An end-to- end Credit Risk Model with Probability of Default

Authors Naveen Kumar V, Praveen Gujjar

Faculty of Management Studies

CMS Business School

JAIN (Deemed-to-be University), Bengaluru, India

Faculty of Management Studies

CMS Business School

JAIN (Deemed-to-be University), Bengaluru, India

Abstract

This study provides a composite framework for measuring credit risk with particular emphasis on the probability of default (PD) as a reference. We examine how various factors come together to impact PD and propose an integrated framework that combines macroeconomic variables, borrower-level variables, and market sentiment. The results provide key determinants of PD and have implications for risk management and policy-making.

Keywords: Credit Risk, Probability of Default, Risk Management, Financial Regulations, Credit Indices, Loan Pricing

1. Introduction

Quantification of credit risk in terms of probability of default (PD) is essential to the making of financial decisions. PD gives a numerical figure of the probability that a borrower would default on his dues. The present study attempts to encompass many aspects of credit risk under one construct and highlight the interdependence of these macroeconomic variables, borrower attributes, and market sentiment.

1.1 Study Objectives

- To determine the most impactful factors of PD.

- To construct a composite model for PD forecasting.

- To analyze the implications of the proposed model for credit risk management.

2. Literature Review

2.1 Background

Literature considers PD as a pillar of credit risk modeling. Research works like Altman (1968) and Merton (1974) have established the foundation for quantitative models, focusing on financial ratios and market-based methods.

2.2 Recent Developments

Recent research integrates machine learning models and big data analysis to enhance the precision of PD forecasting. Credit index contribution, such as the Loan-Only Credit Default Index (LCDX), is particularly emphasized for syndicated loan markets when it comes to forecasting (Ashcraft & Santos, 2009).

2.3 Literature Gaps

Even with advancement, existing models are lacking in terms of integrating an overall framework taking into consideration macroeconomic trends and borrower-specific risks. The present study attempts to plug this research gap by proposing a composite construct.

3. Methodology

3.1 Data Sources

The study employs the following data:

- Macroeconomic data: GDP growth, interest rates, inflation, and employment rates.

- Borrower-specific data: Balance sheets, credit scores, debt-to-income ratios, and payment behavior.

- Market sentiment: Imputed from credit spreads, indices such as LCDX, and trading volumes.

Data were gathered from publicly available finance databases such as Thomson Reuters, Markit, and World Bank economic data indicators.

3.2 Model Framework

The composite model incorporates:

- Logistic Regression: For binary default results.

- Principal Component Analysis (PCA): For dimensionality reduction and finding principal variables.

- Machine Learning Algorithms: Random forests and gradient boosting for sensitivity analysis.

3.3 Hypotheses

- H1: Macroeconomic variables have a significant impact on PD.

- H2: Borrower-specific variables are predictive of PD.

- H3: Market sentiment gives preliminary cues on PD changes.

4. Results

4.1 Descriptive Statistics

Information contains:

- Macroeconomic Indicators: GDP growth of -2% to 6% per annum during the study period, with average inflation of 3%.

- Borrower Data: Average debt-to-income of 35%, median credit score of 680.

- Market Sentiment: LCDX spreads of 200 to 600 basis points.

4.2 Key Findings

- Macroeconomic Variables: A 1% gain in GDP growth reduces PD by 0.5%.

- Borrower-Specific Variables: Leverage and volatility of cash flows are both positively related to PD ($p < 0.01$).

- Market Sentiment: An increase of 10 basis points in credit spreads forecasts an increase of 0.3% in PD.

4.3 Model Performance

Composite model exhibits:

- Accuracy: 87%.

- Precision: 82%.

- Recall: 85%.

Comparative analysis indicates the model is 15-20% better than conventional methods as far as predictive accuracy is concerned.

4.4 Sensitivity Analysis

Subgroup analysis indicates that the model is very effective for high-risk borrowers with a 90% accuracy level in forecasting defaults for this subgroup.

5. Data and Models

Table 1: Macroeconomic Indicators Summary

Indicator	**Mean**	**Std. Dev.**	**Min**	**Max**
GDP Growth (%)	2.5	1.8	-2.0	6.0
Inflation (%)	3.0	0.9	1.5	4.5
Unemployment (%)	5.2	1.2	3.8	7.4

Table 2: Borrower Financial Characteristics

Metric	Mean	Std. Dev.	Min	Max
Debt-to-Income Ratio	35%	10%	20%	50%
Credit Score	680	50	600	750
Leverage Ratio	40%	15%	20%	60%

Table 3: LCDX Spreads

Period	Mean Spread (bps)	Std. Dev.	Min Spread (bps)	Max Spread (bps)
2010-2015	350	120	200	500
2016-2020	420	140	250	600

Table 4: Model Inputs and Weights

Variable	Weight (%)

GDP Growth	25
Leverage Ratio	30
Credit Spreads	20
Cash Flow Volatility	15
Inflation	10

Table 5: Logistic Regression Coefficients

Variable	Coefficient	Std. Error	p-value
GDP Growth (%)	-0.5	0.1	<0.01
Leverage Ratio	1.2	0.3	<0.01
Credit Spreads (bps)	0.03	0.01	0.02

Table 6: Model Accuracy Metrics

Metric	Composite Model	Traditional Model
Accuracy (%)	87	72
Precision (%)	82	68
Recall (%)	85	70

Table 7: Subgroup Analysis - High-Risk Borrowers

Metric	Accuracy (%)
High Leverage	90
Low Credit Score	88

Table 8: Subgroup Analysis - Low-Risk Borrowers

Metric	Accuracy (%)

Low Leverage	80
High Credit Score	85

Table 9: Sensitivity Analysis

Scenario	PD Change (%)
+1% GDP Growth	-0.5
+10 bps Spreads	+0.3

6. Discussion

6.1 Implications for Risk Management

The composite construct provides a multi-dimensional view of credit risk, and banks are able to:

- Improve Decision-Making: Integrate various sources of information to provide better risk estimates.

- Pre-emptive Action: Modify lending policies with respect to forecasted market conditions.

6.2 Regulatory Implications

- Identification of Systemic Risk: Regulators are able to identify potential risks using the model.

- Policy Formulation: Recommendations derived from the model can be used to determine capital adequacy and stress testing frameworks.

6.3 Limitations and Future Research

- Data Scope: The research is based on older, secured loans, and the findings may not be extrapolated to other types of loans.

- Temporal Context: The results are derived from 2010-2020, and future research must examine what follows.

7. Conclusion

This study emphasizes the need for a comprehensive credit risk assessment technique. The suggested composite model's better performance in PD prediction and its importance in financial institutions and regulatory practices are highlighted. In subsequent work, the application scope of the model can be extended to other loan categories and under various economic conditions to check its generalizability.

Acknowledgements

The authors acknowledge the Institute of Risk Management for the research support and colleagues for input.

References

1. Altman, E. (1968). Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy. Journal of Finance.

2. Ashcraft, A. B., & Santos, J. A. C. (2009). Has the CDS Market Lowered the Cost of Corporate Debt? Journal of Monetary Economics.

3. Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. Journal of Finance.