison is performed using the MCS procedure recently proposed by Hansen et al. (2011). The MCS technique is beneficial when several different models are available and it is not apparent which one performs better. The MCS sequence of tests delivers the Superior Set of Models having Equal Predictive Ability using a loss function that discriminates the models. We investigated the impact of different quantile levels, different model re-estimation strategies, and different forecasting periods on the out-of-sample VaR prediction performance of the GARCH family models. To fit the joint VaR and ES models and evaluate them using the FZL function, we construct the SSM for each equity utilizing different symmetric and asymmetric distributional innovations through the GAS technique. The ability of the MCS procedure has been empirically investigated on an out-of-sample analysis of the 1%, 2.5%, and 5% VaR

forecasts based on a panel of daily returns of six South African financial institutions and the JSE stock index. The GARCH family models' VaR forecasting performance was not generally impacted by the quantile level for all daily financial returns, except Capitec and Investec. Also, the VaR prediction performance of the selected GARCH models was not influenced by the varying forecasting periods. The forecasting accuracy of the GARCH models was minimally affected by the different model re-estimation strategies. In general, the risk models displayed homogeneity, and their tail risk is time-invariant and percentile-independent, based on the empirical results from the joint (VaR, ES) models. The homogeneity of the SSM models demonstrates that no particular model significantly outperforms others within the set, consequently relieving the burden on risk managers to select a single best model.

6.3 Limitations

One of the challenges encountered in the study is the acquisition of data. Some of the bankrupt banks lacked data, and some lacked data for four years before declaring bankruptcy. This reduced the number of bankrupt banks included in the study.

One of the limitations was related to using SMOTE. Since different results are obtained depending on the synthetic examples outputted by this technique, one possible solution to minimize this drawback could be achieved by testing the model several times and then averaging the results.

The dataset used for the ML algorithms was selected from US financial data; hence, it is not possible to extrapolate the results to other geographical areas.

The choice of the kernel function and the determination of optimal values of the parameters have a critical impact on the performance of the resulting system. The study used the grid search technique to select the optimal values of the parameters for SVM for the best prediction performance.

The study did not consider all the GARCH models and all probability distributions. This may weaken the findings of the study. Also, we only considered six financial institutions in the empirical research.

The study did not include more systemic risk measures such as Marginal Expected Shortfall (MES), component Expected Shortfall (CES), nonparametric conditional value-at-risk (NCoVaR), Conditional Autoregressive Value-at-risk (CAViaR)

The study did not consider bivariate Dynamic Conditional Correlation (DCC) GARCH

The study did not examine systemic risk spillover during different market conditions

The analysis relied on market data, such as stock returns, which may not fully capture other sources of systemic risk

The study focused on the top six South African banks, which may not represent smaller or non-listed financial institutions. The result may not generalise to the entire financial system.

The study did not consider interest rate regimes, regulatory policy or any macroprudential interventions during the period that may potentially affect systemic risk behaviour.

6.4 Recommendations

The findings of the study will assist the government, policy makers and professionals in the financial and insurance sector to broaden and deepen regulations for both the financial and nonfinancial sectors. There is a need for a regulatory framework to prevent a repetition of the 2008 GFC. This study will also help in the formulation of internal banking models for the computation of regulatory capital.

Also, the findings will help governments to exert more pressure on financial institutions and insurance companies to comply with national regulatory standards. There is a need for banks to be closely monitored for information asymmetries and to check complex product features that could affect the health of the institutions.

The study uses stock prices and stock indices for volatility modelling analysis, which is helpful for investors to deepen their understanding of the risks in the financial industry and be able to invest cautiously. The findings will inform financial institutions on the need to understand and manage model risks better. To mitigate risk in the financial sector, there is a need for rigorous guidelines for developing and validating risk models and for constant monitoring and improvement of these risk models. This will help anticipate emerging risks and how to detect and mitigate them.

The findings of the research will assist risk managers and financial institutions in measuring and tracking their exposure to contagion and its impact on the financial system. The extent of contagion has been identified as one of the drivers for classifying banks as systemically important.

The findings of the study have implications for risk management, portfolio diversification, and investment decision-making. Investors, risk managers and other stakeholders can use the forecasts from the models to assess their decision to purchase, sell or hold shares from investment portfolios which attempt to mirror stock market movements. In addition, the study will help investors to make proper asset-allocation and hedging decisions by understanding the effect of extreme events on the financial return VaR measure. The result of this study will help policymakers to formulate effective hedging strategies that could avoid the adverse impact of extreme events.

The study will help financial regulators, governments, and other stakeholders to establish early warning mechanisms to accurately forecast risk and avoid the spread of risk and, if possible, mitigate the effects of any risks. It also allows risk managers to assess risk levels and implement corresponding risk prevention measures effectively.

The findings of the research will offer practical significance to financial institutions by helping them control their risk accumulation and prevent any excessive contagion of risk. The research has practical implications for investors. For investors who want to protect their investments, they need to diversify their investments. The daily returns volatility modelling in the study indicates effective risk management techniques and will be of great interest to researchers, investors, governments, policymakers and portfolio managers. The study will further provide recommendations to researchers, policy makers and other professionals on the need to backtest volatility risk models.

Based on the findings from the study, the research will assist investors in making investment decisions to reduce risk spillover effects by constructing diversified investment portfolios, especially during periods of high volatility and instability, such as the COVID-19 era. By diversifying assets, the overall risk could be reduced, offering a more resilient investment portfolio.

Technological innovations continuously emerge, enabling new risk-management techniques to help make better risk decisions at a lower cost. Machine learning and big data offer some of the potential impacts. The findings of this research will inform stakeholders of the need to include machine learning and big data in their risk management practices. The use of cheaper and faster computing power will enable risk functions to use reams of customer information to help make better credit decisions, predict operational losses, and monitor portfolios for early signs of risks inherent in them. Also, the use of machine learning methods will help to improve the accuracy of risk models by identifying complex, non-linear patterns in huge datasets. Financial institutions and regulators will need to employ these machine learning models.

In summary, the high predictive power of the ML models developed should encourage bank managers to track the most critical variables identified as drivers of bankruptcy. This will help managers to avoid financial distress by taking timely action rather than waiting for government intervention. Financial regulators may also find the model useful to identify and warn potentially distressed banks.

6.5 Areas for further research

Further research should be done to compare the performance of the models in estimating VaR, CoVaR and Δ CoVaR between tranquil and turbulent conditions, such as COVID periods.

The combination of Extreme Value Theory (EVT) with the GARCH model can be analysed in forecasting intraday VaR and ES.

This study reveals the limitations of combining strategies in the one-day-ahead VaR and ES forecasts in emerging markets. A possible direction for further research is to investigate whether this finding holds for multi-day-ahead forecasts.

The detection of bankruptcy is significant to the health of any economy. With extensive and continuing financial losses being experienced by financial firms and given the increasing difficulty of detecting bankruptcy, it is essential to develop more effective approaches for detecting bankruptcy.

The time series movements of daily financial returns are mainly characterised as non-linear, posing challenges in quantifying and capturing these movements. It is recommended that further studies be conducted with a hybrid of GARCH and deep learning methods to harness the predictive abilities of each technique to generate more accurate forecasts.

Future areas of work may include exploring volatility changes using other processes to combine with GARCH-type models. It seems that the hybrid models are more useful in extreme event forecasting as the structure of the volatility process becomes more complex.

This study should include more systemic risk measures and other financial institutions, insurance companies and real estate firms to gain a broader picture of how systemic risk can impact real economic activity in South Africa. The use of other systemic risk measures will enable other studies to verify consistency in the results. The significance of each systemic risk measure should be evaluated by assessing its ability to predict future economic downturns.

One of the main challenges affecting bankruptcy prediction is the identification of the most essential features that produce high accuracy. Future studies should explore new feature selection techniques.

Combined applications of the improved LightGBM algorithm with other techniques might offer the potential to overcome the drawbacks of each method. To improve bankruptcy prediction, hybrid bankruptcy prediction methods might be a promising solution worth investigating.

The ML algorithms used in the study were only tested on data from the US. The applicability of the methods to other areas should be explored. Other parameter optimisation methods should be used to improve the prediction accuracy of the models further.

Appendix A

Graphs

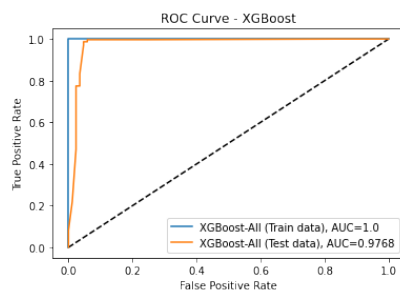


Figure A.1: ROC curve for XGBoost with all features

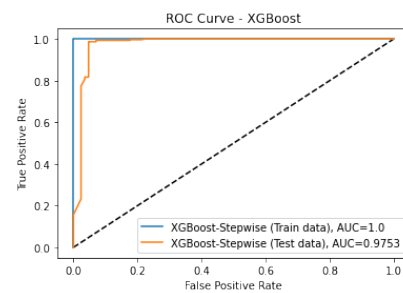


Figure A.2: ROC curve for XGBoost-Stepwise

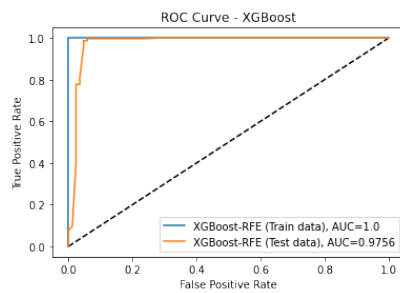


Figure A.3: ROC curve for XGBoost-RFE

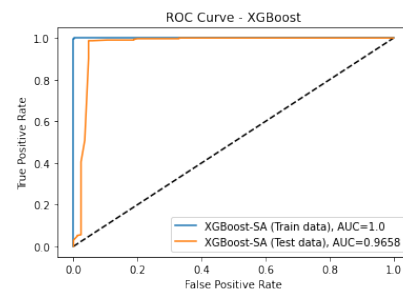


Figure A.4: ROC curve for XGBoost-SA

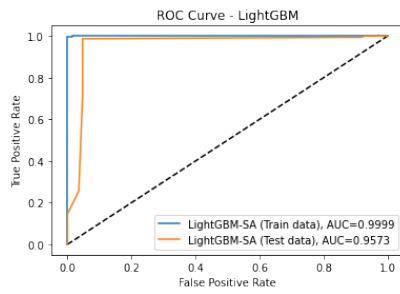


Figure A.5: ROC curve for LightGBM-SA

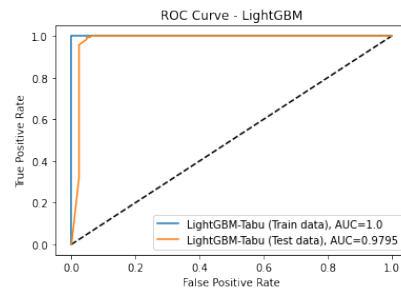


Figure A.6: ROC curve for LightGBM-Tabu

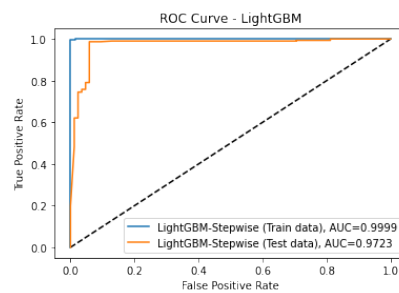


Figure A.7: ROC curve for LightGBM-Stepwise

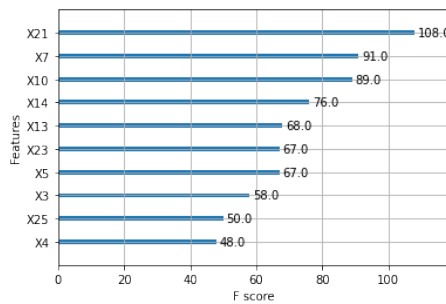


Figure A.8: RFE

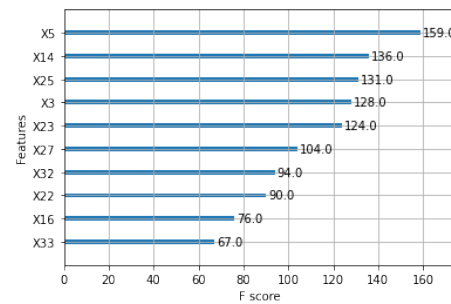


Figure A.9: XGBoost-SA

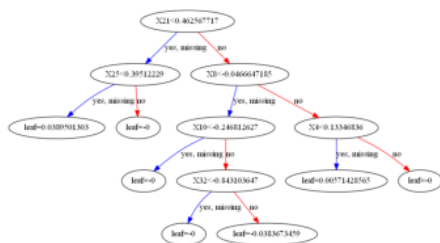


Figure A.10: Tree diagram of LightGBM using all features

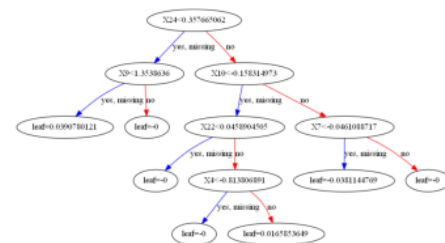


Figure A.11: Tree diagram of LightGBM-RFE



Figure A.12: Tree diagram of LightGBM-SA

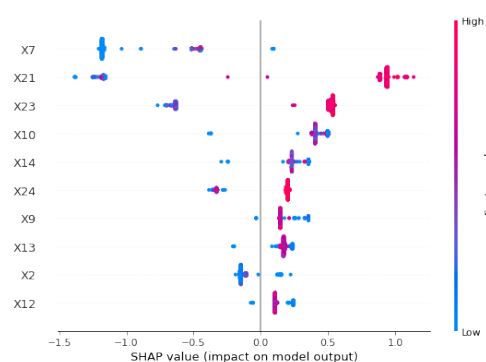


Figure A.13: SHAP Bar Graph XGBoost-RFE

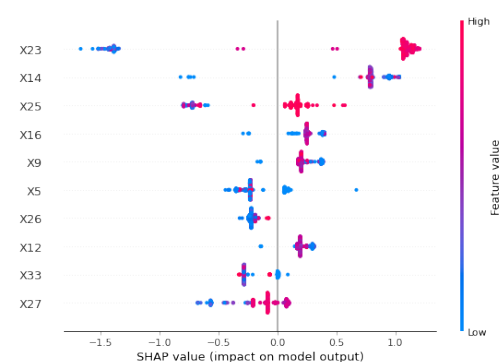


Figure A.14: SHAP Bar Graph XGBoost-SA

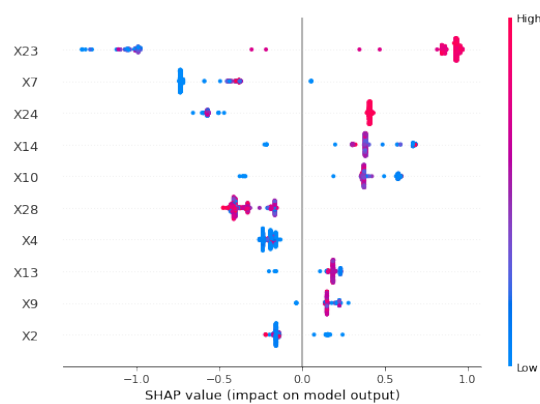


Figure A.15: SHAP Bar Graph XGBoost-Stepwise



Figure A.16: Force plot using XGboost-RFE

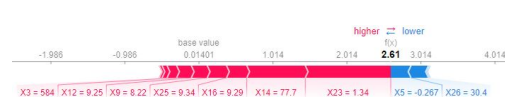


Figure A.17: Force plot using XGboost-SA



Figure A.18: Force plot XGboost using all features

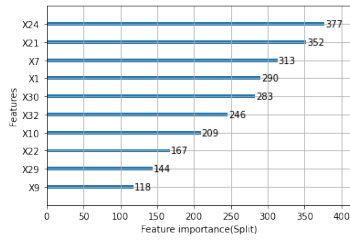


Figure A.19: Feature Importance using GA selected features

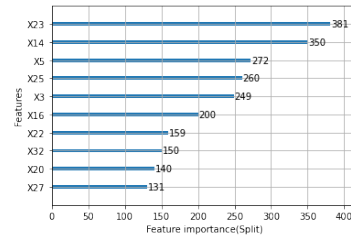


Figure A.20: Feature Importance using SA selected variables

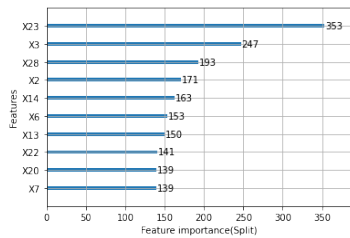


Figure A.21: Feature Importance using stepwise selected features

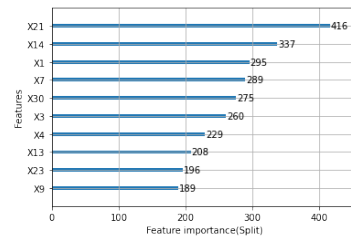


Figure A.22: Feature Importance using Tabu selected features

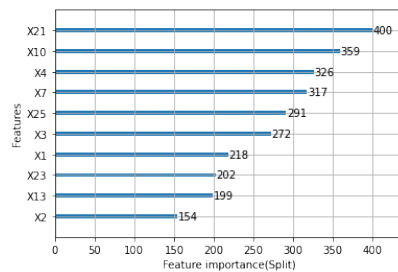


Figure A.23: Feature Importance using LVQ selected variables

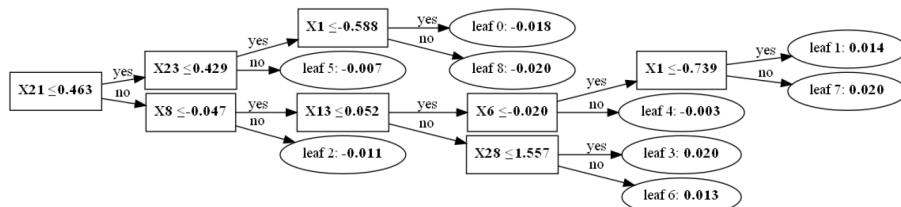


Figure A.24: Tree ensemble model

Appendix B

Tables

Table B.1: Backtest Results of Univariate GAS (VAR, ES) models at 1%

Institution	Distribution	UC	CC	AE	DQ	QL	AD mean	AD Max
JSE	norm	1.92 (0.17)	2.39 (0.30)	1.44	696.68 (0.000)	0.5006	32.85	148.32
	snorm	22.18 (0.00)	37.80 (0.00)	2.70	1351.94 (0.00)	1.5862	58.08	321.24
	std	3.65 (0.06)	4.24 (0.12)	1.62	773.88 (0.00)	0.4553	26.38	125.38
	sstd	4.68 (0.03)	5.34 (0.07)	1.71	857.31 (0.00)	0.4106	22.39	81.67
	ast	28.55 (0.00)	30.57 (0.00)	2.97	1187 (0.00)	0.6745	21.88	155.77
	ast1	28.55 (0.00)	30.57 (0.00)	2.97	1187 (0.00)	0.6745	21.88	155.28
	ald	0.43 (0.51)	0.58 (0.75)	0.81	589.04 (0.00)	0.2106	22.52	62.68
FNB	norm	2.72 (0.10)	3.25 (0.20)	1.53	865.59 (0.00)	0.5072	31.43	108.96
	snorm	8.41 (0.00)	11.61 (0.00)	1.98	1101.99 (0.00)	1.0123	50.03	215.19
	std	4.68 (0.03)	5.34 (0.07)	1.71	883.39 (0.00)	0.5902	33.01	166.95
	sstd	3.65 (0.06)	4.24 (0.12)	1.62	878.08 (0.00)	0.5796	34.22	148.14
	ast	20.19 (0.00)	24.33 (0.00)	2.61	831.57 (0.00)	1.1587	43.67	391.33
	ast1	20.19 (0.00)	24.33 (0.00)	2.61	831.57 (0.00)	1.1587	43.67	391.33
	ald	0.97 (0.33)	1.08 (0.58)	0.72	526.55 (0.00)	0.2422	29.94	84.54
Standard	norm	3.65 (0.06)	4.75 (0.09)	1.62	907.48 (0.00)	0.4608	26.79	103.13
	snorm	22.18 (0.00)	25.99 (0.00)	2.70	1529.58 (0.00)	1.2143	44.23	193.40
	std	2.72 (0.10)	3.25 (0.20)	1.53	898.60 (0.00)	0.4276	26.18	100.78
	sstd	7.07 (0.01)	7.88 (0.02)	1.89	1446.40 (0.00)	0.4689	23.43	69.13
	ast	22.18 (0.00)	23.85 (0.00)	2.70	1475.37 (0.00)	0.7865	28.30	122.93
	ast1	22.18 (0.00)	23.85 (0.00)	2.70	1475.37 (0.00)	0.7865	28.30	122.93
	ald	1.76 (0.18)	1.85 (0.40)	0.63	561.97 (0.00)	0.1636	21.60	44.75

Table B.2: Backtest Results of Univariate GAS (VAR, ES) models at 1%

Institution	Distribution	UC	CC	AE	DQ	QL	AD mean	AD Max
Absa	norm	9.84 (0.00)	10.30 (0.01)	2.07	1054.88 (0.00)	0.6395	29.71	154.68
	snorm	26.36 (0.00)	36.24 (0.00)	2.88	1491.71 (0.00)	1.6374	56.30	256.74
	std	9.84 (0.00)	10.82 (0.00)	2.07	1016.45 (0.00)	0.5855	27.09	152.04
	sstd	14.66 (0.00)	15.91 (0.00)	2.34	1298.55 (0.00)	0.6332	26.03	125.42
	ast	33.11 (0.00)	33.12 (0.00)	3.15	1631.03 (0.00)	1.1688	36.52	177.21
	ast1	33.11 (0.00)	33.12 (0.00)	3.15	1631.03 (0.00)	1.1688	36.52	177.21
	ald	0.43 (0.51)	0.58 (0.75)	0.81	551.31 (0.00)	0.2152	23.32	71.49
Capitec	norm	24.24 (0.00)	97.72 (0.00)	2.79	1632.84 (0.00)	0.8037	27.97	275.24
	snorm	1.25 (0.26)	7.26 (0.03)	1.35	548.52 (0.00)	1.0771	78.38	427.94
	std	3.65 (0.06)	4.24 (0.12)	1.62	718.72 (0.00)	0.7540	45.03	211.24
	sstd	7.07 (0.01)	7.74 (0.02)	1.89	878.93 (0.00)	0.6889	35.17	191.74
	ast	42.93 (0.00)	44.52 (0.00)	3.51	2294.45 (0.00)	1.5135	42.62	162.91
	ast1	42.93 (0.00)	44.52 (0.00)	3.51	2294.45 (0.00)	1.5135	42.62	162.91
	ald	0.97 (0.33)	1.08 (0.58)	0.72	322.79 (0.00)	0.3010	38.24	163.52
Nedbank	norm	7.07 (0.01)	10.58 (0.01)	1.89	938.76 (0.00)	0.5355	26.98	131.10
	snorm	34.48 (0.00)	46.88 (0.00)	3.24	1741.69 (0.00)	2.0050	61.49	263.21
	std	8.41 (0.00)	11.61 (0.00)	1.98	962.97 (0.00)	0.5436	26.17	136.96
	sstd	9.84 (0.00)	10.82 (0.00)	2.07	1131.83 (0.00)	0.6315	29.33	124.05
	ast	14.66 (0.00)	14.88 (0.00)	2.34	1286.57 (0.00)	0.8652	36.06	164.94
	ast1	14.66 (0.00)	14.88 (0.00)	2.34	1286.57 (0.00)	0.8652	36.06	164.94
	ald	0.43 (0.51)	0.58 (0.75)	0.81	580.75 (0.00)	0.2514	27.77	65.24
Investec	norm	11.37 (0.00)	11.73 (0.00)	2.16	1528.02 (0.00)	0.6516	29.06	91.71
	snorm	28.55 (0.00)	29.43 (0.00)	2.97	1539.94 (0.00)	1.2447	41.2787	245.23
	std	16.43 (0.00)	16.59 (0.00)	2.43	1446.00 (0.00)	0.7721	30.83	125.97
	sstd	18.28 (0.00)	19.73 (0.00)	2.52	1538.65 (0.00)	0.8665	33.53	139.83
	ast	8.41 (0.00)	8.97 (0.00)	1.98	908.12 (0.00)	0.6861	33.57	166.23
	ast1	8.41 (0.00)	8.97 (0.00)	1.98	908.12 (0.00)	0.6861	33.57	166.23
	ald	0.31 (0.58)	0.62 (0.73)	1.17	750.54 (0.00)	0.2913	22.74	74.12

Table B.3: Backtest results of univariate GAS (VAR, ES) models testing it at 2.5%

Institution	Distribution	UC	CC	AE	DQ	QL	AD mean	AD Max
JSE	norm	0.002 (0.96)	0.116 (0.94)	1.01	500.75 (0.000)	0.3736	11.84	72.09
	snorm	3.49 (0.06)	17.54 (0.00)	1.37	686.82 (0.00)	0.9142	24.89	151.41
	std	0.02 (0.88)	1.37 (0.50)	0.97	553.03 (0.00)	0.3747	12.32	63.19
	sstd	2.87 (0.09)	5.42 (0.07)	1.33	742.97 (0.00)	0.3663	8.78	42.96
	ast	17.36 (0.00)	22.48 (0.00)	1.87	1043.77 (0.00)	0.5758	10.87	78.59
	ast1	17.36 (0.00)	22.48 (0.00)	1.87	1043.77 (0.00)	0.5758	10.87	78.59
	ald	11.61 (0.00)	11.87 (0.00)	0.43	358.73 (0.00)	0.2330	14.52	38.27
FNB	norm	0.12 (0.73)	1.36 (0.51)	0.94	565.96 (0.000)	0.3813	13.19	54.46
	snorm	0.96 (0.33)	10.27 (0.01)	1.19	636.58 (0.00)	0.6254	18.84	102.90
	std	0.38 (0.54)	2.16 (0.34)	1.12	605.20 (0.00)	0.4481	13.51	83.04
	sstd	0.002 (0.96)	1.45 (0.48)	1.01	565.49 (0.00)	0.4334	14.42	74.11
	ast	5.67 (0.02)	5.82 (0.05)	1.48	496.78 (0.00)	0.7382	18.37	188.29
	ast1	5.67 (0.02)	5.82 (0.05)	1.48	496.78 (0.00)	0.7382	18.37	188.29
	ald	4.95 (0.03)	5.47 (0.06)	0.61	408.73 (0.00)	0.2589	12.19	48.84
Standard	norm	0.38 (0.54)	0.40 (0.82)	1.12	655.76 (0.000)	0.3671	10.55	51.70
	snorm	5.67 (0.02)	11.28 (0.00)	1.48	848.63 (0.00)	0.7524	18.67	92.89
	std	0.64 (0.42)	2.54 (0.28)	1.15	691.65 (0.00)	0.3633	10.10	51.67
	sstd	0.38 (0.54)	2.16 (0.34)	1.12	874.72 (0.00)	0.3964	11.67	37.25
	ast	8.31 (0.00)	11.95 (0.00)	1.59	965.04 (0.00)	0.5905	13.30	62.81
	ast1	8.31 (0.00)	11.95 (0.00)	1.59	965.04 (0.00)	0.5905	13.30	62.81
	ald	1.84 (0.16)	2.65 (0.27)	0.76	525.33 (0.00)	0.2199	7.76	29.41

Table B.4: Backtest results of univariate GAS (VAR, ES) models testing it at 2.5%

Institution	Distribution	UC	CC	AE	DQ	QL	AD mean	AD Max
Absa	norm	0.38 (0.54)	1.54 (0.46)	1.12	677.90 (0.000)	0.4644	14.18	75.25
	snorm	8.31 (0.00)	12.85 (0.00)	1.59	804.23 (0.00)	0.9496	22.52	121.83
	std	1.35 (0.25)	1.35 (0.51)	1.23	711.58 (0.00)	0.4569	12.70	75.55
	sstd	6.50 (0.01)	9.81 (0.01)	1.51	907.62 (0.00)	0.4959	11.35	63.57
	ast	9.28 (0.00)	13.09 (0.00)	1.62	932.09 (0.00)	0.7680	17.52	88.13
	ast1	9.28 (0.00)	13.09 (0.00)	1.62	932.09 (0.00)	0.7680	17.52	88.13
	ald	0.54 (0.46)	0.91 (0.63)	0.86	557.66 (0.00)	0.2662	9.06	42.64
Capitec	norm	1.80 (0.18)	45.28 (0.00)	1.26	627.52 (0.000)	0.5147	14.15	131.04
	snorm	0.88 (0.35)	7.47 (0.02)	0.83	304.60 (0.00)	0.6259	27.28	200.45
	std	0.002 (0.96)	1.45 (0.48)	1.01	465.09 (0.00)	0.5220	17.92	104.01
	sstd	0.002 (0.96)	0.12 (0.94)	1.01	529.33 (0.00)	0.4858	16.53	94.90
	ast	16.08 (0.00)	16.27 (0.00)	1.84	1186.32 (0.00)	0.9522	19.44	81.11
	ast1	16.08 (0.00)	16.27 (0.00)	1.84	1186.32 (0.00)	0.9522	19.44	81.11
	ald	8.52 (0.00)	8.87 (0.01)	0.50	250.09 (0.00)	0.2835	16.86	88.38
Nedbank	norm	0.96 (0.33)	1.84 (0.40)	1.19	641.36 (0.000)	0.4069	11.33	64.35
	snorm	9.28 (0.00)	34.73 (0.00)	1.62	874.88 (0.00)	1.1186	26.29	124.65
	std	0.38 (0.54)	1.54 (0.46)	1.12	658.09 (0.00)	0.4305	12.91	68.69
	sstd	1.35 (0.25)	3.50 (0.17)	1.23	732.71 (0.00)	0.4769	13.36	62.80
	ast	2.30 (0.13)	4.72 (0.09)	1.30	774.67 (0.00)	0.5927	16.33	81.28
	ast1	2.30 (0.13)	4.72 (0.09)	1.30	774.67 (0.00)	0.5927	16.33	81.28
	ald	0.54 (0.46)	0.91 (0.63)	0.86	550.50 (0.00)	0.2972	10.49	39.71
Investec	norm	0.18 (0.67)	0.23 (0.89)	1.08	843.84 (0.000)	0.4702	14.91	46.41
	snorm	5.67 (0.02)	6.91 (0.03)	1.48	846.02 (0.00)	0.7681	19.17	116.49
	std	2.30 (0.13)	2.33 (0.31)	1.30	859.84 (0.00)	0.5567	15.16	63.69
	sstd	0.96 (0.33)	2.99 (0.22)	1.19	845.66 (0.00)	0.5993	18.04	70.20
	ast	0.29 (0.59)	1.44 (0.49)	0.90	521.07 (0.00)	0.4575	17.54	80.20
	ast1	0.29 (0.59)	1.44 (0.49)	0.90	521.08 (0.00)	0.4575	17.54	80.20
	ald	0.06 (0.81)	0.13 (0.94)	1.05	692.77 (0.00)	0.3332	10.17	43.80

Bibliography

- Abdalla, Suliman Zakaria Suliman (2012). "Modelling exchange rate volatility using GARCH models: Empirical evidence from Arab countries". In: *International Journal of Economics and Finance* 4.3, pp. 216–229.
- Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi (2015). *Networks, shocks, and systemic risk*. Tech. rep. National Bureau of Economic Research.
- Acerbi, Carlo (2002). "Spectral measures of risk: A coherent representation of subjective risk aversion". In: *Journal of Banking & Finance* 26.7, pp. 1505–1518.
- Acerbi, Carlo and Balazs Szekely (2014). "Back-testing expected shortfall". In: *Risk* 27.11, pp. 76–81.
- Acerbi, Carlo and Dirk Tasche (2002). "On the coherence of expected shortfall". In: *Journal of banking & finance* 26.7, pp. 1487–1503.
- Acharya, Viral et al. (2009). "The financial crisis of 2007-2009: Causes and remedies". In: *Restoring financial stability: how to repair a failed system*, pp. 1–56.
- Acharya, Viral V and Anjan V Thakor (2016). "The dark side of liquidity creation: Leverage and systemic risk". In: *Journal of Financial Intermediation* 28, pp. 4–21.
- Acharya, Viral V et al. (2012). "How to calculate systemic risk surcharges". In: *Quantifying systemic risk*. University of Chicago Press, pp. 175–212.
- (2017). "Measuring systemic risk". In: *The review of financial studies* 30.1, pp. 2–47.
- Adrian, Tobias and Markus K Brunnermeier (2008). "CoVaR: A method for macroprudential regulation". In: *Federal Reserve Bank of New York Staff Reports* 348.
- Agarwal, Vineet and Richard Taffler (2008). "Comparing the performance of market-based and accounting-based bankruptcy prediction models". In: *Journal of Banking & Finance* 32.8, pp. 1541–1551.
- Ahmed, Rizwan Raheem et al. (2018). "Mean reversion in international markets: evidence from GARCH and half-life volatility models". In: *Economic research-Ekonomska istraživanja* 31.1, pp. 1198–1217.
- Ahmed, Shamima et al. (2022). "Artificial intelligence and machine learning in finance: A bibliometric review". In: *Research in International Business and Finance* 61, p. 101646.
- Akaike, H (1973). "Information theory as an extension of the maximum likelihood principle. Á In: Petrov, BN and Csaki, F". In: *Second International Symposium on Information Theory. Akademiai Kiado, Budapest*, pp. 276Á281.

- Al-Helali, Baligh et al. (2021). "A new imputation method based on genetic programming and weighted KNN for symbolic regression with incomplete data". In: *Soft Computing* 25.8, pp. 5993–6012.
- Alabadla, Mustafa et al. (2022). "Systematic review of using machine learning in imputing missing values". In: *Ieee Access* 10, pp. 44483–44502.
- Alaka, Hafiz A et al. (2018). "Systematic review of bankruptcy prediction models: Towards a framework for tool selection". In: *Expert systems with applications* 94, pp. 164–184.
- Alberg, Dima, Haim Shalit, and Rami Yosef (2008). "Estimating stock market volatility using asymmetric GARCH models". In: *Applied Financial Economics* 18.15, pp. 1201–1208.
- Albrecht, Andreas A (2006). "Stochastic local search for the feature set problem, with applications to microarray data". In: *Applied Mathematics and Computation* 183.2, pp. 1148–1164.
- Alfaro, Rodrigo A and Mathias Drehmann (2009). "Macro stress tests and crises: what can we learn?" In: *BIS Quarterly Review December*.
- Ali, Aida, Siti Mariyam Shamsuddin, and Anca L Ralescu (2013). "Classification with class imbalance problem". In: *Int. J. Advance Soft Compu. Appl* 5.3.
- Allen, Linda, Turan G Bali, and Yi Tang (2012). "Does systemic risk in the financial sector predict future economic downturns?" In: *The Review of Financial Studies* 25.10, pp. 3000–3036.
- Alnuaimi, Amer FAH and Tasnim HK Albaldawi (2024). "Concepts of statistical learning and classification in machine learning: An overview". In: *BIO Web of Conferences*. Vol. 97. EDP Sciences, p. 00129.
- Alter, Adrian and Andreas Beyer (2014). "The dynamics of spillover effects during the European sovereign debt turmoil". In: *Journal of Banking & Finance* 42, pp. 134–153.
- Altman, Edward I (1968). "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy". In: *The journal of finance* 23.4, pp. 589–609.
- Altman, Edward I, Robert G Haldeman, and Paul Narayanan (1977). "ZETATM analysis A new model to identify bankruptcy risk of corporations". In: *Journal of banking & finance* 1.1, pp. 29–54.
- Altman, Edward I and Edith Hotchkiss (1993). *Corporate financial distress and bankruptcy*. Vol. 1998. New York: John Wiley & Sons.
- Altman, Edward I, Giancarlo Marco, and Franco Varetto (1994). "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)". In: *Journal of banking & finance* 18.3, pp. 505–529.
- Angelini, Eliana, Giacomo Di Tollo, and Andrea Roli (2008). "A neural network approach for credit risk evaluation". In: *The quarterly review of economics and finance* 48.4, pp. 733–755.
- Ardia, David, Kris Boudt, and Leopoldo Catania (2016). "Generalized autoregressive score models in R: The GAS package". In: *arXiv preprint arXiv:1609.02354*.

- (2018). “Downside risk evaluation with the R package GAS”. In: *R journal* 10.2, pp. 410–421.
- Arndt, Channing et al. (2020). “Impact of Covid-19 on the South African economy”. In: *Southern Africa-Towards Inclusive Economic Development Working Paper* 111, pp. 1–37.
- Artzner, Philippe et al. (1999). “Coherent measures of risk”. In: *Mathematical finance* 9.3, pp. 203–228.
- Asteriou, Dimitrios and SG Hall (2011). *Applied econometrics: Palgrave Macmillan*.
- Atiya, Amir F (2001). “Bankruptcy prediction for credit risk using neural networks: A survey and new results”. In: *IEEE Transactions on neural networks* 12.4, pp. 929–935.
- (2020). “Why does forecast combination work so well?” In: *International Journal of Forecasting* 36.1, pp. 197–200.
- Aydilek, Ibrahim Berkan and Ahmet Arslan (2013). “A hybrid method for imputation of missing values using optimized fuzzy c-means with support vector regression and a genetic algorithm”. In: *Information Sciences* 233, pp. 25–35.
- Aziz, M and H Dar (2006). “Predicting corporate bankruptcy-where we stand?, Corporate Governance Journal, vol. 6, nr 1”. In.
- Azzalini, Adelchi (1985). “A class of distributions which includes the normal ones”. In: *Scandinavian journal of statistics*, pp. 171–178.
- Azzalini, Adelchi and Antonella Capitanio (2003). “Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 65.2, pp. 367–389.
- Baek, Jinwoo and Sungzoon Cho (2003). “Bankruptcy prediction for credit risk using an auto-associative neural network in Korean firms”. In: *2003 IEEE International Conference on Computational Intelligence for Financial Engineering, 2003. Proceedings. IEEE*, pp. 25–29.
- Balcaen, Sofie and Hubert Ooghe (2006). “35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems”. In: *The British Accounting Review* 38.1, pp. 63–93.
- Baldi, Pierre et al. (2000). “Assessing the accuracy of prediction algorithms for classification: an overview”. In: *Bioinformatics* 16.5, pp. 412–424.
- Banking Supervision, Basel Committee on (2016). *Minimum capital requirements for market risk*.
- Banulescu, Georgiana-Denisa and Elena-Ivona Dumitrescu (2015). “Which are the SIFIs? A Component Expected Shortfall approach to systemic risk”. In: *Journal of Banking & Finance* 50, pp. 575–588.
- Barboza, Flavio, Herbert Kimura, and Edward Altman (2017). “Machine learning models and bankruptcy prediction”. In: *Expert systems with applications* 83, pp. 405–417.
- Barr, Michael S et al. (2018). “Big data in finance: highlights from the big data in finance conference hosted at the University of Michigan October 27-28, 2016”. In.
- Barroso, Marta and Juan Laborda (2022). “Digital transformation and the emergence of the Fintech sector: Systematic literature review”. In: *Digital Business* 2.2, p. 100028.

- Başarır, Çağatay and Cengiz Toraman (2014). "Financial stability analysis in banking sector: a stress test method". In: *Muhasebe ve Finansman Dergisi* 62, pp. 129–144.
- Bauwens, Luc and Sébastien Laurent (2005). "A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models". In: *Journal of Business & Economic Statistics* 23.3, pp. 346–354.
- Beaver, William H (1966). "Financial ratios as predictors of failure". In: *Journal of accounting research*, pp. 71–111.
- Becerra, Victor M, Roberto KH Galvão, and Magda Abou-Seada (2005). "Neural and wavelet network models for financial distress classification". In: *Data Mining and Knowledge Discovery* 11, pp. 35–55.
- Begley, Joy, Jin Ming, and Susan Watts (1996). "Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models". In: *Review of accounting Studies* 1, pp. 267–284.
- Bekkar, Mohamed, Hassiba Khelouane Djemaa, and Taklit Akrouf Alitouche (2013). "Evaluation measures for models assessment over imbalanced data sets". In: *J Inf Eng Appl* 3.10.
- Benavoli, Alessio, Dario Azzimonti, and Dario Piga (2020). "Skew Gaussian processes for classification". In: *Machine Learning* 109.9, pp. 1877–1902.
- Benoit, Sylvain et al. (2017). "Where the risks lie: A survey on systemic risk". In: *Review of Finance* 21.1, pp. 109–152.
- Berg, Daniel (2007). "Bankruptcy prediction by generalized additive models". In: *Applied Stochastic Models in Business and Industry* 23.2, pp. 129–143.
- Berger, Allen N and Christa HS Bouwman (2013). "How does capital affect bank performance during financial crises?" In: *Journal of financial economics* 109.1, pp. 146–176.
- Bergstra, James and Yoshua Bengio (2012). "Random search for hyper-parameter optimization". In: *The journal of machine learning research* 13.1, pp. 281–305.
- Bergstra, James et al. (2011). "Algorithms for hyper-parameter optimization". In: *Advances in neural information processing systems* 24.
- Bernanke, Ben S (2013). *Stress Testing Banks: What Have We Learned?: a speech at the "Maintaining Financial Stability: Holding a Tiger by the Tail" financial markets conference sponsored by the Federal Reserve Bank of Atlanta, Stone Mountain, Georgia, April 8, 2013*. Tech. rep.
- Bernard, Jason et al. (2017). "Learning style Identifier: Improving the precision of learning style identification through computational intelligence algorithms". In: *Expert Systems with Applications* 75, pp. 94–108.
- Billio, Monica et al. (2012). "Econometric measures of connectedness and systemic risk in the finance and insurance sectors". In: *Journal of financial economics* 104.3, pp. 535–559.
- Bisias, Dimitrios et al. (2012). "A survey of systemic risk analytics". In: *Annu. Rev. Financ. Econ.* 4.1, pp. 255–296.

- Black, Fischer and Myron Scholes (1973). "The pricing of options and corporate liabilities". In: *Journal of political economy* 81.3, pp. 637–654.
- Blaschke, Winfrid et al. (2001). "Stress testing of financial systems: an overview of issues, methodologies, and FSAP experiences". In.
- Blasques, Francisco, Siem Jan Koopman, et al. (2014). *Maximum likelihood estimation for correctly specified generalized autoregressive score models: Feedback effects, contraction conditions and asymptotic properties*. Tech. rep. Tinbergen institute discussion paper.
- Blazsek, Szabolcs and Hector Hernández (2018). "Analysis of electricity prices for Central American countries using dynamic conditional score models". In: *Empirical Economics* 55.4, pp. 1807–1848.
- Board, Financial Stability (2012). *Basel Committee on Banking Supervision.(BCBS).(2011). Assessment of the macroeconomic impact of higher loss absorbency for global systemically important banks, 10.10. 2011.*
- Board, Financial Stability et al. (2009). "Guidance to assess the systemic importance of financial institutions, markets and instruments: initial considerations". In: *Report to G20 finance ministers and governors*.
- Bollerslev, Tim (1986). "Generalized autoregressive conditional heteroskedasticity". In: *Journal of econometrics* 31.3, pp. 307–327.
- (1987). "A conditionally heteroskedastic time series model for speculative prices and rates of return". In: *The review of economics and statistics*, pp. 542–547.
- Borio, Claudio, Mathias Drehmann, and Kostas Tsatsaronis (2014). "Stress-testing macro stress testing: does it live up to expectations?" In: *Journal of Financial Stability* 12, pp. 3–15.
- Boukhatem, Jamel and Fatma Ben Moussa (2018). "The effect of Islamic banks on GDP growth: Some evidence from selected MENA countries". In: *Borsa Istanbul Review* 18.3, pp. 231–247.
- Bousquet, Olivier, Stéphane Boucheron, and Gábor Lugosi (2003). "Introduction to statistical learning theory". In: *Summer school on machine learning*. Springer, pp. 169–207.
- Boyacioglu, Melek Acar, Yakup Kara, and Ömer Kaan Baykan (2009). "Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey". In: *Expert Systems with Applications* 36.2, pp. 3355–3366.
- Bradley, Andrew P (1997). "The use of the area under the ROC curve in the evaluation of machine learning algorithms". In: *Pattern recognition* 30.7, pp. 1145–1159.
- Braione, Manuela and Nicolas K Scholtes (2016). "Forecasting value-at-risk under different distributional assumptions". In: *Econometrics* 4.1, p. 3.
- Brédart, Xavier, Eric Séverin, and David Veganzones (2021). "Human resources and corporate failure prediction modeling: Evidence from Belgium". In: *Journal of Forecasting* 40.7, pp. 1325–1341.

- Brownlees, Christian and Robert F Engle (2017). "SRISK: A conditional capital shortfall measure of systemic risk". In: *The Review of Financial Studies* 30.1, pp. 48–79.
- Brownlees, Christian, Christina Hans, and Eulalia Nualart (2021). "Bank credit risk networks: Evidence from the Eurozone". In: *Journal of Monetary Economics* 117, pp. 585–599.
- Brunnermeier, Markus K (2009). "Deciphering the liquidity and credit crunch 2007–2008". In: *Journal of Economic perspectives* 23.1, pp. 77–100.
- Buchanan, Bonnie G (2019). "Artificial intelligence in finance". In.
- Burnham, Kenneth P and David R Anderson (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.
- Cajueiro, Daniel O and Benjamin M Tabak (2005). "Testing for time-varying long-range dependence in volatility for emerging markets". In: *Physica A: Statistical Mechanics and its Applications* 346.3–4, pp. 577–588.
- Callaghan, Michael et al. (2013). "Global Cooperation Among G20 Countries: Responding to the Crisis and Restoring Growth". In: *Global Cooperation Among G20 Countries: Responding to the Crisis and Restoring Growth*. Springer, pp. 1–21.
- (2014). *Global Cooperation Among G20 Countries: Responding to the Crisis and Restoring Growth*. Springer.
- Campbell, Sean D (2005). "A review of backtesting and backtesting procedures". In.
- Canbas, Serpil, Altan Cabuk, and Suleyman Bilgin Kilic (2005). "Prediction of commercial bank failure via multivariate statistical analysis of financial structures: The Turkish case". In: *European journal of operational research* 166.2, pp. 528–546.
- Cele, Simphiwe K and Nhlanhla BW Mlitwa (2024). "The global COVID-19 impact on the financial services industry in South Africa". In: *South African Journal of Information Management* 26.1, pp. 1–11.
- Cetina, J, I Lelyveld, and K Anand (2015). *Making supervisory stress tests more macroprudential: Considering liquidity and solvency interactions and systemic risk*. Tech. rep. BCBS Working Paper.
- Chambers, Michele and Thomas W Dinsmore (2014). *Advanced analytics methodologies: Driving business value with analytics*. Pearson Education.
- Chan-Lau, Jorge A et al. (2009). "Assessing the systemic implications of financial linkages". In: *IMF global financial stability report* 2.
- Chang, Chia-Lin et al. (2019). "Choosing expected shortfall over VaR in Basel III using stochastic dominance". In: *International Review of Economics & Finance* 60, pp. 95–113.
- Chang, Chih-Chung and Chih-Jen Lin (2011). "LIBSVM: a library for support vector machines". In: *ACM transactions on intelligent systems and technology (TIST)* 2.3, pp. 1–27.
- Charitou, Andreas, Evi Neophytou, and Chris Charalambous (2004). "Predicting corporate failure: empirical evidence for the UK". In: *European accounting review* 13.3, pp. 465–497.

- Chawla, Nitesh V et al. (2002). "SMOTE: synthetic minority over-sampling technique". In: *Journal of artificial intelligence research* 16, pp. 321–357.
- Chen, Hsueh-Ju, Shaio Yan Huang, and Chin-Shien Lin (2009). "Alternative diagnosis of corporate bankruptcy: A neuro fuzzy approach". In: *Expert Systems with Applications* 36.4, pp. 7710–7720.
- Chen, Mark A, Qinxu Wu, and Baozhong Yang (2019). "How valuable is FinTech innovation?" In: *The Review of financial studies* 32.5, pp. 2062–2106.
- Chen, Mu-Yen (2011). "Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches". In: *Computers & Mathematics with Applications* 62.12, pp. 4514–4524.
- Chen, Tianqi and Carlos Guestrin (2016). "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.
- Chen, Tianqi et al. (2015). "Xgboost: extreme gradient boosting. R package version 04-2". In: *OS Independent*.
- Chen, Tsung-Kang et al. (2023). "Bankruptcy prediction using machine learning models with the text-based communicative value of annual reports". In: *Expert Systems with Applications* 233, p. 120714.
- Chen, Wei-Sen and Yin-Kuan Du (2009). "Using neural networks and data mining techniques for the financial distress prediction model". In: *Expert systems with applications* 36.2, pp. 4075–4086.
- Chen, Yiyun, Lu Zhang, and Kunyi Li (2016). "The Review of Network Theory Applied to the Financial System." In: *WHICEB*, p. 70.
- Chicco, Davide and Giuseppe Jurman (2020). "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation". In: *BMC genomics* 21, pp. 1–13.
- Chicco, Davide, Niklas Tötsch, and Giuseppe Jurman (2021). "The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation". In: *BioData mining* 14.1, pp. 1–22.
- Chitiga-Mabugu, Margaret et al. (2021). "Economic and distributional impact of COVID-19: Evidence from macro-micro modelling of the South African economy". In: *South African Journal of Economics* 89.1, pp. 82–94.
- Choudhury, Arkopal and Michael R Kosorok (2020). "Missing data imputation for classification problems". In: *arXiv preprint arXiv:2002.10709*.
- Chowdhury, Abdullahi, Shahriar Kaisar, and Ranesh Naha (2023). "Bankruptcy prediction for imbalanced dataset using oversampling and ensemble machine learning methods". In: *AIP Conference Proceedings*. Vol. 2968. 1. AIP Publishing LLC, p. 040003.
- Christoffersen, Peter F (1998). "Evaluating interval forecasts". In: *International economic review*, pp. 841–862.

- Chung, Hyung-Min Michael and Kar Yan Tam (1993). "A comparative analysis of inductive-learning algorithms". In: *Intelligent Systems in accounting, finance and management* 2.1, pp. 3–18.
- Cieri, Richard M, Lyle G Ganske, and Heather Lennox (1998). "Breaking up is hard to do: Avoiding the solvency-related pitfalls in spinoff transactions". In: *Bus. Law.* 54, p. 533.
- Claessens, Stijn, Rudiger Dornbusch, and Yung Chul Park (2001). "Contagion: Why crises spread and how this can be stopped". In: *International financial contagion*. Springer, pp. 19–41.
- Claessens, Stijn and Kristin Forbes (2001). "International financial contagion: An overview of the issues and the book". In: *International financial contagion*, pp. 3–17.
- Claessens, Stijn et al. (2010). *A safer world financial system: Improving the resolution of systemic institutions*. International Center for Monetary and Banking Studies Geneva.
- Claessens, Stijn et al. (2018). "Fintech credit markets around the world: size, drivers and policy issues". In: *BIS Quarterly Review September*.
- Cleary, Sean and Greg Hebb (2016). "An efficient and functional model for predicting bank distress: In and out of sample evidence". In: *Journal of Banking & Finance* 64, pp. 101–111.
- Climent, Francisco, Alexandre Momparler, and Pedro Carmona (2019). "Anticipating bank distress in the Eurozone: An extreme gradient boosting approach". In: *Journal of business research* 101, pp. 885–896.
- Cohen, Jacob (1968). "Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit." In: *Psychological bulletin* 70.4, p. 213.
- Cont, Rama, Artur Kotlicki, and Laura Valderrama (2020). "Liquidity at risk: Joint stress testing of solvency and liquidity". In: *Journal of Banking & Finance* 118, p. 105871.
- Cornett, Marcia Millon et al. (2011). "Liquidity risk management and credit supply in the financial crisis". In: *Journal of financial economics* 101.2, pp. 297–312.
- Cortes, Corinna and Mehryar Mohri (2003). "AUC optimization vs. error rate minimization". In: *Advances in neural information processing systems* 16.
- Cox, David R et al. (1981). "Statistical analysis of time series: Some recent developments [with discussion and reply]". In: *Scandinavian Journal of Statistics*, pp. 93–115.
- Cox, Raymond AK and Grace W-Y Wang (2014). "Predicting the US bank failure: A discriminant analysis". In: *Economic Analysis and Policy* 44.2, pp. 202–211.
- Creal, Drew, Siem Jan Koopman, and André Lucas (2012). "Generalized Autoregressive Score Models". In: *AENORM* 20, p. 75.
- (2013). "Generalized autoregressive score models with applications". In: *Journal of Applied Econometrics* 28.5, pp. 777–795.
- Cummins, J David and Mary A Weiss (2014). *Systemic risk and the US insurance sector*.
- Danenas, Paulius and Gintautas Garsva (2015). "Selection of support vector machines based classifiers for credit risk domain". In: *Expert systems with applications* 42.6, pp. 3194–3204.

- Dang, Uyen (2011). "The CAMEL rating system in banking supervision. A case study". In.
- Danielsson, Jon et al. (2005). "Sub-additivity re-examined: the case for Value-at-Risk". In.
- Dash, Manoranjan and Huan Liu (1997). "Feature selection for classification". In: *Intelligent data analysis* 1.1-4, pp. 131–156.
- Davis, E Philip and Dilruba Karim (2008). "Comparing early warning systems for banking crises". In: *Journal of Financial stability* 4.2, pp. 89–120.
- Davis, Jesse and Mark Goadrich (2006). "The relationship between Precision-Recall and ROC curves". In: *Proceedings of the 23rd international conference on Machine learning*, pp. 233–240.
- Davis, Jesse et al. (2005). "View Learning for Statistical Relational Learning: With an Application to Mammography." In: *IJCAI*. Citeseer, pp. 677–683.
- De Menezes, Fortunato S et al. (2017). "Data classification with binary response through the Boosting algorithm and logistic regression". In: *Expert Systems with Applications* 69, pp. 62–73.
- De Miguel, LJ et al. (1993). "A comparison between statistical and neural network based methods for predicting bankfailures". In: *Proceedings of the IIIth International Workshop on Artificial Intelligence in Economics and Management, Portland (USA)*.
- Deakin, Edward B (1972). "A discriminant analysis of predictors of business failure". In: *Journal of accounting research*, pp. 167–179.
- Denkowska, Anna and Stanisław Wanat (2020). "Dependencies and systemic risk in the European insurance sector: New evidence based on Copula-DCC-GARCH model and selected clustering methods". In: *Entrepreneurial Business and Economics Review* 8.4, pp. 7–27.
- Devar, Sai Abhishek Abhishek (2019). *Multivariate time series pattern recognition using machine learning and deep learning methods*. The University of Texas at Arlington.
- Dickey, David A and Wayne A Fuller (1979). "Distribution of the estimators for autoregressive time series with a unit root". In: *Journal of the American statistical association* 74.366a, pp. 427–431.
- Diebold, Francis X and Robert S Mariano (2002). "Comparing predictive accuracy". In: *Journal of Business & economic statistics* 20.1, pp. 134–144.
- Ding, Jie, Vahid Tarokh, and Yuhong Yang (2018). "Model selection techniques: An overview". In: *IEEE Signal Processing Magazine* 35.6, pp. 16–34.
- Ding, Zhuanxin, Clive WJ Granger, and Robert F Engle (1993). "A long memory property of stock market returns and a new model". In: *Journal of empirical finance* 1.1, pp. 83–106.
- Douglas, Ella, David Lont, and Tom Scott (2014). "Finance company failure in New Zealand during 2006–2009: Predictable failures?" In: *Journal of Contemporary Accounting & Economics* 10.3, pp. 277–295.

- Du, Zaichao and Juan Carlos Escanciano (2017). "Backtesting expected shortfall: accounting for tail risk". In: *Management Science* 63.4, pp. 940–958.
- Du Jardin, Philippe (2010). "Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy". In: *Neurocomputing* 73.10-12, pp. 2047–2060.
- Duffie, Darrell (2013). "Systemic risk exposures: a 10-by-10-by-10 approach". In: *Risk topography: Systemic risk and macro modeling*. University of Chicago Press, pp. 47–56.
- Durante, Fabrizio and Carlo Sempì (2015). *Principles of copula theory*. CRC press.
- Ecer, Fatih (2013). "Comparing the bank failure prediction performance of neural networks and support vector machines: The Turkish case". In: *Economic research-Ekonomska istraživanja* 26.3, pp. 81–98.
- Ehigiamusoe, Kizito Uyi and Hooi Hooi Lean (2019). "Do economic and financial integration stimulate economic growth? A critical survey". In: *Economics* 13.1, p. 20190004.
- Eichengreen, Barry et al. (2012). "How the subprime crisis went global: Evidence from bank credit default swap spreads". In: *Journal of International Money and Finance* 31.5, pp. 1299–1318.
- Eling, Martin and David Antonius Pankoke (2016). "Systemic risk in the insurance sector: a review and directions for future research". In: *Risk Management and Insurance Review* 19.2, pp. 249–284.
- Elith, Jane, John R Leathwick, and Trevor Hastie (2008). "A working guide to boosted regression trees". In: *Journal of animal ecology* 77.4, pp. 802–813.
- Elliott, Matthew, Benjamin Golub, and Matthew O Jackson (2014). "Financial networks and contagion". In: *American Economic Review* 104.10, pp. 3115–3153.
- Elloumi, Fathi and Jean-Pierre Gueyié (2001). "Financial distress and corporate governance: an empirical analysis". In: *Corporate Governance: The international journal of business in society*.
- Ellul, Andrew and Marco Pagano (2019). "Corporate leverage and employees' rights in bankruptcy". In: *Journal of Financial Economics* 133.3, pp. 685–707.
- Embrechts, Paul et al. (2014). "An academic response to Basel 3.5". In: *Risks* 2.1, pp. 25–48.
- Engle, Robert (2002). "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models". In: *Journal of Business & Economic Statistics* 20.3, pp. 339–350.
- (2009). *Anticipating correlations: a new paradigm for risk management*. Princeton University Press.
- Engle, Robert F and Gary GJ Lee (1993). "A permanent and transitory component model of stock return volatility". In: *University of California at San Diego, Economics Working Paper Series*.
- Engle, Robert F and Simone Manganelli (2004). "CAViaR: Conditional autoregressive value at risk by regression quantiles". In: *Journal of business & economic statistics* 22.4, pp. 367–381.

- Engle, Robert F and Tianyue Ruan (2019). "Measuring the probability of a financial crisis". In: *Proceedings of the National Academy of Sciences* 116.37, pp. 18341–18346.
- Engle, Robert F and Jeffrey R Russell (1998). "Autoregressive conditional duration: a new model for irregularly spaced transaction data". In: *Econometrica*, pp. 1127–1162.
- Erdogan, Birsan Eygi (2013). "Prediction of bankruptcy using support vector machines: an application to bank bankruptcy". In: *Journal of Statistical Computation and Simulation* 83.8, pp. 1543–1555.
- Erdogan Erten, Gamze, Sinem Bozkurt Keser, and Mahmut Yavuz (2021). "Grid search optimised artificial neural network for open stope stability prediction". In: *International Journal of Mining, Reclamation and Environment* 35.8, pp. 600–617.
- Erer, Elif and Deniz Erer (2024). "The domino effect of Silicon Valley Bank's bankruptcy and the role of Fed's monetary policy". In: *Borsa Istanbul Review* 24.3, pp. 573–591.
- Escanciano, Juan Carlos and Pei Pei (2012). "Pitfalls in backtesting historical simulation VaR models". In: *Journal of Banking & Finance* 36.8, pp. 2233–2244.
- Faris, Hossam et al. (2020). "Improving financial bankruptcy prediction in a highly imbalanced class distribution using oversampling and ensemble learning: a case from the Spanish market". In: *Progress in Artificial Intelligence* 9.1, pp. 31–53.
- Fawcett, Tom (2006). "An introduction to ROC analysis". In: *Pattern recognition letters* 27.8, pp. 861–874.
- Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth (1996). "The KDD process for extracting useful knowledge from volumes of data". In: *Communications of the ACM* 39.11, pp. 27–34.
- Fernández, Carmen and Mark FJ Steel (1998). "On Bayesian modeling of fat tails and skewness". In: *Journal of the american statistical association* 93.441, pp. 359–371.
- Fissler, Tobias and Johanna F Ziegel (2016). "Higher order elicibility and Osband's principle". In: .
- Fissler, Tobias, Johanna F Ziegel, and Tilmann Gneiting (2015). "Expected shortfall is jointly elicitable with value at risk-implications for backtesting". In: *arXiv preprint arXiv:1507.00244*.
- Fitzpatrick, CA, Peter Bull, and Oliver Dupriez (2018). "Machine learning for poverty prediction: A comparative assessment of classification algorithms". In: *Wired at: www.github.com*.
- Fitzpatrick, Trevor and Christophe Mues (2016). "An empirical comparison of classification algorithms for mortgage default prediction: evidence from a distressed mortgage market". In: *European Journal of Operational Research* 249.2, pp. 427–439.
- Flood, Mark D et al. (2016). "The application of visual analytics to financial stability monitoring". In: *Journal of financial stability* 27, pp. 180–197.
- Fonseca, Eduardo et al. (2017). "Acoustic scene classification by ensembling gradient boosting machine and convolutional neural networks". In: *Virtanen T, Mesáros A, Heittola T, Diment A, Vincent E, Benetos E, Martínez B, editors. Detection and Classification of Acoustic Scenes and Events 2017 Workshop (DCASE2017); 2017 Nov 16; Munich,*

- Germany. Tampere (Finland): Tampere University of Technology; 2017. p. 37–41. Tampere University of Technology.
- Forbes, Kristin J and Roberto Rigobon (2002). “No contagion, only interdependence: measuring stock market comovements”. In: *The journal of Finance* 57.5, pp. 2223–2261.
- Frankel, Jeffrey and George Saravelos (2012). “Can leading indicators assess country vulnerability? Evidence from the 2008–09 global financial crisis”. In: *Journal of International Economics* 87.2, pp. 216–231.
- Frankel, Jeffrey A and Andrew K Rose (1996). “Currency crashes in emerging markets: An empirical treatment”. In: *Journal of international Economics* 41.3–4, pp. 351–366.
- Friedman, Jerome H (2001). “Greedy function approximation: a gradient boosting machine”. In: *Annals of statistics*, pp. 1189–1232.
- Friedrichs, Frauke and Christian Igel (2005). “Evolutionary tuning of multiple SVM parameters”. In: *Neurocomputing* 64, pp. 107–117.
- Fulmer, John G et al. (1984). “A bankruptcy classification model for small firms”. In: *Journal of commercial bank lending* 66.11, pp. 25–37.
- Gai, Prasanna and Sujit Kapadia (2010). “Contagion in financial networks”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 466.2120, pp. 2401–2423.
- García, Vicente, José Salvador Sánchez, and Ramón Alberto Mollineda (2012). “On the effectiveness of preprocessing methods when dealing with different levels of class imbalance”. In: *Knowledge-Based Systems* 25.1, pp. 13–21.
- Gaspar, Paulo, Jaime Carbonell, and José Luís Oliveira (2012). “On the parameter optimization of Support Vector Machines for binary classification”. In: *Journal of integrative bioinformatics* 9.3, pp. 33–43.
- Geng, Bingzhen, Yang Liu, and Yimiao Zhao (2024). “Value-at-Risk-and Expectile-based Systemic Risk Measures and Second-order Asymptotics: With Applications to Diversification”. In: *arXiv preprint arXiv:2404.18029*.
- Gerlein, Eduardo A et al. (2016). “Evaluating machine learning classification for financial trading: An empirical approach”. In: *Expert Systems with Applications* 54, pp. 193–207.
- Géron, Aurélien (2019). *Machine Learning avec Scikit-Learn: Mise en oeuvre et cas concrets*. Dunod.
- Giacomini, Raffaella and Halbert White (2006). “Tests of conditional predictive ability”. In: *Econometrica* 74.6, pp. 1545–1578.
- Girardi, Giulio and A Tolga Ergün (2013). “Systemic risk measurement: Multivariate GARCH estimation of CoVaR”. In: *Journal of Banking & Finance* 37.8, pp. 3169–3180.
- Global Financial System, Committee on the (2000). *Stress Testing by Large Financial Institutions: Current Practice and Aggregation Issues*.
- Glosten, Lawrence R, Ravi Jagannathan, and David E Runkle (1993). “On the relation between the expected value and the volatility of the nominal excess return on stocks”. In: *The journal of finance* 48.5, pp. 1779–1801.

- Gneiting, Tilmann (2011). "Making and evaluating point forecasts". In: *Journal of the American Statistical Association* 106.494, pp. 746–762.
- Gogas, Periklis, Theophilos Papadimitriou, and Anna Agravetidou (2018). "Forecasting bank failures and stress testing: A machine learning approach". In: *International Journal of Forecasting* 34.3, pp. 440–455.
- Gola, Carlo and Alessandro Roselli (2009). *The UK banking system and its regulatory and supervisory framework*. Springer.
- Golub, Gene H and Charles F Van Loan (2013). *Matrix computations*. JHU press.
- Gómez-Puig, Marta and Simón Sosvilla-Rivero (2014). "Causality and contagion in EMU sovereign debt markets". In: *International Review of Economics & Finance* 33, pp. 12–27.
- González-Rivera, Gloria, Tae-Hwy Lee, and Santosh Mishra (2004). "Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood". In: *International Journal of forecasting* 20.4, pp. 629–645.
- Gordy, Michael B and Sandeep Juneja (2010). "Nested simulation in portfolio risk measurement". In: *Management Science* 56.10, pp. 1833–1848.
- Graham, John R et al. (2023). "Employee costs of corporate bankruptcy". In: *The Journal of Finance* 78.4, pp. 2087–2137.
- Gray, D and A Jobst (2010). "Lessons from the financial crisis on modeling systemic risk and sovereign risk". In: *Lessons from the financial crisis*. RISK Books, London.
- Gray, Dale, Cristián Echeverría, and Leonardo Luna (2006). "A measure of the risk of bank insolvency in Chile". In: *Financial Stability Report, Central Bank of Chile, second semester*.
- Gray, Dale F (2009). "Modeling financial crises and sovereign risks". In: *Annu. Rev. Financ. Econ.* 1.1, pp. 117–144.
- Gu, Qiong, Li Zhu, and Zhihua Cai (2009). "Evaluation measures of the classification performance of imbalanced data sets". In: *Computational Intelligence and Intelligent Systems: 4th International Symposium, ISICA 2009, Huangshi, China, October 23-25, 2009. Proceedings* 4. Springer, pp. 461–471.
- Haldane, Andrew G (2013). *Rethinking the financial network*. Springer.
- Han, Hui, Wen-Yuan Wang, and Bing-Huan Mao (2005). "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning". In: *Advances in Intelligent Computing: International Conference on Intelligent Computing, ICIC 2005, Hefei, China, August 23-26, 2005, Proceedings, Part I* 1. Springer, pp. 878–887.
- Han, Jiawe and Micheline Kamber (2001). "Data mining concepts and techniques San Francisco Moraga Kaufman". In.
- Hanley, James A and Barbara J McNeil (1982). "The meaning and use of the area under a receiver operating characteristic (ROC) curve." In: *Radiology* 143.1, pp. 29–36.
- Hanley, John A et al. (1989). "Receiver operating characteristic (ROC) methodology: the state of the art". In: *Crit Rev Diagn Imaging* 29.3, pp. 307–335.

- Hannan, Edward J and Barry G Quinn (1979). "The determination of the order of an autoregression". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 41.2, pp. 190–195.
- Hansen, Bruce E (1994). "Autoregressive conditional density estimation". In: *International Economic Review*, pp. 705–730.
- Hansen, LP (2013). *Challenges in identifying and measuring systemic risk. Risk Topography: Systemic Risk and Macro Modeling*.
- Hansen, Peter R and Asger Lunde (2005). "A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?" In: *Journal of applied econometrics* 20.7, pp. 873–889.
- Hansen, Peter R, Asger Lunde, and James M Nason (2011). "The model confidence set". In: *Econometrica* 79.2, pp. 453–497.
- Hansen, Peter Reinhard (2005). "A test for superior predictive ability". In: *Journal of Business & Economic Statistics* 23.4, pp. 365–380.
- Harvey, Andrew C (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*. Vol. 52. Cambridge University Press.
- Hasan, Md Kamrul et al. (2021). "Missing value imputation affects the performance of machine learning: A review and analysis of the literature (2010–2021)". In: *Informatics in Medicine Unlocked* 27, p. 100799.
- Hastie, Trevor et al. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.
- Helbing, Dirk (2018). "Societal, economic, ethical and legal challenges of the digital revolution: from big data to deep learning, artificial intelligence, and manipulative technologies". In: *Towards digital enlightenment: Essays on the dark and light sides of the digital revolution*. Springer, pp. 47–72.
- Hinkley, David V and Nagesh S Revankar (1977). "Estimation of the Pareto law from underreported data: A further analysis". In: *Journal of Econometrics* 5.1, pp. 1–11.
- Hol, Suzan (2007). "The influence of the business cycle on bankruptcy probability". In: *International transactions in operational research* 14.1, pp. 75–90.
- Hotta, Luiz Koodi, Edimilson C Lucas, and Helder P Palaro (2008). "Estimation of VaR using copula and extreme value theory". In: *Multinational Finance Journal* 12.3/4, pp. 205–218.
- Houari, Rima et al. (2014). "Handling missing data problems with sampling methods". In: *2014 International conference on advanced networking distributed systems and applications*. IEEE, pp. 99–104.
- Hsieh, Nan-Chen and Lun-Ping Hung (2010). "A data driven ensemble classifier for credit scoring analysis". In: *Expert systems with Applications* 37.1, pp. 534–545.
- Hsu, Chih-Wei, Chih-Chung Chang, Chih-Jen Lin, et al. (2003). *A practical guide to support vector classification*.

- Huang, Fu-yuan (2008). "A particle swarm optimized fuzzy neural network for credit risk evaluation". In: *2008 Second International Conference on Genetic and Evolutionary Computing*. IEEE, pp. 153–157.
- Huang, Guang-Bin, Qin-Yu Zhu, and Chee-Kheong Siew (2006). "Extreme learning machine: theory and applications". In: *Neurocomputing* 70.1-3, pp. 489–501.
- Huang, Xin, Hao Zhou, and Haibin Zhu (2009). "A framework for assessing the systemic risk of major financial institutions". In: *Journal of Banking & Finance* 33.11, pp. 2036–2049.
- Hurd, Thomas R et al. (2014). "Illiquidity and insolvency: a double cascade model of financial crises". In: *Available at SSRN* 2424877.
- Hutter, Frank, Holger H Hoos, and Kevin Leyton-Brown (2011). "Sequential model-based optimization for general algorithm configuration". In: *Learning and Intelligent Optimization: 5th International Conference, LION 5, Rome, Italy, January 17-21, 2011. Selected Papers* 5. Springer, pp. 507–523.
- Imbierowicz, Björn and Christian Rauch (2014). "The relationship between liquidity risk and credit risk in banks". In: *Journal of Banking & Finance* 40, pp. 242–256.
- International Settlements, Bank for (2011). *Basel III: a global regulatory framework for more resilient banks and banking systems*. BIS.
- Iturriaga, Félix J López and Iván Pastor Sanz (2015). "Bankruptcy visualization and prediction using neural networks: A study of US commercial banks". In: *Expert Systems with applications* 42.6, pp. 2857–2869.
- Jabeur, Sami Ben (2017). "Bankruptcy prediction using partial least squares logistic regression". In: *Journal of Retailing and Consumer Services* 36, pp. 197–202.
- Jabeur, Sami Ben et al. (2021). "CatBoost model and artificial intelligence techniques for corporate failure prediction". In: *Technological Forecasting and Social Change* 166, p. 120658.
- Ji, Qiang, Elie Bouri, and David Roubaud (2018). "Dynamic network of implied volatility transmission among US equities, strategic commodities, and BRICS equities". In: *International Review of Financial Analysis* 57, pp. 1–12.
- Jin, Ruoming and Gagan Agrawal (2003). "Communication and memory efficient parallel decision tree construction". In: *Proceedings of the 2003 SIAM international conference on data mining*. SIAM, pp. 119–129.
- Jobst, Mr Andreas A and Mr Dale F Gray (2013). *Systemic contingent claims analysis: Estimating market-implied systemic risk*. International Monetary Fund.
- John, George H, Ron Kohavi, and Karl Pfleger (1994). "Irrelevant features and the subset selection problem". In: *Machine learning proceedings 1994*. Elsevier, pp. 121–129.
- Jones, M Chris and MJ1959820 Faddy (2003). "A skew extension of the t-distribution, with applications". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 65.1, pp. 159–174.

- Jones, Stewart, David Johnstone, and Roy Wilson (2017). "Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks". In: *Journal of Business Finance & Accounting* 44.1-2, pp. 3–34.
- Jordan, Dan J et al. (2010). "Predicting bank failures: Evidence from 2007 to 2010". In: *Available at SSRN* 1652924.
- Jorion, Philippe (1997). "Value at risk: the new benchmark for controlling market risk". In: (No Title).
- Jovanović, Dejan, Mirjana Todorovic, and Milka Grbic (2017). "Financial indicators as predictors of illiquidity". In: *Available at SSRN* 2945488.
- Jung, Kichul et al. (2020). "Evaluation of nitrate load estimations using neural networks and canonical correlation analysis with k-fold cross-validation". In: *Sustainability* 12.1, p. 400.
- Junior, Peterson Owusu and Imhotep Alagidede (2020). "Risks in emerging markets equities: Time-varying versus spatial risk analysis". In: *Physica A: Statistical Mechanics and its Applications* 542, p. 123474.
- Jurman, Giuseppe, Samantha Riccadonna, and Cesare Furlanello (2012). "A comparison of MCC and CEN error measures in multi-class prediction". In.
- Kainulainen, Laura et al. (2011). "Ensembles of local linear models for bankruptcy analysis and prediction". In: *Case Studies In Business, Industry And Government Statistics* 4.2, pp. 116–133.
- Kaisar, Shahriar and Abdullahi Chowdhury (2022). "Integrating oversampling and ensemble-based machine learning techniques for an imbalanced dataset in dyslexia screening tests". In: *ICT Express* 8.4, pp. 563–568.
- Kallestrup, René, David Lando, and Agatha Murgoci (2016). "Financial sector linkages and the dynamics of bank and sovereign credit spreads". In: *Journal of Empirical Finance* 38, pp. 374–393.
- Kaminsky, Graciela, Saul Lizondo, and Carmen M Reinhart (1998). "Leading indicators of currency crises". In: *Staff Papers* 45.1, pp. 1–48.
- Kaminsky, Graciela L and Carmen M Reinhart (1999). "The twin crises: the causes of banking and balance-of-payments problems". In: *American economic review* 89.3, pp. 473–500.
- Kang, Hyun (2013). "The prevention and handling of the missing data". In: *Korean journal of anesthesiology* 64.5, pp. 402–406.
- Kay, Jim W, D Mike Titterington, Statistics DM Titterington, et al. (1999). *Statistics and neural networks: advances at the interface*. Oxford University Press on Demand.
- Ke, Guolin et al. (2017). "Lightgbm: A highly efficient gradient boosting decision tree". In: *Advances in neural information processing systems* 30.
- Kells, Stuart (2020). "Impacts of COVID-19 on corporate governance and assurance, international finance and economics, and non-fiction book publishing: some personal reflections". In: *Journal of Accounting & Organizational Change* 16.4, pp. 629–635.

- Kim, Hong Sik and So Young Sohn (2010). "Support vector machines for default prediction of SMEs based on technology credit". In: *European Journal of Operational Research* 201.3, pp. 838–846.
- Kim, Kyoung-jae (2003). "Financial time series forecasting using support vector machines". In: *Neurocomputing* 55.1-2, pp. 307–319.
- Kim, Myoung-Jong, Dae-Ki Kang, and Hong Bae Kim (2015). "Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction". In: *Expert Systems with Applications* 42.3, pp. 1074–1082.
- Kirkos, Efsthathios (2015). "Assessing methodologies for intelligent bankruptcy prediction". In: *Artificial Intelligence Review* 43, pp. 83–123.
- Kliestik, Tomas et al. (2018). "Bankruptcy prevention: new effort to reflect on legal and social changes". In: *Science and Engineering Ethics* 24, pp. 791–803.
- Koh, Hian Chye and Sen Suan Tan (1999). "A neural network approach to the prediction of going concern status". In: *Accounting and Business Research* 29.3, pp. 211–216.
- Kohavi, Ron et al. (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection". In: *Ijcai*. Vol. 14. 2. Montreal, Canada, pp. 1137–1145.
- Koller, Daphne and Mehran Sahami (1996). *Toward optimal feature selection*. Tech. rep. Stanford InfoLab.
- Kotz, Samuel, Tomasz Kozubowski, and Krzysztof Podgorski (2012). *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Springer Science & Business Media.
- Kuhe, David A (2018). "Modeling volatility persistence and asymmetry with exogenous breaks in the Nigerian stock returns". In: *CBN Journal of Applied Statistics (JAS)* 9.1, p. 7.
- Kuhn, Max, Kjell Johnson, et al. (2013). *Applied predictive modeling*. Vol. 26. Springer.
- Kumar, Mishra Aswini et al. (2012). "Analyzing soundness in Indian banking: A CAMEL approach". In: *Research Journal of Management Sciences* ISSN 2319, p. 1171.
- Kumar, Mohan, Uma Moorthy, and William Perraudin (2003). "Predicting emerging market currency crashes". In: *Journal of Empirical Finance* 10.4, pp. 427–454.
- Kumar, P Ravi and Vadlamani Ravi (2006). "Bankruptcy prediction in banks by fuzzy rule based classifier". In: *2006 1st International Conference on Digital Information Management*. IEEE, pp. 222–227.
- (2007). "Bankruptcy prediction in banks and firms via statistical and intelligent techniques—A review". In: *European journal of operational research* 180.1, pp. 1–28.
- Kumar, Runjun D et al. (2015). "Statistically identifying tumor suppressors and oncogenes from pan-cancer genome-sequencing data". In: *Bioinformatics* 31.22, pp. 3561–3568.
- Kupiec, Paul H et al. (1995). *Techniques for verifying the accuracy of risk measurement models*. Vol. 95. 24. Division of Research and Statistics, Division of Monetary Affairs, Federal ...

- Kvamme, Håvard et al. (2018). "Predicting mortgage default using convolutional neural networks". In: *Expert Systems with Applications* 102, pp. 207–217.
- Kwiatkowski, Denis et al. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" In: *Journal of econometrics* 54.1-3, pp. 159–178.
- Lahmiri, Salim et al. (2020). "Performance assessment of ensemble learning systems in financial data classification". In: *Intelligent Systems in Accounting, Finance and Management* 27.1, pp. 3–9.
- Langkamp, Diane L, Amy Lehman, and Stanley Lemeshow (2010). "Techniques for handling missing data in secondary analyses of large surveys". In: *Academic pediatrics* 10.3, pp. 205–210.
- Langley, P (1994). "Selection of relevant features in machine learning In Proceedibgs of the AAAI Fall Symposium on Relevance". In: *AAAI, September*.
- Laporta, Alessandro G, Luca Merlo, and Lea Petrella (2018). "Selection of value at risk models for energy commodities". In: *Energy Economics* 74, pp. 628–643.
- Lazar, Emese and Ning Zhang (2019). "Model risk of expected shortfall". In: *Journal of Banking & Finance* 105, pp. 74–93.
- Le, Hong Hanh and Jean-Laurent Viviani (2018). "Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios". In: *Research in International Business and Finance* 44, pp. 16–25.
- Lee, Kidong, David Booth, and Pervaiz Alam (2005). "A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms". In: *Expert Systems with Applications* 29.1, pp. 1–16.
- Lee, Sangjae and Wu Sung Choi (2013). "A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis". In: *Expert Systems with Applications* 40.8, pp. 2941–2946.
- Lee, Young-Chan (2007). "Application of support vector machines to corporate credit rating prediction". In: *Expert Systems with Applications* 33.1, pp. 67–74.
- Leukes, Chanelle and Jones Odei Mensah (2019). "Systemic risk contribution of financial institutions in South Africa". In: *African Review of Economics and Finance* 11.2, pp. 188–218.
- Levy-Yeyati, Eduardo, Maria Soledad Martinez Peria, and Sergio L Schmukler (2010). "Depositor behavior under macroeconomic risk: evidence from bank runs in emerging economies". In: *Journal of Money, Credit and Banking* 42.4, pp. 585–614.
- Li, Hui and Jie Sun (2009). "Predicting business failure using multiple case-based reasoning combined with support vector machine". In: *Expert systems with applications* 36.6, pp. 10085–10096.
- Lin, Jie et al. (2020). "Data-driven missing data imputation in cluster monitoring system based on deep neural network". In: *Applied Intelligence* 50.3, pp. 860–877.

- Lin, Shih-Wei et al. (2008). "Particle swarm optimization for parameter determination and feature selection of support vector machines". In: *Expert systems with applications* 35.4, pp. 1817–1824.
- Lin, Wei-Yang, Ya-Han Hu, and Chih-Fong Tsai (2011). "Machine learning in financial crisis prediction: a survey". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42.4, pp. 421–436.
- Liu, Lanbiao, Chen Chen, and Bo Wang (2022). "Predicting financial crises with machine learning methods". In: *Journal of Forecasting* 41.5, pp. 871–910.
- Liu, Yang et al. (2022). "Anticipating financial distress of high-tech startups in the European Union: A machine learning approach for imbalanced samples". In: *Journal of Forecasting* 41.6, pp. 1131–1155.
- Liu, Yingbo et al. (2015). "Research on the Matthews correlation coefficients metrics of personalized recommendation algorithm evaluation". In: *International Journal of Hybrid Information Technology* 8.1, pp. 163–172.
- Lohmann, Christian, Steffen Möllenhoff, and Thorsten Ohliger (2023). "Nonlinear relationships in bankruptcy prediction and their effect on the profitability of bankruptcy prediction models". In: *Journal of Business Economics* 93.9, pp. 1661–1690.
- Longstaff, Francis A (2010). "The subprime credit crisis and contagion in financial markets". In: *Journal of financial economics* 97.3, pp. 436–450.
- Lu, Zudi, Hai Huang, and Richard Gerlach (2010). "Estimating value at risk: From JP Morgans standard-EWMA to skewed-EWMA forecasting". In: *University of Sydney*.
- Lucas, André and Xin Zhang (2016). "Score-driven exponentially weighted moving averages and value-at-risk forecasting". In: *International Journal of Forecasting* 32.2, pp. 293–302.
- Lundberg, Scott M and Su-In Lee (2017). "A unified approach to interpreting model predictions". In: *Advances in neural information processing systems* 30.
- Lyngdoh, Gideon A et al. (2022). "Prediction of concrete strengths enabled by missing data imputation and interpretable machine learning". In: *Cement and concrete composites* 128, p. 104414.
- Ma, Xiaojun et al. (2018). "Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning". In: *Electronic Commerce Research and Applications* 31, pp. 24–39.
- Makatjane, Katleho (2024). "Forecasting Time-varying Value-at-Risk and Expected Short-fall Dependence: A Markov-switching Generalized Autoregressive Score Copula Approach". In: *Austrian Journal of Statistics* 53.2, pp. 81–98.
- Manguzvane, Mathias Mandla and John Weistrass Muteba Mwamba (2020). "GAS Copula models on who's systemically important in South Africa: Banks or Insurers?" In: *Empirical Economics* 59.4, pp. 1573–1604.

- Manguzvane, Mathias Mandla and Sibusiso Blessing Ngobese (2023). "A component expected shortfall approach to systemic risk: An application in the South African financial industry". In: *International Journal of Financial Studies* 11.4, p. 146.
- Mariano, Roberto S and Daniel Preve (2012). "Statistical tests for multiple forecast comparison". In: *Journal of econometrics* 169.1, pp. 123–130.
- Martin, Daniel (1977). "Early warning of bank failure: A logit regression approach". In: *Journal of banking & finance* 1.3, pp. 249–276.
- Martínez, I (1996). "Forecasting company failure: neural approach versus discriminant analysis: An application to Spanish insurance companies". In: *Intelligent Systems in Accounting and Finance, Huelva*, pp. 169–185.
- Martino, Pierluigi et al. (2020). "An introduction to blockchain, cryptocurrency and initial coin offerings". In: *New frontiers in entrepreneurial finance research*. World Scientific, pp. 181–206.
- Matthews, Brian W (1975). "Comparison of the predicted and observed secondary structure of T4 phage lysozyme". In: *Biochimica et Biophysica Acta (BBA)-Protein Structure* 405.2, pp. 442–451.
- Mayes, DG and H Stremmel (2014). "The effectiveness of capital adequacy measures in predicting bank distress: SUERF Study 2014/1". In: *Brussels: Larcier*.
- McNeil, Alexander J and Rüdiger Frey (2000). "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach". In: *Journal of empirical finance* 7.3-4, pp. 271–300.
- McNeil, Alexander J, Rüdiger Frey, and Paul Embrechts (2015). *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press.
- Mcskassy, Sofus and Foster Provost (2005). "Suspicion scoring based on guilt-by-association, collective inference, and focused". In.
- Mello, Rodrigo F and Moacir Antonelli Ponti (2018). *Machine learning: a practical approach on the statistical learning theory*. Springer.
- Meng, Qi et al. (2016). "A communication-efficient parallel algorithm for decision tree". In: *Advances in Neural Information Processing Systems* 29.
- Mensah, Jones Odei and Paul Alagidede (2017). "How are Africa's emerging stock markets related to advanced markets? Evidence from copulas". In: *Economic Modelling* 60, pp. 1–10.
- Merton, Robert C (1974). "On the pricing of corporate debt: The risk structure of interest rates". In: *The Journal of finance* 29.2, pp. 449–470.
- Miguel, Luis J de et al. (1993). "Application of the Fuzzy ARTMAP neural network architecture to bank failure predictions". In: *New Trends in Neural Computation: International Workshop on Artificial Neural Networks, IWANN'93 Sitges, Spain, June 9–11, 1993 Proceedings* 2. Springer, pp. 720–725.
- Min, Jae H and Young-Chan Lee (2005). "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters". In: *Expert systems with applications* 28.4, pp. 603–614.

- Min, Sung-Hwan, Jumin Lee, and Ingoo Han (2006). "Hybrid genetic algorithms and support vector machines for bankruptcy prediction". In: *Expert systems with applications* 31.3, pp. 652–660.
- Nassani, Abdelmohsen A et al. (2025). "Financial integration and economic growth: impact of renewable energy investments, technology transfer, and climate change on Europe and central Asian economies". In: *Financial Innovation* 11.1, p. 41.
- Nelsen, Roger B (2007). *An introduction to copulas*. Springer science & business media.
- Nelson, Daniel B (1991). "Conditional heteroskedasticity in asset returns: A new approach". In: *Econometrica: Journal of the econometric society*, pp. 347–370.
- Nier, Erlend et al. (2007). "Network models and financial stability". In: *Journal of Economic Dynamics and Control* 31.6, pp. 2033–2060.
- Nolde, Natalia and Johanna F Ziegel (2017). "Elicitability and backtesting: Perspectives for banking regulation". In.
- Odom, Marcus D and Ramesh Sharda (1990). "A neural network model for bankruptcy prediction". In: *1990 IJCNN International Joint Conference on neural networks*. IEEE, pp. 163–168.
- Ohlson, James A (1980). "Financial ratios and the probabilistic prediction of bankruptcy". In: *Journal of accounting research*, pp. 109–131.
- Olmeda, Ignacio and Eugenio Fernandez (1997). "Hybrid classifiers for financial multi-criteria decision making: The case of bankruptcy prediction". In: *Computational Economics* 10, pp. 317–335.
- Ortega, James M and James M Ortega (1987). *Matrix theory: A second course*. Springer Science & Business Media.
- Owusu Junior, Peterson, Imhotep Paul Alagidede, and Aviral Kumar Tiwari (2021). "On the elicibility and risk model comparison of emerging markets equities". In: *Mathematical and Computational Applications* 26.3, p. 63.
- Paliwal, Mukta and Usha A Kumar (2009). "Neural networks and statistical techniques: A review of applications". In: *Expert systems with applications* 36.1, pp. 2–17.
- Papadimitriou, Theophilos et al. (2013). "Forecasting the insolvency of US banks using support vector machines (SVMs) based on local learning feature selection". In: *International Journal of Computational Economics and Econometrics* 3.1-2, pp. 83–90.
- Papíková, Lenka and Mário Papík (2022). "Effects of classification, feature selection, and resampling methods on bankruptcy prediction of small and medium-sized enterprises". In: *Intelligent Systems in Accounting, Finance and Management* 29.4, pp. 254–281.
- Park, Min Sue et al. (2021). "Explainability of machine learning models for bankruptcy prediction". In: *Ieee Access* 9, pp. 124887–124899.
- Patel, Vaishali R and Rupa G Mehta (2011). "Impact of outlier removal and normalization approach in modified k-means clustering algorithm". In: *International Journal of Computer Science Issues (IJCSI)* 8.5, p. 331.

- Patton, Andrew J (2006). "Modelling asymmetric exchange rate dependence". In: *International economic review* 47.2, pp. 527–556.
- Patton, Andrew J, Johanna F Ziegel, and Rui Chen (2019). "Dynamic semiparametric models for expected shortfall (and value-at-risk)". In: *Journal of econometrics* 211.2, pp. 388–413.
- Pérez-Godoy, María Dolores et al. (2010). "Analysis of an evolutionary RBFN design algorithm, CO2RBFN, for imbalanced data sets". In: *Pattern Recognition Letters* 31.15, pp. 2375–2388.
- Phillips, Peter CB and Pierre Perron (1988). "Testing for a unit root in time series regression". In: *biometrika* 75.2, pp. 335–346.
- Pierret, Diane (2015). "Systemic risk and the solvency-liquidity nexus of banks". In: *Available at SSRN* 2346606.
- Pilch, Bartłomiej (2021). "An analysis of the effectiveness of bankruptcy prediction models—an industry approach". In: *Folia Oeconomica Stetinensia* 21.2, pp. 76–96.
- Piramuthu, Selwyn, Harish Ragavan, and Michael J Shaw (1998). "Using feature construction to improve the performance of neural networks". In: *Management Science* 44.3, pp. 416–430.
- Pollard, DB (1973). *GENERALISED INVERSE OF MATRICES AND ITS APPLICATIONS-RAO, CR AND MITRA, SK.*
- Poon, Ser-Huang and Clive W J Granger (2003). "Forecasting volatility in financial markets: A review". In: *Journal of economic literature* 41.2, pp. 478–539.
- Powers, David MW (2020). "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation". In: *arXiv preprint arXiv:2010.16061*.
- Pramodh, C and Vadlamani Ravi (2007). "Modified great deluge algorithm based auto associative neural network for bankruptcy prediction in banks". In: *International journal of computational intelligence research* 3.4, pp. 363–371.
- Provost, Foster J, Tom Fawcett, Ron Kohavi, et al. (1998). "The case against accuracy estimation for comparing induction algorithms." In: *ICML*. Vol. 98, pp. 445–453.
- Purvinis, Ojaras, Povilas Šukys, and Rūta Virbickaitė (2005). "Research of possibility of bankruptcy diagnostics applying neural network". In: *Inžinerinė ekonomika* 1, pp. 16–22.
- Qamruzzaman, Md (2023). "Does financial innovation foster financial inclusion in Arab world? examining the nexus between financial innovation, FDI, remittances, trade openness, and gross capital formation". In: *PloS one* 18.6, e0287475.
- Rafiei, F Mokhtab, SM Manzari, and S Bostanian (2011). "Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence". In: *Expert systems with applications* 38.8, pp. 10210–10217.
- Ragavan, Harish et al. (1993). "Complex concept acquisition through directed search and feature caching". In: *IJCAI*. Citeseer, pp. 946–951.

- Ranka, Sanjay and Vineet Singh (1998). "CLOUDS: A decision tree classifier for large datasets". In: *Proceedings of the 4th knowledge discovery and data mining conference*. Vol. 2. 8.
- Rao, CR and Y Wu (2005). "Linear model selection by cross-validation". In: *Journal of Statistical Planning and Inference* 128.1, pp. 231–240.
- Raschka, Sebastian (2015). *Python machine learning*. Packt publishing ltd.
- Ravi, Vadlamani et al. (2008a). "A semi-online training algorithm for the radial basis function neural networks: applications to bankruptcy prediction in banks". In: *Advances in banking technology and management: Impacts of ICT and CRM*. IGI Global, pp. 243–260.
- Ravi, Vadlamani et al. (2008b). "Soft computing system for bank performance prediction". In: *Applied soft computing* 8.1, pp. 305–315.
- Rebonato, Riccardo (2010). *Coherent stress testing: A Bayesian approach to the analysis of financial stress*. John Wiley & Sons.
- Reboredo, Juan C and Andrea Ugolini (2015). "Systemic risk in European sovereign debt markets: A CoVaR-copula approach". In: *Journal of International Money and Finance* 51, pp. 214–244.
- Refaeilzadeh, Payam, Lei Tang, Huan Liu, et al. (2009). "Cross-validation." In: *Encyclopedia of database systems* 5, pp. 532–538.
- Rockafellar, R Terry, Johannes O Royset, and Sofia I Miranda (2014). "Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk". In: *European Journal of Operational Research* 234.1, pp. 140–154.
- Romano, Joseph P and Michael Wolf (2005). "Stepwise multiple testing as formalized data snooping". In: *Econometrica* 73.4, pp. 1237–1282.
- Rose, Andrew K and Mark M Spiegel (2010). "Cross-country causes and consequences of the 2008 crisis: International linkages and American exposure". In: *Pacific Economic Review* 15.3, pp. 340–363.
- Rossi, André Luis Debiaso and André CPLF de Carvalho (2008). "Bio-inspired optimization techniques for svm parameter tuning". In: *2008 10th Brazilian symposium on neural networks*. IEEE, pp. 57–62.
- Saafi, Sami, Meriem Bel Haj Mohamed, and Makram Ben Doudou (2016). "Causal nexus between financial integration and economic growth: Does nonlinearity matter?" In: *Journal of Economic Integration*, pp. 817–854.
- Safi, Salah Al-Deen, Pedro A Castillo, and Hossam Faris (2022). "Cost-sensitive meta-heuristic optimization-based neural network with ensemble learning for financial distress prediction". In: *Applied Sciences* 12.14, p. 6918.
- Saldías, Martín (2013). "Systemic risk analysis using forward-looking distance-to-default series". In: *Journal of Financial Stability* 9.4, pp. 498–517.
- Salo, Fadi et al. (2018). "Data mining techniques in intrusion detection systems: A systematic literature review". In: *IEEE Access* 6, pp. 56046–56058.

- Sánchez-Lasheras, Fernando et al. (2012). "A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy". In: *Expert Systems with Applications* 39.8, pp. 7512–7523.
- Saqlain, Syed Muhammad et al. (2019). "Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines". In: *Knowledge and Information Systems* 58, pp. 139–167.
- Schmitz, Stefan W, Michael Sigmund, and Laura Valderrama (2019). "The interaction between bank solvency and funding costs: A crucial effect in stress tests". In: *Economic Notes: Review of Banking, Finance and Monetary Economics* 48.2, e12130.
- Schuermann, Til (2014). "Stress testing banks". In: *International Journal of Forecasting* 30.3, pp. 717–728.
- Schwarz, Gideon (1978). "Estimating the dimension of a model". In: *The annals of statistics*, pp. 461–464.
- Sedunov, John (2016). "What is the systemic risk exposure of financial institutions?" In: *Journal of Financial Stability* 24, pp. 71–87.
- Seeger, Matthias (2004). "Gaussian processes for machine learning". In: *International journal of neural systems* 14.02, pp. 69–106.
- Seiffert, Chris et al. (2007). "Mining data with rare events: a case study". In: *19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)*. Vol. 2. IEEE, pp. 132–139.
- Serrano-Cinca, Carlos and Begoña Gutiérrez-Nieto (2013). "Partial least square discriminant analysis for bankruptcy prediction". In: *Decision support systems* 54.3, pp. 1245–1255.
- Serre, D (2000). "Matrices: Theory and applications. 2002". In: *Graduate texts in mathematics*.
- Sevim, Cuneyt et al. (2014). "Developing an early warning system to predict currency crises". In: *European Journal of Operational Research* 237.3, pp. 1095–1104.
- Shah, Jaymeen R and Mirza B Murtaza (2000). "A neural network based clustering procedure for bankruptcy prediction". In: *American Business Review* 18.2, p. 80.
- Shapley, Lloyd S (1953). "Stochastic games". In: *Proceedings of the national academy of sciences* 39.10, pp. 1095–1100.
- Shapley, Lloyd S et al. (1953). "A value for n-person games". In.
- Shi, Yin and Xiaoni Li (2019). "A bibliometric study on intelligent techniques of bankruptcy prediction for corporate firms". In: *Heliyon* 5.12.
- Shibata, Ritei (1976). "Selection of the order of an autoregressive model by Akaike's information criterion". In: *Biometrika* 63.1, pp. 117–126.
- Shin, Kyung-Shik, Taik Soo Lee, and Hyun-jung Kim (2005). "An application of support vector machines in bankruptcy prediction model". In: *Expert systems with applications* 28.1, pp. 127–135.

- Silva, Thiago Christiano, Sergio Rubens Stancato Souza, and Benjamin Miranda Tabak (2017). "Monitoring vulnerability and impact diffusion in financial networks". In: *Journal of Economic Dynamics and Control* 76, pp. 109–135.
- Silva, Thiago Christiano et al. (2017). "A comparison of DEA and SFA using micro-and macro-level perspectives: Efficiency of Chinese local banks". In: *Physica A: Statistical Mechanics and its Applications* 469, pp. 216–223.
- Sinkey Jr, Joseph F (1975). "A multivariate statistical analysis of the characteristics of problem banks". In: *The Journal of Finance* 30.1, pp. 21–36.
- Snoek, Jasper, Hugo Larochelle, and Ryan P Adams (2012). "Practical bayesian optimization of machine learning algorithms". In: *Advances in neural information processing systems* 25.
- Sokolova, Marina, Nathalie Japkowicz, and Stan Szpakowicz (2006). "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation". In: *AI 2006: Advances in Artificial Intelligence: 19th Australian Joint Conference on Artificial Intelligence, Hobart, Australia, December 4-8, 2006. Proceedings* 19. Springer, pp. 1015–1021.
- Son, Hwijae et al. (2019). "Data analytic approach for bankruptcy prediction". In: *Expert Systems with Applications* 138, p. 112816.
- Song, Yuegang et al. (2021). "Does economic integration lead to financial market integration in the Asian region?" In: *Economic Analysis and Policy* 69, pp. 366–377.
- Subasi, Abdulhamit and M Ismail Gursoy (2010). "EEG signal classification using PCA, ICA, LDA and support vector machines". In: *Expert systems with applications* 37.12, pp. 8659–8666.
- Subramoney, Stephanie Danielle, Knowledge Chinhamu, and Retius Chifurira (2021). "Value at Risk estimation using GAS models with heavy tailed distributions for cryptocurrencies". In: *International Journal of Finance & Banking Studies* 10.4, pp. 40–54.
- Summer, Martin (2003). "Banking regulation and systemic risk". In: *Open economies review* 14, pp. 43–70.
- Sun, Jie et al. (2014). "Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches". In: *Knowledge-based systems* 57, pp. 41–56.
- Sun, Yanmin et al. (2007). "Cost-sensitive boosting for classification of imbalanced data". In: *Pattern recognition* 40.12, pp. 3358–3378.
- Suthar, Bhavisha, Hemant Patel, and Ankur Goswami (2012). "A survey: classification of imputation methods in data mining". In: *International Journal of Emerging Technology and Advanced Engineering* 2.1, pp. 309–12.
- Swicegood, Philip and Jeffrey A Clark (2001). "Off-site monitoring systems for predicting bank underperformance: a comparison of neural networks, discriminant analysis, and professional human judgment". In: *Intelligent Systems in Accounting, Finance & Management* 10.3, pp. 169–186.

- Syarif, Iwan, Adam Prugel-Bennett, and Gary Wills (2016). "SVM parameter optimization using grid search and genetic algorithm to improve classification performance". In: *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 14.4, pp. 1502–1509.
- Taffler, Richard J (1983). "The assessment of company solvency and performance using a statistical model". In: *Accounting and Business Research* 13.52, pp. 295–308.
- Tam, Kar Yan (1991). "Neural network models and the prediction of bank bankruptcy". In: *Omega* 19.5, pp. 429–445.
- Tam, Kar Yan and Melody Y Kiang (1992). "Managerial applications of neural networks: the case of bank failure predictions". In: *Management science* 38.7, pp. 926–947.
- Taskinsoy, John (2022). "Stress Testing Financial Systems: Macro and Micro Stress Tests, Basel Standards and Value-at-Risk as Financial Stability Measures". In: *Basel Standards and Value-at-Risk as Financial Stability Measures (February 11, 2022)*.
- Taylor, James W (2019). "Forecasting value at risk and expected shortfall using a semi-parametric approach based on the asymmetric Laplace distribution". In: *Journal of Business & Economic Statistics* 37.1, pp. 121–133.
- Theodossiou, Panayiotis (1998). "Financial data and the skewed generalized t distribution". In: *Management science* 44.12-part-1, pp. 1650–1661.
- Thomson, Mary E et al. (2019). "Combining forecasts: Performance and coherence". In: *International Journal of Forecasting* 35.2, pp. 474–484.
- Tian, Yingjie, Yong Shi, and Xiaohui Liu (2012). "Recent advances on support vector machines research". In: *Technological and economic development of Economy* 18.1, pp. 5–33.
- Tobias, Adrian and Markus K Brunnermeier (2016). "CoVaR". In: *The American Economic Review* 106.7, p. 1705.
- Troster, Victor et al. (2019). "Bitcoin returns and risk: A general GARCH and GAS analysis". In: *Finance Research Letters* 30, pp. 187–193.
- Tsai, Chih-Fong (2008). "Financial decision support using neural networks and support vector machines". In: *Expert Systems* 25.4, pp. 380–393.
- Tsai, Chih-Fong, Yu-Feng Hsu, and David C Yen (2014). "A comparative study of classifier ensembles for bankruptcy prediction". In: *Applied Soft Computing* 24, pp. 977–984.
- Tsai, Chih-Fong and Jhen-Wei Wu (2008). "Using neural network ensembles for bankruptcy prediction and credit scoring". In: *Expert systems with applications* 34.4, pp. 2639–2649.
- Tsai, Chih-Fong et al. (2021). "Combining feature selection, instance selection, and ensemble classification techniques for improved financial distress prediction". In: *Journal of Business Research* 130, pp. 200–209.
- Tsakonas, Athanasios et al. (2006). "Bankruptcy prediction with neural logic networks by means of grammar-guided genetic programming". In: *Expert Systems with Applications* 30.3, pp. 449–461.

- Tseng, Fang-Mei and Yi-Chung Hu (2010). "Comparing four bankruptcy prediction models: Logit, quadratic interval logit, neural and fuzzy neural networks". In: *Expert systems with applications* 37.3, pp. 1846–1853.
- Van Buuren, Stef and Karin Groothuis-Oudshoorn (2011). "mice: Multivariate imputation by chained equations in R". In: *Journal of statistical software* 45, pp. 1–67.
- Van Hulse, Jason and Taghi Khoshgoftaar (2009). "Knowledge discovery from imbalanced and noisy data". In: *Data & Knowledge Engineering* 68.12, pp. 1513–1542.
- Van Hulse, Jason, Taghi M Khoshgoftaar, and Amri Napolitano (2007). "Experimental perspectives on learning from imbalanced data". In: *Proceedings of the 24th international conference on Machine learning*, pp. 935–942.
- Vapnik, Vladimir (1999). *The nature of statistical learning theory*. Springer science & business media.
- (2013). *The nature of statistical learning theory*. Springer science & business media.
- Veganzones, David and Eric Severin (2021). "Corporate failure prediction models in the twenty-first century: a review". In: *European Business Review* 33.2, pp. 204–226.
- Vellido, Alfredo, Paulo JG Lisboa, and J Vaughan (1999). "Neural networks in business: a survey of applications (1992–1998)". In: *Expert Systems with applications* 17.1, pp. 51–70.
- Vinals, Jose (2012). *Macrofinancial Stress Testing-Principles and Practices*, IMF, Monetary and Capital Markets Department.
- Von Luxburg, Ulrike and Bernhard Schölkopf (2011). "Statistical learning theory: Models, concepts, and results". In: *Handbook of the History of Logic*. Vol. 10. Elsevier, pp. 651–706.
- Wang, Dehua, Yang Zhang, and Yi Zhao (2017). "LightGBM: an effective miRNA classification method in breast cancer patients". In: *Proceedings of the 2017 international conference on computational biology and bioinformatics*, pp. 7–11.
- Weiss, Gary M, Kate McCarthy, and Bibi Zabar (2007). "Cost-sensitive learning vs. sampling: Which is best for handling unbalanced classes with unequal error costs?" In: *Dmin* 7.35–41, p. 24.
- Wen, Fenghua et al. (2020). "Asymmetric relationship between carbon emission trading market and stock market: Evidences from China". In: *Energy Economics* 91, p. 104850.
- Wen, Fenghua et al. (2021). "Dynamic volatility spillovers and investment strategies between the Chinese stock market and commodity markets". In: *International Review of Financial Analysis* 76, p. 101772.
- West, David, Scott Dellana, and Jingxia Qian (2005). "Neural network ensemble strategies for financial decision applications". In: *Computers & operations research* 32.10, pp. 2543–2559.
- West, Robert Craig (1985). "A factor-analytic approach to bank condition". In: *Journal of Banking & Finance* 9.2, pp. 253–266.
- White, Halbert (2000). "A reality check for data snooping". In: *Econometrica* 68.5, pp. 1097–1126.

- Wiginton, John C (1980). "A note on the comparison of logit and discriminant models of consumer credit behavior". In: *Journal of Financial and Quantitative Analysis* 15.3, pp. 757–770.
- Wirnkar, Alphonsius D and Muhammad Tanko (2008). "CAMELS and banks performance evaluation: The way forward". In: *Retrieved May 24*, p. 2011.
- Wong, Bo K, Vincent S Lai, and Jolie Lam (2000). "A bibliography of neural network business applications research: 1994–1998". In: *Computers & Operations Research* 27.11–12, pp. 1045–1076.
- Wu, Chunchi and Xu-Ming Wang (2000). "A neural network approach for analyzing small business lending decisions". In: *Review of Quantitative Finance and Accounting* 15, pp. 259–276.
- Xu, Wei et al. (2014). "Financial ratio selection for business failure prediction using soft set theory". In: *Knowledge-Based Systems* 63, pp. 59–67.
- Yadav, Miklesh Prasad et al. (2023). "The domino effect: analyzing the impact of Silicon Valley Bank's fall on top equity indices around the world". In: *Finance Research Letters* 55, p. 103952.
- Yao, Jingxiu and Martin Shepperd (2020). "Assessing software defection prediction performance: Why using the Matthews correlation coefficient matters". In: *Proceedings of the evaluation and assessment in software engineering*, pp. 120–129.
- Yu, Keming and Jin Zhang (2005). "A three-parameter asymmetric Laplace distribution and its extension". In: *Communications in Statistics—Theory and Methods* 34.9-10, pp. 1867–1879.
- Yuan, Sheng-Fa and Fu-Lei Chu (2007). "Fault diagnostics based on particle swarm optimisation and support vector machines". In: *Mechanical Systems and Signal Processing* 21.4, pp. 1787–1798.
- Zakoian, Jean-Michel (1994). "Threshold heteroskedastic models". In: *Journal of Economic Dynamics and control* 18.5, pp. 931–955.
- Zayernyuk, Victor Makarovich et al. (2015). "Stress testing as a tool for assessing systemic risk organizations Russian banking sector". In: *Mediterranean Journal of Social Sciences* 6.3, S3.
- Zhang, Qingwen et al. (2018). "Improvement of Makkink model for reference evapotranspiration estimation using temperature data in Northwest China". In: *Journal of Hydrology* 566, pp. 264–273.
- Zhang, Wenjie, Hao Quan, and Dipti Srinivasan (2018). "Parallel and reliable probabilistic load forecasting via quantile regression forest and quantile determination". In: *Energy* 160, pp. 810–819.
- Zhang, Zhichao et al. (2016). "Determinants of financial distress in large financial institutions: Evidence from US bank holding companies". In: *Contemporary Economic Policy* 34.2, pp. 250–267.

- Zhao, Huimin, Atish P Sinha, and Wei Ge (2009). "Effects of feature construction on classification performance: An empirical study in bank failure prediction". In: *Expert Systems with Applications* 36.2, pp. 2633–2644.
- Zhou, Qiyun et al. (2023). "Examining the trade-offs in clean energy provision: Focusing on the relationship between technology transfer, renewable energy, industrial growth, and carbon footprint reduction". In: *Heliyon* 9.10.
- Zhu, Dongming and John W Galbraith (2010). "A generalized asymmetric Student-t distribution with application to financial econometrics". In: *Journal of Econometrics* 157.2, pp. 297–305.
- (2011). "Modeling and forecasting expected shortfall with the generalized asymmetric Student-t and asymmetric exponential power distributions". In: *Journal of Empirical Finance* 18.4, pp. 765–778.
- Zhu, Qiuming (2020). "On the performance of Matthews correlation coefficient (MCC) for imbalanced dataset". In: *Pattern Recognition Letters* 136, pp. 71–80.
- Zhuo, Li et al. (2008). "A genetic algorithm based wrapper feature selection method for classification of hyperspectral images using support vector machine". In: *Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Classification of Remote Sensing Images*. Vol. 7147. SPIE, pp. 503–511.