WISSENSCHAFTLICHE BEITRÄGE
AUS DEM TECTUM VERLAG

Reihe Wirtschaftswissenschaften

Band 74

Bernd Galler

Scarce Data based Credit Risk Assessment

Evaluating Missing Data Methods in PD-estimation by Logistic Regression

Tectum Verlag

Contents

1 Introduction 1
- 1.1 Problem . . . 2
- 1.2 Process of Investigation . . . 7

2 Credit Risk Assessment 9
- 2.1 EAD . . . 10
- 2.2 LGD . . . 11
- 2.3 PD . . . 12
  - 2.3.1 Estimating PD . . . 14

3 Missing Data Theory 27
- 3.1 Classification of and Causes for Missing Data . . . 27
- 3.2 Goals of Missing Data Methods . . . 30
- 3.3 Mechanism of Missingness . . . 31
- 3.4 Missing Data Analysis . . . 37
- 3.5 Missing Data Methods . . . 42
  - 3.5.1 Case Deletion . . . 42
  - 3.5.2 Mean Imputation . . . 44
  - 3.5.3 Median Imputation . . . 46
  - 3.5.4 Empiric Distribution Based Imputation . . . 47
  - 3.5.5 Multiple Imputation . . . 49
  - 3.5.6 Maximum Likelihood . . . 56

4 Evidence from Financial Statements of Bank Customers 67
- 4.1 Data . . . 69
  - 4.1.1 Original Data Description . . . 69
  - 4.1.2 Data Selection . . . 70
  - 4.1.3 Selected Data Description . . . 73
- 4.2 Reference Model Ratios (RMR) . . . 79
  - 4.2.1 Calculation of Financial Ratios . . . 79
  - 4.2.2 Data Cleansing and Data Transformation . . . 83
  - 4.2.3 Univariate Discriminatory Power . . . 84
  - 4.2.4 Economic Hypotheses . . . 86

4.2.5 Feature Selection 87
4.3 MDM Data Preparation 93
4.3.1 Generating Synthetic Complete and Missing Data . 93
4.3.2 Transformation of Reference Ratio Distributions . . 99
4.4 MDM Evaluation 100
4.4.1 Implementation 101
4.4.2 β-Estimates 109
4.4.3 Standard Errors 125
4.4.4 PD Estimates 141

5 Discussion **155**
5.1 Influence of MDMs on the Assessment of Credit Risk . . . 157
5.2 Efficient Data Use of Particular MDMs 159
5.3 Performance of MDMs 160
5.3.1 Regression Coefficients 161
5.3.2 Standard Errors 165
5.3.3 Probabilites of Default 167
5.3.4 Concluding remarks on RQ 3a and RQ 3b 170
5.4 Limitations 172
5.5 Future Research 174

6 Conclusion **175**