

# Contents

<b>Figures</b>	<b>xvii</b>
<b>Tables</b>	<b>xix</b>
<b>Variables</b>	<b>xxi</b>
<b>Abbreviations</b>	<b>xxiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation and Objectives . . . . .	1
1.2 Course of Investigation . . . . .	6
<b>2 Data and LGD Estimation</b>	<b>8</b>
2.1 Data . . . . .	8
2.2 LGD Estimation . . . . .	10
2.3 Prediction Accuracy Measurements . . . . .	17
<b>3 Heterogeneities among LGD Distributions: The Modality Defines the Best Estimation Method</b>	<b>19</b>
3.1 Fundamentals and Research Questions . . . . .	19
3.2 LGD Estimation and Cluster Analysis . . . . .	21
3.2.1 Data . . . . .	21
3.2.2 LGD Estimation Methods . . . . .	23
3.2.3 Clustering and LGD Distribution Analysis . . . . .	24
3.3 Comparative Analysis . . . . .	26
3.3.1 Model Comparison Procedure . . . . .	26
3.3.2 Hyperparameter Tuning . . . . .	27
3.3.3 Empirical Results . . . . .	31
3.4 Robustness Checks . . . . .	35
3.4.1 Inclusion of Enterprise-specific Variables . . . . .	36
3.4.2 Clustering based on Loan-specific Variable . . . . .	38
3.4.3 Logarithmic Transformation of the Positively Skewed Unimodally Distributed LGDs . . . . .	41
3.4.4 Non-European Credit Portfolios . . . . .	43
3.5 Interim Results . . . . .	45

<b>4</b>	<b>Tuning White Box Model with Black Box Models: Transparency in LGD Modeling</b>	<b>46</b>
4.1	Fundamentals and Research Questions . . . . .	46
4.2	Model . . . . .	50
4.3	Monte Carlo Experiment . . . . .	52
4.3.1	Data . . . . .	53
4.3.2	Optimized Regression Model . . . . .	53
4.3.3	Model Comparison . . . . .	55
4.4	Empirical Framework . . . . .	57
4.4.1	Data . . . . .	57
4.4.2	Competitive Methods . . . . .	60
4.4.3	Empirical Setup . . . . .	61
4.5	Empirical Results . . . . .	62
4.5.1	Variable Selection for Optimized Linear Regression . . . . .	62
4.5.2	In-sample Results . . . . .	64
4.5.3	Comparative Analysis . . . . .	69
4.6	Robustness Checks . . . . .	72
4.6.1	Change in the Split Ratio . . . . .	72
4.6.2	European Credit Portfolio . . . . .	74
4.7	Interim Results . . . . .	75
<b>5</b>	<b>Machine Learning-Based Variable Selection for Clustered Credit Risk Modeling</b>	<b>77</b>
5.1	Fundamentals and Research Questions . . . . .	77
5.2	Optimized Clustered Model . . . . .	80
5.3	Empirical Framework . . . . .	82
5.3.1	Data . . . . .	83
5.3.2	Competitive Modeling Approaches . . . . .	83
5.3.3	Empirical Setup . . . . .	85
5.4	Empirical Results . . . . .	85
5.4.1	Variable Importance Measure . . . . .	85
5.4.2	Cluster Analysis . . . . .	87
5.4.3	Comparative Analysis . . . . .	90
5.5	Robustness Check . . . . .	92
5.6	Interim Results . . . . .	97
<b>6</b>	<b>Conclusion</b>	<b>98</b>
6.1	Summary . . . . .	98
6.2	Outlook . . . . .	100

<b>References</b>	<b>102</b>
<b>Appendix</b>	<b>A-1</b>
Appendix A (Chapter 2) . . . . .	A-1
Appendix B (Chapter 3) . . . . .	A-5
Appendix C (Chapter 4) . . . . .	A-22
Appendix D (Chapter 5) . . . . .	A-35