Moody's Risk Management Services
Rare Event Modeling and Validation Through Time:

The case of corporate credit analysis

Roger M. Stein
Managing Director
Credit Risk Analytics
Moody’s Risk Management Services
steinr@moodys.com
The first principle is that you must not fool yourself -- and you are the easiest person to fool.

- Richard Feynman
Overview

• Brief introduction to the RiskCalc default model
• Discussion of validation and backtesting in finance
• Differences between validating market- and credit-related models
• A validation approach for sparse data sets
• Examples of problems that arise from violating the approach
• Conclusion
• Research Co-authors
  • V. Dhar
  • S. Keenan
  • J. Sobehart
The corporate credit problem

- What is the probability of “default” (PD) within a specified period of time?
- Uses of PD’s
  - Regulators
    - Basel, National bank regulators
  - Securitization
    - Collateralized Loan Obligations
- Credit Process
  - Decisioning (yes/no)
  - Monitoring (work-out, remedial action)
  - Provisioning
  - Pricing
  - Incentive compensation
- Related problems
  - Recovery (loss given default)
  - Correlation of default rates and arrival times
  - State transition
RiskCalc™ modeling approach

- **Transform**
  - Ratios transformed from unwieldy broad distributions to more uniform and predictive variants
  - Micro-modeling used to capture useful aspects of behavior and to decompose problem

- **Model**
  - Transformed variables weighted statistically to produce default scoring model

- **Map**
  - Model output (score) converted to PD by non-parametrically mapping into historical population default estimates
The components of Moody’s modeling approach

- Structural model (Merton variant)
  - Distance to distress
- Rating information (where available)
  - Moody’s rating or quantitative rating estimate
- Financial statement information
  - Leverage, Liquidity, Size, Profitability, etc.
- Non-linear statistical regression
  - Simple neural network
- Mapping result to empirical probability of default (PD) and adjusting for prior probabilities
- Extensive validation
  - Out-of-sample / out-of-time (walk forward analysis)
  - Multidimensional metrics
Distance to Distress: Equity as a Call Option

1. Calculate the firm’s obligations (CL, LTD)
2. Use equity information to estimate:
   • a) market value of the firm’s assets (MVA)
   • b) volatility of assets
   This is done with a variant of the Merton model:
   - Market Equity = Present Value (Residual Value of the Firm)
   - Stock Volatility = Leveraged Volatility of Assets
3. Calculate
   Distance to Distress = \( \frac{(MVA - (1/2 \text{ LTD} + CL))/(\text{volatility} \times MVA)}{\text{MVA}} \)
Mapping score to PD
Predictive power of financial statement data

Historical Default Rate

ROA percentile
Univariate Performance of Variables

Distance to Default percentile

ROA percentile

Size percentile

Tot. Liab / Tot. Assets percentile

Rating (or estimate) percentile

WC / Total Assets percentile

ROE percentile

Equity Growth percentile

Equity Volatility percentile

Moody’s Risk Management Services
Non-Linear Relationships:
ROA vs. Distance to Distress
Heuristic overview of the model

Risk Components:
- Credit Quality
- Profitability
- Liquidity
- Capital Structure
- Market Presence
- Market Equity

Model Variables:
- Agency Rating
- Return on Assets
- Return on Net Worth
- Working Capital / Total Assets
- Leverage Ratio
- Firm Size
- Stock Volatility
- Stock Return
- Distance to Distress
- Industry Averages

Variable Weights:

Default Risk Score

Default Risk Score = f(\sum w_j variable_j)
Validation Methodology
Moody’s view of the spectrum of validation

**Development**  
**Data Poor**

- Anecdotal cases
- Validating on small samples of “training” cases  
  - “number right”
- Validating on out-of-sample data  
  - “number right”

**Certification**  
**Data Rich**

- Bootstrapping out-of-sample out-of-time data  
  - “number right”
- Bootstrapping out-of-sample out-of-time data  
  - higher order statistics
- Bootstrapping out-of-sample out-of-time data  
  - higher order statistics w/ cost function

Moody’s Q is currently here
Validation

“...the area of validation will prove to be a key challenge for banking institutions in the foreseeable future.”

“Credit Risk Modeling Practices and Applications,” Basle Committee on Banking and Supervision, Basle, April, 1999, p. 50.
The components of our current approach to validation

- How to measure and calculate performance statistics
  - How to sample available data
  - How to use the data to achieve robust statistics
- What types of statistics to measure
  - Simple (hits vs. misses)
  - Measures of goodness based on geometry
  - Measures of information content and association based on entropy
  - Other measures (forthcoming)
Validation in Finance

• Backtesting dominates market research
  – Identify interesting relationship
  – Evaluate the (risk-adjusted) “profitability” of the relationship through simulated trading on historical data

• Backtesting requires long time series of relatively high frequency

• Backtesting is often not appropriate for lower frequency data or rare/long term events since not enough data exists to both build a model and test it
  • If more data are saved for testing, models tend to be mis-specified
  • If more data are used to parameterize a model, tests loose power: too few examples exist for meaningful inferences
Findings: Market character
Stop level and market character in favorable systems

Close-up of Cond.omni.60.08.01 for short trades on DM
Findings: Market character
Stop level and market character in random systems

Looser stops avoid minor retracements when Cond.e.v.v.10.9.1 obtains (long)
Unlike in the trading problem, corporate credit often involves separating “goods” from “bads”
Model Performance: Cumulative Accuracy

Cumulative Accuracy Profile

Population sorted by Model Score

Defaults (%)

Distribution of Populations

Decision axis

Population

High Risk

Low Risk

Moody’s Risk Management Services
But the distribution of “interesting” cases for the default problem is sparse

- **Data Set**
  - Moody’s Default and Ratings databases, Compustat, IDC
  - over 14,000 U.S. non-financial corporations
  - over 1,400 defaults
  - 1980 through 1999

- **Firm years**
  - Model Fitting: ~ 100,000
  - Validation: ~ 65,000

- **Population default rate:**
  - Actual: 1.6%
  - Sample: 1.1%
Power and Sample Size Related

80% w/in red range for sample size = 100

80% w/in blue range for sample size = 1000
Model Validation and Performance

- **Walk forward and K-fold methods**
  - Training sample versus validation sample
  - Out-of-sample and out-of-time validation

- **Empirical validation versus comparable tools**
  - Power statistics are sample biased
  - Performance can be truly assessed relative to a benchmark
  - Multi-dimensional performance measures

- **Use of large datasets**
  - Documented performance on large out-of-sample datasets
  - Testing that the model is not “overfitted”
Training set of firms taken at $t_0$

Validation set of original firms in training sample but taken at $t_j$

Validation set of new firms not in training sample and taken at $t_j$

Resampling Performance Statistics

Non-Defaulters

Defaulters

Training sample

<table>
<thead>
<tr>
<th>Non-Defaulter</th>
<th>Defaulter</th>
</tr>
</thead>
<tbody>
<tr>
<td>86% [15%]</td>
<td>14% [5%]</td>
</tr>
<tr>
<td>28% [15%]</td>
<td>72% [5%]</td>
</tr>
</tbody>
</table>

Cross-validation sample

<table>
<thead>
<tr>
<th>Non-Defaulter</th>
<th>Defaulter</th>
</tr>
</thead>
<tbody>
<tr>
<td>66% [15%]</td>
<td>15% [5%]</td>
</tr>
<tr>
<td>33% [15%]</td>
<td>67% [5%]</td>
</tr>
<tr>
<td>Model Name</td>
<td>Year</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Fitzpatrick</td>
<td>(32)</td>
</tr>
<tr>
<td>Beaver</td>
<td>(67)</td>
</tr>
<tr>
<td>Altman</td>
<td>(68)</td>
</tr>
<tr>
<td>Lev</td>
<td>(71)</td>
</tr>
<tr>
<td>Wilcox</td>
<td>(71)</td>
</tr>
<tr>
<td>Deakin</td>
<td>(72)</td>
</tr>
<tr>
<td>Edmister</td>
<td>(72)</td>
</tr>
<tr>
<td>Blum</td>
<td>(74)</td>
</tr>
<tr>
<td>Taffler</td>
<td>(74)</td>
</tr>
<tr>
<td>Libby</td>
<td>(75)</td>
</tr>
<tr>
<td>Diamond</td>
<td>(76)</td>
</tr>
<tr>
<td>Altman, Haldeman and Narayanan</td>
<td>(77)</td>
</tr>
<tr>
<td>Marais</td>
<td>(79)</td>
</tr>
<tr>
<td>Damboleana and Khoury</td>
<td>(80)</td>
</tr>
<tr>
<td>Ohlson</td>
<td>(80)</td>
</tr>
<tr>
<td>Taffler</td>
<td>(82, 83)</td>
</tr>
<tr>
<td>El Hennawy and Morris</td>
<td>(83a)</td>
</tr>
<tr>
<td>Moyer</td>
<td>(84)</td>
</tr>
<tr>
<td>Taffler</td>
<td>(84)</td>
</tr>
<tr>
<td>Zmijewski</td>
<td>(84)</td>
</tr>
<tr>
<td>Zavgren</td>
<td>(85)</td>
</tr>
<tr>
<td>Casey and Bartczak</td>
<td>(85)</td>
</tr>
<tr>
<td>Peel and Peel</td>
<td>(88)</td>
</tr>
<tr>
<td>Barniv and Raveh</td>
<td>(89)</td>
</tr>
<tr>
<td>Boothe and Hutchinson</td>
<td>(89)</td>
</tr>
<tr>
<td>Gupta, Rao, and Bagchi</td>
<td>(90)</td>
</tr>
<tr>
<td>Kease and McGuiness</td>
<td>(90)</td>
</tr>
<tr>
<td>Keasey, McGuiness and Short</td>
<td>(90)</td>
</tr>
<tr>
<td>Shumway</td>
<td>(96)</td>
</tr>
<tr>
<td>Moody’s RiskCalc Public Firm</td>
<td>(00)</td>
</tr>
<tr>
<td>Moody’s RiskCalc Private Firm</td>
<td>(00)</td>
</tr>
<tr>
<td>Median</td>
<td>40</td>
</tr>
</tbody>
</table>
Actual accuracy rate: 65% vs. 70%

N=25, wrong=43%

N=50, wrong=36%

N=100, wrong=18%

N=1000, wrong=2%
Examples of faulty inferences due to violations of transitivity of universe of.
Size Bias Makes Model Estimation, Testing, Difficult
Quick Ratio
Quick Ratio = (Current Assets - Inventory)/Current Liabilities

Higher rated companies have lower liquidity...

... yet lower liquidity implies higher default rates for both public and private companies.
Different universes:
All Models Do Better on Bigger Firms

- Rated
- Unrated
- Private
RiskCalc Validation Results
Relative Model Performance

Selected CAP curves

Models for public firms

Default Rate

Distance to Distress
Perfect Model (Ideal CAP)

Naïve Model (Random CAP)

Performance Differential

Model being evaluated

$\mathcal{M}$

Moody’s Risk Management Services
Accuracy Ratio (out-of-sample)

- Reduced Z’-score model
- Moody’s Public Firm
- Merton model variant
- Hazard model
- Z-score model
- ROA
Some reading

- RiskCalc documents Available at [www.moodysrms.com](http://www.moodysrms.com) to download Adobe Acrobat files
  - Navigate to “research”

- Some validation readings
We have found that validation can be done even with sparse data but is difficult particularly with sparse data. It is useful to carefully design validation experiments that test a model in simulated real-world environments controlling for time and universe. Meaningful benchmarks (not straw-men) are usually necessary for reference. Many validation tests are sensitive to the exact sample chosen: observed performance differences may be due to sampling issues particularly with rare events. There is little that we can do to increase the power in sparse data for validation. The best we can do is to acknowledge limitations and understand bounds.
Rare Event
Modeling and Validation
Through Time:
The case of corporate credit analysis

Roger M. Stein
Managing Director
Credit Risk Analytics
Moody’s Risk Management Services
steinr@moodys.com