



Contact	Phone
New York	
Eric Falkenstein	1.212.553.1653
Andrew Boral	
Lea V. Carty	

RISKCALC™ FOR PRIVATE COMPANIES: MOODY'S DEFAULT MODEL

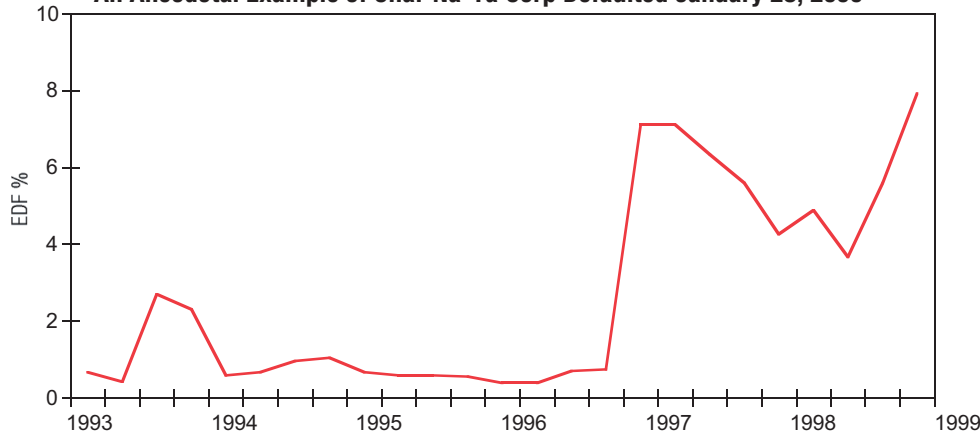
Rating Methodology Summary

This report describes and documents Moody's version of its RiskCalc™ default model for private firms. RiskCalc™ analyzes financial statement data to produce default probability predictions for corporate obligors - particularly those in the middle market. We discuss the model's derivation in detail, analyze its accuracy, and provide context for its application. The model's key advantage derives from Moody's unique and proprietary middle market private firm financial statement and default database (Credit Research Database), which comprises 28,104 companies and 1,604 defaults. Our main insights and conclusions are:

- The relationship between financial variables and default risk varies substantially between public and private firms. An important consequence of this is that default models based on public firm data and applied to private firms will likely misrepresent actual default risk.
- RiskCalc™ generates 1- and 5-year expected default frequencies, as well as a mapping into Moody's standard rating categories. Hence, the meaning of the model output is easily understood and amenable to benchmarking and quantitative portfolio risk management techniques.
- Comprehensive testing and validation suggest that RiskCalc's predictive power is superior to that of other publicly available benchmark models and is robust across non-financial industry sectors, and over time.
- RiskCalc™ was developed to achieve maximum predictive power with the smallest number of inputs. It requires just 10 financial ratios & indicators computed from 17 basic financial inputs.
- RiskCalc's predictive power derives, in part, from its meticulous transformation of input financial ratios, which are highly 'nonnormally' distributed, as well as the large number of defaulting private firms used in its estimation.

Exhibit 0.1

Moody's Default Model For Private Companies: An Anecdotal Example of Chai-Na-Ta Corp Defaulted January 28, 2000



The company, headquartered in British Columbia, is the world's largest producer of North American ginseng. The company farms, processes and distributes the root. Coming under continued pressure from depressed ginseng prices, the firm defaulted. The previous eight quarters of data were used for each RiskCalc value.

continued on page 3

Our goal for RiskCalc is to develop a benchmark that will facilitate transparency for what has traditionally been a very opaque asset class - commercial loans. Transparency, in turn, will facilitate better risk management, capital allocation, loan pricing, and securitization. RiskCalc is currently being used within Moody's to support its ratings of loan securitizations. To learn more, please visit www.moodyrms.com.

Author
Eric Falkenstein

Editor
Crystal Carrafiello

Associate Analyst
Andrew Boral

Production Associate
John Tzanos

© Copyright 2000 by Moody's Investors Service, Inc., 99 Church Street, New York, New York 10007. All rights reserved. ALL INFORMATION CONTAINED HEREIN IS COPYRIGHTED IN THE NAME OF MOODY'S INVESTORS SERVICE, INC. ("MOODY'S"), AND NONE OF SUCH INFORMATION MAY BE COPIED OR OTHERWISE REPRODUCED, REPACKAGED, FURTHER TRANSMITTED, TRANSFERRED, DISSEMINATED, REDISTRIBUTED OR RESOLD, OR STORED FOR SUBSEQUENT USE FOR ANY SUCH PURPOSE, IN WHOLE OR IN PART, IN ANY FORM OR MANNER OR BY ANY MEANS WHATSOEVER, BY ANY PERSON WITHOUT MOODY'S PRIOR WRITTEN CONSENT. All information contained herein is obtained by MOODY'S from sources believed by it to be accurate and reliable. Because of the possibility of human or mechanical error as well as other factors, however, such information is provided "as is" without warranty of any kind and MOODY'S, in particular, makes no representation or warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability or fitness for any particular purpose of any such information. Under no circumstances shall MOODY'S have any liability to any person or entity for (a) any loss or damage in whole or in part caused by, resulting from, or relating to, any error (negligent or otherwise) or other circumstance or contingency within or outside the control of MOODY'S or any of its directors, officers, employees or agents in connection with the procurement, collection, compilation, analysis, interpretation, communication, publication or delivery of any such information, or (b) any direct, indirect, special, consequential, compensatory or incidental damages whatsoever (including without limitation, lost profits), even if MOODY'S is advised in advance of the possibility of such damages, resulting from the use of or inability to use, any such information. The credit ratings, if any, constituting part of the information contained herein are, and must be construed solely as, statements of opinion and not statements of fact or recommendations to purchase, sell or hold any securities. **NO WARRANTY, EXPRESS OR IMPLIED, AS TO THE ACCURACY, TIMELINESS, COMPLETENESS, MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OF ANY SUCH RATING OR OTHER OPINION OR INFORMATION IS GIVEN OR MADE BY MOODY'S IN ANY FORM OR MANNER WHATSOEVER.** Each rating or other opinion must be weighed solely as one factor in any investment decision made by or on behalf of any user of the information contained herein, and each such user must accordingly make its own study and evaluation of each security and of each issuer and guarantor of, and each provider of credit support for, each security that it may consider purchasing, holding or selling. Pursuant to Section 17(b) of the Securities Act of 1933, MOODY'S hereby discloses that most issuers of debt securities (including corporate and municipal bonds, debentures, notes and commercial paper) and preferred stock rated by MOODY'S have, prior to assignment of any rating, agreed to pay to MOODY'S for appraisal and rating services rendered by it fees ranging from \$1,000 to \$1,500,000. PRINTED IN U.S.A.

Table of Contents

<i>Summary</i>	1
<i>Introduction</i>	9
<i>What We Will Cover</i>	10
<i>Section I: The Current Credit Risk Toolbox</i>	10
<i>Applications</i>	11
<i>Consumer Bureau Scores</i>	11
<i>Business Report Scores</i>	12
<i>Public Firm Models</i>	12
<i>Agency Ratings</i>	13
<i>Hazard Models</i>	13
<i>Exposure Models</i>	13
<i>Portfolio Models</i>	13
<i>Where RiskCalc Fits In</i>	13
<i>Section II: Past Studies And Current Theory Of Private Firm Default</i>	14
<i>Empirical Work</i>	14
<i>Theory Of Default Prediction: Comparing The Approached</i>	15
<i>Traditional Credit Analysis: Human Judgement</i>	15
<i>Structural Models</i>	17
<i>The Merton Model and KMV</i>	18
<i>The Gambler's Ruin</i>	18
<i>Equity Value vs. Cash Flow Measures: Is One Better?</i>	19
<i>Nonstructural Models</i>	19
<i>Appendix 2</i>	21
<i>Merton Model and Gambler's Ruin</i>	21
<i>Section III: Data</i>	22
<i>Public Company Data: Moody's Default Database and Compustat</i>	22
<i>Private Company Data: Moody's Credit Research Database (CRD)</i>	23
<i>Data Composition</i>	23
<i>Geographic Distribution of Data</i>	24
<i>Industrial Composition of Data</i>	24
<i>Financial Statement Quality</i>	25
<i>Sales Size</i>	26
<i>Financial Institutions' Internal Risk Ratings</i>	26
<i>Summary of Databases</i>	27
<i>Section IV: Univariate Ratios As Predictors Of Default: The Variable Selection Process</i>	27
<i>The Forward Selection Process</i>	28
<i>Statistical Power and Default Frequency Graphs</i>	30
<i>Profitability Ratios</i>	32
<i>Leverage Ratios</i>	34
<i>Size</i>	35
<i>Liquidity Ratios</i>	36
<i>Activity Ratios</i>	37
<i>Sales Growth</i>	37
<i>Growth vs. Levels</i>	38
<i>Means vs. Levels</i>	40
<i>Audit Quality</i>	40
<i>Risk Factors We Do Not Use</i>	41
<i>Conclusion</i>	42

<i>Appendix 4A</i>	42
<i>Calibration and Power</i>	42
<i>Power and Default Frequency</i>	44
<i>Testing the Need for Recalibration</i>	46
<i>Section V: Similarities And Differences Between Public And Private Companies</i>	46
<i>Distribution of Ratios</i>	47
<i>Public/Private Default Rates By Ratio</i>	49
<i>Ratios and Ratings</i>	52
<i>Section VI: Transformations And Functional Form</i>	54
<i>Transformations</i>	54
<i>Functional Forms</i>	57
<i>Appendix 6A: Transformations Of Input Ratios</i>	59
<i>Appendix 6B: RiskCalc Schema</i>	60
<i>Input Fields</i>	60
<i>Ratios</i>	61
<i>Ratio Calculation</i>	61
<i>Intermediate Output - The Unadjusted Probability of Default</i>	62
<i>Final Output: 1-Year DP & 5-Year DP</i>	62
<i>Mapping into Moody's Rating Symbols</i>	62
<i>Supplemental Information</i>	62
<i>An Example of Relative Contributions</i>	63
<i>Section VII: Mapping To Default Rates And Moody's Ratings</i>	63
<i>Definitions Of Default</i>	63
<i>Prediction Horizon</i>	64
<i>Default Rate And Moody's Mapping</i>	64
<i>Moody's Mapping Preliminaries</i>	65
<i>Issue 1 - Unrated Firm Default Rates</i>	66
<i>Issue 2 - Readjusting Moody's Default Rates for Withdrawals</i>	67
<i>Issue 3 - Time Horizon for the Mapping to Moody's Ratings</i>	67
<i>Issue 4 - Time Period for Sample</i>	67
<i>Summary of Default Rate Calibration and Ratings Mapping Method</i>	68
<i>Appendix 7A: Perceived Risk Of Private Vs. Public Firm Debt</i>	68
<i>Section VIII: Model Validation</i>	70
<i>The Lessons of Z-Score</i>	71
<i>Testing the Alternatives</i>	73
<i>RiskCalc Tests</i>	74
<i>Walk-Forward Tests on Compustat</i>	75
<i>Out-of-Sample Tests on the CRD</i>	76
<i>Final In-Sample CAP plots</i>	77
<i>Miller Tests</i>	78
<i>Parameter Stability and Input Correlation</i>	78
<i>Correlation Over Time</i>	79
<i>Industry Performance</i>	80
<i>Appendix 8A: Accuracy Ratios And Conditional Entropy Ratios</i>	80
<i>Appendix 8B: Information Entropy Ratios</i>	82
<i>Section IX: Conclusion</i>	82
<i>References</i>	83

Table Of Exhibits

<i>Exhibit 0.1 - Moody's Default Model For Private Firms: An Anecdotal Example Of Chai-Na-Ta Corp Defaulted January 28, 2000</i>	.1
<i>Exhibit 1.1 - Number Of Firms By Asset Size, 1996 IRS Return Data</i>	.11
<i>Exhibit 1.2 - Credit Scoring Application Chart</i>	.12
<i>Exhibit 2.1 - Empirical Studies Of Corporate Default: Year Published And Sample Count</i>	.14
<i>Exhibit 3.1 - Time Distribution Of Financial Statements And Defaults</i>	.23
<i>Exhibit 3.2 - Borrower Counts By Number Of Yearly Observations</i>	.24
<i>Exhibit 3.3 - Geographic Distribution Of Borrowers</i>	.24
<i>Exhibit 3.4 - Industrial Composition Of CRD Borrowers</i>	.25
<i>Exhibit 3.5 - Distribution Of Financial Statement Quality</i>	.25
<i>Exhibit 3.6 - Distribution Of Financial Statement By Sales Size Group</i>	.26
<i>Exhibit 3.7 - Distribution Of Institutions' Internal Risk Ratings</i>	.26
<i>Exhibit 3.8 - Database Summary</i>	.27
<i>Exhibit 4.1 - One-Year Default Rates By Alpha-Numeric Ratings, 83 - 99</i>	.30
<i>Exhibit 4.2 - Power Vs. Probability Of Default</i>	.31
<i>Exhibit 4.3 - Profit Measures, 5-Year Cum. Prob. Of Default, Public Firms, 80 - 99</i>	.32
<i>Exhibit 4.4 - Leverage Measures, 5-Year Probability Of Default, Public Firms</i>	.34
<i>Exhibit 4.5 - Size Measures, 5-Year Cum. Prob. Of Default, Public Firms, 80 - 99</i>	.35
<i>Exhibit 4.6 - Liquidity Measures, 5-Year Cum. Prob. Of Default, Public Firms, 80 - 99</i>	.36
<i>Exhibit 4.7 - Activity Measures, 5-Year Cum. Prob. Of Default, Public Firms, 80 - 99</i>	.37
<i>Exhibit 4.8 - Sales Measures, 5-Year Cum. Prob. Of Default, Public Firms, 80 - 99</i>	.38
<i>Exhibit 4.9 - Growth Vs. Levels, 5-Year Cum. Prob. Of Default, Public Firms, 80 - 99</i>	.38
<i>Exhibit 4.10 - Public Firms, 80 - 98, Probit Model Estimating Future 5-Year Cum. Default Trends and Levels</i>	.39
<i>Exhibit 4.11 - Mean Vs. Latest Levels, 5-Year Cum. Prob. Of Default, Public Firms, 80 - 99</i>	.40
<i>Exhibit 4.12 - Public Firms, 80 - 98, Probit Model Estimating Future 5-Year Cumulative Default</i>	.40
<i>Exhibit 4.13 - Change In Auditor, Public Firms, 80 - 99</i>	.41
<i>Exhibit 4.14 - Public Firms, 5-Year Cumulative Default Rate, 80 - 99, Audit Quality</i>	.41
<i>Exhibit 4.15 - Private Firms, 5-Year Cumulative Default Rate, 80 - 99, Audit Quality</i>	.41
<i>Exhibit 4.16 - RiskCalc Inputs And Ratios</i>	.42
<i>Exhibit 4.A.1 - Cap Plots Graphically Present Information On Statistical Power</i>	.44
<i>Exhibit 4.A.2 - Cap Plots Vs. Default Frequency</i>	.45
<i>Exhibit 5.1 - Histograms, Public Vs. Private Firms</i>	.48
<i>Exhibit 5.2 - Public Vs. Private Firms 5-Year Cumulative Default Frequency</i>	.50
<i>Exhibit 5.3 - Total Assets, 5-Year Cum. Prob. Of Default</i>	.51
<i>Exhibit 5.4 - Cash/Assets, 5-Year Cum. Prob. Of Default</i>	.51
<i>Exhibit 5.5 - Retained Earnings/Assets, 5-Year Cum. Prob. Of Default</i>	.52
<i>Exhibit 5.6 - Relative Market Value, (Market Value In \$ Millions/S&P 500)</i>	.52
<i>Exhibit 5.7 - Median Ratios By Grouping Over Time</i>	.53
<i>Exhibit 6.1 - Regression Of Model Residuals On Explanatory Variables, Public Firms, 80 - 99</i>	.56
<i>Exhibit 6.2 - Impact Of Sales Growth Using Nonametric Transformations Vs. Percentiles And Their Squares</i>	.56

<i>Exhibit 6.3 - Sales Growth Probability Of Default - Transformation Function</i>	58
<i>Exhibit 6.A.1 - Transformation Functions For The Ratios Used In RiskCalc</i>	59
<i>Exhibit 7.1 - Annual Default Rate Estimates</i>	66
<i>Exhibit 7.2 - Moody's 5-Year, Smoothed, Cumulative Default Rates, Unadjusted For Withdrawals, 83 - 99</i>	68
<i>Exhibit 7.A.1 - Banks Vs. Debt Values</i>	69
<i>Exhibit 7.A.2 - P/Es Of Public And Private As Suggested By Acquisition Prices</i>	69
<i>Exhibit 7.A.3 - Debt Charge Off History</i>	70
<i>Exhibit 8.1 - Accuracy Ratios And Cumulative Accuracy Profiles (Cap Plots)</i>	70
<i>Exhibit 8.2 - Accuracy Ratios For Out-Of-Sample Tests On Private Firm Data</i>	71
<i>Exhibit 8.3 - Z-Score And Liabilities/Assets Prior To Default, Compustat, 80 - 99</i>	72
<i>Exhibit 8.4 - Accuracy Ratios On Public And Private Firms, 1- And 5-Year Horizons</i>	73
<i>Exhibit 8.5 - The Improper Linear Model (Ni/A - L/A) Performance</i>	74
<i>Exhibit 8.6 - Compustat Rolling Forward Tests Of Alternative Models</i>	75-76
<i>Exhibit 8.7 - Out Of Sample Performance of Approach on CRD, Private Companies 94 - 99, 1-Year Cumulative Default</i>	76
<i>Exhibit 8.8 - Out -Of-Sample Tests On Crd, Accuracy Ratios</i>	77
<i>Exhibit 8.9 - In Vs. Out-Of-Sample Performance Of RiskCalc</i>	77
<i>Exhibit 8.10 - 122 Nonfinancial Firms Rated B As Of 12/31/92, 18 Defaulters Within 5 Years</i>	78
<i>Exhibit 8.11 - Correlation Matrix Of Inputs Used In RiskCalc. Ratios Vs. Transformed Ratios. Private Firms, 89 - 99</i>	79
<i>Exhibit 8.12 - Parameter Stability Of The RiskCalc* Algorithm, Compustat Data-</i>	79
<i>Exhibit 8.13 - Riskcalc Industry Performance, Crd, 5-Year Cumulative Default Horizon</i>	80-81

Introduction

This document describes RiskCalc™,¹ Moody's proprietary model for estimating private firm default risk. RiskCalc is the next generation of Moody's RiskScore, and contains improvements in several dimensions: power, comprehensiveness, simplification of data requirements, and calibration to default probabilities. The following is a self-contained primer on the theory and estimation of commercial default risk underlying this model.

The RiskCalc algorithm uses 9 financial ratios and firm size, adjusts these inputs to linearize the problem, and then estimates the final collection of transformed inputs within a probit model. The output is then mapped into an Default Probability (DP) at 1-year and 5-year horizons. A final mapping to an estimated Moody's rating is based on the 5-year DP.

The model is estimated on U.S. and Canadian private firms, and tested on private and public firm data. It is not intended for finance, insurance, and real estate industries. While not formally tested upon countries outside North America, limited testing on other regional exposures suggests the general applicability remains robust.

Two key facts underlie the usefulness of RiskCalc:

- It is specifically designed for private firms.
- It ties credit scores directly to default probabilities, which is a critical component for determining pricing and enabling securitization.

RiskCalc is the most statistically powerful model available for private firm default modeling because it is estimated on private rather than public firms. Public and private firms are different in important ways. Private firms are typically smaller, with lower leverage, higher retained earnings, higher short-term debt, higher current ratios, and lower inventories than public firms with similar risk. While models fit to public companies can be useful when applied to private firms, the relationship between certain ratios and default probability display markedly different behavior between public and private firms.

Second, by tying RiskCalc to a default probability, it moves quantitative tools from merely monitoring trends to affecting pricing directly and enabling securitization. Further, the fact that the mean default rate for the entire middle market company segment, as determined by the model, is Ba2 as opposed to B2 implies that there exist substantial opportunities for balance-sheet collateralized loan obligations (CLOs) because post-CLO capital allocation could be well below that required when keeping entire portfolios of private firm loans on banks' books. By tying the output to a default rate, this model can also assist in the building of internal capital models within banks, in line with the new Basel capital directives.

Additional benefits of RiskCalc include:

- transparency,
- integration with underwriting and deal capture software (e.g., FAMAS),
- Moody's commitment to maintaining and improving the model,
- RiskCalc's acceptance and understanding within Moody's structured finance group that rates pools of private loans for collateralized loan obligations, and
- various complementary supporting information on the drivers of a final score.

Like all new technologies, RiskCalc is a supplement to, not a substitute for, good judgement. Many factors not reflected in balance sheets and income statements are relevant to gauging loan risk. The score produced by RiskCalc alone cannot answer the deeper question as to whether the credit adds value from a portfolio and relationship perspective. However, what RiskCalc can do is efficiently summarize one portion of the problem (financial statements) so that an analyst can focus her expertise more productively.

¹ RiskCalc is a trademark of Moody's Risk Management Services, Inc. Moody's currently provides two RiskCalc models, RiskCalc for Private Companies, and RiskCalc for Public Companies, where the distinction is the existence of liquid equity prices.

What We Will Cover

This report is organized as follows:

Section I provides an overview of current credit scoring tools, discusses where Moody's RiskCalc fits into the array, and highlights some of RiskCalc's common applications.

Section II examines past studies and current theory on commercial loan defaults. The section highlights the low number of defaults in previous studies, which can explain why commercial loan default model advancement has been relatively slow. It also outlines the main approaches to modeling credit risk so the reader can see how RiskCalc relates to these approaches.

Section III discusses the datasets used in RiskCalc's estimation and testing, focusing upon Moody's unique private firm dataset.

Section IV highlights the variable selection process, one of the most important steps in default prediction.

Section V discusses the major differences between public and private companies, and the different relation to default for the same financial ratios in these two universes.

Section VI explores the inner workings of RiskCalc, including an appendix which walks the user through the model.

Section VII details the important technical assumptions necessary to map RiskCalc's output to default rates and to Moody's ratings.

Section VIII assesses RiskCalc's statistical power through several tests on public and private firm data.

Section IX presents our conclusions.

To the extent that some of the references in this report may be unfamiliar at first, we have attempted to provide enough context and explanation to allow attentive readers, regardless of their academic or professional background, to capture its key points.

Section I: The Current Credit Risk Toolbox

Commercial lending is certainly nothing new, and neither are the statistical models designed to study its risks. Written records from Sumer circa 3000 B.C. indicate that interest rates were between 15 and 33%, which implies that commercial lending rivals other pursuits as the world's oldest profession (Durant (1917)). In terms of analyzing the risks involved, lenders have examined accounting ratios since at least the 19th century (Dev (1974)). But the modern era of commercial default prediction really begins with the work of Beaver and Altman in the late 1960s.

Even though statistical models were outlined 30 years ago, middle market lending is still primarily a subjective process. There are no benchmarks in commercial lending with wide usage, which helps explain why middle market portfolios are infrequently securitized and considered unusually opaque assets (Bergson (1995)). The financial statements of a borrower are invariably analyzed prior to the issuance of a loan, but the interpretation of this information varies from analyst to analyst.

In contrast, consumer lending has undergone a significant transformation over the past 30 years. Today, a bureau score that captures at least 90% of the measurable risk inherent in a consumer relationship, can be purchased for a few dollars. This score, which has become an essential input to any underwriting or portfolio analytic process, can help segregate pools of customers whose expected loss varies by as much as 10% of notional balances.

How did consumer credit modeling leapfrog commercial credit modeling? Data. Hundreds of thousands of bad credit card debts and millions of goods ones make for confident inference. In contrast, if one examines 30 of the major academic papers on commercial default models over the past 30 years, the median number of defaulting companies used in these studies is 40.

Moody's databases are sufficiently large (over 1,500 private firm defaults, 1,400 nonfinancial public firm defaults) to make a significant leap in the estimation and testing of commercial default models as applied to private firms. It is not hyperbole to assert that the move from 40 to 1,500 defaults enables a shift to the next level in model accuracy and reliability.

APPLICATIONS

There are several well-known quantitative lending tools available, some complimentary, some direct competitors. The primary determinant of which tool is appropriate in any given situation is the size of the firm to which the model is to be applied. RiskCalc for Private Companies,² for its part, occupies the niche of middle market private firms, generally with assets greater than \$100,000 (i.e., about 2 million firms in the U.S.).

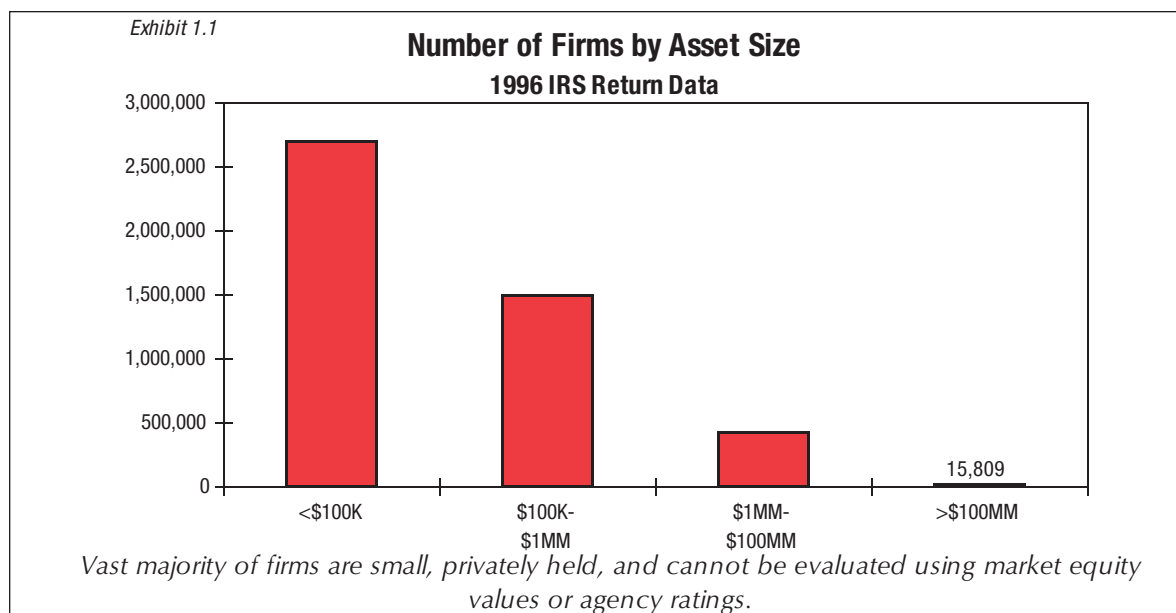


Exhibit 1.1 shows that according to 1996 US tax return data there were 2.5 million companies we would characterize as small, with less than \$100,000 in assets. There are approximately 1.5 million firms in the region between the small and middle market, with \$100,000 to \$1 million in assets, and 300,000 in the middle market category, with \$1 million to \$100 million in assets. There are a mere 16,000 large corporates, with greater than \$100 million in assets. Only around 9,000 companies are publicly listed in the U.S., and only 6,500 companies worldwide have Moody's ratings.

Therefore, the vast majority of firms must be evaluated using either financial statements, business report scores, or credit bureau information—not market equity values or agency ratings. Exhibit 1.2 shows how various credit tools are applied, going from small to large as we go from left to right.

The relation between the amount of popular press for these tools and their actual impact on credit decisions is weak. Consumer models are by far the most prevalent and deeply embedded quantitative tools in use, while their examination in academic and trade journals is rare. In contrast, academic journals frequently cite hazard models, and trade journals frequently discuss refinements of portfolio models, even though these models are used far less frequently than business report scores. This does not imply that these approaches should be judged on their obscurity, just that there can be large differences between a model's popularity in the press and its actual usage.

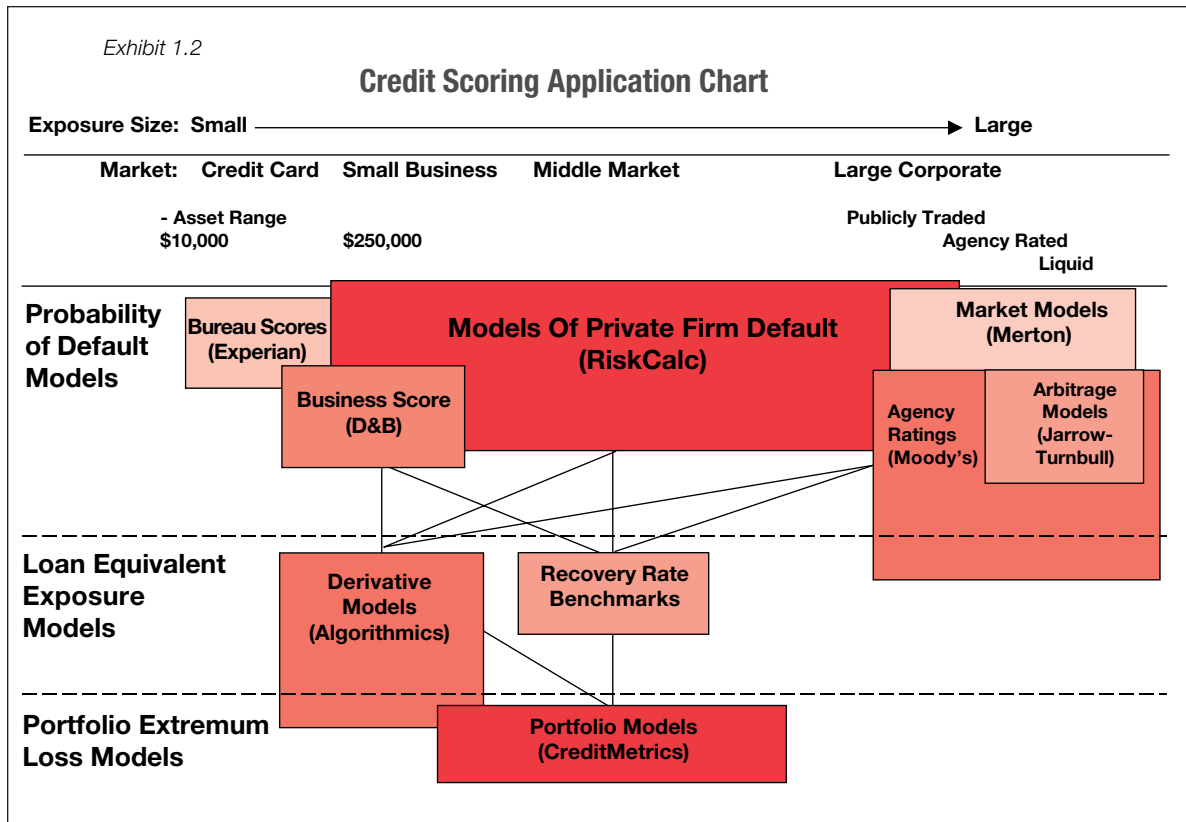
The sections that follow provide a brief overview of each model and the situations in which they are commonly applied.

Consumer Bureau Scores

Bureau scores are aimed at evaluating consumer credit, such as credit card and auto loans. Standardized consumer scores developed in the 1980s as credit bureaus focused on information such as delinquency, debt burden, inquiries, and inactivity to assess credit quality. Literally millions of exposures and hundreds of thousands of 'bads' underlie the development of these models. These transparent and validated risk measures have enabled securitization of consumer debt, by providing investors with objective and comparable information across portfolios. The success of bureau scores at integrating with both underwriting and portfolio analysis place them as the gold standard of quantitative scores, not so much for their accuracy or power, but for their ability to provide efficient, consistent, cheap, and meaningful scores to lenders and investors.

² RiskCalc is Moody's trademark for its quantitative scoring models. A model for public firms is currently available, and development is ongoing for similar but significantly different models focused upon different countries and industries.

Exhibit 1.2



Bureau scores are not simply applicable to credit cards and boat loans, but to small businesses as well. Many start-up companies are really extensions of an individual and share the individual's characteristics. Bureau scores are relevant for small businesses up to at least \$100,000 in asset size, and perhaps \$1 million.

The major bureaus include Equifax, Experian, TRW and Trans Union, and have all used the consulting of Fair, Isaac Inc. to help develop their models.

Business Report Scores

Dun & Bradstreet and Experian provide Business Report Scores, based primarily on liens, court actions, creditor petitions, company age, and size of the company. These scores are directed mainly at assessing suppliers and purchasers of trade credit. For a nominal fee, a company can ascertain the credit risk of purchasers and separate out potential borrowers demonstrating an inability or unwillingness to pay, irrespective of personal credit quality of the principal owners or the firm's balance sheet.

Public Firm Models

The most popular public firm model is the Merton model, which is based on options theory. Equity of the firm is considered a call option on the value of the assets of the firm, where the strike price is some proportion of the liabilities. In efficient markets, presumably the equity value and its volatility, which are directly observable, combined with information on the level of liabilities, provide sufficient information to estimate default probabilities.

Extensions of this model, such as Moody's public firm model,³ use a variant of the Merton model as an input in conjunction with other data, such as financial ratios, and, given valuable equity information, will have different weightings on the financial ratios vis-à-vis a private firm default model (i.e., less weight on the financial ratios). Most importantly, we found that by adding information such as profitability, in conjunction with a 'distance to default' measure based on the Merton approach, we can significantly improve upon a model that uses a stricter interpretation of the Merton framework.⁴ This approach is limited to firms with sufficiently liquid equity prices (so that one can estimate volatility and the current market value), and generally refers to less than 10,000 firms in the U.S.

³ Moody's Public Firm Risk Model: A Hybrid Approach to Modeling Short Term Default Risk. Moody's Investors Service. 2000.

⁴ See Sobehart and Keenan (1999) for a discussion of why adding financial statement data to the Merton approach can generate better default predictions.

Agency Ratings

Agency ratings are opinions based on extensive human analysis of both the quantitative and qualitative performance of a firm. Companies with agency-rated debt tend to be large and publicly traded. Moody's primary business is providing credit opinions on financial obligations for investors: traditional Aaa to C ratings. These ratings are well-accepted by the investment community, and extend not only to commercial firms but municipal, sovereign, and other obligors. These ratings cover approximately 6,500 firms worldwide and 3,000 in the US. The credit opinions are statements about loss given default and default probability, specifically expected loss, and thus act as combined default prediction and exposure models.⁵

Hazard Models

Hazard models refer to a subset of agency-rated firms, those with liquid debt securities. For these firms, the spread on their debt obligations relative to the risk-free rate can be observed, and by applying arbitrage arguments, a 'risk-neutral' default rate can be estimated. The universe of U.S. firms with such information numbers approximately 500 to 3,000, depending on how much liquidity one deems necessary for reliable estimates. While some of these models are similar to the Merton model, their unifying feature is that they are calibrated to bond spreads, not default rates. Examples include the models of Unis and Madan, Longstaff and Schwartz, Duffie and Singleton, and Jarrow and Turnbull.

Exposure Models

These models estimate credit exposure conditional on a default event, and thus are complements to all of the above models. That is, they are statements about how much is at risk in a given facility, not the probability of default for these facilities. These are important calculations for lenders who extend 'lines of credit', as opposed to outright loans, as well as for derivatives such as swaps, caps, and floors.

Exposure models also include estimations of the recovery rate, which vary by collateral type, seniority, and industry. For individual credits, these are usually rules of thumb based on statistical analysis (e.g., 50% recovery for loans secured by real estate, 5% of notional for loan equivalency of a 5-year interest rate swap).

Much has been written about the approach in recent textbooks on derivatives, and guideline recovery rate assumptions are mainly set through consultants with industry experience, Moody's published research, and other industry surveys.

Portfolio Models

Like exposure models, these are complements to obligor rating tools, not substitutes. Given probabilities of default and the exposure for each transaction in a portfolio, a summing up is required, and this is not straightforward due to correlations and the asymmetry of debt payoffs. One needs to use the correlations of these exposures and then calculate extrema of the portfolio valuation, such as the 99.9% adverse value of the portfolio.

The models are more complicated than simple equity portfolio calculations because of the very asymmetric nature of debt returns as opposed to equity returns. Examples include CreditMetrics, CreditRisk+, and rating agency standards for evaluating diversification in CDOs (collateralized debt obligations).⁶ Some derivative exposure models are similar to portfolio models; in fact, optimally one would like to simultaneously address portfolio volatility and derivative exposure.

WHERE RISKCALC FITS IN

RiskCalc targets the middle market class of borrowers. Other models aimed at this segment include Altman's Z-score model. These models use financial statement data to predict default. The lower bound for applicability is around \$100,000 in asset size, and extends up to publicly traded companies. Importantly, size does not define the upper limit. The existence of market equity information - a new and important source of information that should not be excluded - does establish the upper boundary. The lower bound comes from general experience in the credit field by examining the performance of scoring models on firms of various size. As market value information is valuable and not reflected in financial statements, these private firm default models are sub-optimal for those companies with traded equity.

RiskCalc is at the heart of the commercial credit evaluation process. It is for firms too large to be considered a simple extension of an individual, yet without publicly traded equity information. It generates essential input for portfolio variability calculations and, when combined with facility information, can be used to estimate an expected loss.

⁵ Moody's ratings also target financial stability and have a long-term horizon, which make them more stable than typical quantitative models.

⁶ Rating Cash Flow Transactions Backed by Corporate Debt. 1995 Update. April 7, 1995. Moody's Investors Service.

Transparency of bank portfolios is currently poor relative to examining the value of the assets of traditional nonfinancial corporations. In financial institutions, the margin between incoming and outgoing cash flow is so thin and the leverage so high that small differences in asset quality affect their solvency, and thus the solvency of financial systems. The difference between a pool of commercial loans with different credit quality is significant at a level too fine for current methods of analysis to be of much practical benefit.

Benchmarks, such as those generated by RiskCalc, provide a way for market participants to evaluate the credit quality of different financial institutions more effectively. Moody's commitment to providing default models not just for the US, but for many countries with large complex financial organizations, will hopefully make it possible for meaningful evaluation of this least transparent portion of bank portfolios.

Section II: Past Studies And Current Theory Of Private Firm Default

For every complex problem, there is a solution that is simple, neat, and wrong.

~ HL Mencken

A model can not be evaluated without an understanding of the state of the art. What models are "good" can only be determined in the context of feasible or common alternatives, as most models are "statistically significant." Though references to prior work are made throughout this text, we believe that it is useful to give a brief overview at the outset of some of the empirical and theoretical literature in this field. More comprehensive reviews of this literature can be found in Morris (1999), Altman (1993), and Zavgren (1984).

EMPIRICAL WORK

All statistical models of default use data such as firm-specific income and balance sheet statements to predict the probability of future financial distress of the firm. Lenders have always known that financial ratios vary predictably between investment-grade and speculative-grade borrowers. This knowledge manifests itself in many useful rules of thumb, such as not lending to firms with leverage ratios above 70%.

Exhibit 2.1

Empirical Studies of Corporate Default: Year Published and Sample Count

	Year	Defaults	Non-Defaults
Fitzpatrick	(32)	19	19
Beaver	(67)	79	79
Altman	(68)	33	33
Lev	(71)	37	37
Wilcox	(71)	52	52
Deakin	(72)	32	32
Edmister	(72)	42	42
Blum	(74)	115	115
Taffler	(74)	23	45
Libby	(75)	30	30
Diamond	(76)	75	75
Altman, Haldeman and Narayanan	(77)	53	58
Marais	(79)	38	53
Dambolena and Khoury	(80)	23	23
Ohlson	(80)	105	2,000
Taffler	(82, 83)	46	46
El Hennawy and Morris	(83a)	22	22
Moyer	(84)	35	35
Taffler	(84)	22	49
Zmijewski	(84)	40	800
Zavgren	(85)	45	45
Casey and Bartczak	(85)	60	230
Peel and Peel	(88)	35	44
Barniv and Raveh	(89)	58	142
Boothe and Hutchninson	(89)	33	33
Gupta, Rao, and Bagchi	(90)	60	60
Kease and McGuiness	(90)	43	43
Keasey, McGuiness and Short	(90)	40	40
Shumway	(96)	300	1,822
Moody's RiskCalc for Public Companies ⁷	(00)	1,406	13,041
Moody's RiskCalc for Private Companies	(00)	1,621	23,089
Median		40	45

Small samples inhibit the search for a truly superior default model.

⁷ Sobehart and Stein (2000)

More recently, Beaver (1967) found that several ratios differed significantly between failed and viable firms, especially cash flow/net worth and debt/net worth.⁸ Beaver documented differences in common ratios such as debt/net worth and cash flow/assets between failed and viable firms increased as the time to failure shortened (i.e., as failure neared, the firms became more measurably dissimilar). Altman (1968) extended this analysis to a multivariate model, as it seemed natural that by using a set of these informative variables - all individually powerful but not perfectly correlated - one could create a model better than with any single ratio alone. Exhibit 2.1 lists a selection of some of the major work published in this area, along with the sample size used in their studies.

Survey work has consistently given the edge to Altman's Z-score, or at least declared a tie when other models have challenged it. Therefore, Altman's Z-score has developed benchmark status in the academic literature and among accounting and financial analysis textbooks.⁹ However, it should be kept in mind that proving that models are statistically significant - that is, better than random guessing - can be accomplished with 40 defaults. Discriminating shades of gray, such as the difference between two statistically significant models, on the other hand, is not feasible with 40 defaults. In this case Altman's Z-score, the first multivariate model, has persisted in the literature because 'ties' are usually awarded to the older and more established model.

While each researcher has been constrained by small samples, a pattern among the variable selections suggests the importance of four main variables - profit, leverage, size, and liquidity - though not all were used in each study.

We found that Shumway's model worked best in the sense of separating future defaulters from non-defaulters among all the published models listed above, and therefore used it as one of the benchmarks for testing. The power of Shumway's model is likely related to the large number of defaults he used, 300, which dwarfs the numbers used in other studies. Another attractive quality of Shumway's model is its parsimony: three variables. Altman's Z-score is also used in our benchmarking, as it is clearly the most prominent in the historical literature.

THEORY OF DEFAULT PREDICTION: COMPARING THE APPROACHES

Ad hoc data mining is appropriately eyed with skepticism, which suggests models based on theory are preferred. Yet, one cannot ignore the miserable track record of structural (i.e., theoretical) models in economics. Most models, therefore, attempt a solution that embodies the best of both approaches without the corresponding limitations.

RiskCalc's algorithm is best described as nonstructural and parsimonious. Below, we explain why we think this approach is optimal by examining the strengths and weaknesses of three types of competing approaches: traditional human judgement, structural models, and nonstructural models.

Traditional Credit Analysis: Human Judgement

In some sense, the debate between models and human judgement is besides the point (one should always use both), yet it is useful to acknowledge that judgmental models alone, from purely subjective to the most elaborate "expert-rule" systems, dominate commercial loan analysis. This does not mean that quantitative information is not used. But, since several quantitative inputs are examined, the final judgmental distillation of this information is not transparent or easily validated. This tendency toward noncomparability is reflected in how easy it is within an average bank, or even in Yahoo! or on CNBC, to examine various historical data on a particular firm's equity value, balance sheet, income statement, and management, yet how hard it is to find the firm's percentile ranking for these same items.

Clearly, users like vertical information (i.e., time series) and have less use for horizontal (i.e., cross sectional) information. From an empirical perspective, this is understandable. It would be very difficult for a researcher to determine risk given a relative ranking of, say, the liabilities to assets ratio, since data that relates defaults to such a ratio is relatively difficult to get (though this report somewhat fills that gap). Analyzing a company over time, on the other hand, allows the researcher to at least say whether a company's risk has gone up or down.

⁸ One of the first known analyses of financial ratios and defaults was by Fitzpatrick (1928)

⁹ e.g., Lovie and Lovie (1986), Casey and Bartczak (1985), Zavgren (1984), Bing, Mingly, and Watts (1998)

Still, cross-sectional analysis is useful. Credit analysts know that to gauge risk you have to know relative values; that is, leverage can only be determined to be high if the leverage ratio's average value is known. Given that some of these ratios vary systematically by industry, and also that analysts often specialize in a particular industry, these benchmarks are usually taken from peer groups. This allows the analyst to say whether a company has high leverage and is therefore, on this dimension, risky. Since for virtually any industry some ratios exist which have significantly different means, analysts naturally object to stating that a ratio is "high" when it is perfectly average for that industry. The main job of a credit analyst has and always will be the determination of risk relative to the portfolio they are analyzing, as absolute risk is often affected by factors that are too difficult to forecast.

The bottom line is that data are generally presented in a way that facilitates assessing a firm's trend, with a highly selective base for inter-company comparison. The problem is that without a multivariate model, one is often constrained to compare, sequentially, individual ratios, which often leads to ambiguous results. For example, if a firm is rated Aa in liquidity, Baa in profitability, B in leverage, the net result is unclear.

The value of quantitative models over judgement is not purely a scale economy argument. There is no debate as to which method is cheaper, faster, and more consistent between institutions. Yet, many presume that given enough time most sufficiently intelligent and experienced analysts would outperform any model. The superiority of quantitative models may exist for cases where it does not pay enough to individually analyze loans (consumer), or where one has complex "option" information (as in a Merton model). However, with financial statements, the situation should be different. For example, quantitative models invariably focus upon a more restricted set of information than is available to an analyst, which presumably creates an advantage for the analyst.

Certainly, many analysts are better than many models, and some analysts are better than all models. But what we want to know is whether a model developed with significant numbers of defaults is better than an average analyst.

What must first be considered is that humans have limitations and biases just as models do. One of the more active fields in academic finance, "behavioral finance," uses documented psychological biases to explain anomalies to traditional "rational agent" theory. These biases include the following: people tend to overestimate the precision of their knowledge (Alpert and Raiffa (1982)); their overconfidence increases with the importance of the task; and finally, they recall information related to their successes more easily than information related to their failures (Barber and Odean (1999)). The bottom line is that individuals are often, perhaps even systematically, miscalibrated.

Empirical evidence in favor of quantitative models vs. judgement as applied to lending comes from Libby (1975). He asked 16 loan officers from small banks and 27 loan officers from large banks to judge which 30 of 60 firms would go bankrupt within three years of the financial statements with which they were presented. The loan officers requested five financial ratios on which to base their judgements. While they were correct 74% of the time, this was inferior to such simple alternatives as the liabilities/assets ratio.

Outside of lending, there are many examples in which models outperformed the experts, including: evaluating grad school applicants, future student GPA, future faculty ratings, and radiology diagnostics (Dawes and Corrigan (1974)).

Why might this be the case? That is, why might statistical models dominate judgement in prediction? Paul Meehl, in his classic 1954 book *Clinical Versus Statistical Prediction*, reviewed evidence that while humans are good at finding important variables, they are not as good at integrating such diverse information sources optimally (Meehl (1954)). Several subsequent reviews corroborate his initial findings (Sawyer (1966)). For example, everyone knows that SAT scores and GPA help predict future student success, but how many know the optimal weights? Without knowledge of the distributional characteristics of GPA and SAT scores, or knowledge of studies indicating their validity as predictors of success, it is not surprising that a statistical weighting scheme does better than a human in these circumstances.

Another reason why quantitative models may outperform judgement in default forecasting is that analysis is usually not focused upon a strict default objective. As opposed to quantitative models that are judged solely on their calibration and power, human analysis is also focused upon presenting a compelling explanation, and focuses more deeply on explaining individual assessments as opposed to broad statistical

performance. As banks have not historically kept sufficient data to really test and calibrate their analysts' judgements in a statistical manner, it would be unsurprising if their judgement was not optimized to statistical objectives.¹⁰ Improving inductive reasoning requires continual feedback, and unfortunately in most lending institutions such feedback is anecdotal, not statistical (Nisbet, Krantz, Jepson, and Fong (1982)).

While there are strong reasons to believe that quantitative models are invaluable for producing consistent and accurate predictions, this does not imply that judgement is not useful. The final competitive advantage will always remain with a judgmental process. The key is for this judgement to focus upon areas where it adds the most value, as opposed to an undefined and unrestricted scope of analysis. Quantitative models should be support tools, not decision-making tools. Giving humans, and hence a judgmental model, the final say, however, is different than the current state where judgmental models dominate as they do.

A decision-making process that uses both quantitative information and judgement in a judicious way has the following characteristics:

First, the quantitative information used is not presented as a smorgasbord of ratios and risk factors, but focuses upon one composite number. When several numbers are relied upon in the final judgement, the aggregation of this information has so many possibilities that any one person's summary judgement is basically subjective because it is not transparent or comparable between individuals.

Second, judgement is focused on exceptions, where quantitative scores are extreme, but extenuating circumstances are present. While a final number is the initial focus, the exception process focuses on what factors are driving the result. For example, things known to be 'outside the model', such as knowledge that an export market has recently crashed or that a major competitor recently went out of business, should affect one's outlook as to the future viability of a company. An expert rules system should help the analyst to focus upon those factors that make for valid exceptions. Moody's currently provides a platform for developing an expert rule system within a lending institution that helps underwriters focus upon key risk factors, including subjective factors, and integrating them in a consistent way.¹¹

Judgemental models and expert rules systems provide a useful way to integrate information about a company so that more complex refinements of any single score can be made in a disciplined manner. For statistical validation purposes, however, they are ambiguous, and in general inferior to more statistical methods.

Structural Models

People tend to like structural models. A structural model is usually presented in a way that is consistent and completely defined so that one knows exactly what's going on. Users like to hear 'the story' behind the model, and it helps if the model can be explained as not just statistically compelling, but logically compelling as well. That is, it should work, irrespective of seeing any performance data. Clearly, we all prefer theories to statistical models that have no explanation.

Our choice, however, is not "theory vs. no theory," but particular structural models vs. particular non-structural models. Straw man versions of either can be constructed, but we have to examine the best of both to see how they compare, and what one might imply for the other.

The most popular structural model of default today is the Merton model, which models the equity as a call option on the assets where the strike price is the value of liabilities. This maps nicely into the well-developed theory of option pricing. However, there is another structural model, that of the gambler's ruin, which predates the Merton model. In the gambler's ruin, equity and the mean cash flow are the reserve, and a random cash flow exhausts this cushion with a certain probability.

Lower volatility or larger reserve implies lower default rates in both the Merton and gambler's ruin models. The distinction between the two is that between cash flow volatility and market asset volatility - that is, between a default trigger point that is market-value based versus one that is cash-flow based.

Both of these models are explained below, and both are useful for thinking about which variables are relevant to a model that predicts company default.

¹⁰ See section 3 for a discussion of bank data collection issues

¹¹ See Antonov (2000) for a discussion of expert-rule systems and how they can assist credit decisioning

The Merton Model and KMV

In Merton's original formulation (1973), debt has an unambiguous maturity, and the option value is computed with this singular date. When the market value of the firm's assets fall below a certain level, the firm will default. In the upside, the equity owners keep the residual value, just like an equity option.

Under the Merton model, the firm's future asset value has a probability distribution characterized by its expected value and standard deviation. The number of standard deviations the future value of assets is away from the default point is the 'distance to default'. The greater the value of the firm, and the smaller its volatility, the lower the probability of default. This model is described mathematically in Appendix 2A.

KMV's popular implementation of this model makes some useful adjustments to Merton's formulation.¹² The first adjustment addresses the trigger point of default, since the staggered debt maturities that companies actually have imply that the simple Merton formulation is ambiguous in practice. A firm can remain current on its debt even though technically insolvent (liabilities > assets), it can forestall and, with luck, avoid bankruptcy, even though the liability holders would like to liquidate.

In view of this complication, KMV uses ½ the value of long term debt plus current liabilities as a proxy for the 1-year default point, a formulation based on empirical analysis. Thus, in their formulation, the default point is not total liabilities as in the Merton model, but current liabilities 1/2 long term liabilities.

This adjustment is consistent with the distribution of recovery rates on defaulted bonds. As opposed to having the highest recovery rates close to 100% and declining exponentially to zero, as implied by a strict interpretation of the Merton model, the mode recovery rate is more likely to be in the 50-60% range rather than the 90-100% range. That is, if the trigger point was total debt, the most probable recovery rate would be in the highest range; it is not, which suggests the trigger is well below the amount of total debt.¹³

A final adjustment is made in mapping the distance from default into a probability. In most cases, the probabilities calculated from the Merton model are much too low, less than 0.1%, which is probably quite rare (i.e., an Aa designation is for a very select group) which is not consistent with reality (of between 0.01% and 10.0% default rates). Thus, KMV maps their initial output into actual defaults using historical data, as opposed to using the standard normal probability tables.

The adjustments suggest that the Merton model is more of a guideline than a rule for estimating a quantitative model. The final transformation from standard normal probabilities into empirical probabilities implies that even the strongest proponents of the approach do not take the Merton model literally.

The Gambler's Ruin

Turning our attention to the gambler's ruin model of Wilcox (1971), we find a similar, but less well-known approach in which the value of equity is a reserve, and cash flows either add to or drain from this reserve. In the case of a bankruptcy, the reserve is used up.

The "gambler's ruin" name comes from initial applications of this approach to gambling. Assuming you approached a roulette wheel with N dollars, and bet \$1, with 50:50 probability of receiving \$2 or \$0, what is the probability of losing all your money after X bets? This statistical problem has been well known for years, and intuitively captures the default scenario for a firm.

Wilcox set up a model where cash flow was a two-state Markov process, with either positive or negative values, and the reserve is the value of book equity. One then computes the probability of default given the Markov process as applied to cash flows. In this model the "distance to default" is the sum of book equity and expected cash flow, divided by the cash flow volatility. Several researchers have extended this approach,¹⁴ adding more states to the Markov process and adjusting the drift in the cash flows to account for inflation.

One extension of the gambler's ruin problem relevant to the Merton model is that a firm's book equity is not the total reserve, as examined by Scott (1981). This adaptation recognizes the fact that companies do not go bankrupt because they run out of cash, but because people lose faith in them. If a firm's book equity is exhausted through losses and there remains market equity value, equity holders will have cause to infuse more book equity into the company to stave off bankruptcy that would otherwise extinguish their market value.

¹²We do not presume to know the precise method by which KMV calculates their EDFs. The general representation is taken from public conference records and published material.

¹³Direct bankruptcy costs, such as legal bills, imply that 99% recovery would not be the mode even if the Merton Model were strictly true, yet empirical estimates of these costs are all well below the 40% that corresponds to the recovery rate averages we observe.

¹⁴Wilcox (1971, 1973) and Santomero and Vinso (1977), Vinso (1979) Benishay (1973) and Katz, Lilien and Nelson (1985).

This is a financing dynamic that is predictable given the fact that the average market/book ratio is 3:1, indicating that even if a firm lost all its book value, it still remains valuable. Further, approximately 10% of all firms have negative net worth, yet the annual default rate on the entire population is around 1.5% per year. The use of market equity as a cushion can explain why so often technically insolvent institutions avoid bankruptcy. In fact, we estimate that even firms with both negative net worth and negative cash flow usually do not default. This is sensible only if we recognize that not all the firm value is reflected in the balance sheet.

Equity Value vs. Cash Flow Measures: Is One Better?

These two structural models boil down to a univariate truism: if either market equity goes to zero or if cash flow stays negative, the firm will fail. Under both models, prediction of the key event is based primarily on a targeted ratio. For the Merton model, this ratio uses primarily equity information, and for the Gambler's Ruin model, cash flow information is used. The question is: which is better?

Just as an altimeter predicts plane crashes with certainty, market equity values go to zero prior to firm bankruptcy. "Successful" anecdotes are a necessary consequence of these models, but they do not help one gauge whether statistically these approaches are optimal. In most cases the prediction is too late to do any good, just as an altimeter reading of "10 feet" is a warning without usefulness to the pilot.

What is more interesting and useful is a model that predicts default well before the equity value closes in on zero. That is, the most useful model addresses not the characteristics of firms that are imminently failing, but rather those at higher risk of failing at some intermediate period in the future. The optimal functional form for this problem is primarily an empirical question.

In prediction, one must always distinguish between proximate and ultimate objectives. For example, many central bankers view long run GDP growth as their ultimate objective, but policy action is influenced not so much by the latest GDP growth as the latest interest rate behavior. The variable of interest is not targeted directly because it is considered to have a lagged and variable response to other measures one can measure and control.

Applying the same concept to default prediction, we see that market equity levels or EBITDA/interest are at some level identical to default, in that if these are below certain levels for a certain period default is a necessity. They are necessary contemporaneous correlates with firm failure, which is quite different than saying they are sufficient predictors of firm failure. If instead of using these ultimate correlates with failure we instead use an array of correlated variables (e.g., book leverage, size, and current ratio), we may find that this broader set of variables better predicts default.

One can believe individual ratios and the models that utilize them accurately describe the default experience, yet for empirical reasons believe that over-reliance on one targeted variable itself is sub-optimal. If the model is misspecified, a multivariate approach based not solely on the "distance to default" would be better.

Still, when limited in choice to only one set of variables, it is interesting to note the sentiment that the grass is always greener on the other side. That is, while many credit analysts are becoming excited about using equity information, many equity analysts are looking more at accounting information. This alternative view is from the "behavioral finance" school, which finds that markets, like individuals, are prone to over and under-react, and due to liquidity constraints do not imply arbitrage opportunities exist. These irrational tendencies are reflected in the time patterns of ratios like the book/market or price/earnings.

Merton's model has benefited from better data on asset volatility relative to cash flow volatility (there are usually only a handful of relevant cash flow observations, which makes volatility estimation difficult). For private firms, however, market value information is not available by definition. One solution is to infer what we think these values would be, based on industry, size and other financial variables. In practice, a Merton model applied to private firms looks very much like a nonstructural model that was formed using vague intuition from the Merton model.

Nonstructural Models

In the 1970s, it may have been reasonable to be optimistic about the future practical importance of structural models in economics and finance; today it would almost be naïve. The bottom line is that economic and financial models have turned out to be much less useful in an empirical forecasting role than initially thought. It would be without precedence for an economic structural model, absent a feasible arbitrage mechanism (e.g., Black-Scholes) or a tautology of statistics (e.g., diversification), to work well in predicting default.¹⁵

¹⁵Exceptions include: Covered interest rate parity, many options pricing formula, or Markowitzian portfolio mathematics.

Economics has a clear and dismal track record with structural models. The most quantitative of the social sciences, it remains a poor second cousin to the physical sciences in terms of prediction. To give an example, the charge of an electron has been theoretically and empirically estimated to 11 digits (Dirac's number), while most economic debates concern whether a coefficient is significantly different than zero (e.g., is beta positively correlated with future stock returns?). This does not mean economists do not have significant valuable understanding, only that precise prediction is not feasible. Economics will forever be far different than physics in terms of the "laws" it produces.

Consider the following economic models, all of which at one time were purported to give specific, practical information for tactical decisions (i.e., they were not "wrong" in the sense that Marxism was 'wrong'). They are all associated with decades of popularity, Nobel prizes, and hundreds of scholarly journal articles. All are now considered of little practical value for prediction:

- Phillips curve
- Baumol-Tobin transactions demand for money
- Large Keynesian macroeconomic models
- The quantity theory of money
- The capital asset pricing model
- IS-LM curves
- Life cycle hypothesis of saving
- Input-output models

While these models are still studied for important theoretical understanding, only a small fringe of researchers use them for concrete predictions in a way consistent with the original hopes for these approaches. Many more lesser-known economic models have undergone similar life cycles: initial ability to explain an important stylized fact or generate statistically significant coefficients leads to optimism in potential practical applications, a variety of refinements within the paradigm are found to make the model work even better, but eventually a much simpler model, though less structural, is found to work just as well if not better for practical purposes.

In economics, "naïve" models consistently outperform more sophisticated models (see Zarnowitz (1979), or Sims (1982)). "Naïve" does not mean uninformed or arbitrary, but parsimonious and informed by theory. For example, a naive model for predicting inflation is to use last year's inflation rate, plus a weighting on the recent rate of increase in this rate. For time series, a naive model is an autoregressive function. This is often refined to include a moving average term in the error, so that it goes from a simple autoregressive (AR) process to an autoregressive moving average (ARMA) process. Further refinements often include differencing the series in question so that it becomes stationary (ARIMA). "Naïve" models therefore include somewhat subtle refinements, yet to someone with a strong statistical background and experience with the specific problem, these are all very elementary models.

In a study of macroeconomic forecasts Federal Reserve economist, Stephan Rees stated that "increased sophistication provides no improvement in economic forecast accuracy." (McNees (1995)). Or, in the words of the econometrician Arnold Zellner:

I do not know of a complicated model ... that performs well in explanation and prediction and have challenged many audiences to give me examples. So far, I have not heard about a single one ... it appears useful to start with a well understood, sophisticatedly simple model and check its performance empirically in explanation and prediction.¹⁶

This is precisely our approach: nonstructural, well understood, sophisticatedly simple.

Conceptually, the approach taken by both the private firm default modeling effort described here and the public firm model, is highly similar. The essence of our approach to nonstructural models is to be extremely mindful of overfitting the data, trying a variety of explanatory variables that are informed by experience and theory, transforming the inputs appropriately, and then estimating a model. For example, experience points to using variables like liquidity ratios, retained earnings, and size, while theory points to leverage ratios and profitability. Coefficients and transformations are meticulously tested for robust out-of-sample performance, highly correlated explanatory variables are excluded, the sign on coefficients are checked for how they fit with intuition, and one's alternative is always a simpler model than the one under consideration.

¹⁶Journal of Economic Perspectives, Spring 1999, p. 234.

This process is outlined more specifically in Section 6. Like a good structural model our approach should work irrespective of the data for the following simple reasons. First, we have a sufficient number of defaults to be confident that we are finding 'real' relationships. Secondly, the algorithm basically adds together several unbiased predictors of default, in that univariate default probabilities for each ratio-the transformation functions-are unbiased predictors of default. For technical reasons discussed in later sections, this implies that it is highly probable that it will outperform simple univariate models and also that it is robust to suboptimal weightings on these inputs, because they are all normalized to a similar distribution and directional relationship with the predicted variable.

Appendix 2

Merton Model and Gambler's Ruin

In practice there are several different Merton variants used by practitioners, and the following is meant to describe, in general, the gist of the algebraic formula that underlie these approaches. In order to calculate default probability using the Merton model for a firm with traded equity, the market value of equity and its volatility, as well as contractual liabilities, are observed. Using an options approach, the market value of equity is the result of the Black-Scholes formula:

where

$$d_1 = \frac{\ln\left(\frac{A}{e^{-rL}}\right) + \frac{1}{2}\sigma_A^2 t}{\sigma_A\sqrt{t}}$$

$$d_2 = d_1 - \sigma_A\sqrt{t}$$

The volatility of equity can be found as the function of

$$\sigma_E = g(A, \sigma_A; L, r, t) = \frac{N(d_1) A \sigma_A}{E}$$

where:

E=market value of equity

L=book value of liabilities

A=market value of assets

t=time horizon

r=risk free borrowing and lending rate

σ_A =volatility of asset value

σ_E =volatility of equity value

N=cumulative normal distribution

Therefore, there are two equations and two unknowns: the market value of assets (A) and the volatility of assets (σ_A). We can observe the value of equity (E), the volatility of equity (σ_E), the book value of liabilities, (L), interest rate (r), and the time horizon (t). A solution can be found using the Newton-Raphson technique, specifically:

$$\begin{pmatrix} \sigma_A' \\ A' \end{pmatrix} = \begin{pmatrix} \sigma_A \\ A \end{pmatrix} + \begin{pmatrix} \delta f / \delta \sigma_A & \delta f / \delta A \\ \delta g / \delta \sigma_A & \delta g / \delta A \end{pmatrix}^{-1} \begin{pmatrix} f(A, \sigma_A) - E \\ g(A, \sigma_E) - \sigma_E \end{pmatrix}$$

Given initial starting values for σ_A , and A, and using numerical derivatives, convergence takes only a few iterations.

Once the market value and the volatility of assets are determined, one needs to determine the expected value of those assets. This expected value ideally comes from an equilibrium market model, such as the capital asset pricing model (CAPM) or a multifactor model (such as Fama and French's three factor model, which includes systematic risk, size, and the book/market ratio). In practice, this latter refinement is of little importance.

From this information, we can determine the number of standard deviations the asset value is from the default point. In this case KMV's "distance to default" is simply:

$$KMV \text{ Model Distance to Default} = \frac{E(A_t) - L}{A \cdot \sigma_A \sqrt{t}}$$

Distance to default is the distance in standard deviations between the expected value of assets and the value of liabilities at time T. E(A) implies one uses the expected value of assets at the end of some horizon (usually 1 year). If the distance to default=1, the standard normal distribution implies the probability of default is 15% (a one-tailed p-value is used).

This is a complete model of default. Industry, country and size differences should be captured in the inputs.

The gambler's ruin also calculates a distance to default, but in this case, the volatility is based on cash flow. Letting CF be a two dimensional vector of cash flow (high and low) and s_t a two dimensional vector that indicates which element of the vector CF is realized in period t, and RCF_t be a scalar representing the realized cash flow in period t we have:

$$Gambler's \text{ Ruin Distance to Default} = \frac{\text{mean}(CF) + \text{book equity}}{\sigma_{RCF}}$$

$$CF_{t+1} = s_t CF_t$$

$$s_{t+1} = s_t p$$

$$p = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

Where s is the initial state of the cash flow (e.g., high and low), and p represents the transition matrix for cash flows from period to period. The states of the cash flow and its transition matrix are estimated from historical cash flow volatility, which can be simulated using assumed values for CF and p . Book equity can be augmented to market equity.

Both models normalize the reserve that keeps a firm from default by dividing it by the standard deviation of what can detract from the reserve: market value and/or cash flow. Clearly it is easier to estimate the parameters of the equity-based model, as the cash flow model will have significantly fewer datapoints with which to estimate key parameters.

Section III: Data

Moody's default prediction models use large databases to estimate and test their models. Our proprietary database of defaults for public companies, and database of defaults and financial statements for private, middle market companies, give Moody's a unique and privileged standpoint from which to develop and evaluate various models. Below, we describe these databases.¹⁷

Public Company Data: Moody's Default Database and Compustat

Moody's uses Compustat's Research Insight for financial statement information. This is then combined with Moody's Default Database, which contains information on 1,975 public US and Canadian companies that have defaulted since 1980. Approximately 1,400 of these companies were used in this study, as we excluded finance, insurance and real estate companies, as well as those instances where we did not have sufficient financial information on the company prior to default. Approximately 28% of these defaults were by Moody's-rated companies.

¹⁷The database is compiled from participating banks, which included: Bank of America, Bank of Montreal, Bank of Hawaii, Banque Nationale du Canada, Bank One, CIBC, CIT Finance, Citizen's, Crestar, First Tennessee, Hibernia National Bank, KeyCorp, People's Heritage/Banknorth, PNC Bank, Regions, Toronto Dominion.

Private Company Data: Moody's Credit Research Database (CRD)¹⁸

Moody's CRD is a proprietary database consisting of financial statement, commercial loan accounting, and default data on predominantly private, middle market corporations provided by financial institutions participating in Moody's Risk Management Services' private firm default research initiative.¹⁹ Because the CRD pools financial statement and default data from a number of commercial lenders, it is well diversified across regions and industries.

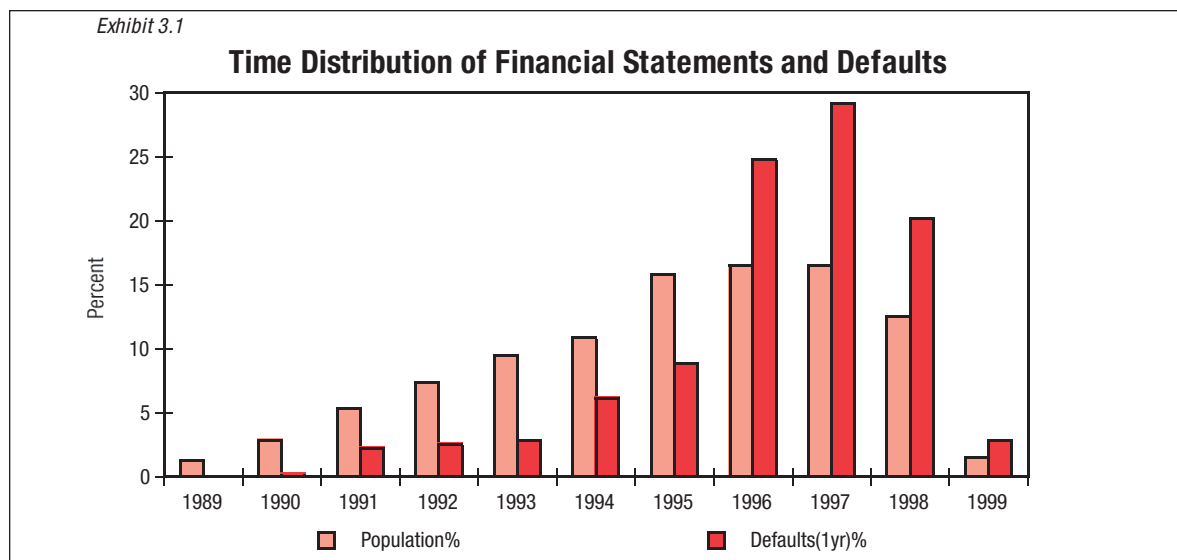
Our goal is to replicate, for our modeling purposes, as closely as possible the same financial data as the contributing lenders have used in arriving at their own credit decisions. To achieve this goal, we identified and corrected a number of potential data biases. One potential bias stems from the fact that relatively few lenders have linked their financial statement databases into their loan accounting systems making it difficult to separate those financial statements associated with actual borrowers from those associated with prospects. Another stems from the fact that many institutions do not continue to spread financial statements on classified borrowers, resulting in our loss of important observations on the financial condition associated with weak borrowers for the critical period prior to default.²⁰

Data Composition

As of May 1, 2000, the CRD contained 1,621 defaults and loan accounting and financial statement data on over 28,000 confirmed borrowers. The lenders in the CRD had \$31 billion outstanding to these borrowers.

We chose a subset of that data to specify and validate the RiskCalc for Private Companies. We excluded finance, insurance, or real estate firms (6000 series SIC code), or if they had total assets below \$100,000. Additionally, we used only fiscal year end financial statements. These criteria reduced the data set to approximately 115 thousand financial statements on 24,710 confirmed borrowers. Fourteen contributors were large regional or money center banks in the US, plus two large Canadian banks.

The CRD spans the period from 1989 to 1999, though the data are concentrated in the period following 1994, with approximately 60% of the financial statement observations coming from 1995 and later. The sharp drop off in 1999 observations is a result of two characteristics of private firm data. First, the majority of private firms have calendar year ends, and second, banks do not receive and spread the vast majority of fiscal year end statements until May of the following year.²¹ Defaults are also concentrated in the more recent years. This bias is a result of the availability of the data and the sampling techniques employed by our participants. Clearly, recent defaults are easier to manually locate than older defaults.²² Additionally, using current loan accounting system data (which usually does not contain archival data) to identify defaults creates a bias towards recent years.²³ These distributions of the financial statement and one-year default observations are shown in Exhibit 3.1.



¹⁸ This section was completed by Moody's Associate Jim Herry, CRD database initiative head.

¹⁹ A small percentage of public firms may be present in the private firm database. Most borrower names were encrypted before the data was submitted. Participant banks include Bank of America, Bank of Montreal, Bank of Hawaii, Banque Nationale du Canada, Bank One, CIBC, CIT Finance, Citizen's, Crestar, First Tennessee, Hibernia National Bank, KeyCorp, People's Heritage/Banknorth, PNC Bank, Regions, Toronto Dominion.

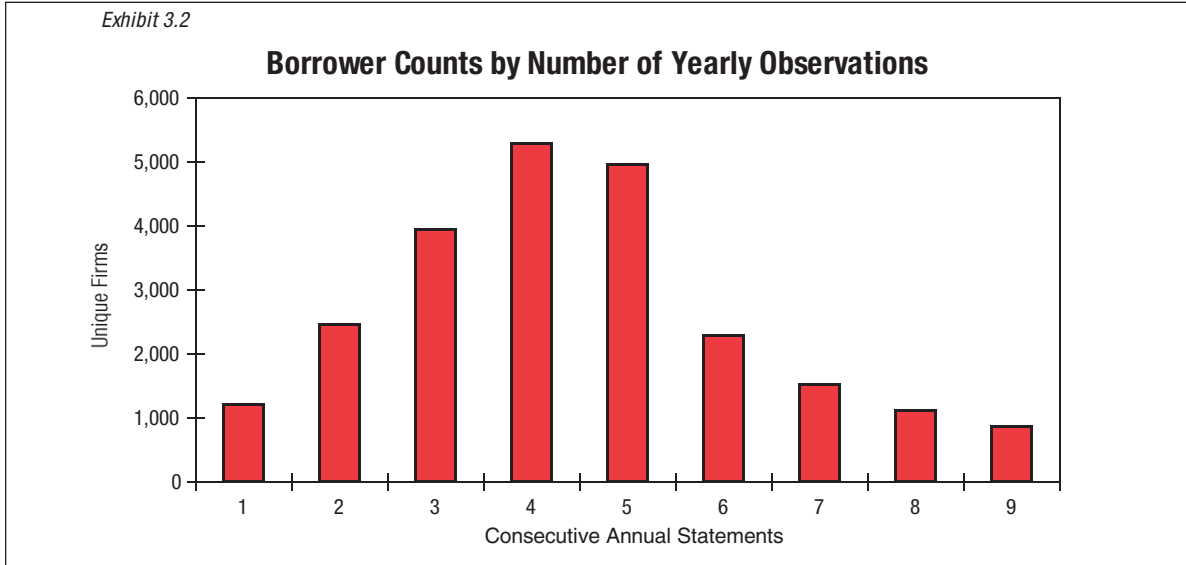
²⁰ Many institutions transfer credits to special asset groups once a credit is placed in any of the regulator criticized asset categories. Once there, many institutions do not continue to spread the financial statements associated with these high risk borrowers, or the borrowers no longer submit them.

²¹ The submission period for this version of the CRD closed on March 1, 2000.

²² Many institutions do not update financial statements on problem credits, so obtaining old defaults usually entails researching through dated credit files with incomplete financial statements.

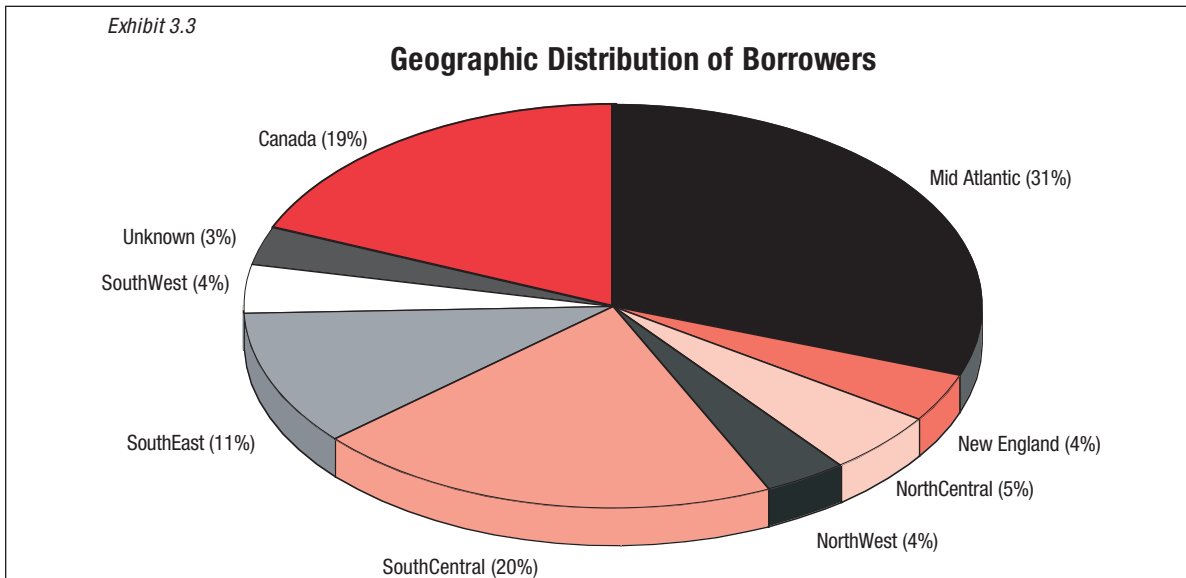
²³ We attempted to reduce this bias by reviewing loan accounting system delinquency counter fields (e. g. times 90 days past due, etc.) and reviewing "charge-off" bank data to identify defaults no longer on the books.

In the subject CRD test sample, 1,209 borrowers have only one financial statement observation, while 23,501 borrowers have multiple observations for different lengths of time. Multiple observations allow us to evaluate the extent to which trends in financial ratios help predict defaults. The distribution of the number of observations per borrower is presented in Exhibit 3.2 below.



Geographic Distribution of Data

The borrower's state or province was determined using the loan accounting system data.²⁴ Forty-five percent of the borrowers were domiciled on the East Coast of the United States. The national breakdown is 78% US and 19% Canadian, with 3% unknown. The distribution of borrowers by location is presented in Exhibit 3.3 below.

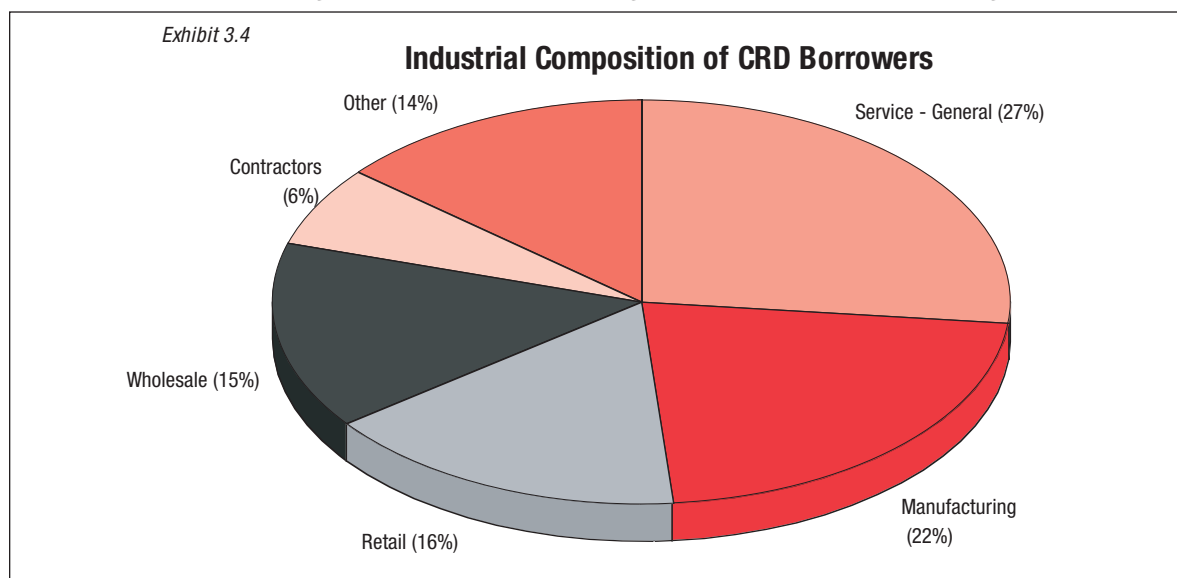


Industrial Composition of Data

We could associate approximately 70% of the CRD borrowers with a four-digit SIC code. For analytical simplicity, we rolled SIC codes up into 14 broad industry groups based on SIC categories.

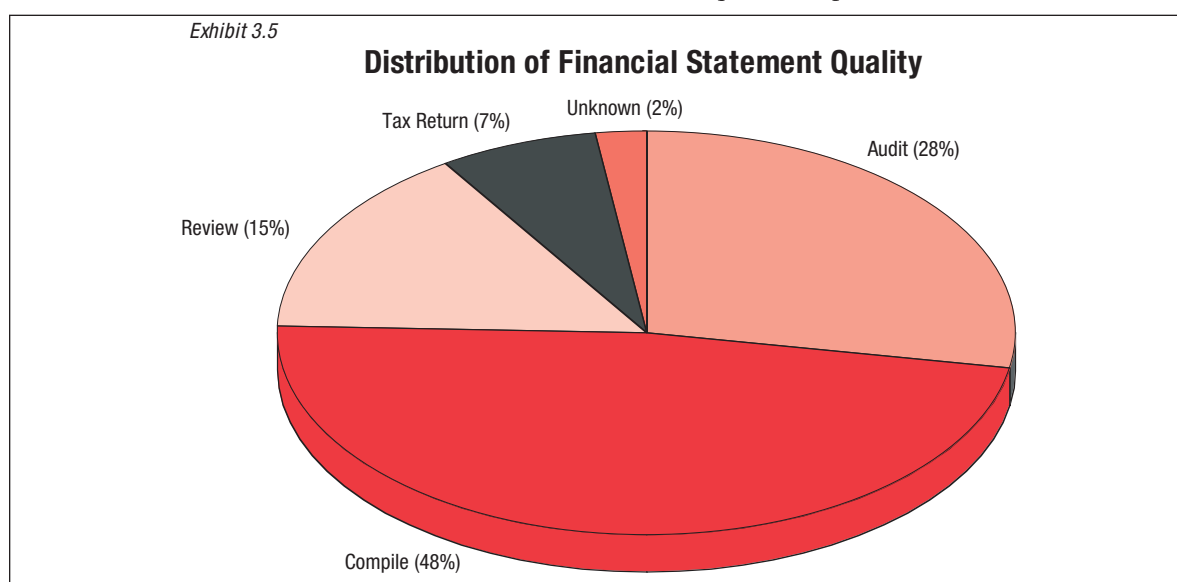
²⁴In the June 1999 version of the CRD, we were not able to identify the state or province of 26% of the borrowers. The improvement in this version of the CRD is due to the introduction of matched loan accounting system data.

The distribution of the borrowers is presented in Exhibit 3.4. Note that this graphic represents the entire CRD confirmed borrower base. Borrowers categorized in Holding Companies, Financial, and Real Estate, totaling 11%, were not used in the subject model's specification. Approximately 72% of the confirmed borrowers were categorized in the Manufacturing, Retail, Service, or Wholesaling sectors.



Financial Statement Quality

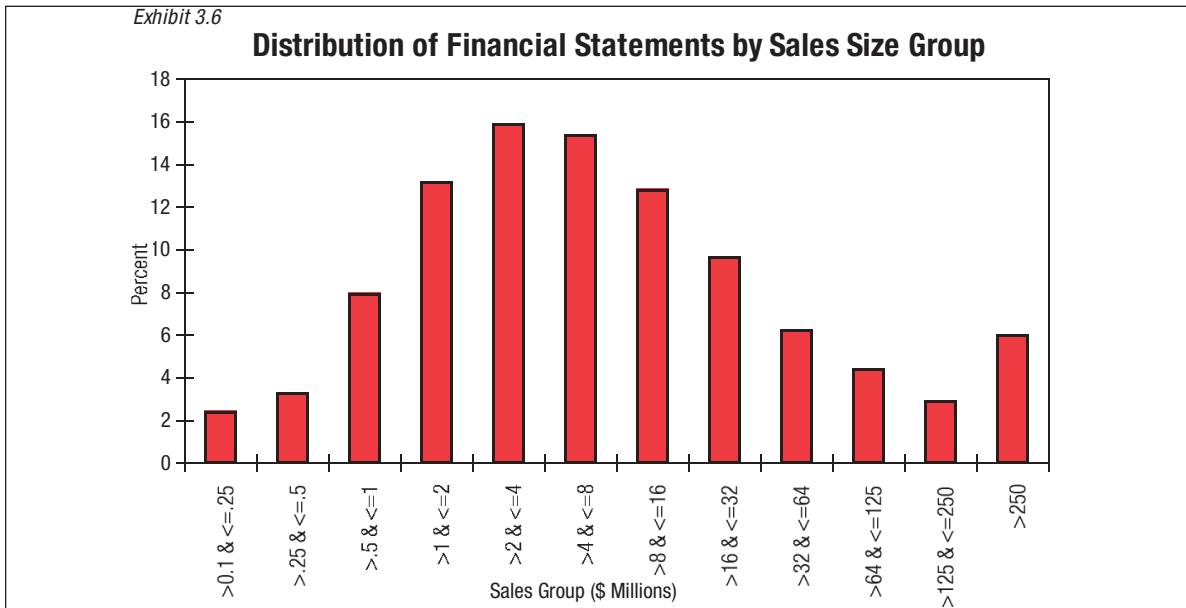
A feature of the CRD data set is that it contains the same private firm data lenders use for their credit approval and management decisions. This is a clear differentiating factor from the Compustat database which contains predominately audited financial statements. The participants' data included a statement type identifier on 98% of the financial statements. The statement quality designations varied by institution and statement spreading software. The original data set had over 500 different statement quality types. Our staff examined those designations and grouped them into 5 broad statement categories: audit, compile, review, tax return, and unknown. The distribution of statement qualities is presented in Exhibit 3.5.



Audit is the highest level of financial statement quality, but it is the costliest, and for many firms exceeds the benefits derived. The next level of financial statement quality is the 'review', which is an expression of limited assurance. The accounting communicates that the financial statements appear to be consistent with Generally Accepted Accounting Principles. A 'compilation' is basically the outside construction of financial statement numbers so they are presented in a consistent format, yet management's presentations of the numbers that underlie this compilation are taken at face value. Finally, the 'tax return' is the creation of financial statement numbers from the tax return for the company.

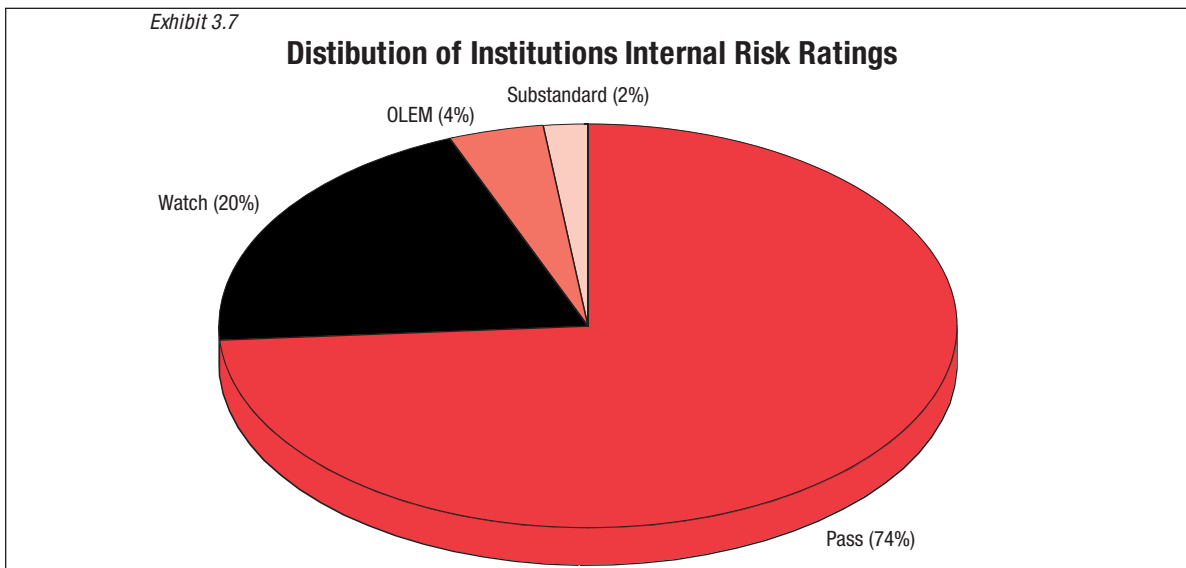
Sales Size

Commercial lenders grant credit to private firms which are substantially smaller than those borrowers served by "corporate" banking divisions. The distribution of sales size in the CRD data set is representative of the distribution of our participants' commercial borrowers. From Exhibit 3.6 we see that approximately 14% of the confirmed borrowers had sales less than or equal to \$1 million, approximately 44% had sales greater than \$1 million and less than or equal to \$8 million, approximately 29% had sales greater than \$8 million and less than or equal to \$64 million, and the remaining 13% had sales above \$64 million. None of the participants provided a means to positively identify public firms in their data set, so we were not able to exclude all public firms. Consequently, we suspect that some of the larger sales observations are actually from public firms.



Financial Institutions' Internal Risk Ratings

The loan accounting system data included the institutions' internal borrower risk rating. All institutions had the regulatory criticized asset classifications corresponding to Other Loans Especially Mentioned (OLEM), Substandard, Doubtful and Loss. Additionally, all institutions had a separate "Watch" or "Caution" category, where credits are usually not originated, but represent a deterioration that bears extra attention. The number of "pass" grade categories used by any one institution ranged from a low of 3 to a high of 8 with the median being 5 pass grades. The distribution of the internal ratings is presented in Exhibit 3.7. Note that due to the wide variation in pass grade scales and the associated definitions, all pass grades have been placed in one category.



Summary of Databases

The bottom line to any database used for default modeling is how many non-defaults and defaults were available for estimating the default prediction parameters. Below we list the key database statistics described above, where all data is from nonfinancial firms.

Exhibit 3.8

Database Summary				
Time Span	Unique	Firms	Unique Firm Defaults	Financial Statements
Private Firms	1989-99	24,718	1,621	115,351
Public Firms	1980-99	15,805	1,529	130,019

Section IV: Univariate Ratios As Predictors Of Default: The Variable Selection Process

Men who wish to know about the world must learn about it in its particular details.

~ Heraclitus

Financial ratios are related to firm failure the way that the speed of a car is related to the probability of crashing: there's a correlation, it's nonlinear, but there's no point at which failure is certain. Failure depends upon other variables, such as the environment and driver skill. When modeling such phenomena, one must, therefore, build in a tolerance for extreme values, as these observations do not necessarily relate to failure - as evidenced by their prevalence.

Yet all is not noise: higher leverage, lower profitability, firm size, and liquidity are all related to default in ways most credit analysts would expect. What is not clear is the shape of the relationship (what default rates correspond to which input ratio levels), which ratios are most powerful (e.g., profitability: net income or EBIT?), and how correlations affect the relative weightings assigned to these various ratios in a multivariate context.

The selection of variables and their transformations are often the most important part of modeling default risk. While some distinctions are relatively insignificant—such as whether one uses the quick ratio or current ratio for liquidity, assets or sales for size—the inclusion or exclusion of certain variables can make a major difference. The purpose of analyzing the various ratios individually as we do in this section is to demonstrate the univariate power of many popular candidates for inclusion in the model.

Dawes and Corrigan (1974) argue that in empirical prediction 'the whole trick is to know what variables to look at and then know how to add.' While this position may be a bit extreme, they are on to something. Once one has the most important explanatory variables and those variables are appropriately normalized, the problem often displays a 'flat maximum'. Many different sets of coefficients produce an output nearly the same as what would come from an optimal model. This result was originally pointed out by Wilks (1938), who examined the situation where there was positive correlation between predictors. For example, the correlation between Z_1 and Z_2 , where $Z_1 = X + 2Y$ and $Z_2 = 2X + Y$ and X and Y are both independent and normally distributed, is 0.8. That is, two very different coefficient weightings yield surprisingly similar results; finding X and Y is more important than determining their weights.

Virtually all potential explanatory variables are ratios in that they are normalized for size, which is included as a separate factor. This is because 'high profitability' is obviously different depending on whether one is analyzing a multinational conglomerate or a regional warehouse. Company size varies by several orders of magnitude across firms and this makes items like total net income more correlated with size than true profitability as usually conceived. Further, by using ratios one can avoid differences from time variation in the value of money (this is why size is normalized by a price index while other ratios are not). Each explanatory characteristic can have several different representations (EBIT vs. EBITDA), and explanatory variables can be related to more than one risk factor (retained earnings/assets is related to leverage and profitability).

Chen and Shimerda (1981) list over a hundred ratios cited in the financial distress and other literature, more than anyone has time to analyze systematically. This highlights the main problem of financial ratios: there are too many of them. For reasons addressed below, all of them cannot be used, so one must find a way to achieve the optimal subset.

The most transparent way to observe this process of winnowing down potential variables is to examine graphs that show the power of individual ratios. That is,

1. Rank firms by a ratio such as the net operating margin.
2. Divide into 50 groups and examine the future firm default rate for those groups.
3. Smooth the default rates to remove noise.
4. Examine the resulting graphs.

This section takes you through the variable selection process in this manner.

The Forward Selection Process

Without a structural model that dictates the inputs of a model, there are two main methods for selecting the appropriate variables. The first is forward selection. Start with those independent variables that have the highest univariate correlation and then add those with lower correlation until additional variables have no additional significance.²⁵ The second is backward elimination. In this approach one starts with all the variables, then reduces all of the insignificant variables.

For this problem, forward selection was preferred because the hundred variables precluded the backward selection process. The graphs below illustrate the forward selection process, highlighting those variables that are most powerfully related to default.

Good variables for inclusion in a model tend to be those variables that have a conditionally monotone relationship with default probabilities. Monotone means the relationship is always increasing or always decreasing, not decreasing over part of its range and increasing over another part. Conditionality refers to the independence of the relationship; if the relation between liquidity and default is negative for small firms and the exact the opposite for large firms (i.e., conditional upon size), such instability does not bode well for out-of-sample prediction (Kranz (1972)). These are guidelines rather than rules. For example, sales growth's relation to default is not monotonic, but it adds value to the model and it is included.²⁶

Adding regressors always increases fit (e.g., R^2), but also always increases the variance of the predicted variable. In the words of Khaneman and Tversky, "a paradoxical situation occurs in estimation where high correlation among inputs increases confidence while decreasing validity" (Khaneman and Tversky (1982), p. 65). Adding net profit margin, gross profit margin, net income/equity and EBIT/interest appears at first like a great way to totally capture the many aspects of 'true' profitability. Other manifestations of 'factor overload' include looking at means, trends and levels over many different horizons for the same variable. While common, such an approach is simply not optimal. The high degree of correlation between these measures will worsen out-of-sample performance and obscure interpretability of the real drivers of default.

Psychologists such as Khaneman have documented that additional information adds linearly to one's confidence, even though after a certain point the inflated standard errors from collinearity worsen out-of-sample prediction. It may seem prudent to 'look at everything', and this definitely allows better ex-post anecdotal explanation,²⁷ but it is a statistical fact that additional information, after a certain point, just adds confusion in the form of worse predictability. For example, in a multivariate model with two explanatory variables, x_1 and x_2 , the correlation between these two inputs affects the standard error on their coefficients proportional to $1/(1-\text{corr}(x_1, x_2))$. The higher the correlation between x_1 and x_2 , the higher the standard error on the coefficients. From a prediction, not an explanation, standpoint, there is a trade-off to adding more explanatory variables.

While highly correlated inputs are not helpful, the other extreme, complete independence among inputs, is not necessary. In fact a modest degree of positive correlation can be a good thing, as reflected in the Wilks example using $2X+Y$. While he found that independent models with different weighting generated highly correlated outputs, this correlation only rises if the correlation of two regressors, X and Y is positive.

²⁵ Given the sequential nature of the selection process the assumptions underlying the t-tests are violated, and so the t-tests themselves highly suspect. Nonetheless this is a common rule-of-thumb for deciding whether to use additional variables.

²⁶ In general, growth or trend information tended to be the only inputs with nonmonotonic relationships that appeared stable, or where not the result of perverse meanings (e.g., see discussion of NI/book equity in this section).

²⁷ With many different inputs, eventually one is guaranteed the ability to 'explain' what eventually happens.

It is a common misconception that correlation among regressors biases coefficients. The problem is not one of bias, but of inflated variances of coefficients. Instead of getting within 20% of the true coefficient, you get within 30%. Hopefully, the added imprecision of the coefficient estimates is outweighed by the extra information from the additional variables. This is the dilemma of adding more variables, one that is biased more and more against the addition of new variables as the set of used variables increases.

Another issue with adding many explanatory variables is the 'wrong sign' problem. If relations between two explanatory variables and the dependent variable are individually positive, in a multivariate context one may have a negative coefficient. For example, if you use both net income/assets and EBIT/assets in a multivariate model of default, both are correlated with default and with each other. In a statistical estimation, invariably one has a coefficient sign consistent with its univariate relation, while the other (the weaker univariate predictor) has an opposite sign.²⁸

This result implies meaningless 'what if' analysis (If EBIT is higher, a wrong sign problem could attribute then higher probability of default). The result also generally has poor out-of-sample properties because it places greater demands on the model by requiring that correlations among the predictor variables in the derivation sample must also exist in the prediction sample (Zavgren (1983)). For these reasons, we are not only seeking powerful predictors, but parsimony.

The variable selection process therefore consists of the following exercise. First, find the most powerful ratios that reflect the most obvious risk factors: profitability, leverage, firm size, and liquidity. Then, sequentially add ratios and see if they add statistical significance to the group. Usually, the more powerful risk factor ratio, such as net income/assets, when used with a similar, correlated measure, such as net sales margin, will generate coefficients where the more powerful ratio has a positive coefficient and the less powerful ratio has a negative coefficient, for reasons mentioned above. We do not use the additional ratio if it contains a 'wrong sign' or if it is statistically insignificant. This is the stepwise process of variable selection: suggested by the univariate power, validated in a multivariate context.

Factor analysis also provides useful insights for analyzing the right number of explanatory variables to use. One or two ratios determined to be representative of a factor grouping are selected from each factor. Chen and Shimerda (1981) looked at five studies by researchers who used factor analysis and concluded that the obtained factors were very close to seven factors identified by Pinches, Mingo and Caruthers (1973). In another study, Libby (1975) uses factor analysis to reduce a fourteen variable set to five factors, selecting one variable from each factor. The reduced five variable set predicted nearly as accurately on the derivation sample and more accurately on the prediction sample than the initial fourteen variables, demonstrating how too many variables may lead to overfitting. Gombola and Ketz (1983) performed factor analysis on a set of forty variables using data on 119 companies over a 19-year period. They concluded that for all but a few years there were eight significant factors. Seven were substantially similar to the seven found by Pinches, Mingo and Caruthers.

This all suggests that a model of approximately seven well-selected variables is optimal for the purpose of predicting default/bankruptcy for private firm default models. Seven is not a magic number; it is just a useful intuitive point of reference to note that the number is seven, not 2 or 20. One should also remember that the sample sizes in the above-cited studies were much smaller than ours, which will bias the number of identified factors downward.

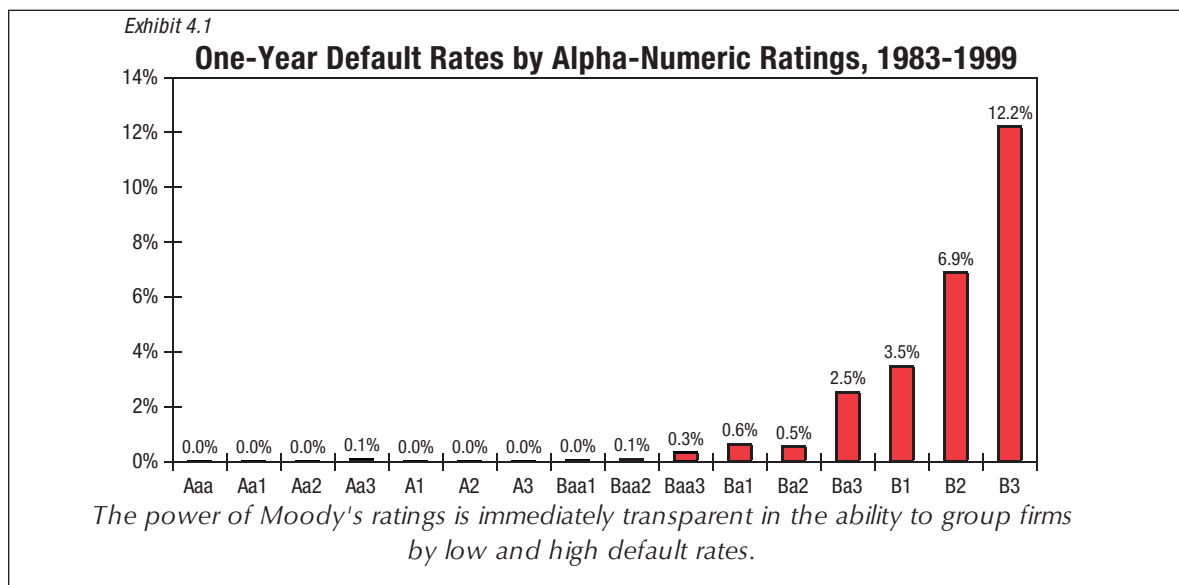
²⁸For a model, where $y = \beta_1 x_1 + \beta_2 x_2$, where

$$\beta_1 = \frac{\rho_{1,y} - \rho_{1,2}\rho_{2,y}}{1 - \rho_{2,y}^2}$$

The coefficient can only be negative if $\rho_{1,2} > \rho_{1,y} / \rho_{2,y}$. That is, if the correlation between regressors is greater than between the regressors and the predicted variables. For example, if $\rho_{1,y} = .3$, $\rho_{2,y} = .8$ and $\rho_{1,2} = .4$, then the sign of β_1 will be 'wrong', i.e., it will be negative in a multivariate context but positive in a univariate relation. See Ryan (1996 p.132).

Statistical Power and Default Frequency Graphs

Prior to examining the univariate ratios and their relation to default prediction, it is useful to explain some fundamental statistical concepts. As much of the evidence and explanation on variable selection and testing in subsequent chapters is graphical, it is essential to know how various graphs relate to statistical measures of 'good' and 'better' models for predicting default. The main graphical tools we will use are default frequency graphs, in which the frequency of default is on the y-axis and ratio level or its percentile is on the x-axis. For example, Moody's ratings are widely recognized as powerful predictors of default. This is often demonstrated by the information in Exhibit 4.1.



Lower ratings have higher default probabilities, and this relation is nonlinear in that the default rate rises exponentially for the lowest rated companies. Any powerful predictor should be able to demonstrate a similar pattern. When ranked from low to high, the metric should show a trend in the future default rate, the steeper the better; if not, the metric is uncorrelated with default.

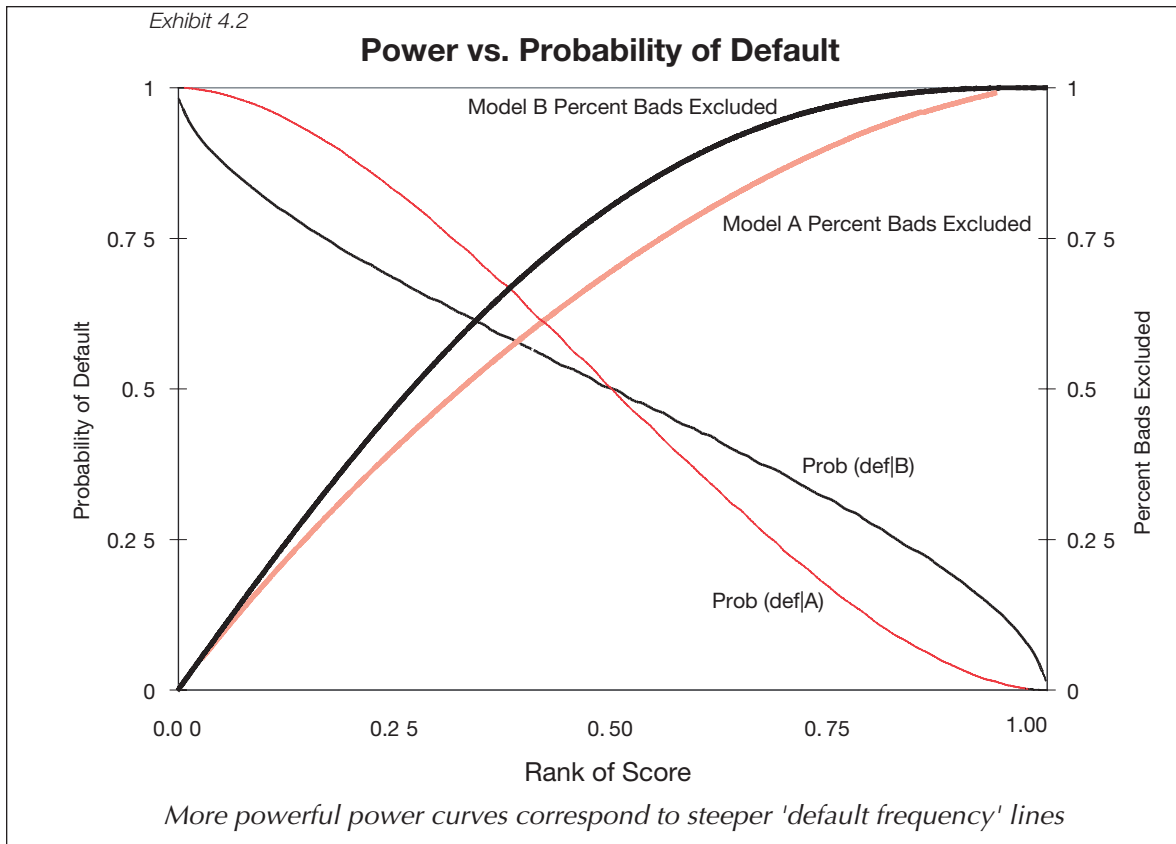
This may seem a simple point, but it underlies much of modern empirical financial research. For example, the initial findings relating firm size and stock return (Banz 1980) or book/equity and return (Fama and French 1992) both could be displayed in such a simple graph. If a relationship can only be seen using obscure statistical metrics, but not a two-dimensional graph, it is often - and appropriately - viewed skeptically.

A model may be powerful, but not calibrated, and vice versa. For example, a model that predicts that all firms have the same default probability as the population average is calibrated in that its expected default rate for the portfolio equals its predicted default rate. It is probable, however, that a more powerful model for predicting default would include some variable(s), such as liabilities/assets, that would be able to help discriminate to some degree bads from goods.

A powerful model, on the other hand, can distinguish goods and bads, but if it predicts an aggregate default rate of 10% as opposed to a true population default rate of 1%, it would not be calibrated. More commonly, the model may simply rank firms from 1 to 10 or A to Z, in which case no calibration is implied as default rates are not generated.

Calibration can, therefore, be appropriately considered a separate issue from power, though both are essential characteristics of a useful credit scoring tool. The driver of variable selection is one of power and is discussed in this section and the next; calibration is discussed in the section on mapping model outputs to default rates (section 7).

Power curves and default frequency graphs are related in a fundamental way. A model with greater power will be able to generate a more extreme default frequency graph. Consider the two models presented below. Model B is more powerful, and this is reflected in a power curve that is always above Model A.



For the power curves (the thick lines), the vertical axis is the cumulative probability of 'bads' (right-hand side), and the horizontal axis is the cumulative probability of the sample. The higher the line for any particular horizontal value, the more powerful the model (i.e., the more Northwesterly the line, the better). In this case, Model B dominates Model A over the entire range of evaluation. That is, Model B, at every cut-off, excludes a greater proportion of 'bads'.

The white lines portray the estimated default probability for various percentiles of scores generated by the models. The percentile of the score is graphed along the horizontal axis and the probability of default for that percentile is on the vertical axis (left-hand side). Steepness of the default frequency lines and more traditional graphs of statistical power are really two sides of the same coin, thus if we observe a steeper line, this corresponds, generally, to a more powerful statistical predictor. The actual probability of default for each percentile of Model B's scores is either higher or lower than for Model A. The more powerful Model B generates more useful information, since the more extreme forecasts are closer to the 'true' values of 0 and 1, subject to the constraint of being consistent (i.e., a forecast of 8% default rate is really 8%).

There is in fact a mathematical relation between the power curves and default frequency graphs, such that one can go from a default frequency graph to a power curve. Given a default frequency curve, one can generate a power curve. Given a power curve and a sample mean default rate (so one can set the mean correctly) one can generate a default frequency curve. Power, calibration, and the relation between power curves and default frequency graphs are discussed further with an example in Appendix 4A.

The statistical power of a model - its ability to discriminate good and bad obligors - constrains its ability to produce default rates that approximate Moody's standard rating categories. For example, Aaa through B ratings span year-ahead default rates of between 0.01% and 6.0%, and many users think that all rating systems should map into these grading categories with approximately similar proportions, that is, some B's, some Ba's, and some Aa's. Yet, it is not so straightforward.

For example, someone may map their model into the Moody's universe by noting that since 20% of Moody's ratings correspond to B1-rated companies and below, the lowest 20% of their model should be rated B1 or below. This does not guarantee that the new 'mapped' scores will produce 6% annual default rates as we see with Moody's ratings. In order to generate such a high default rate (compared to the average 1.2% corporate default rate), such a model would need equivalent statistical power as Moody's ratings. Without power, default rate predictions can not deviate from the overall mean expected default rate; the more power, the more they can deviate from the mean.

Thus, a model that does not have equivalent granularity to Moody's ratings is not necessarily biased. Indeed, given the high power of Moody's ratings, many mappings that are too similar to Moody's ratings distribution are probably biased (i.e., their higher quality credits probably default at higher rates than Moody's, and their lower quality credits probably default at lower rates than Moody's).

To generate the default frequency graphs below, we divided firm-year observations into 50 groups, and examined whether or not default occurred in 90 days to 5 years from the statement date. The subsequent default rate for firms with net income/asset ratio within each group was calculated. The result was then smoothed to reduce noise using a Hodrick-Prescott filter.²⁹

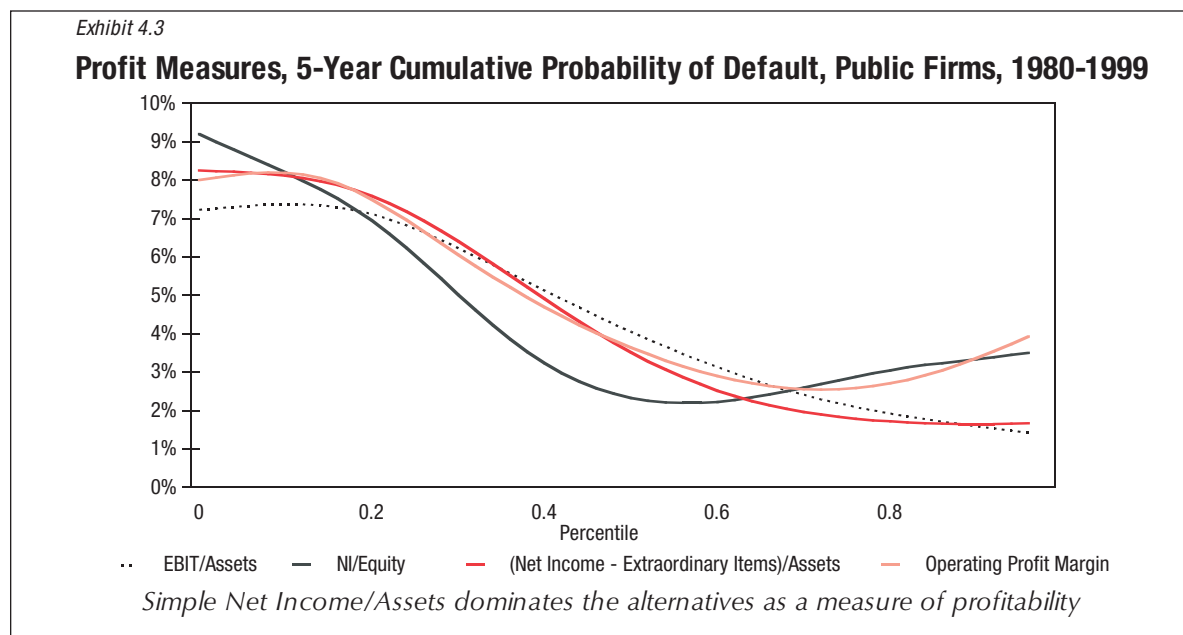
All the graphs below were estimated using Compustat financial data on publicly traded firms, matched with Moody's default database, over the 1980 - 1999 period. We are primarily using this data for these charts because we wish to keep as much of our private dataset as proprietary as possible, and the demonstration of the points made in the variable selection process translate directly into the private firm case, unless otherwise noted in the text.

The 5-year cumulative default probability was chosen for two reasons. First, it generated a larger number of useable defaults than the 1-year probability, which allowed better estimation because of the greater number of observations. Secondly, and most importantly, firms have a mortality curve. Very few firms, of any risk level, default within one year of borrowing. And so prediction of the default rate one year after loan origination, is not particularly interesting. Most of these early firm failures involve fraud, in which case the financial ratios ex ante are of little use.

Predicting annual default rates, like quoting annual yields on bonds, may be the best way to normalize predictions, but this representation should not be the primary target of either estimation or testing because of its limited relevance to lenders. The average time to default, and the average contractual bank loan maturity, is about 4 years. A 5-year cumulative default rate, therefore, is more relevant, and generates more data, than 1-year measures.

Profitability Ratios

Higher profitability should raise a firm's equity value. It also implies a longer way for revenues to fall or costs to rise before losses occur.



²⁹See Hodrick and Prescott (1997). The smoothing parameter is set at 100.

Among all the potential risk factors, there are more profitability ratios than any other factor. The set of profitability measures we display - EBIT/assets, net income/common equity, net income/assets, and operating profit margin³⁰ - are in Exhibit 4.3.

As this is the first of our charts related to variable selection, remember that we are looking for steep lines, as these indicate more powerful predictors of default. One issue addressed in Exhibit 4.3 is whether profitability is best gauged relative to assets or equity. It appears that net income/assets (i.e., NI/A) dominates NI/equity. NI/equity rises slightly more strongly at the low end, yet at the high end something perverse occurs: higher NI/equity implies a higher default above the 60th percentile.

Clearly, this goes against intuition: higher profitability, higher default. This effect, however, is driven more by the denominator, book equity, than the numerator, net income. In fact, 10% of all public companies and 12% of all private companies have negative book equity, which makes NI/equity extremely confusing conceptually (e.g., a negative ratio can result from either negative NI or negative book equity). At the high end of the NI/equity percentile, these companies tended to have very low levels of book equity as opposed to high NI, which inflates the ratio NI/Equity. Using the absolute value of equity does not eliminate the problem, because those values close to zero will generally have very high profitability ratios. Because of NI/assets comparable performance relative to NI/equity and an 'outside the model' knowledge of the perverse effects of negative and near-zero equity, we prefer NI/assets to NI/equity.

Operating profit margin does worse than net income/assets, as it shows less variation in default probability among the higher levels. Unreported tests on the gross profit margin were even worse. Profitability is essentially related to margin (i.e., markup) times quantity. A sales margin abstracts from the quantity dimension, and thus total profitability. Sales margins also vary across industries, which makes it less useful for gauging credit quality on a universe-wide basis. Lending is primarily driven by the need to finance assets, not sales (liabilities/sales is much less stable across firms compared to liabilities/assets). The return on investment, therefore, is better described as the profitability/assets.

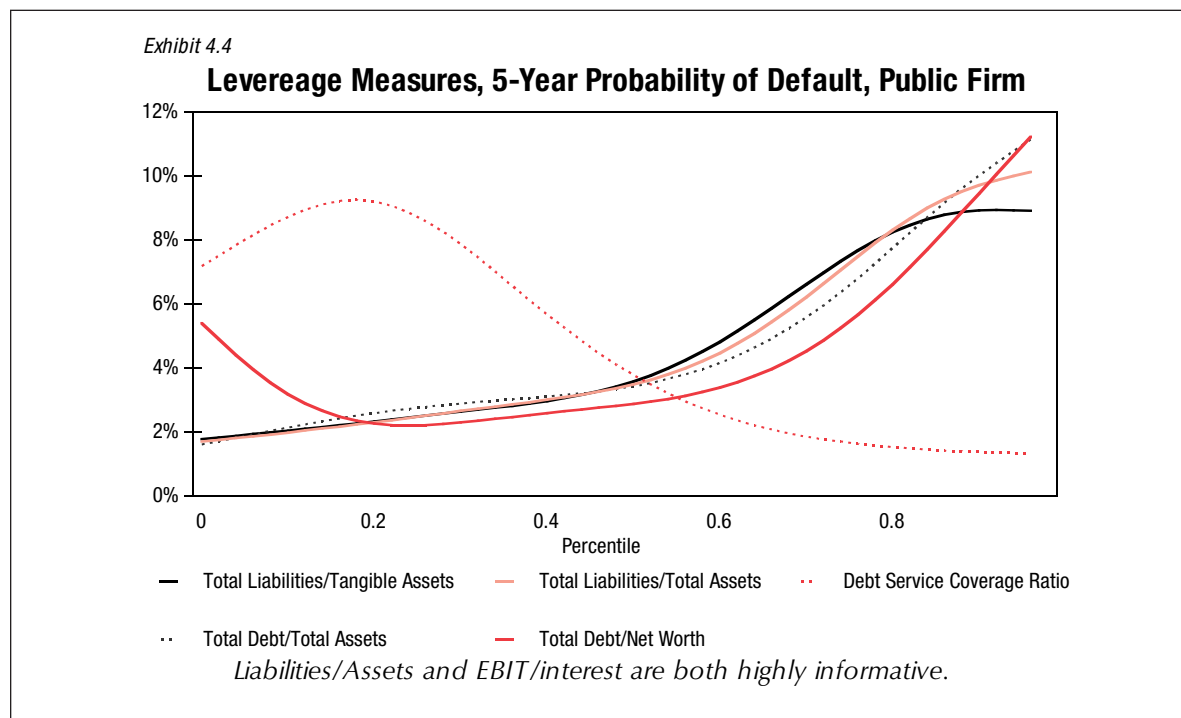
'Cash flow' is less valuable as an univariate predictor than simple net income. In Exhibit 4.3 we see that EBIT/Assets is less steep than NI/Assets, reflecting the greater power of NI/Assets. In fact, when we tested a more sophisticated measure of cash flow, putting in changes in accounts payable and accounts receivable, it did considerably worse. Unreported tests that amended EBIT to EBITDA or to EBITDA - Capital Expenditures made little difference in default prediction performance. It appears that interest, taxes, and capital expenditures are not something to be abstracted from when evaluating profitability/assets. This is the most controversial of the findings in our variable selection process. Our preference is towards relationships that are intuitive, which in practice is often synonymous with customary. Clearly EBIT, EBITDA, or EBITDA - capital expenditures, are more common concepts of profitability for experienced credit analysts than simple 'net income minus extraordinary items.' Yet the data suggest that the traditional accounting focus upon net income, not earnings abstracting from accrual items, is what's most important in measuring firm default risk.

Unreported tests show that net income-extraordinary items/assets dominates NI/A alone, which is consistent with taking the term 'extraordinary' at face value. We therefore use this measure of net income.

³⁰ Operating profit margin = (sales - cost of goods sold - sales, general, and administrative expenses)/sales

Leverage Ratios

In addition to profitability, leverage is a key measure of firm risk. The higher the leverage, or gearing, the smaller the cushion for adverse shocks.



The debt service coverage ratio (a.k.a. interest coverage ratio, EBIT/interest) is highly predictive, although at very low levels of interest coverage, default risk actually decreases. This is due to the fact that extremely low levels of this ratio are driven more by the denominator, interest expense, than by the numerator, EBIT. These low values have, in many cases, slightly negative EBIT, but a much smaller interest expense. The measure's sharp slope over the range 20% to 100% suggests that it is an interesting candidate for the multivariate model and a valuable tool for discriminating between low and high risk firms with 'normal' interest coverage ratios.

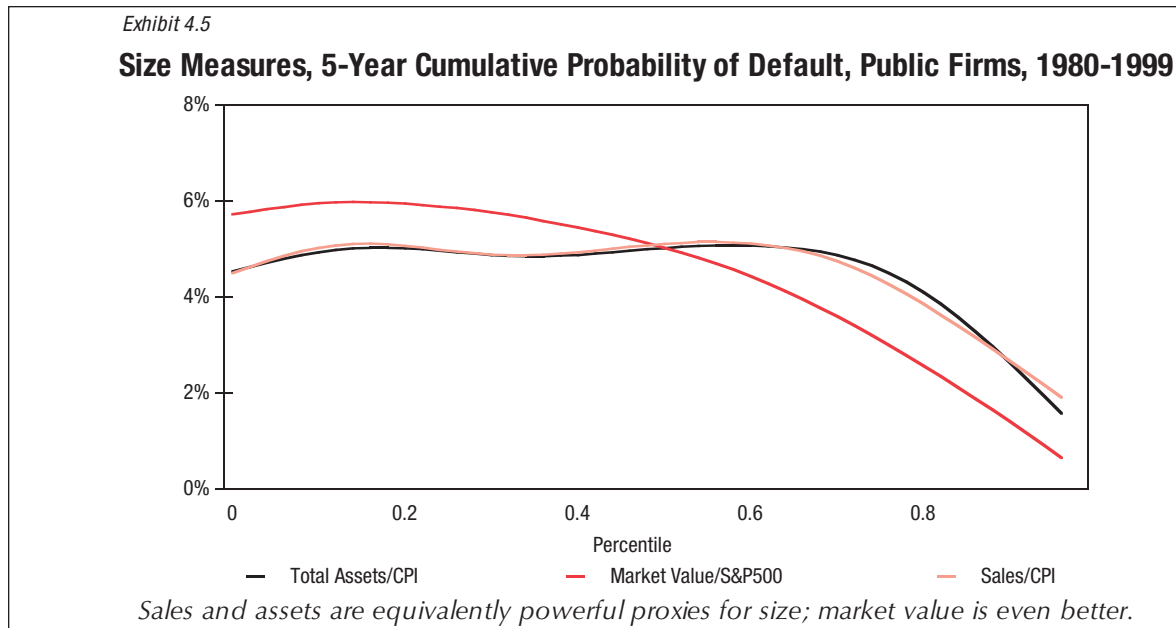
In fact, EBIT/interest turns out to be one of the most valuable explanatory variables in the public firm dataset in a multivariate context, though in the private firm database, its relative power drops significantly. For private firms, it moves from most important to one of the lesser important inputs. It is unclear exactly why this is so, as it could be related to measurement error (e.g., interest expense on unaudited statements is often inconsistent with the amount of liabilities documented).

For the other measures, equity/assets is basically the mirror image of liabilities/assets (L/A) as is expected: they are mathematical complements. We chose L/A as opposed to equity/assets, as it is more common to think of leverage this way. Debt/net worth showed a non-monotonicity, because for the extremely low values, net worth is actually negative, making the ratio negative. Again, it is not helpful when a ratio has this knife-edged interpretability (negative values could be due to low earnings, or high earnings but negative net worth), and we therefore excluded it.

While debt/assets (D/A) does about as well as L/A for public firms, it does considerably worse among private firms, which makes L/A preferred. One does not lose any power by using L/A as opposed to debt/assets on public firms, while it strictly dominates D/A when applied to private firms. The difference between debt and liabilities is that liabilities is a more inclusive term that includes debt, plus deferred taxes, minority interest, accounts payable, and other liabilities.

Many credit analysts use the tangible asset ratio, that is, the ratio of debt to total assets minus intangibles, as they are skeptical of the value of intangibles (in fact, totally unappreciative of their value). It appears, however, that subtracting intangible assets does not help the leverage measure predict default, as ignoring intangibles produces a less steep default frequency slope than simple L/A alone. Intangibles appear really to be worth something.

Size



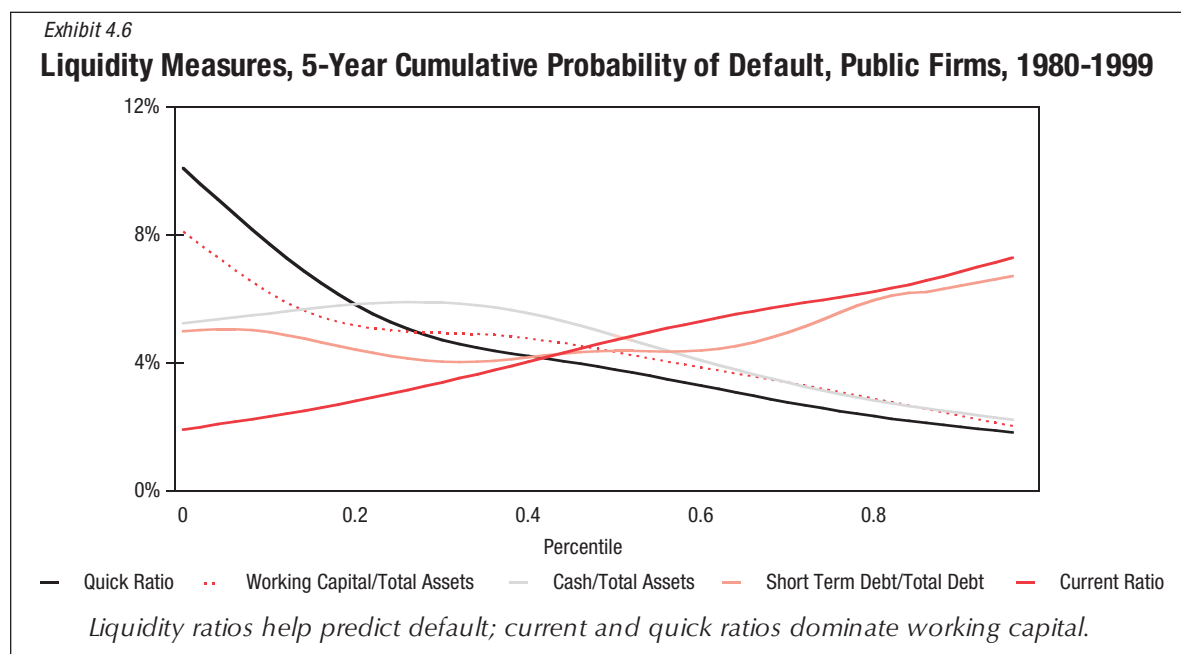
Size is related to volatility, which is inherently related to both the Merton and the Gambler's Ruin structural models. Smaller size implies less diversification and less depth in management, which implies greater susceptibility to idiosyncratic shocks. Size is also related to 'market position', a common qualitative term used in underwriting. For example, IBM is assumed to have a good market position in its industry; not coincidentally, it is also large.

Sales or total assets are almost indistinguishable as reflections of size risk, which makes the choice between the two measures arbitrary. Because assets are used as the denominator in other ratios, we will use it as our main size proxy.

We see that the market value of equity is an even better measure of default risk, and this highlights the usefulness of market value information, which by definition is not available for private firms.

Interestingly, the size effect within Compustat appears to level off at around the 60th percentile. In fact, the effect is slightly positive for smaller firms: the relation is higher size, higher risk. Therefore, bigger is better, but only for the very largest public firms. This result will be addressed in the next section, which compares these effects between the public and private versions. But we should mention here that this is probably the result of a sample bias in Compustat.

Liquidity Ratios



Liquidity is a common variable in most credit decisions - a fact that is brought to mind by the adage that a bank will only lend you money when you don't need it. That is, if you have sufficient current assets, you can pay current liabilities, but neither do you need working capital. Liquidity is also an obvious contemporaneous measure of default, since if a firm is in default, its current ratio must be low.

Yet, just as the cash in your wallet doesn't necessarily imply wealth, a high current ratio doesn't necessarily imply health. It is an empirical issue, whether or not this ratio can predict default with sufficient timeliness, since predicting default in 1 month is not as relevant to underwriters as predicting it 1 to 2 years hence.

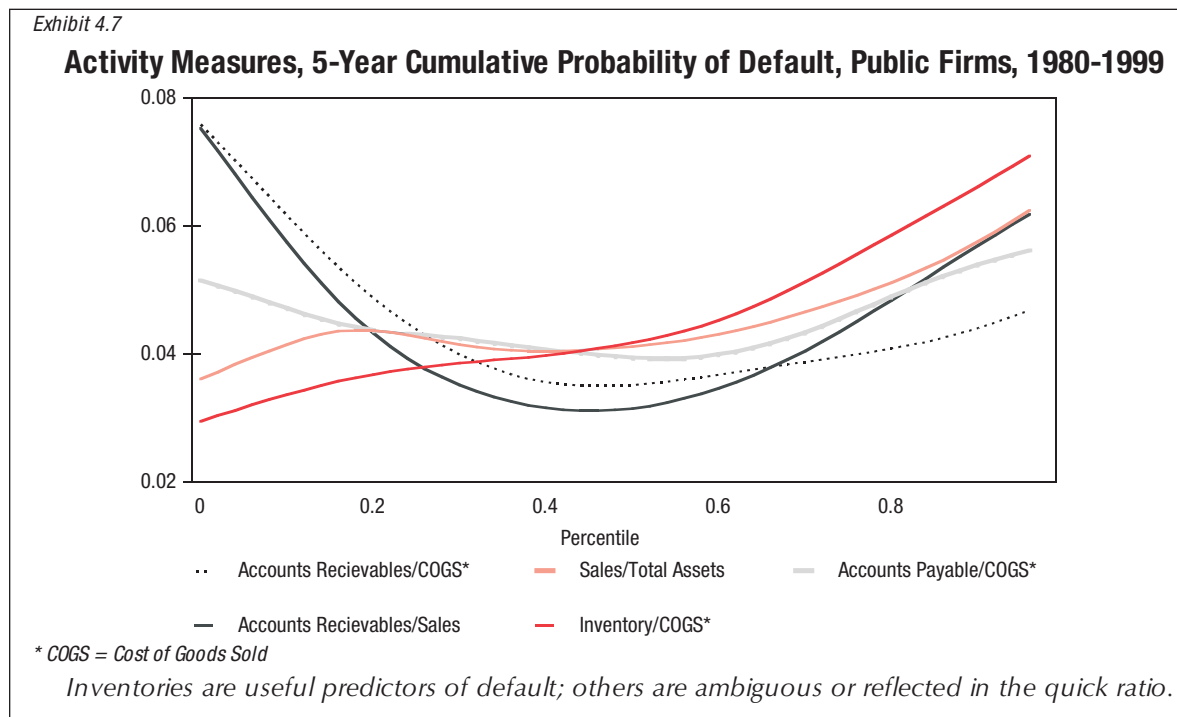
The first point to note looking at Exhibit 4.6 is that the ratio of short-term to long-term debt appears of little use in forecasting. This is of even less relevance for private firms, as banks often put loans with functionally multiyear maturities into 364-day facilities for regulatory purposes.³¹ Second, the quick ratio appears slightly more powerful than the ratio of working capital/total assets (a variable used in Altman's Z-score), though only modestly. Third, the current ratio shows a more linear relation to default, but the basic trendline is, on average, not significantly different than the quick ratio. The quick ratio is simply the current ratio (i.e., current assets/current liabilities), excepting one removes inventories from current assets. Thus the quick ratio is a 'leaner' version of the current ratio that excludes relatively illiquid current assets. As we also use the ratio of inventories/cost of goods sold in the multivariate model, the quick ratio was preferred because it abstracts from the inventories. Clearly, by themselves, the quick ratio and current ratio have roughly similar information.

Cash/assets shows only a modest relation to default. In the private dataset, however, it is the most important single variable. Just as a rich man with many credit cards would not have the same proportion of cash in his wallet relative to his wealth as a poor man with no credit, the relevance of this information is different for public and private companies. For a public company, having cash and equivalents is more wasteful, since access to capital markets implies less of a benefit for such liquidity and the absolute cost of minimizing cash increases linearly with the size of the firm. This highlights again the usefulness of our private dataset, as otherwise we would have neglected this important variable.

³¹ Banks have vastly different regulatory capital requirements for 364-day facilities versus those that are above 364 days, which creates a tendency to put what are ostensibly longer-term commitments into one of these shorter-term categories even though its practical maturity is much longer. Debt maturity for private firms is, therefore, measured with a bias, and this probably contributes to its even weaker relation to default for private firms.

Activity Ratios

Activity ratios have less straightforward relations to risk than other variables, but they do capture important information. In Exhibit 4.7 we see that sales/assets (turnover) is non-monotonic and very flat. There is no good story to explain this at this time. The other ratios show more interesting and powerful relations to future default rates. It is interesting to note that of the different incarnations of Z-score, the one that drops the sales/asset ratio performs better than the one that keeps it; the sales/asset variable degrades model predictability.



The variable we chose from this grouping was inventories/COGS,³² even though accounts payable and accounts receivable are more powerful predictors by themselves. This is because the multivariate model contains the quick ratio, which is current assets minus inventories/current liabilities, and thus much of the accounts payable and accounts receivable information is already "contained in" the quick ratio, while inventories are not. Thus, in spite of its relatively weak univariate performance, inventories/COGS dominates the alternative measures of activity in a multivariate setting.

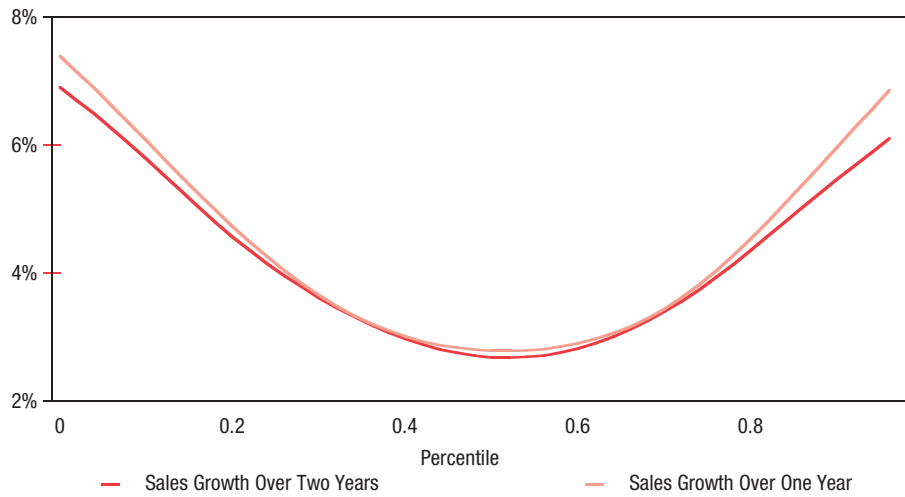
Sales Growth

Sales growth is an interesting variable, in that it displays complicated relationships to future defaults, yet we still use it. Monotonicity is a good thing in modeling, as it implies a stable and real relationship, not just an accident of the sample. Yet the non-monotonicity here is very strong and intuitive, as opposed to what one would see in results from random variation.

³²COGS = cost of goods sold

Exhibit 4.8

Sales Measures, 5-Year Cumulative Probability of Default, Public Firms, 1980-1999



Sales growth an informative though non-monotone variable.

There is a good explanation for what is driving this result. At low levels of sales growth, this metric is symptomatic of a high risk: lower sales imply weaker firm prospects. At high levels of sales growth, this metric is a cause of a higher risk: high sales growth implies the firm is rapidly expanding, probably fueled by financing, and for a significant number of these firms the future will not be as rosy as the prior year. The financing needed to fuel growth will be difficult to accommodate for a significant proportion of these firms.

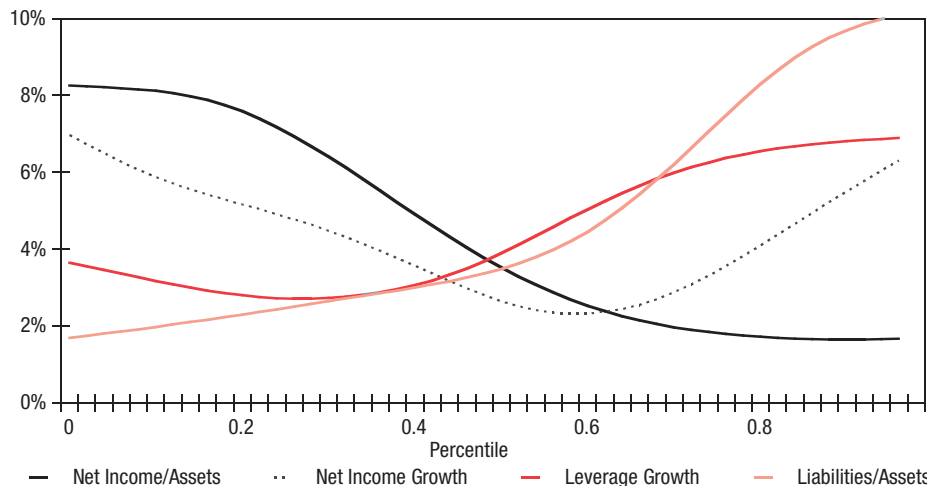
When we look at the two-year growth rate as opposed to the one-year rate, we see that the relation is weaker. Going back several years makes for a better estimate in the standard statistical sense of using more observations, but at a cost of using less relevant information. The dominance of the shorter period is a nice result because it is easier to collect information for two years rather than for three.

Growth vs. Levels

Lenders are more interested in where the firm is going than where it has been. For that purpose, trends are often analyzed. Further, many lenders extensively use projections to underlie their credit underwriting process. While we will not discuss projections here, it is important to note the relative importance of levels and trends. Two important variables are net income/assets and liabilities/assets, and the relative power of these ratio levels and their trends is representative of the relation between levels and trends in other variables.

Exhibit 4.9

Growth vs. Levels, 5-Year Cumulative Probability of Default, Public Firms, 1980-1999



Levels dominate trends as sole sources of information.

One reason why trends are so useful to lenders is that ratios, or any risk metrics, do not dictate pricing or lending decisions. They are often used as inputs to such decisions, but their relation to default rates and their large susceptibility to exceptions make subjective integration of this and other information necessary. However, any deterioration in financial ratios for an existing credit is cause for concern and, therefore, many credit analysts, especially in workout or credit administration groups, often focus on the trends rather than the levels. That is, the ratio level is by itself ambiguous and incomplete information, but a deterioration in an approved credit's financial strength, as reflected in lower earnings, liquidity, or equity, is unambiguously bad.

Despite the undeniable usefulness of trends in monitoring existing accounts, from a pure predictability standpoint, levels dominate trends. In Exhibit 4.9 above illustrates that trends in profitability or leverage are not as powerful as the levels of these ratios. The following probit regressions underscore this result. If we take the truncated ratios in isolation, we can see that the levels are always significantly more powerful predictors of future default.

Exhibit 4.10

Public Firms, 1980-98, Probit Model Estimating Future 5-year Cumulative Default
Trends and Levels

(three separate models were estimated, using only the level and trend of that level within each estimation)

		Coefficient	Z-statistic
Current Ratio	Level	0.111	19.0
	Trend	0.041	5.4
Liab/Assets	Level	0.214	25.4
	Trend	0.003	1.8
NI/Assets	Level	-0.168	-18.6
	Trend	0.000	-0.68

In Exhibit 4.10, we see that when the trend and its corresponding level (e.g., NI/A growth and NI/A), the level dominates the trend in statistical significance.³³ These results suggest that the best single way to tell where a company will end up is where they are now, not the direction they are headed. Prediction, however, is distinct from a contemporaneous correlation, as virtually every firm that fails shows declining trends. For example, almost all failing firms have negative income prior to default, yet negative income is common enough to not even imply that a default is probable, let alone certain. Failing firms also strongly tend to have declining trends in profitability. This is the crux of the variable selection problem, that many-too many-different metrics can help predict failure. The more important issue is which metric dominates, and this is the benefit of multivariate analysis.

In Exhibit 4.9, net income growth displays the same U-shape we saw with sales growth, undoubtedly for similar reasons. At the low end, net income growth is an indicator of weakness; at the high end, it is an indicator of a 'high flyer' in danger of an unanticipated slowdown. It is significant, yet it is less significant than the level of net income/asset.

While ratio trends are secondary in usefulness for modeling, looking at the trend in RiskCalc can serve the needs of a lender for the reasons mentioned above. RiskCalc default projections over time reflect weakening financial strength of the borrower, and this will be reflected in concrete terms by at least some of the key input ratios (i.e., RiskCalc will not deteriorate independent of the ratios). Thus, we are not arguing that trends do not matter, just that trends are less powerful in prediction than levels, and this is why 8 of the 10 inputs refer to current levels, not trends. The trend in RiskCalc itself can serve valuable credit monitoring objectives.³⁴

³³For NI/A growth rates we used $NI/A - NI(-1)/A(-1)$, in order to avoid problems from going from negative to positive values in net income.

³⁴It should be noted that while the trend of RiskCalc can serve many useful purposes, a trend in RiskCalc is not expected to add value to RiskCalc's prediction. That is, a default rate forecast of 4% is just as consistent whether or not the prior RiskCalc estimate was 0.1% or 10%. To the extent trends add information, we tried to capture this information within RiskCalc.

Means vs. Levels

Standard and Poor's periodically publishes averages of many useful accounting ratios for three-year periods, and most underwriters examine the prior three year's statements, though it is unclear how these various year's data are weighted. Thus, it is useful to examine to what degree the mean of the past three years compares to the latest year in predicting default. Using NI/A and L/A we see that for both ratios the most recent level is a more powerful predictor of default than the average of the prior three years. The following probit regression results on these variables demonstrates the same point within a multivariate context.

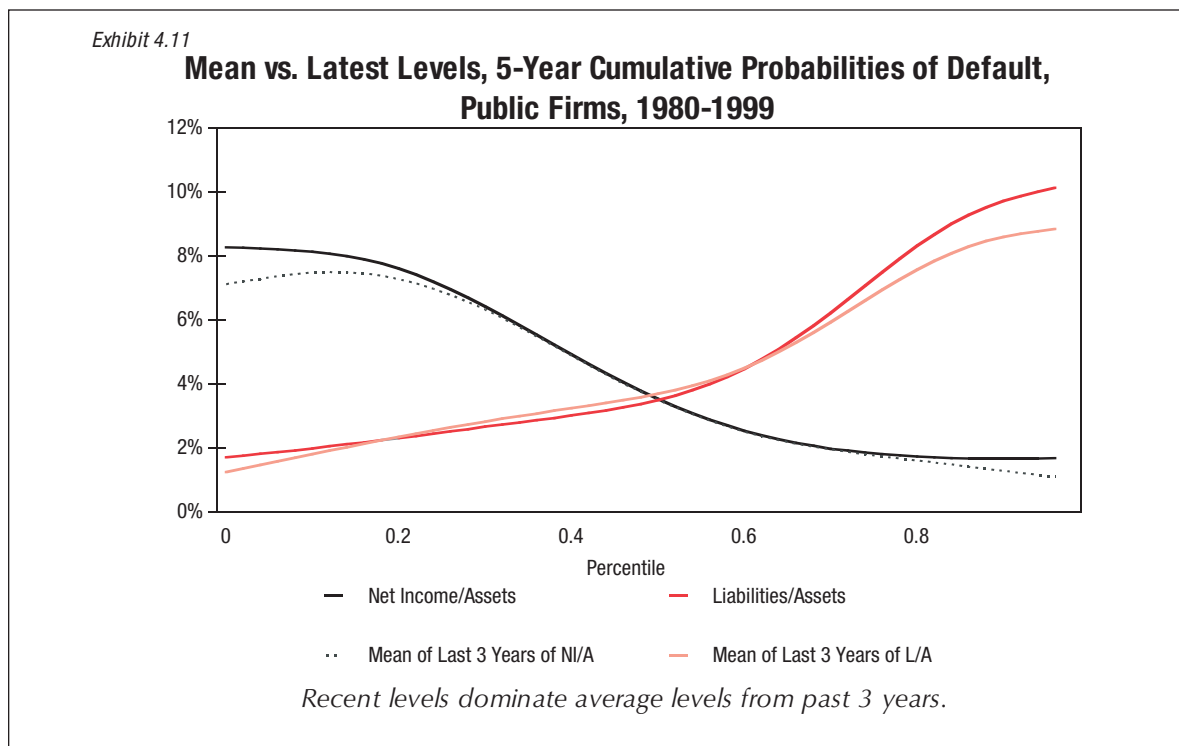


Exhibit 4.12

Public Firms, 1980-98, Probit Model Estimating Future 5-year Cumulative Default

(multivariate probit model with 4 variables)

	Coefficient	z-Statistic
NI/A	-0.36	-15.6
NI/A Mean	0.17	5.2
L/A	0.195	11.3
L/A mean	0.032	1.5

Financial information ages about as well as a hamster, as it appears to weaken after only one year. Unreported tests show that two-year means as opposed to three-year means are, predictably, somewhere in the middle. Again, this has nice implications for gathering data and testing models, since it is always costly to require more annual statements. It also highlights the importance of getting the most recent fiscal year statements, and perhaps not getting too distracted by what happened three years ago.

Audit Quality

A final variable to examine is the audit quality. A change in auditors or, less frequently, and 'adverse opinion' are often discussed as potentially bad signals (Stumpp, 1999). Research has shown that audit information is useful, if less than what we would like (Lennox, 1998). As this information is not ordinal, it is presented in tabular form. The first information we examine is the effect of a change in auditor. Exhibit 4.13 shows that a change in auditor is associated with a higher 5-year cumulative default probability. Those firms that changed auditors had a 50% higher default rate than those firms who did not switch auditors the prior year.

Exhibit 4.13

Change in Auditor, Public Firms, 1980-99

Audit Change	Default Rate	Count
NA	4.72	16,814
No	4.14	101,455
Yes	6.19	8,986

A change in auditor signals a rise in default risk.

Audit quality is the most basic audit information. Indeed, it is by itself, a significant predictor of default. We see in Exhibit 4.14 that firms receiving an unqualified opinion with no additional language have significantly lower future default rates.

Exhibit 4.14

Public Firms, 5-year Cumulative Default Rate, 1980-99, Audit Quality

	Default Rate	Count
Unqualified	3.71	96,232
Qualified	10.24	6,229
Unqualified w/ Additional Language	5.47	21,988

Unqualified audits imply less risk than qualified audits.

For many private firms, no audit is conducted, and this presents a very different view on how to use the audit information (i.e., as opposed to types of audit opinions). For private firms, many audits are company prepared data from tax returns, reviewed internally though unaudited, and directly submitted by the firm. Again, we see that an audited set of financials is symptomatic of greater financial strength, with an approximately 30% difference in the future default rate.

Exhibit 4.15

Private Firms, 5-year Cumulative Default Rate, 1980-99, Audit Quality

	Default Rate	Count
Audited	2.53%	51,032
co-prep	3.32%	16,544
tax return	3.38%	38,633
Reviewed	3.78%	27,224
Direct	3.86%	9,700

Private firms with audited financial statement have lower default rates

While the immediately preceding analysis indicates that audit information is useful, it turns out that audit is highly correlated with size among the smaller firms, and most of its predictive power is subsumed by size in the multivariate regression. Hence, it has not been included as an explanatory variable in RiskCalc. As we gather more data, however, we will be examining this variable closely.

Risk Factors We Do Not Use

Several factors that affect credit are not addressed in RiskCalc. These factors include industry specific information, macroeconomic data, and management quality. The reason these are not included is because they are too difficult to measure consistently, there is not sufficient data to infer a statistical relationship, or they are best left as independent calculations.

Macroeconomic forecasting is fundamentally a time-series issue, and macroeconomists have been studying economic cycles since the 1870s without much progress (Sims, 1982). As macroeconomic forecasting is generally independent of cross-sectional and over-the-cycle default prediction, its difficulty, and the fact that reasonable people have strong and different opinions on this matter, suggests that it is best left 'outside' the model.

Industry variation is something we are hoping to incorporate in the next upgrade to RiskCalc. Currently we simply do not have sufficient data to make industry refinements that are statistically sound. Extrapolation from the public or rated universe is often considered as a potential way around this problem, such as using sector equity indices or industry default rates from Moody's rated universe. While intuitively appealing such adjustments are not so straightforward. For example, retail trade has the highest default rates in Moody's rated universe, but in the Dun & Bradstreet universe of smaller companies, one of the lowest. Which is most representative of middle market firms?

Finally, subjective factors are often considered decisive. A good credit analyst often prides herself at being able to 'look in the eye' of a CFO and discern how the company is really doing. Some problems with this approach were discussed in Section 2 on judgmental models, but the bottom line is that there is no way of compiling such information in a database for statistical validation.

Our basic goal is to create a benchmark for credit risk, and to that end a number that is statistically validated is needed. Subjective information is therefore neglected, not because it is vague or ambiguous to any one person, but because its interpretation varies from person to person.

Conclusion

Many ratios are correlated with credit quality. In fact, too many ratios are correlated with credit quality. Given these variables' correlations with each other, one has to choose a select subset in order to generate a stable statistical model. The final variables and ratios used in RiskCalc are the following:

Exhibit 4.16

RiskCalc Inputs and Ratios	
Inputs (17)	Ratios (10)
Assets (2 yrs)	Assets/CPI
Cost of Goods Sold	Inventories / COGS
Current Assets	Liabilities / Assets
Current Liabilities	Net Income Growth
Inventory	Net Income / Assets
Liabilities	Quick Ratio
Net Income (2 yrs)	Retained Earnings / Assets
Retained Earnings	Sales Growth
Sales (2 yrs)	Cash / Assets
Cash & Equivalents	Debt Service Coverage Ratio
EBIT	
Interest Expense	
Extraordinary Items (2 yrs)	

These ratios were suggested by their univariate power and tested within a multivariate framework on private firm data. As will be discussed in Section 6, the transformation of these variables is based directly on their univariate relations. That is, where they begin and stop adding to the final RiskCalc default rate prediction is not based on their raw level, but their level's correspondence to a univariate default prediction. Once you understand the univariate graphs, especially as laid out in the next section, you will understand how these variables drive the ultimate default prediction within RiskCalc.

Appendix 4A

Calibration And Power

A model may be powerful, but not calibrated, and vice versa. For example, a model that predicts that all firms have the same default rate as the population average is calibrated in that its expected default rate for the portfolio equals its predicted default rate. It is probable, however, that a more powerful model exists in which some variable(s), such as liabilities/assets, help discriminate to some degree bads from goods.

A powerful model, on the other hand, can distinguish goods and bads, but if it predicts an aggregate default rate of 10% as opposed to a true population default rate of 1%, it would be uncalibrated.

This point is best illustrated by an example. Consider the following mechanism that generates default (1 being default, 0 nondefault):

$$y = \begin{cases} 1 & \text{if } y^* < 0 \\ 0 & \text{else} \end{cases}$$

$$y^* = x_1 + x_2 + \epsilon$$

$$x_1, x_2, \epsilon \sim N(0, 1)$$

This is the 'true model', potentially unknown to researchers.

Explanatory data x can be used to predict default in the following way:

$$y^* < 0 \Leftrightarrow x_1 + x_2 + \varepsilon < 0$$

i.e.,

$$y^* < 0 \Leftrightarrow \varepsilon < -(x_1 + x_2)$$

The two optimal models for predicting y conditional upon x_1 or x_1+x_2 are:

$$E(y|x_1) = \text{Prob}(\varepsilon + x_1 < -x_2 | \varepsilon + x_1 \sim N(0, \sqrt{2})) = \Phi\left(\frac{-x_2}{\sqrt{2}}\right)$$

$$E(y|x_1, x_2) = \text{Prob}(\varepsilon < -x_2 - x_1 | \varepsilon \sim N(0, 1)) = \Phi(-x_2 - x_1)$$

These are simply the cumulative normal distribution functions. Note that since y (as opposed to y^*) is a binary variable, 1 or 0, $E(y)$ has the same meaning as default frequency. $E(y)=0.15$ means one would expect 15% of these observations to produce a 1 - a default.

Now assume there are two models that attempt to predict default, A and B:

$$\text{Model A: } \text{Prob}(\text{def}|A) = \Phi\left(\frac{-x_1}{\sqrt{2}}\right) \quad 4A.1$$

$$\text{Model B: } \text{Prob}(\text{def}|B) = \Phi(-x_1 - x_2 + 1) \quad 4A.2$$

Model A is imperfect because it only uses information from x_1 and neglects information from x_2 . Model B is imperfect because while it uses x_1 and x_2 , it has misspecified the optimal predictor by adding a constant of '1' to the argument of the cumulative default frequency.

The implications are the following. Consider two measures of 'accuracy'. For cut-off criteria where both models predict a default rate of at most 50% on approved loans, the actual default rates at the cut-off will be 50% for Model A as intended, but 84% for the miscalibrated Model B.

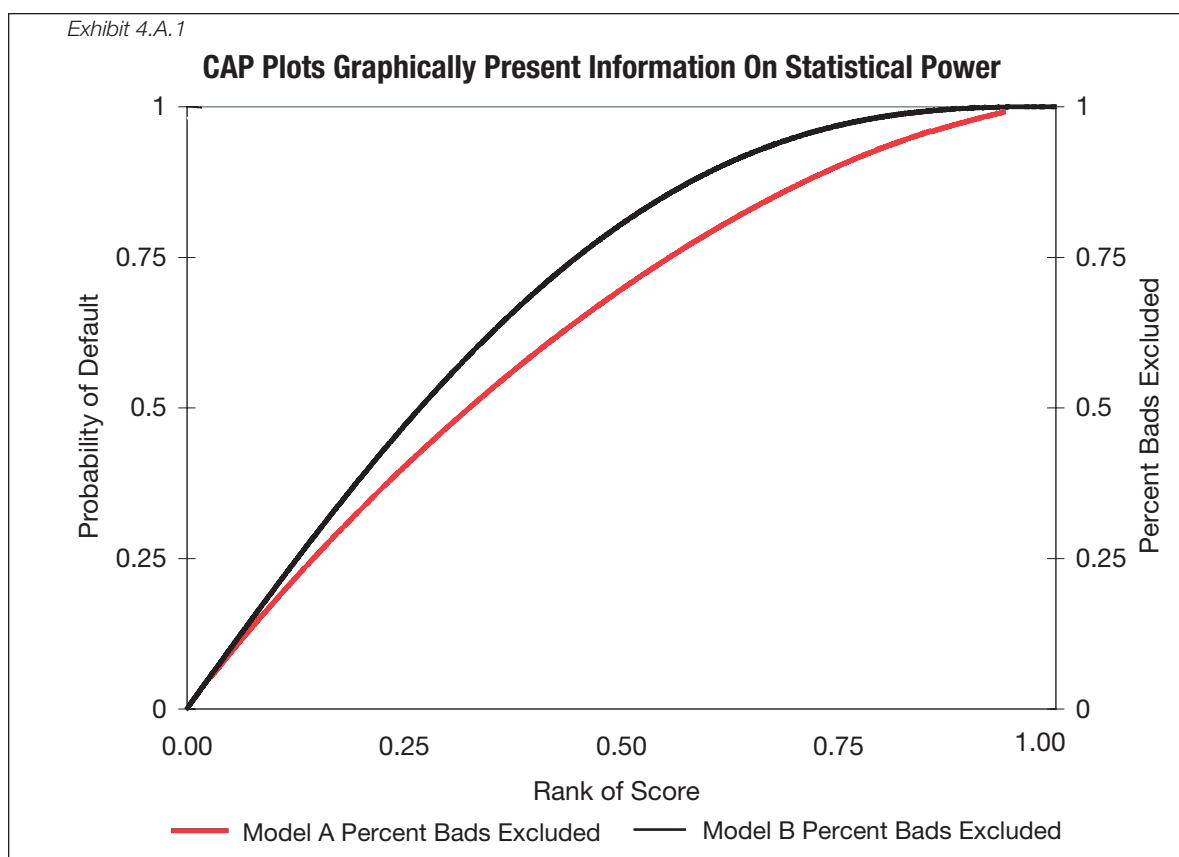
We know this because we know the structure of both models and the true underlying process. For Model A this would be where $x_1=0$, for Model B this would be when $x_1 + x_2=-1$. Thus, Model A is superior by this measure in that it is consistent, while Model B is not. Of course the misspecification in the model could be positive or negative (instead of adding 1 we could have added -1). Consistency between ex ante and ex post prediction is an important measure of score performance, and not all models are calibrated so that predicted and actual default rates are comparable. In this case, model A is consistent, while B is not.

Next, consider the power of these models. A more powerful model excludes a greater percentage of bads at the same level of sample exclusion. Let us examine the case with the following rule: exclude the worst 50% of the sample as determined by the model. What proportion of bads actually get in, i.e., what is the 'type 2' error rate?³⁵ In this example, Model A excludes 69.6% of the bads in the lower 50% of its sample, while the more powerful Model B excludes 80.4% of the bads in its respective lower 50%.

Model B is clearly superior from the perspective of excluding defaults. To see the differences in power graphically we can use a common technique: Cumulative Accuracy Profiles or CAP plots. These are also called 'power curves', 'lift curves', 'receiver operating characteristics' or 'Lorentz curves.'

The vertical axis is the cumulative probability of 'bads' and the horizontal axis is the cumulative proportion of the entire sample. The higher the line for any particular horizontal value, the more powerful the model (i.e., the more northwesterly the line, the better). In this case, Model B dominates Model A over the entire range of evaluation. That is, Model B, at every cut-off, excludes a greater proportions of 'bads'. (Note that at the midpoint of the horizontal axis (50%), the vertical axis for Model B is 80%, just as we determined above.) Thus, a power curve would suggest that Model B is superior in a power sense-the ability to discriminate goods and bads in an ordinal ranking.

³⁵In practice, the use of type 1 and type 2 errors as definitions is subjective. We will use the definition where type 2 error involves "accepting a null hypothesis when the alternative is true." In this case, believing that a firm is 'good' when it is really 'bad'.



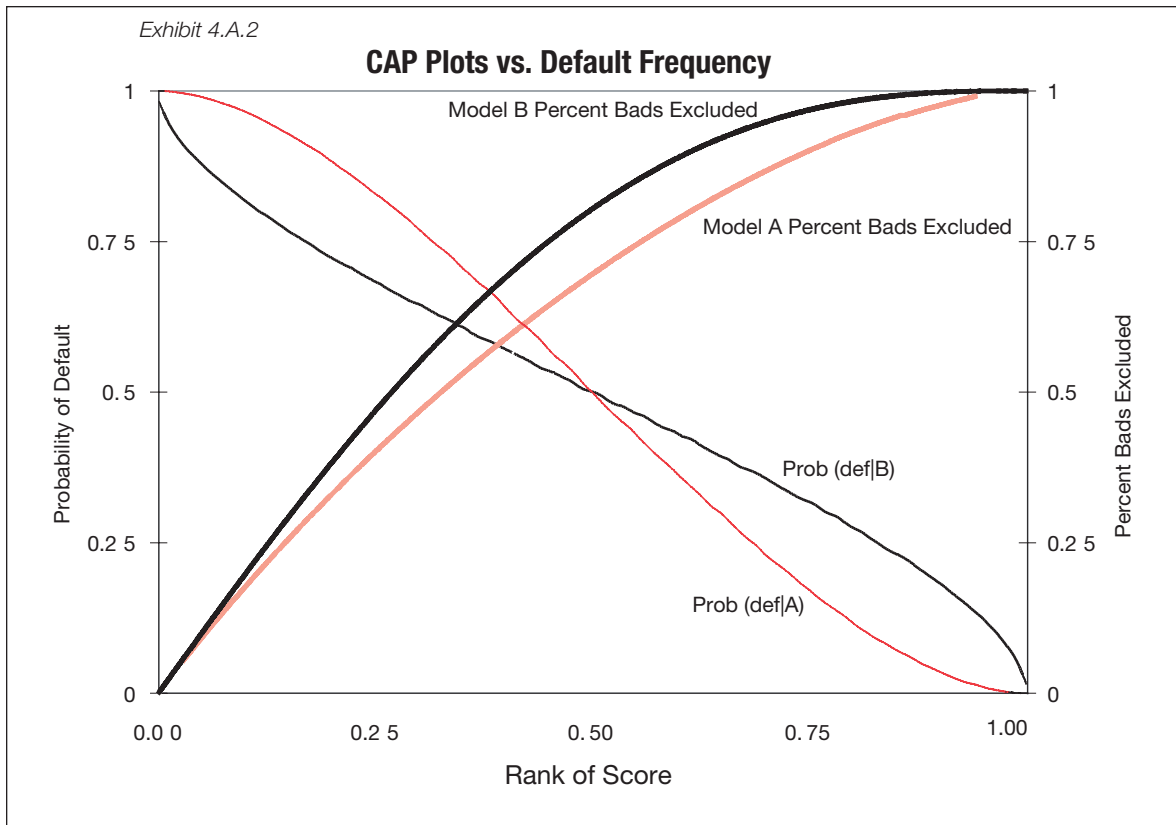
If the outputs of Models A and B were letters (Aaa through Caa), colors (red, yellow, and green), or numbers (integers from 1 to 10), this would be the only dimension on which to evaluate the models. Yet, as these models do have cardinal predictions of default, we can assess their consistency independent of their power, and indeed have a dilemma: A is more consistent, but B is more powerful.

Which to choose? Clearly it depends on the use of the model. If one was trying to determine shades of gray, such as which credits receive more careful examination, then Model B is better: one uses it only to determine relative risk. If one was determining pricing or portfolio loss rates, calibration to default rates would be the key, pointing to Model A.

In practice, however, this is a false dilemma. It is much more straightforward to recalibrate a more powerful model than to add power to a calibrated model. For this reason, and because most models are simply ordinal rankings anyway, tests of power are more important in evaluating models than tests of calibration. This does not imply calibration is not important, only that it is easier to remedy.

Power and Default Frequency

Using the example above, we can see the relation between power curves and default frequency more generally, which is useful for demonstrating two points: 1) how power enhances the ability of a model to produce extreme predictions without compromising consistency and 2) how power curves are related to simpler graphs of default frequencies. A more powerful model will be able to generate more extreme predictions, that is, predictions that deviate significantly from the mean while remaining consistent. In this example, the mean default rate is 50%, so a default prediction of 50% for every credit is consistent - if trivial. A better model would correctly segregate the pools into those firms with probabilities of default of 25% and 75% (note still consistent with a 50% default rate), and better still one that predicts 10% and 90%. This is reflected in Exhibit 4.A.2, which uses the optimal models for using information x_1 and x_1+x_2 (equations 4A.1 and 4A.2 above) to generate CAP plots and default frequency graphs.



The more powerful Model B has the more northwesterly power curve. The white lines are default frequency lines that reflect the actual default rate for various percentiles of scores generated by the models. That is, the percentile of the score is graphed along the horizontal axis and the default frequency for that percentile is on the vertical axis. The actual default frequency for each percentile of Model B's scores is either higher or lower than for Model A. Model B generates more useful information, since its more extreme forecasts are closer to the 'true' values of 0 and 1, while simultaneously being just as consistent in predicting the average value as Model A.

There is in fact a mathematical relation between the power curves and frequency-of-default graphs, such that one can go from a probability graph to a power curve. Given the probability curve one can generate a power curve, while given a power curve and a sample mean default rate (so one can set the mean correctly) one can generate a probability curve. That is, if q is a percentile (say from the 1st percentile to the 100th percentile), and $prob(i)$ is the default frequency associated with that percentile, the curves would have the following relation:

$$power(q) = \frac{\sum_{i=1}^q prob(i)}{\sum_{i=1}^{100} prob(i)}$$

$$prob(q) = 100 \text{ mean}(prob) (power(q) - power(q-1))$$

A dominant power curve is one that is always above the other power curve. The default frequency curve will cross the non-dominant power curve at a single point, and be above it at the 'bad' end of the score (where it predicts the greatest chance of default) and below at the good end. Thus, for models that strictly dominate others, one can observe such domination either through CAP plots (look for the more northwestern line) or the default frequency curves (look for the one that generates a steeper slope), as Model B in Exhibit 4.A.2 shows.

It is often the case that power curves will cross or that their probability curves will cross a couple of times, especially at the very beginning and very end of their ranges. In these cases, which model is superior depends on the relative costs of defaults vs. losing potential revenue. More importantly, all curves are estimates, and so curves that are close are often indistinguishable. This standard error of these curves is mainly a function of the sample size, specifically the minimum of the number of goods or bads in the sample.³⁶

A probability curve is thus fundamentally related to a model's power, and can be aggregated as S-Stat or Information Entropy Ratios, both of which add up the absolute deviation of the model's prediction from its sample mean.³⁷

Testing the Need for Recalibration

Another useful fact to know is that power constrains the range of consistent predictions one can make. More powerful models produced broader ranges of forecasts (e.g., Aaa to Caa as opposed to Ba1 to B1) that are consistent. In our example, the more powerful Model B generates a more extreme output of goods and bads. The worst 18% from Model B have a default rate above 90%, while only 3% for Model A surpasses this level.

Now assume you heard about Model B's performance and were convinced that it was accurate, but still had access only to the less powerful Model A. Could you rig Model A so that its worst 18% would have default rates above 90%? No. This highlights an important constraint on granularity from the power of the model. Even if you know a population contains some observations, which you cannot ex ante identify, with really high default probabilities, a model that is not sufficiently powerful may not be able to generate these 'true rates' and remain consistent. If you took the bottom 18% of Model A and assigned it a default probability of 90% (since you knew this was possible at least to some model, in this case, Model B), that 18% would have the same actual default rate experience as before. In this case, the bottom 18% of Model A generates a mean default rate of only 74% (see Exhibit 4.A.1).

Knowledge that firms with high default rates exist is irrelevant to prediction if you can not statistically segregate these firms, and to do this you need more than anecdotes. A model's power limits the range of its predictions, with more powerful models generating more diverse targeted variables, and less powerful models generating output closer to the population mean.

This is relevant to predicted ratings. A Caa rating corresponds, roughly, to a 15% annual default rate. The lower power of the statistical model as applied to middle market companies implies that, in practice, it is very difficult to generate a significant proportion of commercial obligors with 15% default rates.

Some vendors have been known to generate such very high default rates, and we would suggest the following test in order to assess these predictions. First, take a set of historical data and group it into 50 equally populated buckets (using percentile breakpoints of 2, 4%, ...100%). Then, compare the mean default rate prediction on the x-axis with the actual, subsequent bad rate on the y-axis. More often than not, models will have a relation that is somewhat less than 45% (i.e., slope<1), especially at these very high risk groupings. This implies that the model purports more power than it actually has, and - more importantly - it is miscalibrated and should be adjusted. This sort of test is complicated by the fact that commercial loans default in cyclical fashion, where there might be 8 years with below average defaults. Nonetheless, it is a very useful tool for monitoring and updating quantitative models, especially if one tracks a model over the business cycle.

Section V: Similarities And Differences Between Public And Private Companies

Facts do not cease to exist because they are ignored.

~ Aldous Huxley

Does a leverage ratio of 60% mean the same thing for Bob's Computer Warehouse as it does for IBM? This question is relevant to any lender that evaluates credit exposures that span traditional market segments such as small business, middle market, and large corporate lending. It is also relevant to anyone interested in purchasing credit scores, as these are more often than not estimated on public firms, if not the more exclusive set of agency-rated public firms.

³⁶For example, with 1 million goods yet only 10 bads, only 10 not-perfectly-correlated explanatory variables are needed to completely 'explain' the results.

³⁷See Sobehart, 1999, or Dirk-Emma Baestaens, *Credit Risk Modeling Strategies: The Road to Serfdom? International Journal of Intelligent Systems in Accounting, Finance and Management*, 1999, volume 8, p. 225-235.

In our opinion, there are subtle yet significant differences between public and private firms that are best addressed through direct estimation and testing on private firm data. A model fit using public data will deviate systematically and adversely from a model fit using private data, as applied to private firms.

Since data drive default prediction models, the availability of financial data from Compustat and of ratings and default information from Moody's and S&P, has naturally led to these sources of information being the basis for many academic and private vendor models. One approach to creating a quantitative model is to use the rating (e.g., Baa, A) as the dependent variable and to look at the relative mean ratios for each rating category. These methods include the ordered probit model, and also translating the ratings into numbers (e.g., Aaa=1, Aa1=2 ... C=21), and then regressing the ratios onto these numbers. Another method is to estimate default directly for public firms, using a binary estimation method where the dependent variable is 1 if the firm defaults, 0 otherwise.

Yet, the outstanding question is whether a model estimated on these public firms generates valid inferences about default probability for private firms. This section investigates this issue by examining the ratios of public and private firms and their relations to default. While there are many similarities, there are also some key differences.

The differences highlight the importance of fitting a model targeted for private firms using private firm data. In fact, for some ratios, the rankings by Moody's rating is the opposite of their relation to risk. For example, a higher rating allows the firm the luxury of having a lower liquidity ratio, but it would be incorrect to assume that because higher rated firms have lower risk and also lower liquidity ratios, lower liquidity implies lower risk.

Distribution of Ratios

The juxtaposition of a few key ratio distributions nicely illustrates how public and private firm default risk varies. In Exhibit 5.1, first note that leverage, as measured by liabilities/assets, is generally higher for public firms than for private firms. This point can be displayed using medians, however, and therefore is not the focus of histograms, which are useful for demonstrating differences in the spread of a distribution around its median. What the distribution communicates is the significant number of observations where the L/A ratio is greater than 1. This is true in 8% of public firms and 10% of private firms. The upward tail of this distribution is definitely asymmetric and fat-tailed; most importantly, it is fat-tailed over a region on which underwriting criteria are usually silent-negative net worth.

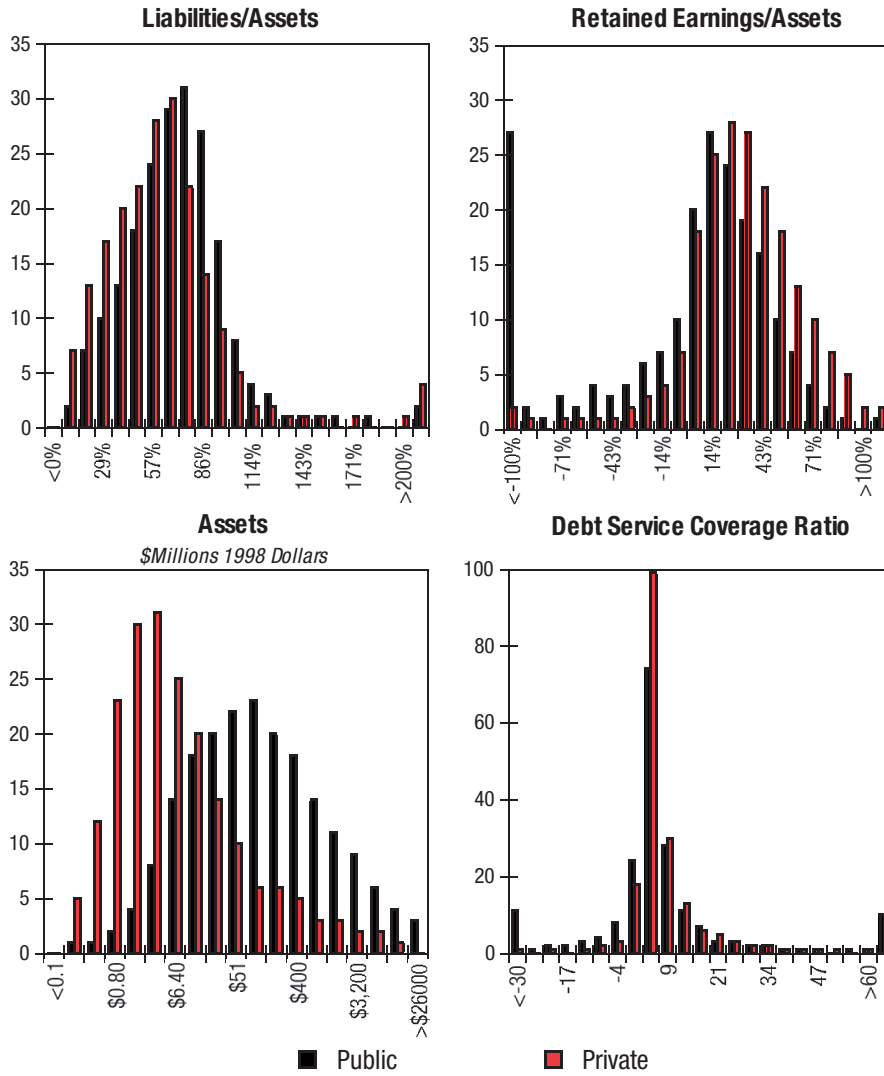
A model needs to be able to handle these unusual ratios in a way that is more robust than standard rules of thumb, which would simply map these firms into Caa1. While lending to these firms may not be prudent, their probabilities of default are not necessarily consistent with extreme ratings such as a Caa. This is borne out in the data. Negative net worth firms default at a rate only 3 times the average firm default rate, which is much lower than the ratio of Caa1 to the average rated firm default rate (around 5).

The ratios net income/assets, retained earnings/assets, and interest coverage have similar properties in that a large portion of observations is outside the range of conventional analysis. It is common to have interest coverage rules using numbers like 0, 1, and 5, yet for many firms these numbers are negative or well above 10. The mere fact that so many firms exist with extreme ratios highlights that extreme values are not fatal. Further, linearly extrapolating a rule based on the 70% of cases that are 'normal' will lead to wildly distorted inferences for the many firms that are not 'normal.'

The ratio that shows a difference between public and private firms that is mainly in its distribution—as opposed to its average—is retained earnings. There are many public firms with highly negative retained earnings: 20% with less than 100% retained earnings/asset ratio. This is because many public firms take special charges but can stay afloat due to their ability to refinance with the capital markets. Retained earnings are a powerful predictor of default for both private and public firms. Yet an adjustment must be made for this vast disparity in the meaning of a significantly negative retained earnings measure.

Exhibit 5.1

Histograms, Public vs. Private Firms (the bars represent percent of the sample in particular groupings)



Most distributions have very fat tails - a large set of outliers to which traditional rules-of-thumb do not comfortably apply.

PUBLIC/PRIVATE DEFAULT RATES BY RATIO

Many of the inputs used for Moody's for Private Companies show similar relations to default in the public and private datasets. Exhibit 5.2 shows default frequency graphs for eight ratios for public and private firms. As in section 3, the line represents the frequency of subsequent 5-year defaults for firms by ratio, which was constructed by bucketing firms into 50 ordinaly ranked groupings, observing their default rate over the 5 years, and then smoothing the relationship. The public firm data are from 1980 - 1999, while the private firm data are from 19989-1999. For these charts, we used the same range, which was constructed using the percentile points of the relevant ratio from Compustat. Therefore, the x-axis spans the same range of values for both public and private firms, unlike the previous probability graphs, which normalized the x-axis by percentiles for each ratio. This allows a direct comparison between the two universes.

The public data in general show more curvature, which reflects the greater power of these ratios in public companies compared to private companies. As we shall see, identical models of any form all show a consistent weakening of power when applied to private vs. public firms (discussed further in section 8). In general, these differences are minor, however, and for these eight ratios the distinction between public and private is not significant.

The broad similarity between the two independent datasets suggests that what we observe are real relationships that transcend institutional boundaries, not simply anomalies of small samples. This evidence supports the belief that ratios can inform prediction, and these predictions can be calibrated and tested.

There are important exceptions, however. The first to note is size. Though public firms range in size from less than \$1 million to \$300 billion, and many have less than \$20 million in assets, they are still much larger than the average middle market firm. The median asset size in Moody's private firm database is \$2 million, versus \$100 million in Compustat.

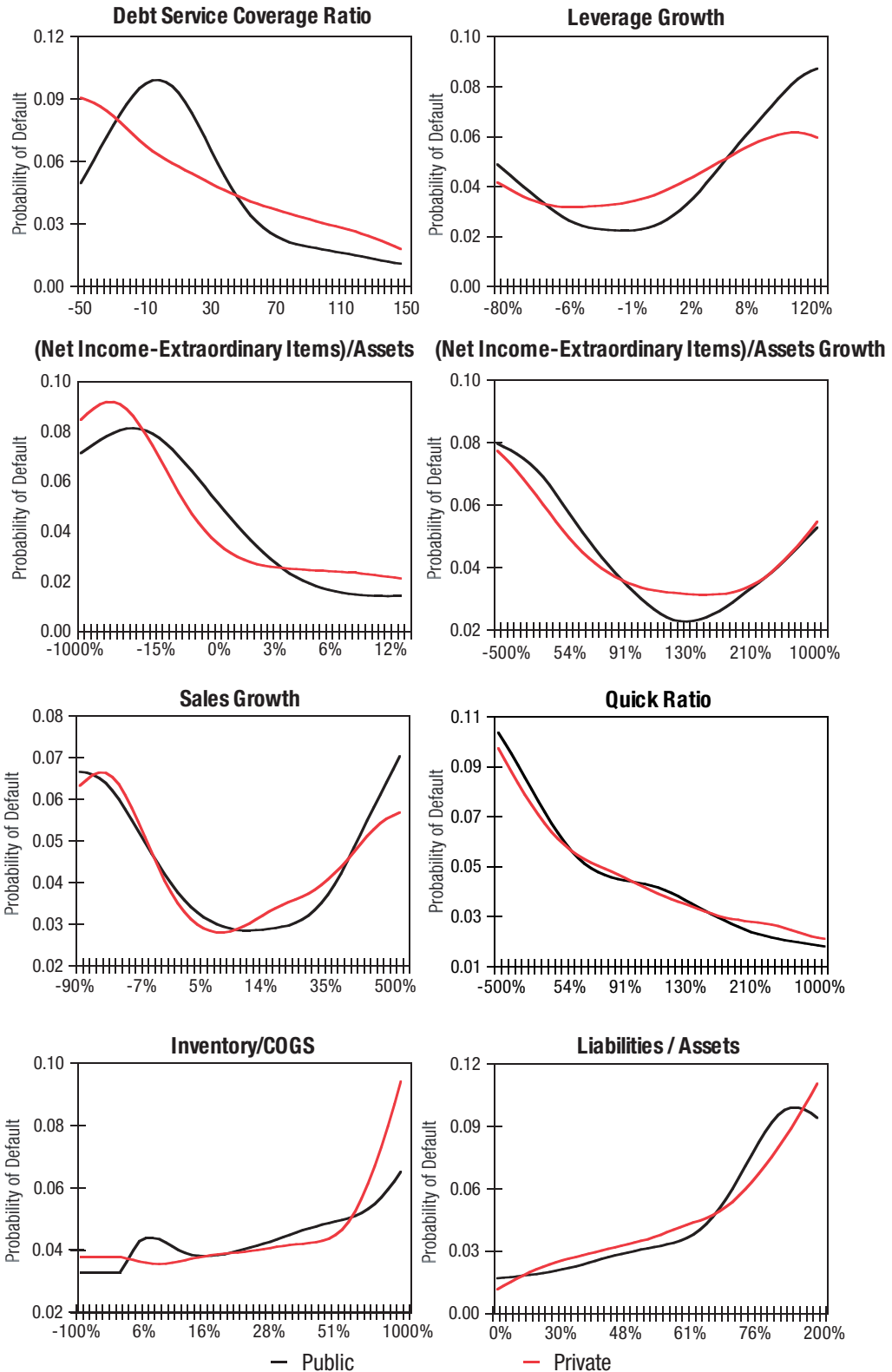
The relation between size and default is not straightforward due, in part, to a selection bias in the Compustat data. Compustat does not include recently listed small-sized firms with spotty financial reporting. A small firm that haphazardly reports its financials, or one that has existed for only a year or two, is usually not listed. If such an unlisted firm defaults or goes bankrupt, it will never be registered. If it survives and adopts more rigorous accounting standards, it will then appear in Compustat with its prior financial statements 'back filled' (e.g., if a firm were added to Compustat in 1987, its financial statements for 1985 and 1986 would be added as well even though they were not available in Compustat prior to 1987). Compustat, therefore, has what is known as a survivorship bias for small firms. Once a small firm is in Compustat, it has already survived perhaps the most crucial period of risk in its existence. Thus, the smaller firms in Compustat are in some sense all winners, and the proportion of 'winners' vs. random firms increases as the size gets smaller, offsetting the naturally higher rate of default for those firms.

For private firm databases, we encounter a similar issue, but in the opposite direction. Many defaults in these databases are added manually, and the selection process has a definite size bias. Larger defaulting firms are more apt to be remembered and recorded. Using only manually collected defaults, one therefore invariably finds that size and default are positively related: bigger size, higher default rate. This is not the data speaking, but the bias in the data.

Fortunately, most of our private firm default data come from an automated process that links financial statements with credit information, mitigating much of the bias. We do not have a sufficient numbers of observations, however, to exclude the manually collected defaults, and so we are careful to make some informed adjustments to the model, even though they are not justified solely by looking at the sample.

Exhibit 5.2

Public vs. Private Firms 5-Year Cumulative Default Frequency
 (Firms ranked by various financial ratios)



For most financial ratios, the implications for default are similar for public and private firms.

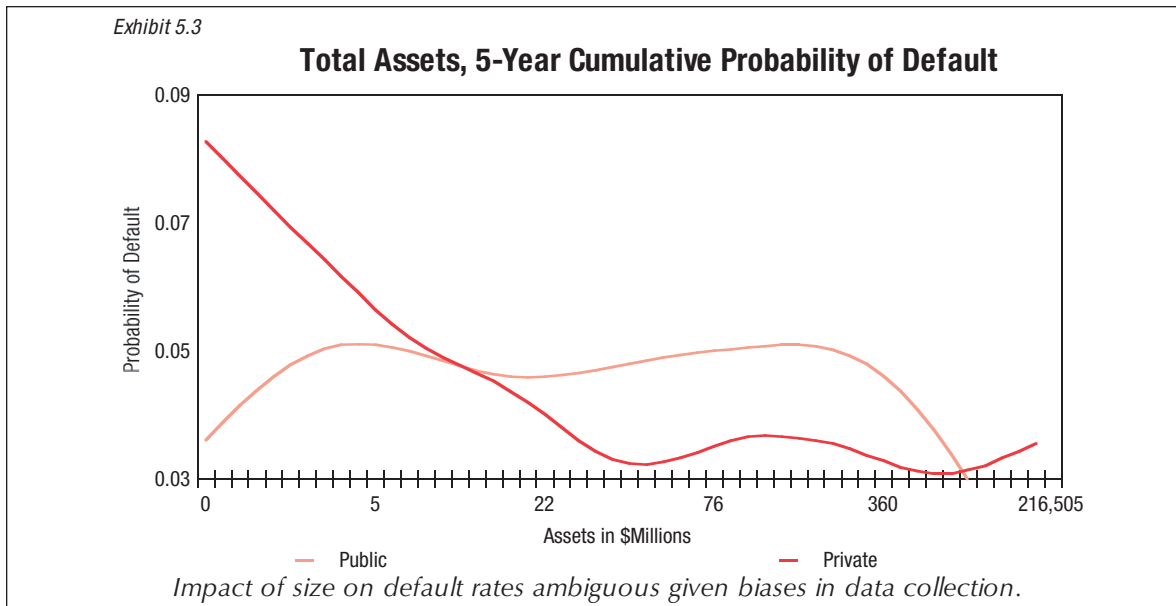
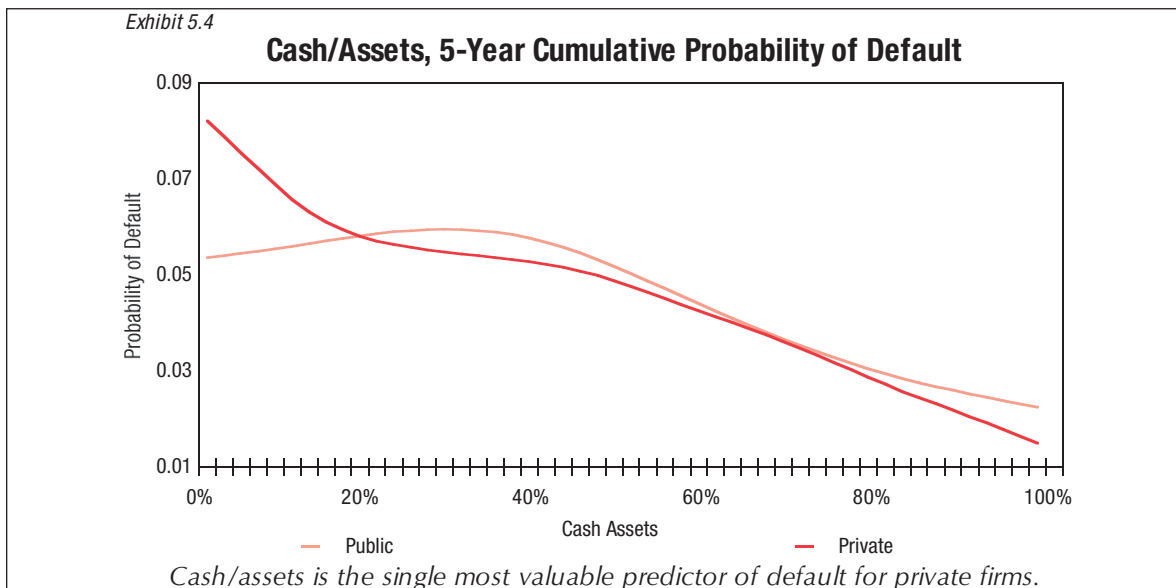


Exhibit 5.3 shows the relationship that underlies the default probabilities of some of the ratios that display marked differences in behavior between public and private firms.³⁸ For public firms, size is positively correlated with default for small firms and negatively related to default for larger firms. Yet, for the private data something different happens. Below \$40 million there is a strong negative relation between default and size, and after \$40 million the relationship flattens, due mainly to those manually added defaults which tended to be larger.

This is an example of how using judgement and more appropriate datasets helps build better models. In small business lending, it is well-known that firms under \$100K generate higher default rates than those in the \$100-\$500K range. Dun & Bradstreet's business report scores correlate size with default in a similar way. Because of the selection bias in the public and private firm databases, it is our best estimate that size is monotonically related to default, though the relationship dampens as one increases in size. It should be noted that there is no simple way to test a model that includes a monotonic size factor given this bias. A model with a commonsensical adjustment for size will actually perform worse in Compustat testing vis-à-vis models that either ignore size or parameterize it consistent with the selection bias that exists.

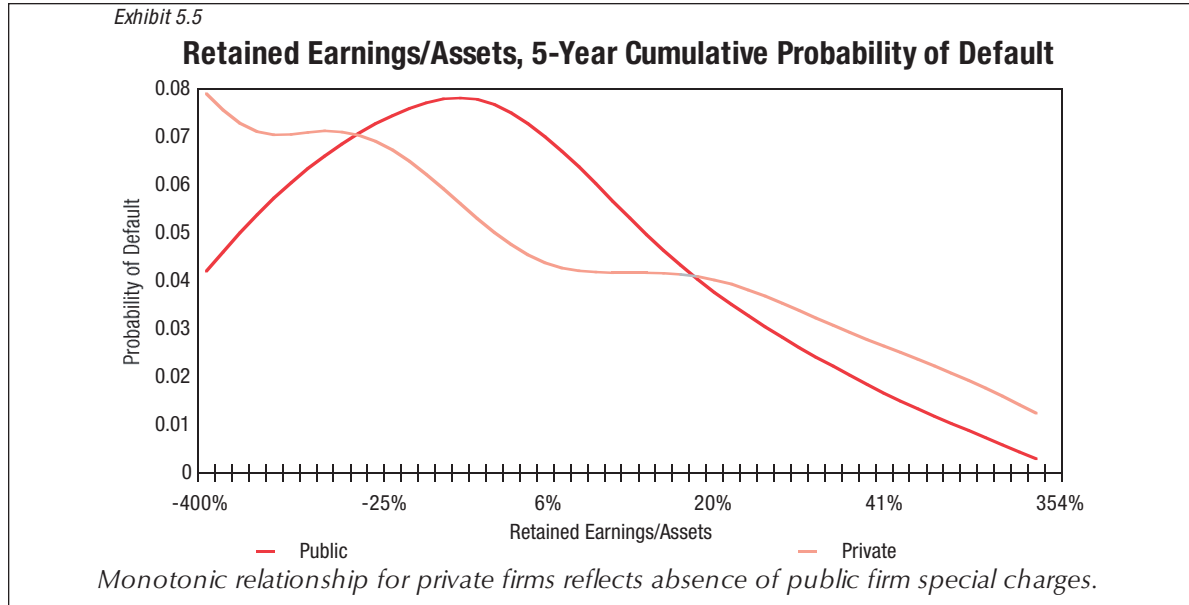
The next significantly different ratio is cash/assets, seen in Exhibit 5.4. For both sets of data, lower cash implies higher default rates, yet the effect dampens considerably at the low-end for the public dataset. For the private dataset, the effect is more dramatic at the low-end. In fact, cash is the single most valuable predictor of default



³⁸This graph shows increasing default rates for the lowest groupings in the Compustat data, as we used the percentile groupings from private firms. In a strict percentile grouping of Compustat the line would basically be flat for the lowest percentiles.

for private firms, while for the public data (in a multivariate context) cash is the least valuable predictor with the same set of data. The default frequency graphs do not illustrate this as well as we see in a multivariate context.

Moving on to retained earnings/ assets, we see that many public firms demonstrate very low ratios, usually due to large extraordinary charges that are uncommon for private firms. Lower retained earnings means higher risk for private firms, as opposed to the somewhat ambiguous relationship for public firms. Exhibit 5.5 shows that the monotonic relationship between retained earnings/assets in the private dataset makes for a more robust and meaningful measure of risk in a model fit to private firms.



Ratios and Ratings

It is important that we not merely observe patterns in the current data, but that we examine whether they show persistence over time. For this purpose, we examined whether the ratios that statistically drive risk, such those used in the RiskCalc model, show any trends in relation to Moody's rating category over time. Further, we compared the results to the unrated sample of public firms, as well as our own private firm database, to see what they imply about fitting a model to ratings and applying it to unrated or even private companies.

We begin by looking again at the relationship between size and risk using Exhibit 5.6, which tracks the relation between relative market equity value and ratings. Size and ratings are very strongly correlated across time. Unrated public firms are also shown to be of lower size than the lowest rated public companies. While the negative relationship between size and ratings is inconsistent with Compustat data (which was found to show little relation between size and default over much of its range due to a probable bias in the dataset), it is consistent with intuition as to what is going on abstracting from dataset bias issues.

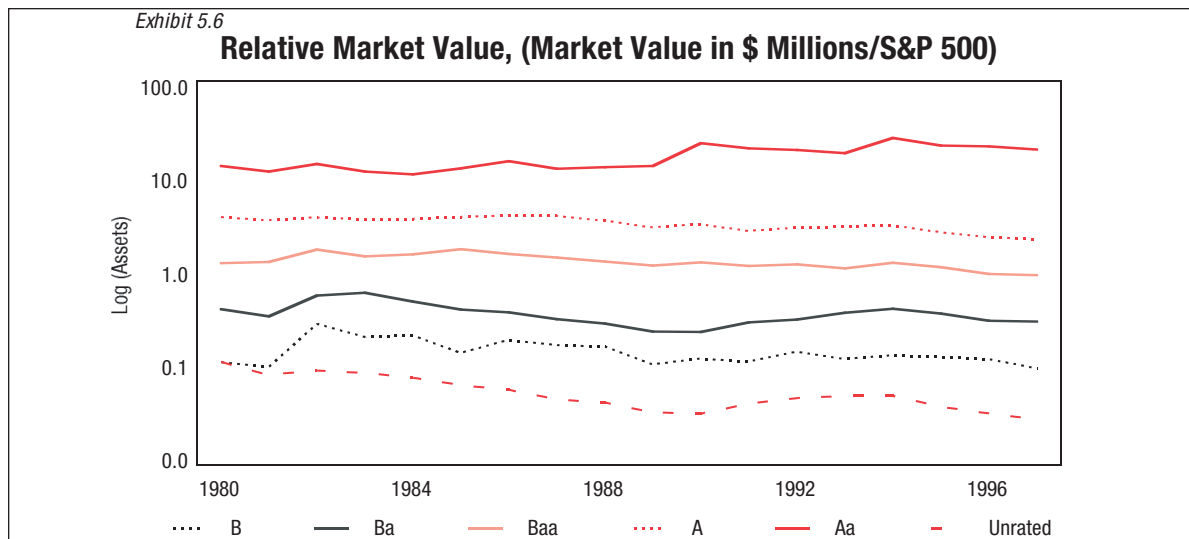
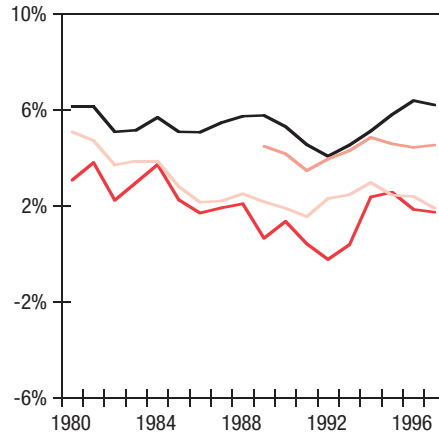
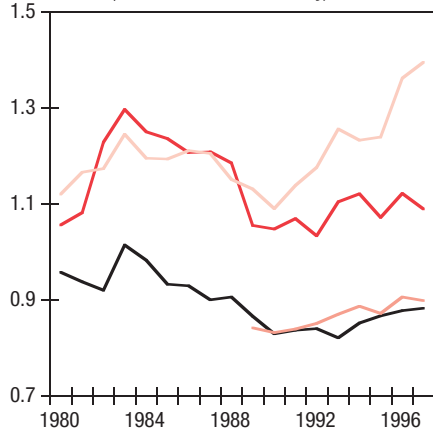


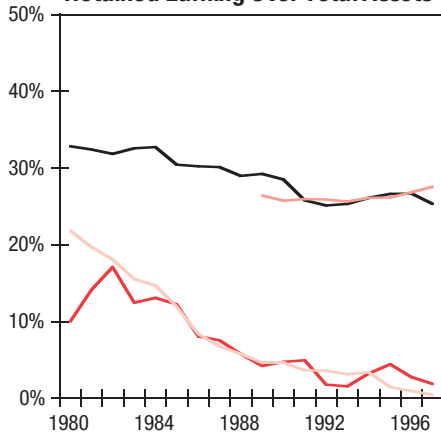
Exhibit 5.7

Median Ratios By Groupings Over Time

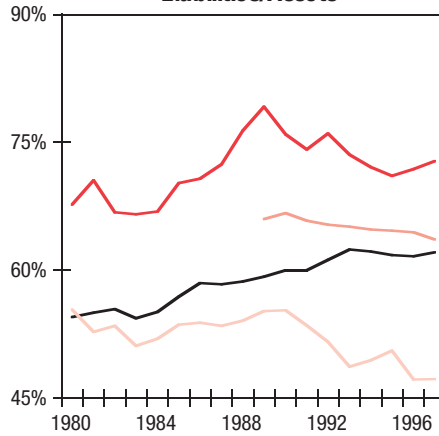
Quick Ratio = (Current Assets - Inventory)/Current Liabilities



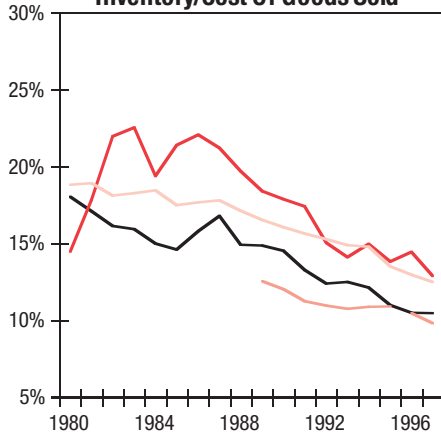
Retained Earnings Over Total Assets



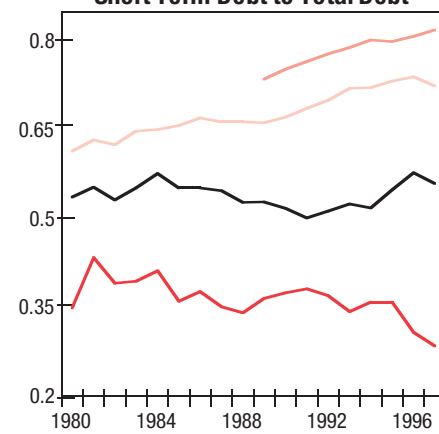
Liabilities/Assets



Inventory/Cost Of Goods Sold



Short Term Debt to Total Debt



- Investment Grade
- Unrated
- Speculative Grade
- Private

Differences between the rated and unrated universes suggest problems for models that are fit only to ratings or public firms.

Looking at the way some key financial ratios evolved over time for rated and unrated public and for private firms, we see an interesting turn of events with the quick ratio. In section 4, the quick ratio was found to be strongly negatively related to default: the higher the quick ratio, the lower the default probability. Yet, looking at the relationship between ratings and quick ratios (see Exhibit 5.7 below), we see an opposite result: lower rated and unrated firms have significantly higher quick ratios than their investment-grade counterparts.

This highlights the difficulty in using a model estimated on ratings to estimate risk. A model estimated on ratings would map higher quick ratios into lower-rated categories, even though from a default perspective they are negatively related to risk. Rated firms have different access to outside credit, which affects their optimal holdings of current assets. Higher-rated firms with access to the commercial paper market also do not need cash or can securitize their receivables. A low quick ratio is therefore an effect of a firm with a high credit rating, not the cause of a high credit rating.

We therefore believe that a model fit to ratings that uses a liquidity ratio should therefore be viewed with suspicion as applied to private firms (unreported charts for other liquidity ratios show a similar result).

Profitability as measured by net income/assets also tells an interesting story. The relation between ratings and this ratio are predictable and stable over time: higher profitability, higher ratings over the entire sample period. However, note that net income/assets fluctuates more for the riskier firms, and that the fluctuations of the speculative grade firms are higher than both the unrated public and private firms. Riskier firms are therefore not only less profitable, but this profitability is less stable over time. A similar relationship also holds with retained earnings, and indeed, retained earnings and profitability are highly correlated in the data.

Liabilities/assets show a predictable pattern between risk ratings, as higher rated companies have lower leverage ratios. Interestingly, private firms have even lower leverage ratios than investment grade companies. Again, this highlights the difficulties of inferring from a ratings model to a model applied to private firms. Clearly, very few private firms deserve investment grade ratings, yet on this dimension it appears most do. Unrated public firms are in the Baa to Ba range on this dimension.

Short-term debt/total debt likewise highlights the bizarre relation of this ratio between the rated, public and private firms. While it shows a nice clear pattern for rated firms—higher short-term debt for the better-rated companies with access to the commercial paper market—the unrated and private firms have ratios consistent with Aaa firms. A model fit to public firms would err by concluding that short-term debt matters for private firms, while a model fit to ratings would err by concluding that it matters in the wrong direction: higher levels of short-term debt are indicative of lower risk.

Finally, inventories/ cost of goods sold shows a distinct trend for the universe as a whole and by ratings category. Computerization of inventories and just-in-time delivery have made a significant and positive effect on the way inventories are managed. This highlights the importance of monitoring relationships over time. While most ratios are relatively stable, this one trends. It is most likely that this trend is unrelated to risk, and instead a reflection of new technology. One should therefore parameterize the model using more recent period data to normalize the inventory ratio, as opposed to simply using the 1980 - 99 sample data - a classic case of layering judgement onto models. Given changes in technology, a 'high' inventory ratio is lower than what it used to be.

Section VI: Transformations And Functional Form

All models are wrong; some are useful.

~ George E.P. Box

Several important modeling assumptions are described in this section. At some level, all assumptions underlying modeling techniques are violated, and the purpose here is not to describe how our model is the inevitable result of certain axioms, but instead to make as clear as possible what the model does and why.

TRANSFORMATIONS

The functional form is often highlighted as the most salient description of the model (is it probit or logit?), but in fact, this distinction is far less important than two other modeling decisions: the variables used for estimation and the transformations of those independent variables. We have already described the variable selection process, essentially a step-forward procedure that started with the most powerful univariate predictors of default for each risk factor (e.g., profitability). The method for transformations is less standard, but very straightforward.

RiskCalc is a generalized linear model of the transformed ratios, where the transformation is derived primarily from each ratio's univariate relation to default.³⁹ As we have seen, many ratios are nonlinear, and sometimes not monotonically related to default (e.g., sales growth). To build upon the powerful univariate information, instead of weighting the ratios, we weight their corresponding individual default frequencies.⁴⁰

One common method of transformation is to standardize the data: remove the mean and divide by standard deviation. Yet, this still leaves very asymmetric and fat-tailed explanatory variables. Most importantly, this imposes a specific functional relationship between the variable and default within a probit model: marginal default risk is proportional to the number of standard deviations from the ratio mean. Other transformations that address nonlinearity, such as polynomial expansions, have the effect of decreasing its transparency (e.g., it is difficult to 'see' if the relation $3.14x - 8.21x^2$ makes sense).

Transparency in the manner in which inputs are transformed for a model is essential for one key reason: it greatly enhances the user's ability to monitor the model's effectiveness and to improve the model as new data roll in. For example, in the 1996 consumer credit debacle when consumer loss rates rose significantly across the board in a nonrecessionary year, modelers were able to adjust their models because they understood exactly how these models were supposed to work - and exactly how they did work. The issue was that for many specifications the model assumed a somewhat linear relation between individual leverage and future 'bad' rates. In fact, the relation was nonlinear, and loss rates rose exponentially as this ratio rose significantly for a large portion of the data sample. A regime shift generated a new experience for borrowers, and ultimately a new documented relationship between the independent variables and the predicted variable.

By comprehending the expected relation, it was straightforward to correct this deficiency when the future turned out differently than expected. The accumulation of such useful modifications is the key to developing a robust model - one that grows in power with experience.

Recognizing that nonlinearities exist yet wishing to use a traditional probit function forces us to address the nonlinearity within transformations. Unfortunately, the transformation function is not obvious: logarithm, exponentiation, standardization (converting to a z-statistic), polynomial expansion? Should one standardize first and then apply the transformation?

To solve this problem, we turned to a nonparametric approach. Nonparametric estimation is a collection of techniques for fitting a curve when there is little a priori knowledge about its shape. Many nonparametric procedures are based on using the ranks of numbers instead of the numbers themselves (Birkes and Dodge (1993)). Most importantly in this context, nonparametric estimation by local averaging works better in univariate regressions as compared to multivariate regressions.⁴¹ Silverman (1985) suggests that 'an initial non-parametric estimate may well suggest a suitable parametric model, but nevertheless will give the data more of a chance to speak for themselves in choosing the model to be fitted.'

Our approach is to use the univariate relationship between individual variables and future default as the basis for the transformation of independent variables. The result is a transformation of the data that largely resembles the univariate default frequency graphs described in section 4. Take, for example, the relation between net income/assets and default. Above 3%, the univariate relation between this ratio and future default rates is flat: higher income does not lower or raise the future default rate. By using this transformed variable in the multivariate model (i.e., its corresponding estimated default probability), NI/A above 3% will likewise have no marginal effect.

This approach is a simple extension of the common technique called 'minimodeling.' Consumer modelers have known for years that there is a bigger difference between 0 and 1, as opposed to 1 and 2, when looking at delinquency rates. Similarly, there's a bigger difference between increasing 10% in leverage from a base of 30% vs. a base of 70%. Calibrating these nonlinearities for individual ratios is the first step in model estimation.

There are two validating tests for the transformation function we chose. In the first, we calculate polynomial expansions of the inputs used in the ultimate model. We then regress the errors of different models on these expansion terms to see if there is any residual information in the original independent variables. If significance exists among the original variables, this implies some of the nonlinear effects were not sufficiently captured in the original model (see Spanos (1986), p. 446). Here we are comparing the approach of using the univariate default frequencies vs. the levels, both within a probit model.

Exhibit 6.1 shows that the coefficients on the square root of the independent variables are significant for the non-transformed (but truncated) inputs, while for the transformed ratios all coefficients become insignificant (private firm data).⁴² This illustrates the degree to which nonlinearity is captured in the transformation process.

³⁹ This is within a generalized linear model because for the probit model the linear component is within a normal cumulative distribution function.

⁴⁰ Going further, one can use orthogonal polynomial terms (Narula, 1979), which uses functions of lower order terms that ensures orthogonality that would otherwise not exist between x and x^2 , but this has the problem of drastically complicating the model.

⁴¹ Hardle, *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, 1990

⁴² Raw ratio values are always truncated at the 2nd and 98th percentile points in estimation and testing of models. This reduces the effect of extreme outliers, and significantly improves the explanatory power of every linear model (estimated or tested).

Exhibit 6.1

Regression of Model Residuals on Explanatory Variables, Public Firms, 1980-99

Variable	Model Residual	
	From Univariate Transform w/in Probit	From Truncated Ratio w/in Probit
	t-Statistic	t-Statistic
Intercept	0.13	4.84
Assets ^{0.5}	-0.05	2.68
L/A ^{0.5}	-0.25	8.49
Inv/COGS ^{0.5}	-0.11	6.74
Leverage growth ^{0.5}	-0.26	-0.96
NI/A ^{0.5}	0.08	-3.97
NI/A growth ^{0.5}	-0.02	-1.74
Quick ratio ^{0.5}	-0.33	6.25
Retained Earnings/Assets ^{0.5}	0.43	-11.76

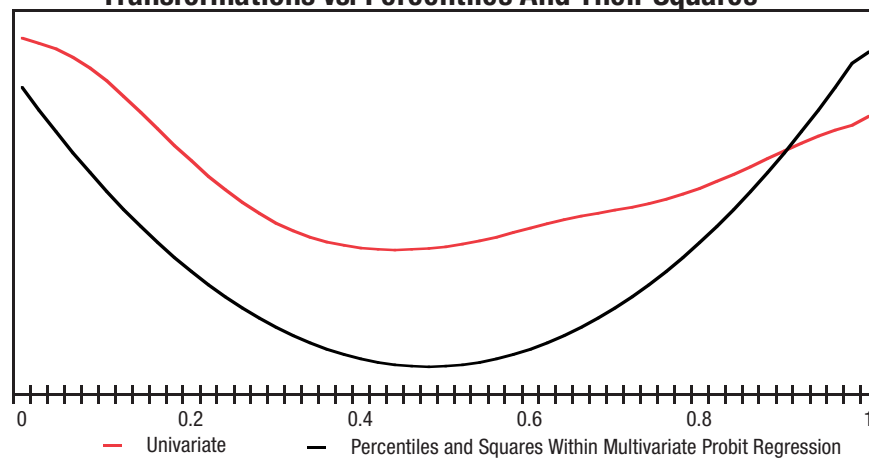
Higher order terms show significance against the residuals from the linear probit model, but not from the probit model that used univariate default frequency transformations.

Another concern is that correlations between independent variables accentuate or mitigate their individual nonlinearity, and the constraint implicit in the transformation to a univariate model may adversely affect the performance of a multivariate model. For example, sales growth shows a 'smile' relation to default, in that both low and high values of sales growth imply high default probabilities. Perhaps in a multivariate context that includes net income growth among other inputs, the marginal effect of sales growth is not a smile but a smirk. In that case, the high default rate associated with negative sales growth could be 'explained' by low net income and interest coverage.

To test this hypothesis for our ultimate set of independent variables, we looked at the estimated univariate contribution to the linear component of the probit model within a polynomial expansion of the percentile values of the explanatory variables. If within a multivariate model we used all the other terms, and various polynomial extensions such as size and size squared, what is the shape of the marginal effect of size? If it became flatter or more curved than the univariate relation, this would suggest that correlations between the independent variables affect their marginal relation to default. Exhibit 6.2 shows one such chart.

Exhibit 6.2

Impact Of Sales Growth Using Nonparametric Transformations vs. Percentiles And Their Squares



The polynomial expansion of the percentile information within a multivariate model allows one to adjust the transformation optimally.

The two lines represent the relative shape of the marginal effect of sales growth on private firms. The black line represents the univariate transform and the white line represents the unrestricted in-sample estimate where we transformed the input ratio into a percentile, and then summed the effect of the linear and nonlinear term within the context of a multivariate model. The absolute levels of these effects is immaterial since their effect is not dependent upon their mean, but their deviation from the mean. The shapes are consistent; if anything the percentiles-and-its-square is more curved.

This result holds for all the independent variables used in the RiskCalc model. The univariate relations to default are highly similar to the marginal relations within a multivariate estimation; the correlations between independent variables suggest that univariate default probabilities are broadly similar to the result from a less structured approach that uses polynomial expansions of these ratios within a multivariate context.

The result is that the transformation is nonparametric - essentially a numerical array that maps one number into another, using linear interpolation between 50 points that evenly span the range of each input. An alternative would be to find a closed form algebraic function that closely approximated the function. While this is possible, it trades efficiency for elegance. The problem is that the errors created by misspecification in the polynomial approximation outweigh the benefits of moving away from a linear interpolation. Further, the 'meaning' of a polynomial expansion only derives from examining the relation of the function over its range graphically, which is exactly what one replaces with the algebraic solution. It may appear more scientific to have a closed form equation than a lookup function, but estimation is more engineering than science, and our bias is towards workable solutions that are intuitive.

Intuitively, the RiskCalc model uses a combination of univariate models-the default frequencies corresponding to each input ratio-within a generalized linear model; the weightings are based on the relative importance of each univariate model. The initial transformation reduces the problem to one amenable to linear modeling, while at the same time capturing the nonlinear effects of the individual ratios.

As all of the individual estimators are unbiased estimates of default, the danger of misspecification is reduced.⁴³ For example, given two ratios with equivalent power in predicting default, if the optimal model were 2:1, which instead was misspecified as 1:1, the model might be suboptimal, but it will still do better than using either variable alone. Further, it will be highly correlated with the unknown optimal model. This is because all the transformed independent variables have similar univariate relations to the dependent variable. An inadvertent extra weighting on one variable at the expense of another may lower power, but the effect is mitigated by the fact that the transference of information is to an input that is by itself related to the dependent variable in the same way. The shape of the transformations used in RiskCalc are displayed in Appendix 6A.

FUNCTIONAL FORMS

The least squares method is the most robust and popular method of regression analysis, so that in general the word regression, if unmodified, refers to ordinary least squares (OLS). OLS is optimal if the population of errors can be assumed to have a normal distribution. If the symmetry of errors assumption is not valid, OLS is consistent but inefficient (i.e., it does not have the lowest standard error). Errors for default prediction have a very asymmetric distribution (in a binary context, given an average 5 year default rate of around 7%, defaults generally are 'wrong' by 0.93, while nondefaults are 'wrong' by 0.07), and so for this reason OLS is inefficient relative to probit, logit, or discriminant analysis in binary modeling.

Default models are ideally suited towards binary choice modeling; the firm fails or does not (1 or 0). Original work by Altman using discriminant analysis (hereafter DA) has in general been replaced by probit and logit models. In one sense, this seems logical since discriminant analysis assumes that the covariance matrix of the bankrupt and nonbankrupt firms is both normally distributed and equal. This assumption is obviously violated (ratio distributions are highly nonnormal, and failed firms have higher variability in their financial ratios). While several studies have found that in practice this violation is immaterial,⁴⁴ our experience suggests that probit and logit are indeed more efficient estimators than DA as theoretically expected (and empirically supported by Lennox 1998). The early popularity of DA owed more to ease of computation-matrix manipulation versus maximum likelihood estimation-than its predictive superiority, and logit and probit are now preferred since most statistical software contain these algorithms as standard packages. DA used to be easier, but now it is actually harder given the extra sensitivity-to-assumptions documentation required by discerning readers.

The choice between logit and probit is less important, as both give very similar results. Logit gained popularity earlier primarily because in the early days of computing anything that simplified the math was preferred. In this case, finding the maximum likelihood is easier for the function than for $[1 + \exp(-x/\beta)]^{-1}$ more than $(2/\beta)^{-1} \exp[-(x/\beta)^2]$. Again, computers have made this a non-issue. We chose probit as

⁴³The transformations are univariate default probability estimates, which are unbiased estimates of default rates associated with the sample.

⁴⁴see Zavgren (1983, p. 147), Amemiya (1985 p. 284), Altman (1993)

opposed to logit, though this is comparable in effect to choosing assets instead of sales as the measure of firm size.

Though DA, logit, and probit generate roughly similar estimation results, there is a very different interpretation that arises from them, which was first explicitly addressed in the default literature by Ohlson (1980). DA is about separating a sample into two groups: default and nondefault. Logit and probit lend themselves more to the interpretation of which observations have a higher probability of belonging in a certain group. With hindsight, loans either pay off completely or they don't; with perfect foresight all the good loans should have been priced the same. If you are thinking about grouping a set into two groups as in DA, one tends to look at the exercise as ex post classification as opposed to ex ante forecasting.

Altman, a proponent of DA, states that the "presumption underlying credit scoring models is that there exists a metric than can divide good credits and bad credits into two distinct distributions" (Caouette, Altman, and Haldeman). Alternatively, one can think of firms having a continuum of intrinsic propensities to failure, just as a poker hand has an intrinsic probability of success. Zen-like, poker hands are neither good nor bad, but instead each has different probabilities of success associated with it.

While the two views are not inconsistent, DA evokes concepts such as how many type 1 and type 2 errors a model produces at various levels of cut-off. Probit looks more at measuring the consistency between expected probabilities and actual probabilities, as well as trying to 'stack the odds' in one's favor as much as possible. In other words, DA targets a bankrupt/nonbankrupt cutoff and the accuracy thereof, while the probit model produces a probability of bankruptcy that can be used for not only decision-making (loan/don't loan), but for estimating expected loss.

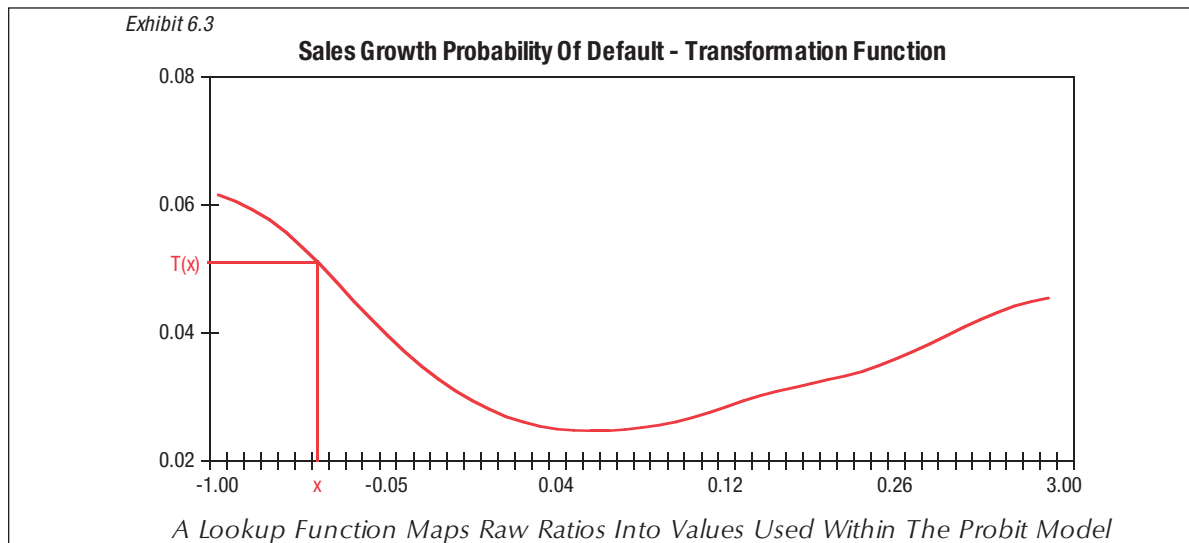
This latter distinction lends itself to pricing decisions, as in RAROC models, and is already frequently used in the consumer world, where there are sometimes 5 different distinctions that imply 5 different loan rates to customers. In contrast, the two-groupings approach of DA implies that differences among pass credits is not the result of the different characteristics of the firms themselves, but from the noise in their

$$Prob(\text{default}|T(x)) = \int_{-\infty}^{-\beta T(x)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \Phi(\beta T(x))$$

estimation. The results from DA are therefore not as amenable to pricing (i.e., pricing appropriate for a firm is independent of an imperfect metric).

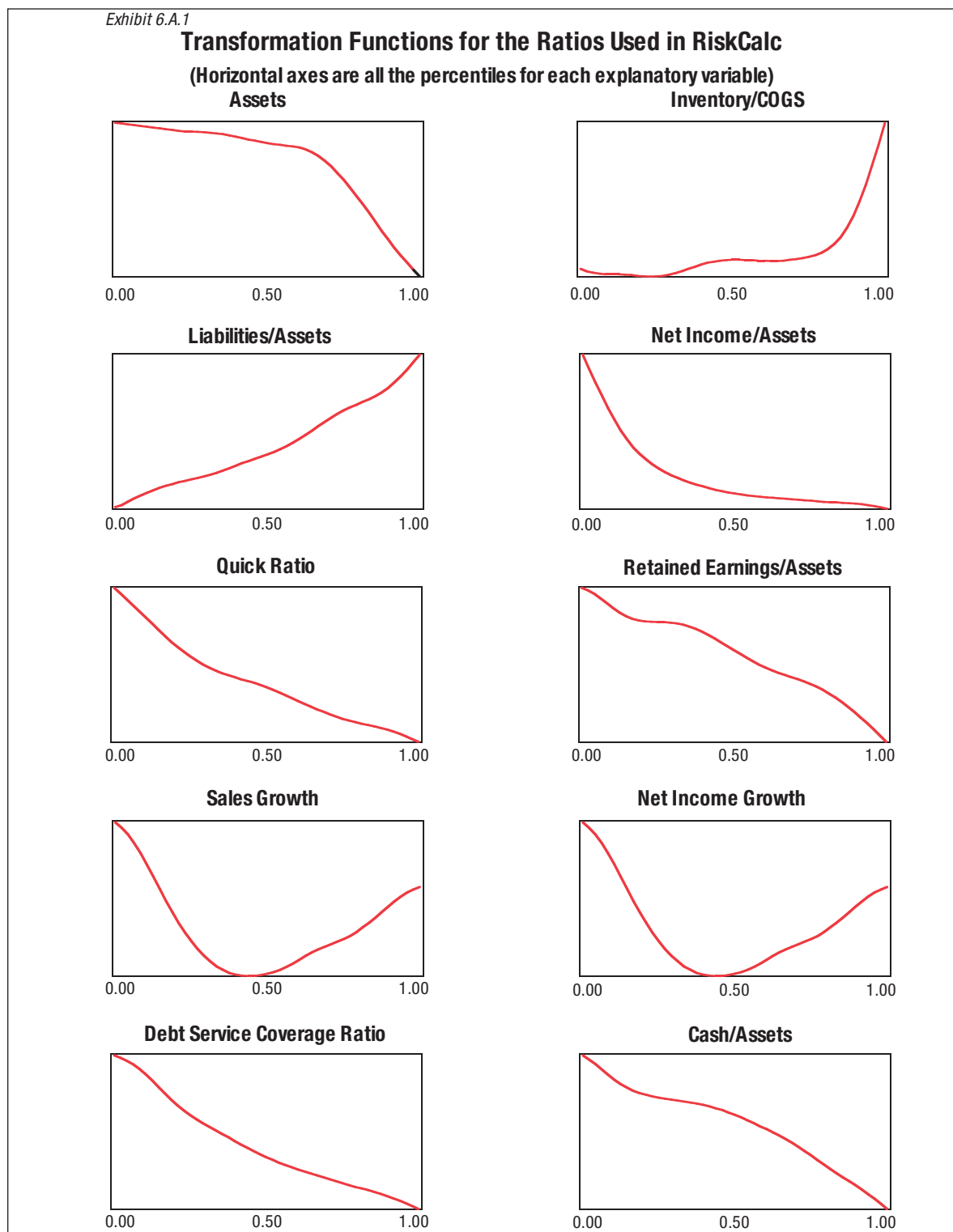
$$x = -0.08 \Rightarrow T(-0.08) = 0.049$$

The model is therefore estimated as follows:



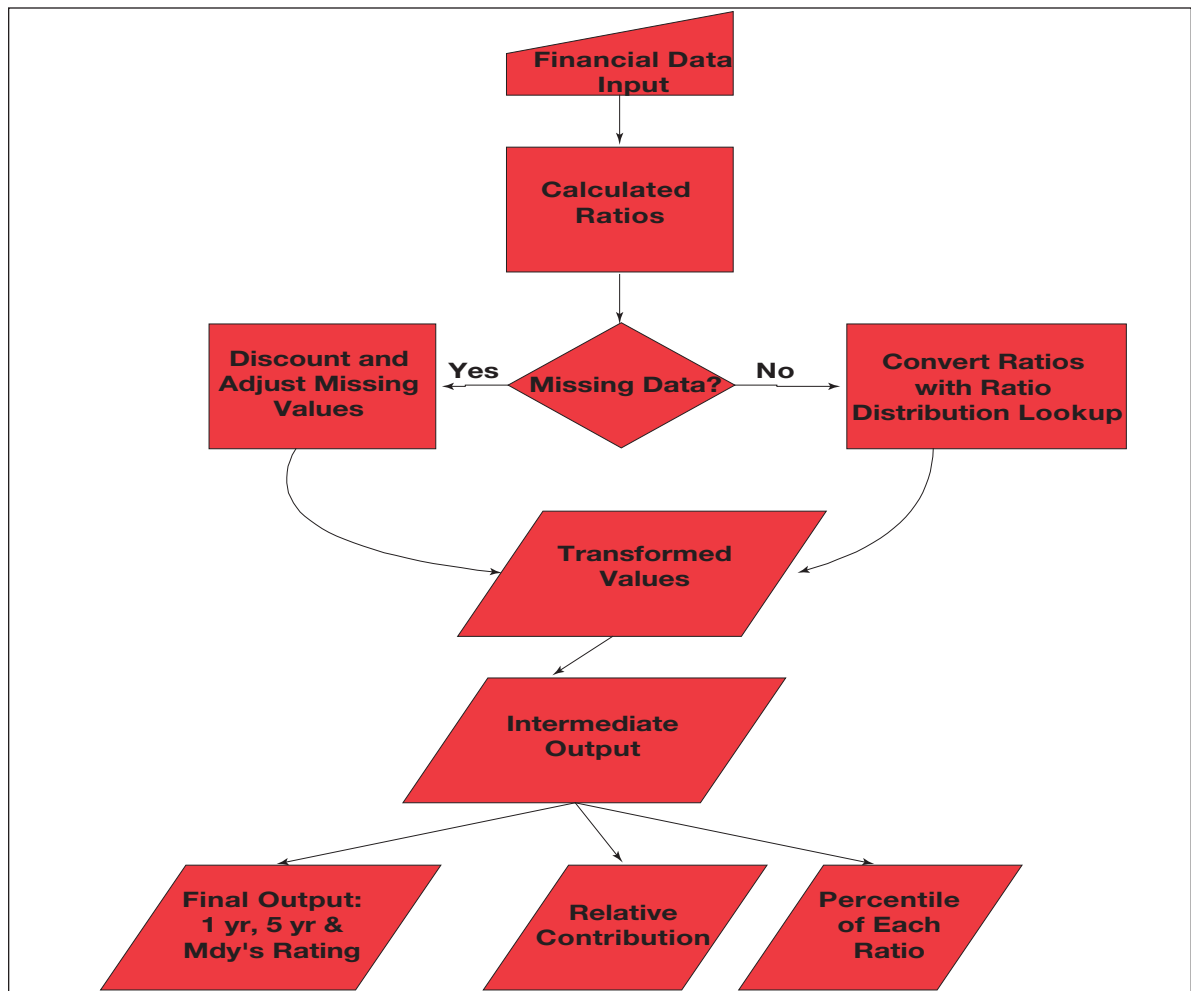
Here, $T(x)$ is a vector of functions that transforms the 10 input ratios, plus a constant term, where the transformations are basically the univariate relationship between each input and the future 5-year default frequency. Thus, for sales growth=-0.08 (i.e., - 8%), we transform the data in the following way:

This is shown for the range of sales growth levels in Exhibit 6.3. This is done for all the input variables.



Appendix 6A: Transformations Of Input Ratios

The shapes of the transformations of the 10 input ratios used in RiskCalc for Private Companies are illustrated below. The vertical axes have different scales, so they are not strictly comparable from one to



another. The total contribution for each input is the product of the transformation (which may result in a value from 0.02 to 0.10, or from 0.03 to 0.05), and the respective coefficient on this transformation.

These charts are similar to the univariate relations to default, except the graphs have been smoothed. The horizontal axes are normalized to be in percentiles of each ratio within our private firm database.

Appendix 6B: RiskCalc Schema

This appendix describes the RiskCalc algorithm in a concise and separate format, and can be used to understand the actual RiskCalc algorithm without reference to the more lengthy text.

The following flow chart shows how RiskCalc is calculated:

Input Fields

Two years of financial data are used. None of the ten input fields are required, since any missing input will assume mean values. Clearly, the more information input, the better the result.

- 1 Assets (total)
- 2 Cost of Goods Sold
- 3 Current Assets

- 4 Current Liabilities
- 5 Inventory
- 6 Liabilities (total)
- 7 Net Income
- 8 Retained Earnings [Net Worth - Book Equity]
- 9 Sales (net)

Ratios	Calculation
Assets	Assets / CPI of the year
Inventory/COGS	Inventory/COGS
Liabilities/Assets	Liabilities /Assets
Net Income/Assets	Net Income /Assets
Net Income Growth	(current Net Income/ Assets)-(prior Net Income/Assets)
Quick Ratio [= (curr. assets - invent)/curr. Liab.]	(Current Assets - Inventory)/Current Liabilities
Retained Earnings/Assets	Retained Earnings/Assets
Debt Service Coverage Ratio	EBIT/interest
Cash/Assets	Cash & Equivalents/Assets
Sales Growth	(current Sales/prior Sales)-1

10 EBIT (Earnings Before Interest and Taxes)

Factor	Relative Contribution
SIZE	14%
Total Assets	14%
PROFITABILITY	23%
Net Income /Assets	9%
Net Income Growth	7%
Interest Coverage	7%
LIQUIDITY/CASH FLOW	19%
Quick Ratio	7%
Cash & Equivalents/Assets	12%
TRADING ACCOUNTS	12%
Inventories / COGS	12%
SALES GROWTH	12%
Sales Growth	12%
CAPITAL STRUCTURE	21%
Liabilities / Assets	9%
Retained Earnings/Assets	12%

- 11 Interest Expense (excluding leases)
- 12 Cash & Equivalents (e.g. marketable securities)
- 13 Extraordinary Items

Ratios

Based on the data from the input fields, the following ratios are calculated. Section 4 outlined the variable selection process. The relative weight is not a strict coefficient, but the effect of the model using a numerical differencing procedure, which approximates the relative effect of each factor, and subcomponent, on the final RiskCalc (due to the nonlinear nature of the model, any effect is only approximated by an particular numerical derivative).

The input ratios by category:

Ratio Calculation

The Consumer Price Index (CPI) is from U.S. Department of Labor, Bureau of Labor Statistics.

Values Used: Transformed Ratios

The ratios are transformed into values used in calculating probabilities of default. The determination of the value used depends on the availability of the input data.

- 1) If ratios are calculated with sufficient data:

If ratios have sufficient input data they are calculated using a linear interpolation of a lookup function. These converted values represent, approximately, the univariate probability of default associated with the values in the CRD over 5 years, and are meant to capture the nonlinear relation between the risk factor and the probability of default. These transformations, like the coefficients, are substantively different than those used in the public firm RiskCalc.

- 2) If ratios are not calculated due to missing data:

The model uses the mean value of the transformed ratios if no value is given. Clearly the less real data, the worse the output of the model. Nonetheless, the model does have power in the face of missing observations, and thus is set up to be robust to missing values.

Intermediate Output - The Unadjusted Probability of Default

5-Year Cutoffs	Corresponding Rating
0.00%	Aaa
0.27%	Aa1
0.39%	Aa2
0.49%	Aa3

An intermediate output, unadjusted probability of default, is calculated from the coefficients and the transformed inputs. It is the Gaussian (standard normal) cumulative distribution function of the sum of the products of the coefficients times their transformations.

Ratios	Relative Contributions		Percentile
	1-Year	5-Year	
Assets	4%	20%	37%
Inventories / Cost of Goods Sold	-2%	-3%	73%
Liabilities / Assets	10%	7%	65%
Net Income Growth	-9%	-5%	36%
Net Income / Assets	-16%	-11%	69%
Quick Ratio	-1%	-4%	59%
Retained Earnings / Assets	0%	0%	49%
Sales Growth	-28%	-21%	55%
Cash/ Assets	21%	23%	21%
Debt Service Coverage Ratio	-9%	-6%	59%

Final Output: 1-Year DP & 5-Year DP

Finally, the intermediate output is transformed into an Default Probability (DP) using another lookup table. One- year and five- year DPs are produced based on an in-sample mapping that corresponds the various output buckets to estimated private firm DPs.

Mapping into Moody's Rating Symbols

This final Default Probability (DP) is mapped into a Moody's rating using a lookup table.

This table uses Moody's 5-year default rates, which are the default rates unadjusted for withdrawals that we observed over the 1983-99 period. The 5-year cutoffs are those rates that represent the floor for the rating.

Example: when RiskCalc's DP for 5 years = 0.29%, this is mapped into the Moody's rating Aa2.

Supplemental Information

There are two useful measures which do not affect the output score, but do help explain it. They are Percentiles and Relative Contributions for each of the nine input ratios (excludes size). The following is an example:

Note that in some cases higher percentiles are more adverse values and vice versa. The percentiles are based on the rank of these input variables within the CRD.

1) Percentiles:

Each input ratio is translated into a percentile.

2) Relative Contributions:

Relative contributions are the proportional contribution of each input ratio to the final RiskCalc. As different coefficients are used for the 1 and 5-year DP calculations, there are different relative contribu-

$$r_i = \frac{T(x_i) - E\{T(x_i)\}}{\sum_{i=1}^{10} \text{abs}[T(x_i) - E\{T(x_i)\}]}$$

tions for each horizon. These outputs are relative to the mean, so that a mean score in all input ratios would show zeros for all inputs, while one with adverse ratios have large positive values (higher DP). The value of these relative contributions is that they allow apples-to-apples comparability between risk drivers as to what is influencing the final result.

An Example of Relative Contributions

Assume the average exam score is 500 for both verbal and math, with an average total exam score of 1000.

If Frederico got a 1100 exam score, based on a 700 verbal, and 400 quantitative, the relative contribution percentages could be calculated as

Sum Absolute Difference = Absolute value(700-500)+abs(400-500)=300

(700-500)/Sum Absolute Difference =200/300= +66% = verbal contribution

(400-500)/Sum Absolute Difference=100/300= -33% = math contribution

These give a sense of the contributors to Frederico's score. In this case, the verbal score contributed twice as much as the math to Frederico's score relative to the average score. The numbers are directly related to each other, and allow one to gauge relative contributions in a straightforward manner for each obligor and each horizon. The absolute value of these numbers will always sum to 1. Mathematically this can be seen as

Where r_i is the relative contribution of ratio i , and abs is the absolute value function.

Section VII: Mapping To Default Rates And Moody's Ratings

Nothing can be done except little by little.

~ Charles Baudelaire

The production of default rates and mapping them into Moody's ratings from a statistical model is another useful characteristic of RiskCalc. As this is a somewhat separate functionality of RiskCalc, it can be evaluated independently of the power of the model. Once a model that produces scores that ordinally rank companies has been constructed, the next step is to map these outputs into default rates and estimated Moody's ratings. This process contains several subtle assumptions, and is much less straightforward than many people realize. Moody's ratings are statements about credit quality that primarily target expected loss (default rate \times loss rate), not default rates. Further, at the higher end, ratings also take into consideration ratings stability. Thus, a mapping into a Moody's rating is never unambiguous in terms of a default

45 Senior implied rates are used in Moody's default studies, and represent the real or estimated rating on a company's senior unsecured debt.

rate prediction. However, given the observed default frequencies with which Moody's "senior implied" ratings are associated, these ratings have developed into a benchmark that is used internally and by the investment community.⁴⁵

The mapping to a Moody's rating is somewhat superfluous given the default predictions generated. That is, if a Moody's rating is more than a default rate estimation, why try to fit a round peg in a square hole? In the transition from internal grades to more quantitative risk metrics, such as loss given default and default probabilities, the existence of a Moody's mapping serves as a useful benchmark for gauging commercial loan risk. Currently most internal loan grades are mapped into a Moody's rating, and ultimately this is the focal point of many comparisons. As this mapping is not straightforward for reasons addressed below, our mapping significantly enhances the usefulness of the RiskCalc product.

DEFINITIONS OF DEFAULT

Ideally, we would like to predict loss rates, that is, the severity of losses associated with those loans that not only default by missing payment but that do not eventually fully pay back principal and accrued interest. Unfortunately, such information is rare. In practice, once a credit is identified as defaulted it moves to a different area within the lending unit (work-out or collections), which makes it impractical to tie the original application data to the eventual recovery amount. Thus, we ignore the recovery issue, and choose a definition of default that is relatively unambiguous, easy to measure, and highly correlated with our target: loss of present value compared to the original loan terms.

The two primary definitions of bad in the commercial loan literature are default and bankruptcy. Bankruptcy undercounts the amounts of defaults, as some defaults are not bankruptcies, while all bankruptcies are defaults. Neither are perfectly correlated with loss.

In this study, we use as a measure of bad an obligor who had any of the following occur:

1. 90 days past due,
2. credit written down (e.g., in the US this means placed in the regulatory classifications of standard, doubtful or loss),
3. classified as non-accrual, or
4. declared bankruptcy.

PREDICTION HORIZON

Once we have determined what we will count as bad, we need to determine whether we wish to measure bad over 1, 5, or 10 years. In analyzing collateral pools for securitizations, Moody's Structured Finance Group targets a 10-year default rate, as many loans and bonds in these structures are outstanding for that duration, and one needs to develop comparable loss forecasts with the contractual maturity of the corresponding senior notes. Moody's Global Credit Analysis describes 'credit opinions' as pertaining to periods from 3 to 7 years in the future. However, most default research refers to 1-year default rates. For example, the B default rate of around 6% refers to a 1-year horizon implicitly, just as an interest rate of 6% refers to the annual interest rate, not the quadrennial or monthly total return.

While translating default rates into annualized numbers is essential for comparability, this does not imply that one must or even should target a one-year default rate in estimation or testing. For RiskCalc, a longer horizon is used for the mapping to Moody's grade because it is more consistent with the horizon objective of Moody's ratings.

The real problem with the 1-year rate is that very few loans go bad within 12 months of origination, and of those that do, fraud is often a factor, in which case a model based on financials wouldn't work well anyway. Moody's documents an average of 5.5 and 16.5 years to default from initial rating coverage for speculative and investment grade bonds, respectively. The mortality curve shows a pronounced, abnormally low default rate in the first year and even second year (see Keenan, Shtogrin, and Sobehart (1999)).

Thus, even Moody's grades, though often discussed and compared in annualized default rate terms, do not imply a year-ahead rate as much as they do an average annualized rate for a randomly seasoned loan in a particular grade. This is a subtle, but important distinction.

RiskCalc generates both 1 and 5-year default rates. The one-year horizon is useful for portfolio managers who have to provision and plan with annual horizons and to anticipate intervention strategies to restructure loans. But this is for monitoring a portfolio of loans with 'average seasoning.' We would suggest that 1-year default rate forecasts are less valuable at origination than the 5-year rate. Models can be estimated and tested on both horizons; yet, at some point, one is going to prioritize one horizon as the more important, and this horizon will receive more attention in the testing process.

The different estimation horizons used for the 1 and 5-year default predictions produce roughly similar ordinal rankings from the different models; the primary difference is in the weightings on the explanatory variables. Inputs like size and retained earnings are not as significant at the 1-year horizon as at the 5-year horizon. Different risk factors are also more salient for monitoring versus origination, and the relative contribution information that accompanies a RiskCalc output will illustrate these differences.

DEFAULT RATE AND MOODY'S MAPPING

Though the output of a probit model is a probability, it does not directly represent a calibrated default probability because the sample default rate used in estimation is often different than the estimated population default rate. The following methods are used to generate a default estimate and a mapping to a Moody's rating.

Moody's Mapping Preliminaries

There are several ways to map a model's output to a default rate. The four main methods are:

- 1) Use the averages of the model output per Moody's grade.

Example: if the average model output for Aaa-rated companies is 0.05% and the average model output for Aa-rated companies is 0.20%, the cut-off for Aa is then some point between 0.05% and 0.20%. Alternatively, if the model is not targeting default rates explicitly, an average score for Aaa might be '3' and for Aa '4.5', implying a cutoff between 3 and 4.5. The cut-off is independent of the default rates for Moody's grades.

- 2) Estimate model on ratings themselves, such as in an ordered probit.

This estimation ranks the grades ordinally, and simultaneously estimates the cutoffs for various grades and the parameter coefficients within the model. It is a straightforward extension of the binary modeling procedure of probit.

- 3) Force the model to have the same proportion of ratings as observed on the rated universe.

In this approach, if 10% of all Moody's-rated corporate firms are rated Aaa, then the model's highest 10% are mapped to Aaa. If the next 5% are rated Aa, then this next percentile defines the range of the grade Aa, etc.

- 4) Use default rate predictions to link with Moody's ratings based on a set of assumed default rates by rating and horizon.

Example: Using the 1-year horizon and a set of Moody's default rates, a default rate forecast of 9% would map into B3.

Given the desirability of spanning Moody's ratings, and the difficulty getting sufficient data to allow one to distinguish between Aaa and Baa1 companies, we use method #1 for estimating cutoffs for these upper investment grade groupings. That is, as the Aaa and Aa annual default rates are approximately 0.01% and 0.03%, respectively, we will never have sufficient data to estimate this distinction empirically using default data. Therefore, we made the judgement call to take the high end of the output range, where default prediction empirically tails off, and use the average RiskCalc scores applied to Aaa, Aa, and A rated companies to determine the cutoffs for these ranges. While this adjustment is somewhat ad hoc, it is an example of the useful judgement that compliments quantitative models. Applying these cutoffs in the

model for private firms produces a maximum estimated Moody's rating for the universe of A1. For companies for which we do have sufficient empirical data on default rates, however, we do not think that any of the first three approaches are optimal for private firms because the differences between the rated and unrated universes are significant, as is borne out in Sections 5. Just as applying a model estimated on rated, or even non-rated public firms, to private firms is suboptimal, so too is a calibration on rated firms suboptimal relative to a model calibrated to private firms. Producing extreme default rates is purely a function of power, and mapping directly to grades without going through default rates implies a power that is not proven; it basically simply asserts equivalent power to Moody's ratings without validating this with actual default rate experience.

This brings us to the final method for mapping to an agency rating - #4. We prefer this last approach because it allows the modeler to do what he does best: opine as to what the rating means in terms of default rates as suggested by examining the targeted data. Further, it is more consistent with a common

Exhibit 7.1

Annual Default Rate Estimates	
Bond Default Rates	
Corporate Bond Default Rate ('83-99):	1.20%
Loan Default Rates:	
Dun & Bradstreet total failure rate ('84-97):	0.97%
FDIC nonconsumer nonperforming ¹ rate estimate ('88-97):	1.5%
Society of Actuaries unrated default rate ('88-94):	1.1%
Average of Loan Default Rates:	1.2%
<i>Recorded default rates for private bank loans are roughly similar to the default rates of rated corporate bonds.</i>	

recalibration technique: given a historical set of data, plot the predicted default rate on the x-axis and the actual default rate on the y-axis for a bucketed set of data. For example, take a dataset from 1997, and rank it from low to high forecast default rates. Average the default rate for 20 buckets, and calculate the actual subsequent default rate for the bucket. By examining these plots and making appropriate adjustments, over time a consistent model will generate points around a 45 degree line, implying that forecasts match actual default rates.

Assuming one chooses the last approach, several assumptions are necessary to generate an estimated Moody's rating:

Issue 1 - Unrated Firm Default Rates

The data presented below suggest that the average bank loan has a default rate comparable to Ba2-rated debt. This is consistent with how banks view their middle market portfolios, as most of their middle market loans are put into Ba1 and Ba3 buckets.

The opaqueness of the middle market often leads outsiders to infer that bank debt is at least as risky as public debt. (See Appendix 7A for empirical evidence). While it may be prudent for investors to assume the worst when evaluating certain credits, one can always assume the worst at the end of the algorithm. Assuming the worst at various steps within the algorithm produces an ultimately confusing output; it is not clear how several conservative adjustments affect the final output: additively, multiplicatively?² Adding intentionally biased judgements within a model is difficult for outsiders to interpret. After all, outsiders may have different and equally valid subjective adjustments.

Exhibit 7.1 documents that the average default rate for loans, as opposed to bonds, from 3 independent studies, was 1.2% annually.

The raw FDIC datapoint is derived aggregate data, and is not supplied directly by the FDIC. The method is the following: Assuming an approximate proportion of consumer loans of 20%, and assuming a conservative 5% nonperforming rate for all consumer loans, and average nonperforming rate of 2.2% for the 1984-96 period, this implies nonconsumer loans (which include middle market loans among other loans) have a nonperforming rate of 1.5%.⁴⁶

An issue with 'bad' rates for unrated companies is that usually they pertain only to bankruptcy. In Moody's rated universe, total defaults were 42% greater than the number of bankruptcies. This implies

⁴⁶All FDIC data are from their Statistics on Banking publication.

that some sort of upward adjustment is necessary to the unrated 'bad' rates to make them comparable with the traditional default rates. If we take the average of 1.2% times 1.42, we get 1.70%.

A final datapoint comes from discussions with leading consultants and experienced industry professionals, who opine a 0.5% annual loss rate through the cycle is a reasonable estimate for middle market portfolios. Assuming an average loss given default rate of 30% for defaulted bank loans, this implies a 1.66% default rate, which is close to our earlier estimate of 1.7%.

Thus, a private firm default rate, through the cycle, is estimated to be around 1.7%, and we will use this as our assumption in the ultimate calibration on RiskCalc. It is expected that this rate will be both too high in expansions and too low in recessions due to the cyclical nature of lending. In fact, an estimate that is 'too low' will be 'right' more often, but good lenders know the significance of one bad year.

It seems most likely that private loans exhibit losses comparable to the Ba2 level, just as the banks assume internally, and that greater transparency (from, say, RiskCalc) could help outsiders gain comfort with this mapping. That is, currently the volatility of the valuations of these portfolios is at least as great as for speculative-grade debt, even though the average default rate is well below the average speculative-grade rate (1.7% vs. 3.8%).

It is necessary to estimate this final population default rate because, invariably, the sample used to estimate and calibrate this model is biased. In many academic studies of defaults, paired-sample tests contain a biased sample where the true population default rate is well below the rate used in the estimation sample (50%), making inferences of probability difficult. Moody's private firm dataset likewise has a different sample default rate as compared to the population default rate due to a straightforward problem. Many banks were able to give us financial statements, but a much smaller percentage were able to link this to credit quality indicators that allowed us to say which were bads. Though we try our best to create a sample that includes only obligors on which we have subsequent credit status information, the low annual default rate (0.4%) suggests that some firms went into default undetected. The model may be accurately calibrated to the sample default rate, but misfire out of this sample because of this oversight.

Therefore, the final mapping is adjusted so that it is consistent with our estimate of the aggregate unrated default rate. In this example, we made forecasts developed from the sample consistent with the population estimate by multiplying the forecasts by 4.25 (i.e., $1.7/0.4$). For example, if an initial forecast of the model, as estimated on the sample with a 0.4% default rate, was 1.0%, its adjusted default rate would be 4.25%.

Generating a 5-year default rate assumption for private loans is not as straightforward. A 5-year default rate is not the geometric product of 1-year survival rates due to withdrawals and the mortality curve of loans where the marginal (i.e., annual) default rate varies over the life of a loan. Moody's corporate universe has 5-year unadjusted default rates of approximately 4 times the 1-year unadjusted default rates, and thus we calculate the 5-year default rate for private firms as 4 times 1.7%, or 6.8%.

Issue 2 - Readjusting Moody's Default Rates for Withdrawals

Statistical models calculate default rates unadjusted for withdrawals. That is, the model assumes a '1' for default and a '0' for nondefault, where a withdrawal of a loan is considered nondefault. In contrast, standard Moody's default tables adjust for withdrawals. To take a simple example, if 100 firms are rated in year 0 and 50 firms withdraw in year 1, and then 10 firms default in year 3, the 5-year default rate as normally represented in Moody's default tables is 20%. It is assumed that any withdrawals that do not default within a year are in essence reinvested into the remaining cohort's outstanding bonds. In contrast, an accurate model would calculate a 10% default rate, as it reflects the probability of firms defaulting, not of firms defaulting conditional upon not withdrawing.

Thus, in order to map a model's output into Moody's ratings, one needs to adjust the Moody's default rates as published in the annual default studies so that they are both referring to the same conception of a default rate. For 1-year defaults, the unadjusted default rates are lower by a factor of 0% to 10%, with smaller adjustments for the higher rated firms with lower default rates (these are proportional adjustments, so a 10% adjustment lowers the default rate from, say, 7.5% to 6.75%). For the 5-year cumulative rate, the adjustments imply lower default rates of between 13% and 43% for the grades Aaa to B3, respectively.

Issue 3 - Time Horizon for the Mapping to Moody's Ratings

One must also decide which horizon to use for the mapping. Does one map over the model's 1-year default rates into Moody's ratings, or its 5-year default rates? This is a potentially material distinction because the difference between the actual and forecast default rates changes over different horizons. For example, the ratio of the B2:Ba2 default rate moves from 2.6 to 2.0 as one moves from 1 to 10 years. A statistical model will probably not exactly match this relative difference for all the obligor forecasts at various horizons, and so any single model could generate different mappings for the same obligor depending upon whether one uses a 1-year or 5-year horizon. This difference is usually not more than one notch, but there is a difference.

We prefer the 5-year as the mapping horizon for three reasons:

Exhibit 7.2
Moody's 5-Year, Smoothed, Cumulative Default Rates, Unadjusted for Withdrawals, 1983-1999

Aaa	0.20%
Aa1	0.35%
Aa2	0.44%
Aa3	0.56%
A1	0.58%
A2	0.62%
A3	0.75%
Baa1	1.31%
Baa2	1.45%
Baa3	3.28%
Ba1	5.73%
Ba2	7.48%
Ba3	13.92%
B1	15.66%
B2	19.52%
B3	23.70%
Caa-C	29.00%

1. The 5-year cumulative default rates have less relative volatility, both for Moody's ratings and for the forecasts of any model.
2. Moody's ratings target a 3 to 7 year horizon.
3. At the 1-year horizon, there are zero defaults in the A3 and above levels, which makes it difficult to statistically map anything into these grades

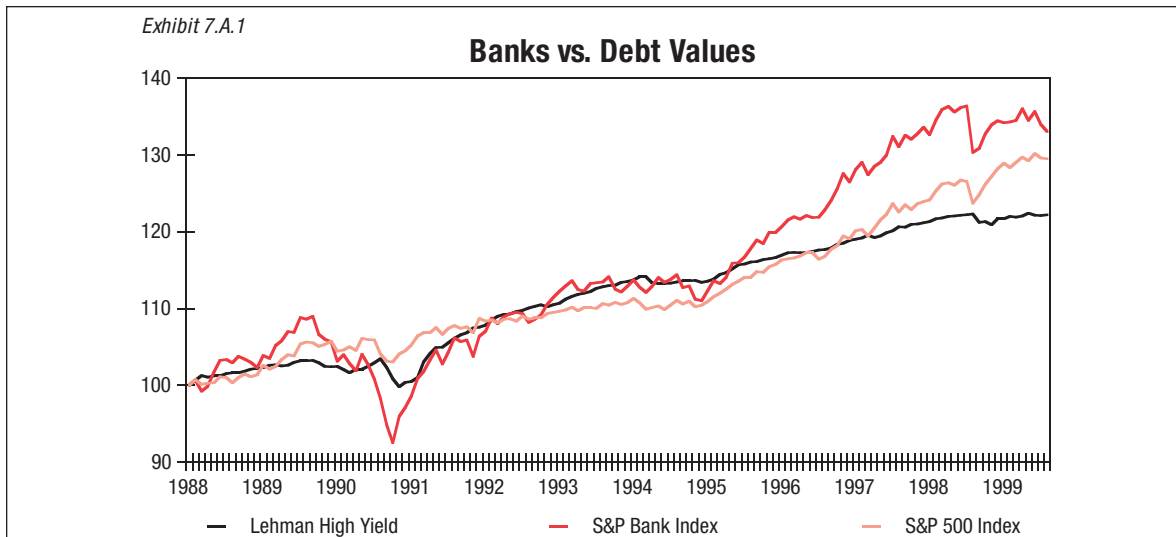
Issue 4 - Time Period for Sample

A final necessary choice is which sample to use to correspond with differing default rates by grade and horizon. For example, the period 1940-1969 experienced extremely low default rates, which many consider atypical. One could reasonably target the 1970-99 period, the 1983-99 period, or even a combination of these samples based on a presumption that they were still abnormally low (e.g., if you expect another Great Depression over the next 30 years).

We chose 1983-99 because it spanned the period when Moody's refined ratings (i.e., Ba2 as opposed to Ba) came into existence as an extension of the 6 basic grades. Also, in the 1980s, a structural shift occurred in the non-investment-grade market such that it became populated not simply by 'fallen angels'-firms originally rated investment-grade - but instead by original issue non-investment-grade companies. As has been discussed, a 'mortality curve' for bond default rates suggests that aging affects marginal default rates by grade, and so default rates can be expected to be different for fallen angels vs. original non-investment-grade securities. The 1983 start to our sample nicely encapsulates the beginning of this structural change in the debt markets, and leads to more meaningful default rates going forward.

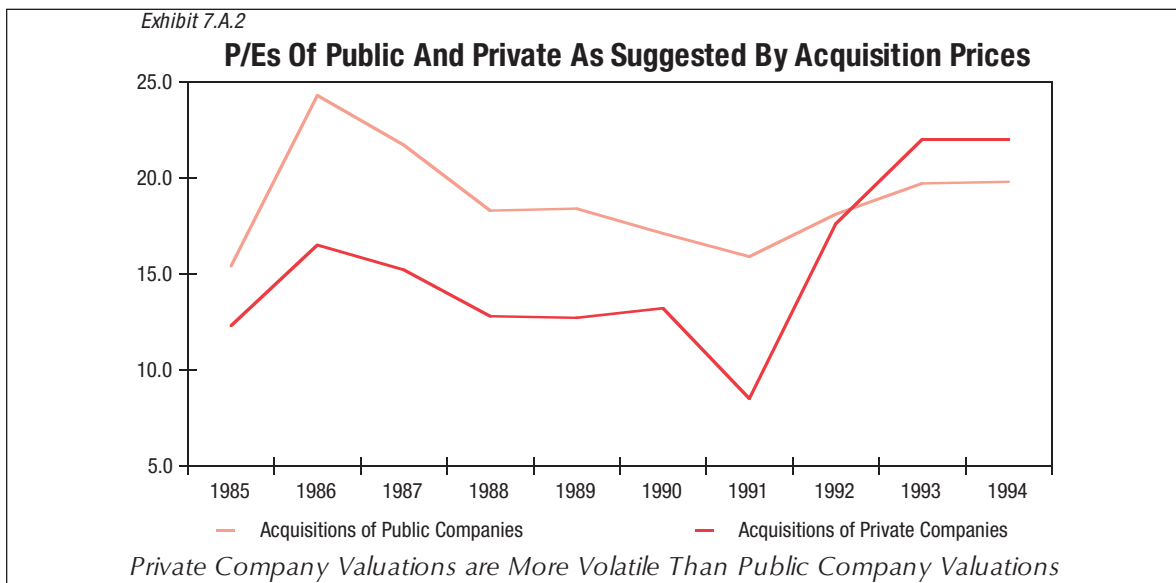
Finally, one must smooth the resulting default rates. Non-monotonicities are the result of small sample variation, as reflected in the fact that these tend to disappear as the default rate horizon and the total number of defaults increase. At the 5-year level, our horizon for mapping, only minor adjustments needed to be made.

The net result is the following:



Summary Of Default Rate Calibration And Ratings Mapping Method

1. Assume a population default rate for private companies of 1.7%.
2. Use a time period to generate default rates representative of Moody's grades: 1983-99.
3. Use 'average of rated scores' to determine granularity of upper investment-grade mapping points (above Baa1).

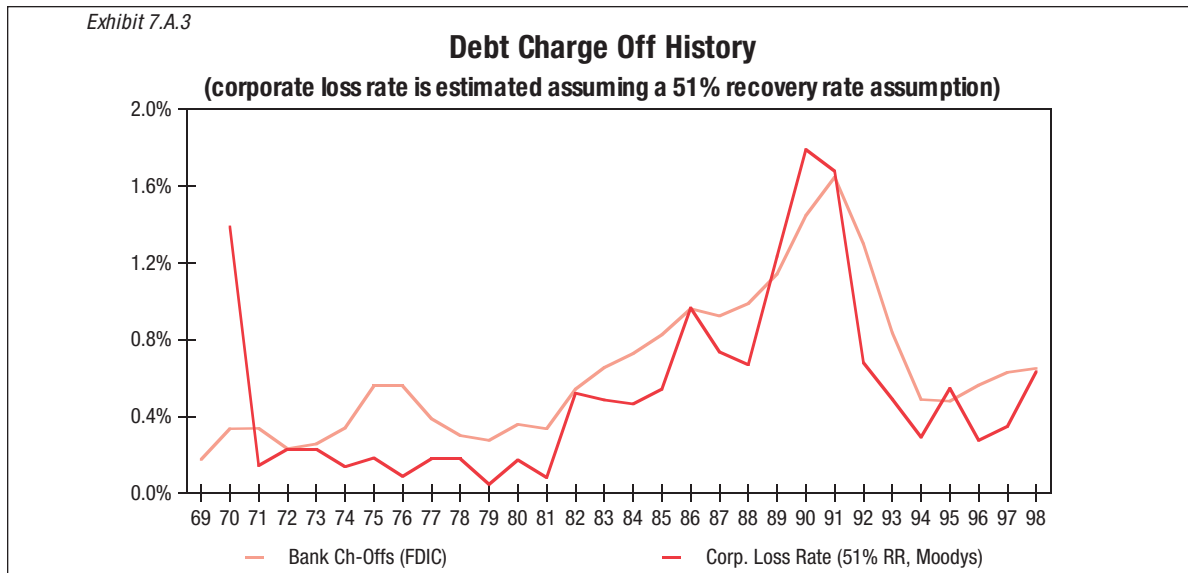


4. Use Moody's default rates unadjusted for withdrawals.
5. Use the 5-year default prediction to map into Moody's ratings.

Appendix 7A: Perceived Risk Of Private Vs. Public Firm Debt

Data in this appendix demonstrate that while private firm debt is perceived by the public as at least as risky as public firm noninvestment-grade debt, in reality, it is likely less so. This is important to document because the average default rate, and its volatility, are key to determining the appropriate mapping of private

Exhibit 7.A.3



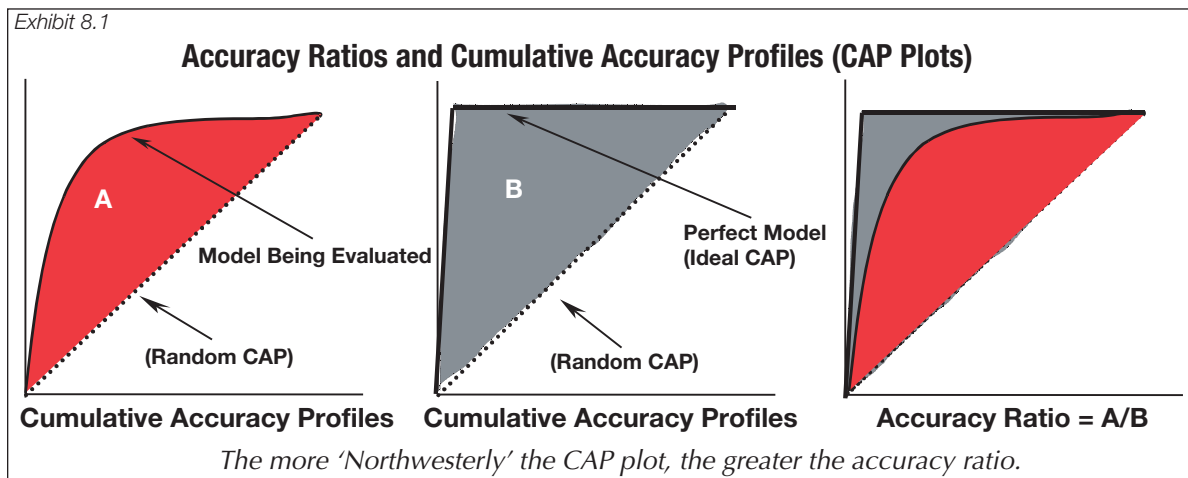
firm loans into Moody's equivalent grades. The difference between perception (private loans are as risky as B1 rated loans) and reality (our best guess is more like Ba2), is a distinction with a significant difference.

Exhibit 7.A.1 shows the variability of the S&P bank index vis-à-vis Lehman's aggregate speculative grade index and the S&P500. Bank equity values are more volatile than a marked-to-market portfolio of speculative grade debt, especially during the 1990 recession. This implies that bank portfolio values are riskier than speculative grade bond portfolios. Interest rate risk of the speculative grade portfolios should be at least as great as the interest rate risk of banks, though there is no published information on bank duration. Thus it is probable that perceived credit quality variation is driving these results.

Every year, Mergerstat Review publishes a table presenting the average P/E multiples for the acquisitions of private companies for which they have data and compare them to the average P/E multiples for the acquisitions of companies that had been publicly traded.⁴⁷ As figure 6.2 demonstrates, public and private firms display similar volatility over the cycle. If anything, the price-earnings ratio, and thus the valuations, of private companies show greater variability over the business cycle, though the limited number of observations makes inference tentative. This points to the greater perceived risk of private companies, which is consistent with the view that bank credit risk is riskier than speculative grade bonds.

Perceived risk appears to be great for private firms, and holders of private firm debt. Yet if we look at the charge-off history of banks, we see comparable charge-off volatility between a bank portfolio and a portfolio of speculative grade loans. Bank charge-offs include things like credit cards and real estate, which have significantly different averages and cyclical properties than the private firm portfolios. Nonetheless,

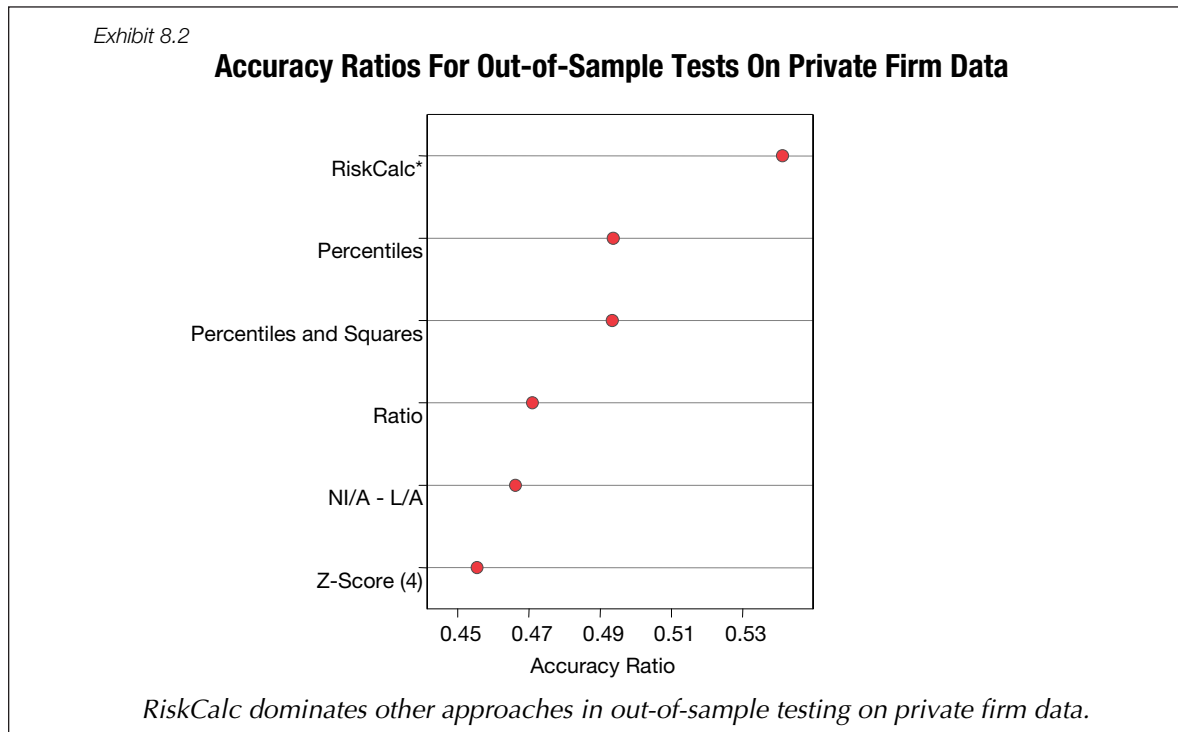
Exhibit 8.1



the resemblance between the two series is suggestive that perhaps, on average, bank loans and speculative grade bonds have similar risk.

The subset of a bank's portfolio that is focused purely on middle market loans is probably less risky than the bank as a whole, as consumer loans and real estate lending have traditionally generated the greatest losses and volatility over the past 25 years. This suggests that middle market lending is less risky than the speculative grade universe, which is consistent with the loss rate estimation data given in section 7.

Section VIII: Model Validation



Prove all things; hold fast that which is good.

~ Thessalonians 5:21

The previous sections have addressed what drives RiskCalc, how it works, how we chose the variables it uses, and why the functional form is what it is. What remains to be addressed is the quality of the model's output: How well does it differentiate good and bad credits? Moody's has been active in developing and applying these tools to quantitative models, and their discussion is elaborated in *Benchmarking Quantitative Default Risk Models: A Validation Methodology*, (Sobehart, Keenan, and Stein, 2000a).

Usually, this information on statistical power is presented solely in graphical form, via what we call cumulative accuracy profiles, or CAP plots, shown below within the context of defining the AR ratio in Exhibit 8.1, and were earlier discussed as power curves in section 4 and appendix 4A. To review, these graphs show the number of defaults excluded given a percentage of the sample excluded. A nice property of these graphs is that they allow one to observe the relative power of one model against another at various cutoff levels. At the low end (to the left), they allow a lender to examine the effect on the risk of their portfolio if they excluded their worst 10%, while in the midrange they allow one to monitor what would happen if one excludes the worst 50%. Most lenders evaluating the usefulness of models are more interested in what would happen if they excluded 10% of their existing customer base vs. 50%, and so one tends to focus on the left half of the CAP plot.

48 While we criticize the Z-score, we should make clear this is not an attack on the competence of Altman or the value of his work. Clearly, he has been a pioneer in default research, and we think he is appropriately recognized worldwide as an expert in the field.

Risk Magazine in 1999 (Shimko (1999)), the only default prediction piece was Altman et al's 1977 piece on the Zeta model-a proprietary extension of the Z-score - which demonstrates its singular credibility in the default prediction literature.

Yet, for all the popularity of Z-score, it is instructive how little the model has been used in practice. While bureau scores and leverage ratio guidelines are essential inputs to consumer or commercial decisions, Z-score never attained such a status. Why? What does this say about the relevance of quantitative models in commercial lending?

The main reason for this weak practical adoption rate is straightforward: the model does not work particularly well. More specifically, it is roughly equivalent in power to simple univariate ratio benchmarks

Exhibit 8.4

	Public		Private	
	5-year	1-year	5-year	1-year
Shumway	0.4255	0.6801	0.3358	0.4782
NI/A - L/A	0.4254	0.6873	0.3378	0.4662
Liab/Assets	0.3690	0.6194	0.3093	0.3986
NI/A	0.3636	0.6145	0.2762	0.4477
Z-score(4)	0.3587	0.6251	0.3250	0.4554

The improper linear model NI/A - L/A is a surprisingly powerful benchmark

such as liabilities/assets or net income/assets, and strictly dominated by something as simple as the additive combination of these two ratios.

This brings forth a profound point: real out-of-sample performance is what ultimately determines long-run acceptance and usage of a model. An intriguing model may gain popularity because of its elegance or ability to explain certain anecdotal events, but if it doesn't work in real time, over a lot of different companies, new users won't be converted and the model will simply fade away as initial proponents move on. There is a Darwinian dynamic to models. In this case, the 'fittest' that survived are simpler, rather than more complex alternatives.

Let us look more closely at the Z-score performance. Exhibit 8.3 shows the difference between ratios and their median values, one to five years prior to default. Note that indeed the Z-scores of defaulters vs. non-defaulters are impressively different. This sort of information is often presented as compelling evidence of the significant statistical power of Z-score. Yet, note that such statistical power is not unique to Z-score. In fact, a simple univariate ratio such as liabilities/assets shows a similar property.

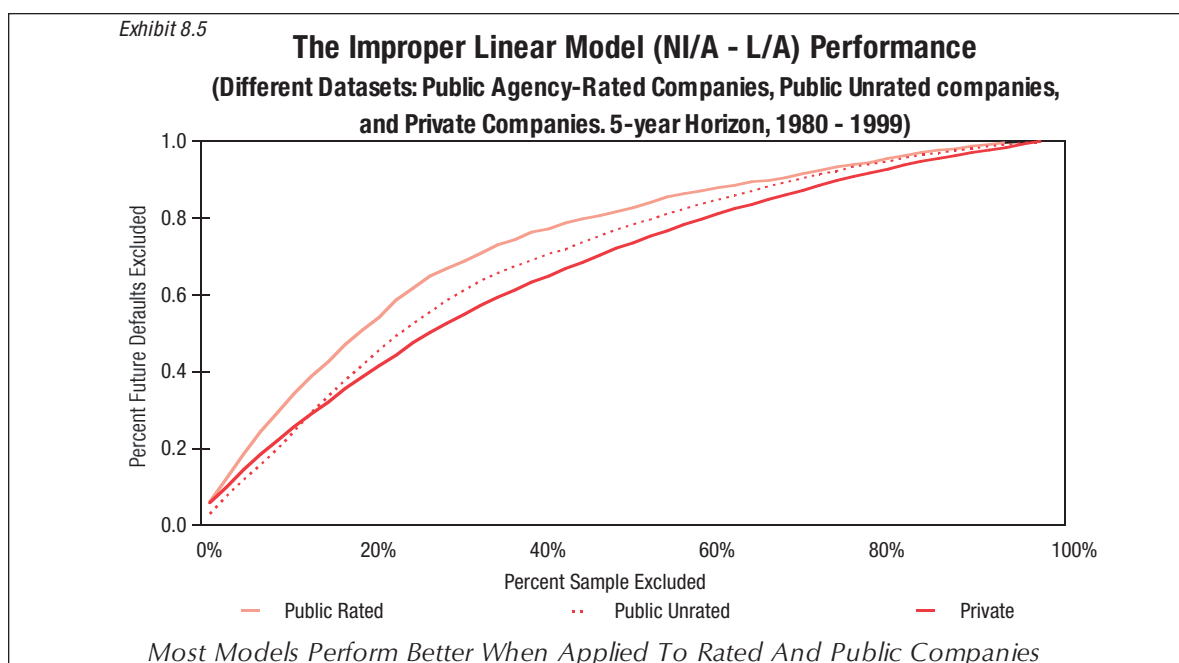
This highlights the important point that mean differences and trends in ratios between defaulters and non-defaulters are common and persistent for many risk metrics, including simple things like univariate ratios. The appearance of these trends does not imply that the model works in the sense of being better than simple alternatives, only that it works in the sense of being better than nothing. It is almost certain that a firm will show declining or negative income and lower and declining market equity prior to default, and thus virtually any risk metric's trend will provide an early warning of default.

The more pertinent question, therefore, is what portion of credits with the worst scores go bad and how consistent the projected expected default frequency is with the actual default frequency.

Testing The Alternatives

In the following tests, we will use the 4 variable Z-score, as it dominates the 5-variable version, as shown in the Appendix 8A.⁴⁹ We will also test Shumway's three variable model, which uses net income/asset, L/A, and the current ratio.⁵⁰ Shumway's model was estimated using a hazard function approach, which was basically an extension of a probit model designed to capture the problems created by using different statements for the same company (i.e., the nonindependence of multiple statements for the same company). Finally, we will test the simple univariate ratios Net Income/Total Assets (NI/A) and Total

49 Z-score (4 variable): $6.56 \cdot WC/A + 3.26 \cdot RE/A + 6.72 \cdot EBIT/A + 1.05 \cdot NW/L$
 Z-score (5 variable): $.717 \cdot WC/A + .847 \cdot RE/A + 3.107 \cdot EBIT/A + .420 \cdot NW/L + 0.998 \cdot S/A$
 50 Shumway = $-6.307 \cdot NI/A + 4.068 \cdot L/A - 0.158 \cdot CA/CL$, see Shumway (1999)



Liabilities/Total Assets (L/A), as well as the simple improper linear model, NI/A - L/A. This is called an 'improper' linear model because its coefficients are not estimated statistically, but instead are based on intuition about how these two ratios relate to default. They are given equal weighting, as opposed to Shumway's more 'proper' weightings that come from statistical analysis.

Most importantly, Exhibit 8.4 shows that the Z-score is dominated by the improper linear model NI/A - L/A. Z-score's comparison to simple univariate ratios is ambiguous, better in some cases, worse in others. In choosing between the Z-score and simple univariate ratios, one would gain transparency by using the univariate ratios and in general not lose statistical power.

Z-score's poor performance has nothing to do with any errors made in Altman's modeling process, but rather is primarily a function of the number of defaults used in the construction of this model. While it is not obvious exactly which sample was used in the estimation of the 4 variable Z-score model we tested, 33 defaults were used in the 1968 study and 53 defaults were in the 1977 Zeta extension. This is simply too few defaults to estimate a model with 4 parameters. Usually, one needs at least a ratio of 15 to 1 for observations/coefficients to outperform an improper linear model. That is, unit weights assigned coefficients based on ex ante theory often outperforms properly estimated statistical models (Schmidt (1971) or Claudy (1972)).

Also note that for both public and private firms, over a 1- and 5-year horizon, the improper linear model and Shumway's model demonstrate quite similar accuracy ratios. This highlights the 'flat maximum' issue explained in section 6 - that models not optimally specified, but which use the best normalized inputs, are often just as good as statistically estimated models, especially when they are positively correlated (as NI/A and - L/A are).

When Altman proposed Z-score, it was presented as a superior alternative to a seemingly straw-man alternative: a multivariate model should outperform a univariate model. Clearly, such simple models are not so easy to beat, and this highlights the problems associated with limited data.

The above accuracy ratios also show a distinct power differential between the 1- and 5-year forecast horizon (i.e. the CAP plots are more northwesterly at the 1-year horizon). It is true for any model that predicting 1 year ahead is 'easier' than predicting 5 years ahead. Yet 1-year prediction is not as useful as a 5-year prediction, since very few loans go default within their first year. As mentioned earlier, 1-year rates are therefore restricted in usefulness to monitoring and work-out, rather than for pricing, decisioning new credits, or evaluating a collateralized debt obligation (CDO). While both horizons are interesting - and indeed RiskCalc was separately optimized on the 1- and 5-year horizon - it is important when comparing

51 As long as these omissions are unbiased, they should not affect the comparisons of models.

CAP plots or Accuracy Ratios that the default horizon for each is the same. Otherwise, virtually any model's 1-year prediction will dominate another model's 5-year prediction.

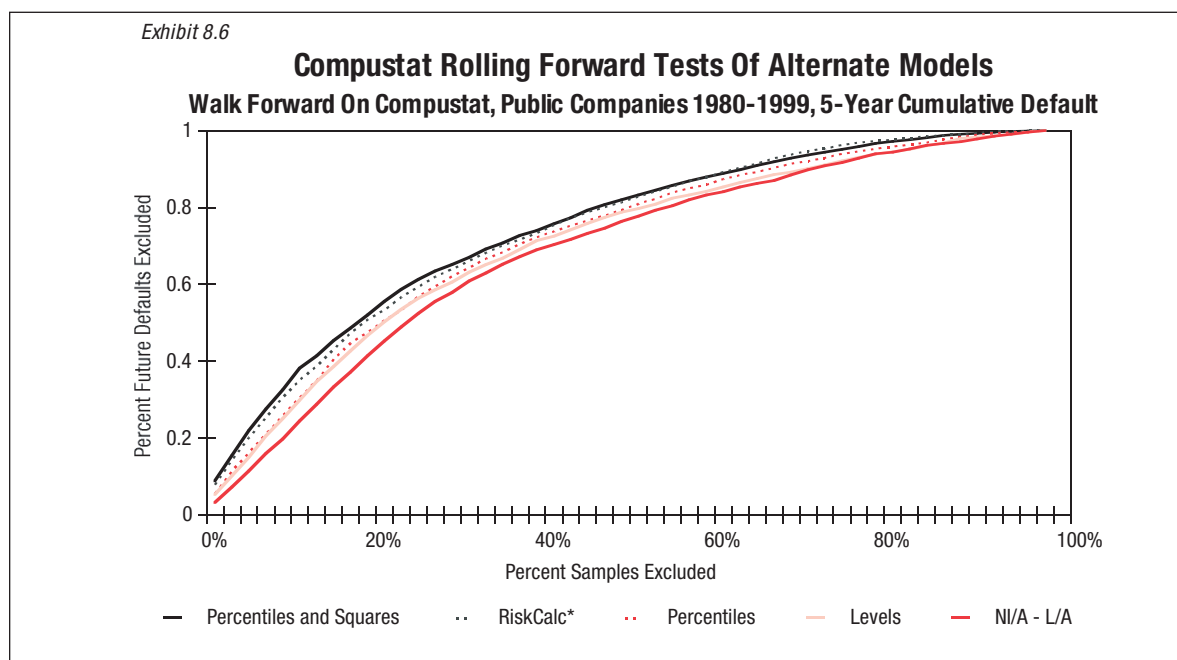
Power degradation also occurs over different sets of firms. Exhibit 8.5 shows the improper model NI/A - L/A as applied to rated companies, public companies that are unrated, and private companies, and this holds true for every model we examined. There is continued degradation of statistical power over these three samples, and this is true for all models. Part of this can be explained by the lower sample default rates for the unrated and private companies. A lower sample default rate makes prediction more difficult irrespective of the true power of the model, a statistical fact first pointed out in this context by Zmijewski (1984).

Yet, some of this loss of power can also be explained by the fact that defaults are measured better for rated firms than unrated public firms, and for public firms vs. private firms. More of the 'goods' in the unrated universes are mislabeled, but unfortunately we do not know which ones.⁵¹ Further, accounting statements are less noisy for rated companies, and this is reflected in the far greater number of audited statements for public companies as opposed to private companies. Adding noise to the input variables clearly weakens the power of any model to predict from them.

This highlights the importance of making inferences between models only when comparing performance on identical samples. The same model will show different performance over different prediction horizons (e.g., 1 vs. 5 years) and different universes (public vs. private). All of the graphs and tables that compare models use identical, intersecting datasets and the same default horizon for this reason.

RISKCALC TESTS

The presentation of proprietary model test results are often, and sometimes justifiably, viewed skeptically by potential consumers and competitors, as there are many minor adjustments to database creation and the modeling of alternatives that can significantly alter performance. This is why each political party has its own pollsters. Yet by publicly presenting these results we generate useful expectations for users as to the discriminatory power of the RiskCalc. The main point we wish to convey is that RiskCalc significantly outperforms the academic benchmarks, as indicated by its relative performance to the surprisingly powerful improper linear model NI/A - L/A. Most importantly, the real advantage of RiskCalc is its estimation on our private dataset, and so by documenting its performance vis-à-vis several attractive alternative models estimated on the same proprietary database, we are comparing RiskCalc to models that someone without access to our dataset would have great difficulty creating. Given our number of defaults the variously estimated commercial loan models are in general not knife-edged, and thus the presentation of comparative results here strongly suggest that while RiskCalc is probably not the ultimate optimal model for predicting default, it is likely that it is very close to that unknown optimal model. When combined with the other characteristics of the model (driving factors, mapping into default rates and Moody's rating), RiskCalc is simply difficult to outperform.



Accuracy Ratios in the Compustat Walk-Forward Tests, 1989-99

	5-year	1-year
Percentiles and their Squares	0.5107	0.7853
RiskCalc*	0.5024	0.7648
Percentiles	0.4586	0.7678
Ratio Level	0.4342	0.6585
NI/A - L/A	0.3931	0.6477

Other than NI/A - L/A, all models used the same 10 input ratios estimated within a probit model. Percentiles used the percentile rank of the various ratios, percentiles and squares used the percentile rank and the square of this input, and ratio level used the ratio levels truncated at the 2% and 98% points. RiskCalc used the univariate transformation process described in the document. Accuracy Ratios are defined above in this section.*

Transformations from pure ratio levels enhance model performance.

Some of the tests are listed below, and all the test results, accuracy ratios and information entropy ratios, are listed in the appendix, as a convenient summary for any applied researchers seeking to compare performance to our work.

Walk-Forward Tests On Compustat

For the first test of the RiskCalc algorithm, we used a rolling-forward approach. This was constructed as follows. Using Compustat data, we estimate the model up to 12/31/89, ignoring not only financial statement data but default information subsequent to 12/31/89. Using this model, we generated forecasts for statement dates in the year 1990. We estimated the model again up through 12/31/90 and generated forecasts for 1991, and so on, rolling forward to 1999. This creates a set of forecasts based on a model whose estimation 'rolls through time,' and realistically presents the performance of the model as it would be applied in real time, since a model usually is re-estimated based on new information.

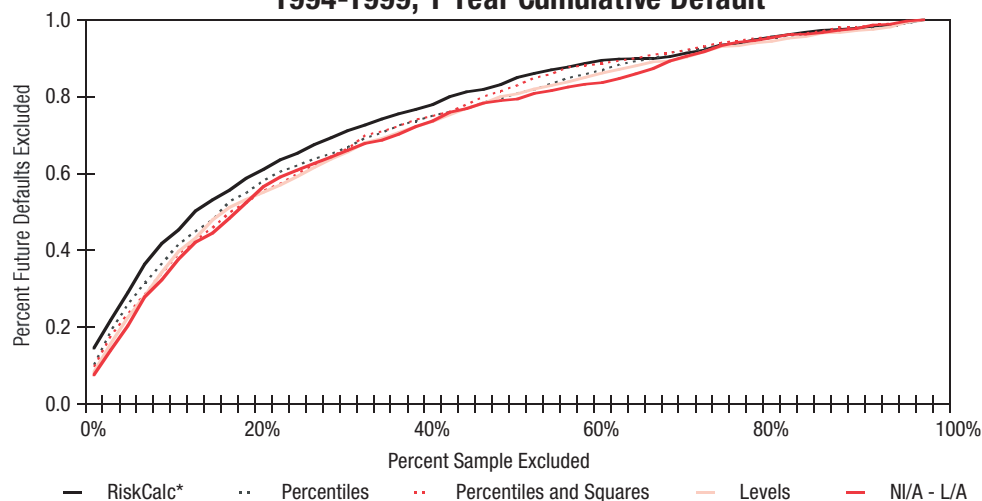
Our public benchmark is the improper linear model, NI/A - L/A. Though the improper linear model is simple, it is not a 'straw-man' alternative, as it was shown to dominate the well-known Z-score model and approximate the performance of the more recently estimated Shumway model (which among academic models is as good as they get). Thus, by the logic of transitivity, any model that outperforms the improper linear model necessarily outperforms all the Z-scores and Shumway's model.

The approach of the new RiskCalc is to use the univariate ratios as the transformations of the independent variables - what we will call RiskCalc* in the tables. It is denoted as RiskCalc* because it refers to the algorithm, as opposed to the final model. For the rolling forward approach, we re-estimate the transforms each year, and then re-estimate the coefficients on these transforms using a probit model. Missing values for independent variables use the mean of these values, that is, the mean of the transform. Unreported tests show this does not significantly change the results, though the power of models using missing data is always slightly lower.

There are an infinite number of alternative models to RiskCalc, and we can never demonstrate performance against all of them. We will, however, compare RiskCalc's performance with three alternatives that are

Exhibit 8.7

Out Of Sample Performance Of Approach On CRD, Private Companies 1994-1999, 1 Year Cumulative Default



The New Approach Significantly Improves Model Performance Vis-A-Vis Other Approaches

Exhibit 8.8

Out-of-Sample Tests on CRD, Accuracy Ratios

	5-year	1-year
RiskCalc*	0.3689	0.5412
Percentiles	0.3657	0.4937
Ratio Level	0.3616	0.4710
Percentiles and their Squares	0.3470	0.4934
NI/A - L/A	0.3378	0.4662

All models were estimated prior to 1995 on Compustat, which makes these out-of-sample tests. Definitions of models are given in Exhibit 8.6 above.

The RiskCalc method is more robust than other transformation methods.

most appealing. All these alternatives use a probit estimation technique and the same set of 10 input ratios. The first uses the truncated levels of the same 10 input variables, as this is most like the approach taken by the essentially linear models in the academic literature. The second approach is nonparametric in that it uses the percentiles of the input values. That is, instead of using an interest coverage ratio of 3.4, we used the number 0.42, which represents its percentile within our private firm database. In the percentiles approach, the ratios were transformed into numbers between 0 and 1 based on the distribution of ratios in Compustat.

Lastly, as we know that the relation between ratios and defaults is often nonlinear, we used the polynomial expansion of percentiles: percentiles and their squares. This model has twice as many inputs as using just the percentiles alone, as it requires coefficients not only for the percentile term (the linear part), but for the percentile's square term (the nonlinear part). It is a more analogous to RiskCalc* in that it allows for a great deal of nonlinearity, but unlike RiskCalc* does not restrict the nonlinearity to closely match its univariate properties. The percentile approaches and RiskCalc* all use 'nonparametric' transformations. All these models used the rolling forward estimation process.

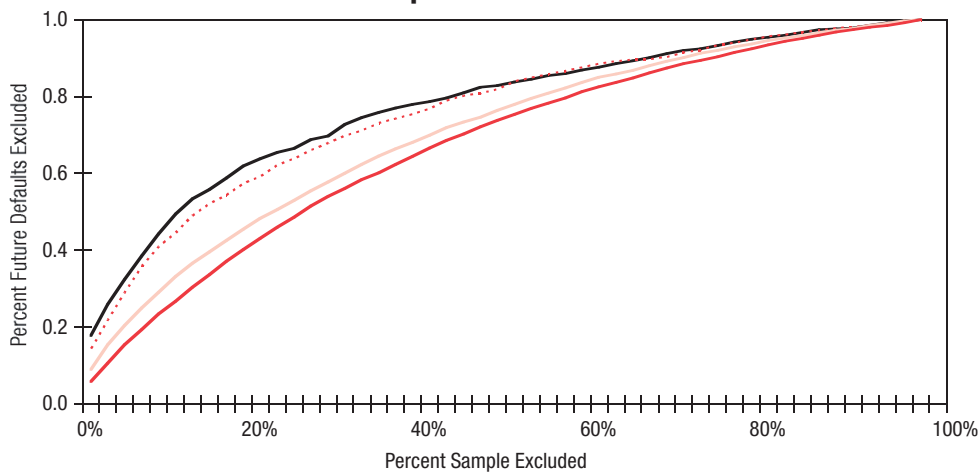
Exhibit 8.6 shows the performance of these 4 approaches, plus the improper linear model, over the period 1990-99.

In Exhibit 8.6 we see that the percentiles and their squares approach dominates the other approaches, with RiskCalc* in second place. The chart illustrates the CAP plot for the 5 year horizon, and the table documents the AR calculated for both the 5 and 1 year horizon. Percentiles alone do slightly better than RiskCalc* at the 1-year horizon, but significantly worse at the more important longer horizon levels. Levels, even truncated levels (which far outperform untruncated levels), do the worst of the statistically optimized models. This is because even truncated levels are too asymmetric and fat-tailed. A normalization of ratios is required so that a model does not place inordinate weight on outliers.

The dominance of percentiles and their squares and RiskCalc* suggests that capturing the nonlinearity is important. Lastly, all nonparametric approaches show significant improvement over the improper linear model, which itself dominates the simple improper linear model, NI/A - L/A (which itself was shown to

Exhibit 8.9

In vs. Out-of-Sample Performance of RiskCalc



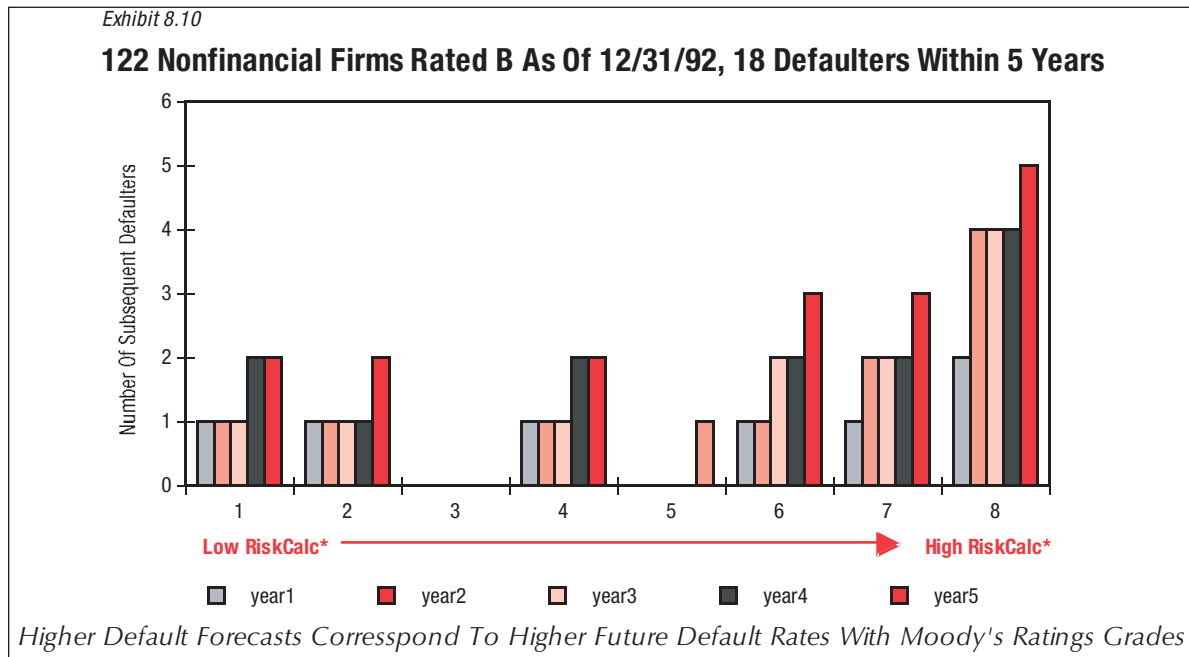
Final in-sample estimation of RiskCalc implies that out-of-sample tests understate true performance

dominate Z-score, which within the academic literature surveys has dominated or equaled the performance of academic models). The great number of defaults we can estimate upon allows us, unlike previous academic researchers, to significantly improve upon the simple improper linear model.

Out-of-Sample Tests On The CRD

We applied these same models to the private firm data. To make the tests as out-of-sample as possible, we estimated the models up to 12/31/95 on Compustat. As most CRD defaults are subsequent to 1995, this is both out-of-universe and out-of-time, and, therefore, a good test of the robustness of these approaches.

In Exhibits 8.7 and 8.8, we see that in this environment, percentiles outperform percentiles and their squares, and both beat the use of levels within a probit estimation. This highlights the dangers of polynomial expansions within models. Polynomial expansions allow one to capture in-sample properties extremely well, but they are more susceptible to 'over-fitting' as slight differences in sample correlations and ranges can accentuate the effect of various nonlinearities. Polynomial expansions of percentiles are more powerful in-sample, if only because they use more degrees of freedom, but less robust in that out-of-sample performance on the CRD is distinctly less impressive.



Using accuracy ratios over the 1 and 5 year horizon, Exhibit 8.8 shows that RiskCalc* dominates these other approaches in out-of-sample performance. This is also displayed in the CAP plot of the 5-year horizon performance in Exhibit 8.7. RiskCalc* is clearly the most powerful under the AR metric at the 1-year horizon, and by a slight degree the most powerful over the longer 5-year horizon. In contrast, Exhibit 8.6 showed that while comparable to the other approaches in the Compustat walk-forward tests, it was not dominant. This suggests that RiskCalc* is not only highly powerful, but more importantly a robust estimation method, which is highly important given the still limited number of defaults in our dataset (relative to consumer modelers, not other commercial modelers), and also the track record of previous models, such as Z-score.

RiskCalc* captures the benefits of nonlinearities while achieving optimal robustness. RiskCalc* is not the 'best fitting' model, as within sample that honor goes to an approach that uses percentiles and their squares (i.e., transformations of input ratios to percentiles, and these percentiles squared). You can generate a higher

Exhibit 8.11

Correlation Matrix of Inputs Used in RiskCalc. Ratios vs. Transformed Ratios. Private Firms, 1989-99

Transformed Ratios										
	Assets	Cash/A	Int. Cov.	Inv/COGS	L/A	NI/A	NI/A gr	Quick	RE/A	SalesGr
Assets	1.00									
Cash/Assets	-0.15	1.00								
Interest Coverage	0.09	0.21	1.00							
Inv/COGS	0.06	0.08	0.03	1.00						
Liab./Assets	-0.11	0.33	0.46	-0.01	1.00					
Net Income/Assets	0.11	0.00	0.78	-0.02	0.29	1.00				
NI/A growth	0.17	-0.12	0.35	0.00	0.12	0.52	1.00			
Quick	-0.15	0.51	0.32	0.05	0.51	0.21	0.09	1.00		
Retained Earnings/Assets	0.12	0.01	0.46	-0.03	0.36	0.52	0.26	0.17	1.00	
Sales Growth	0.12	-0.13	0.24	-0.02	0.07	0.36	0.38	0.04	0.26	1.00
Raw Ratios										
	Assets	Cash/A	Int. Cov.	Inv/COGS	L/A	NI/A	NI/A gr	Quick	RE/A	SalesGr
Assets	1.00									
Cash/Assets	-0.15	1.00								
Interest Coverage	0.02	-0.02	1.00							
Inv/COGS	-0.06	-0.04	-0.01	1.00						
Liab./Assets	0.06	-0.35	-0.14	-0.07	1.00					
Net Income/Assets	0.13	-0.16	0.41	0.00	-0.36	1.00				
NI/A growth	0.02	0.05	0.14	0.00	-0.11	0.43	1.00			
Quick	-0.13	0.70	0.00	-0.02	-0.47	0.01	0.07	1.00		
Retained Earnings/Assets	0.13	-0.15	0.27	0.03	-0.41	0.70	0.12	0.01	1.00	
Sales Growth	-0.06	0.12	0.00	0.00	-0.09	0.00	0.15	0.08	0.01	1.00

Correlation among the transformed variables is generally small and positive.

likelihood ratio (i.e., the correlate of the R2 in ordinary least squares), with the percentiles and their squares approach in-sample. Nonetheless, RiskCalc's use of univariate default frequency transformations is quite close to optimal within sample, and most importantly appears significantly more robust out-of-sample.

Final In-Sample CAP Plots

One of the main benefits of the private firm database (our CRD) is its ability not only to test, but to estimate a model on our targeted universe of private firms. Therefore, as mentioned above, the final model is estimated upon the CRD data. The improvement from estimating on the CRD is suggested (but probably exaggerated) from examining the improvement of the in-sample to out-of-sample performance. Exhibit 8.9 shows the relative performance of the final model used in RiskCalc vs. RiskCalc* (the univariate transform approach that was estimated on Compustat up to 12/95). It is impossible to directly test the final CRD estimated model out-of-sample, as we do not have sufficient time periods to generate the rolling forecasts. However, the per-

Exhibit 8.12

Parameter Stability of the RiskCalc* Algorithm, Compustat Data-

Estimation Period	1-Year		5-Year	
	1980-90	1991-99	1980-90	1991-99
Assets	-2.62	9.54	8.1	18.6
Cash/Assets	13.02	6.11	7.6	3.5
Interest Coverage	5.82	5.54	5.6	2.7
Inv/COGS	4.18	5.52	5.7	8.3
L/A	5.10	5.20	4.0	2.7
NI/A	5.12	6.57	-2.4	3.4
NI Growth	3.83	6.99	2.7	1.7
Quick	3.80	5.59	2.0	5.0
RE/A	7.22	4.33	6.4	5.2
Sales Gr.	4.60	2.56	4.7	3.1

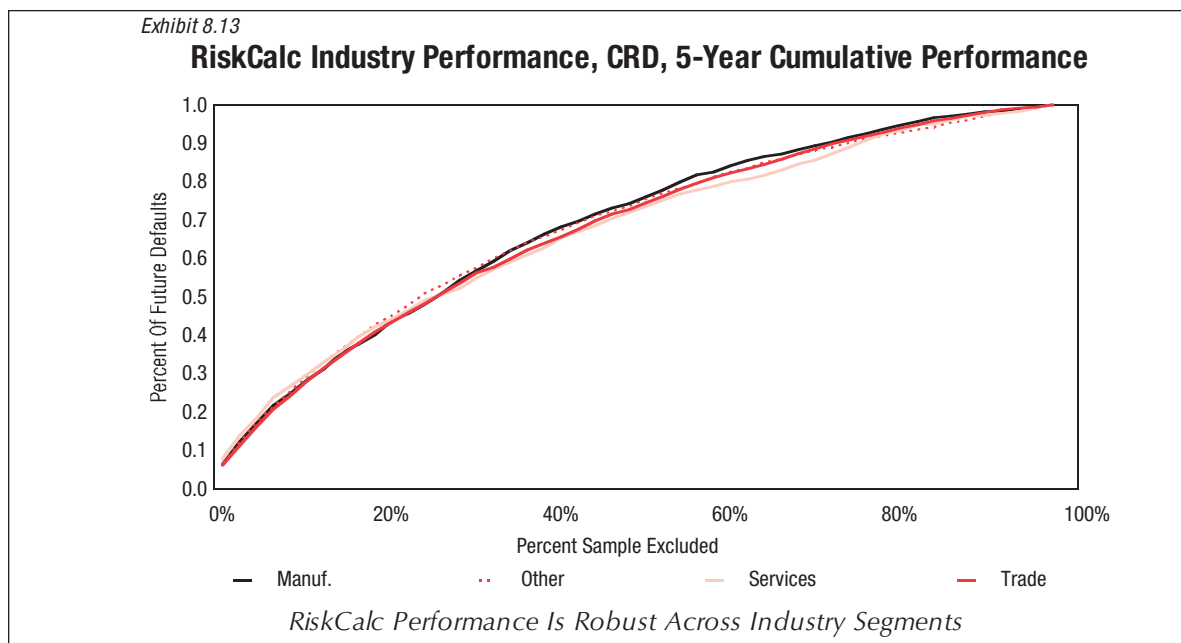
Coefficients are relatively stable over time.

52 In Moody's public model NI/A is included, as EBIT/interest was excluded, and in the context of the variables used in that model, which include market data not used here, it appeared quite robust.

formance of RiskCalc* across samples and its performance vis-à-vis alternatives within the rolling forecasts on Compustat both give one reasonable assurance that the result is both powerful and robust.

Miller Tests

Finally, we tested RiskCalc using the Miller Risk Advisors approach (Risk Magazine, 1998). In order to demonstrate the power of KMV's public firm model, Ross Miller took firms rated B as of December 1990, and looked ahead to see if KMV's model added information within this class of borrowers. That is, he ranked the firms from high to low within this rating to see where the defaults accumulated. Updating this result to 1992, we did the same thing. With only 18 defaulters, this test is not statistically significant, yet it is instructive, and certainly consistent with the assertion that RiskCalc is a powerful tool for assessing default risk.



RiskCalc, which excludes useful equity information, was found to add information to broad ratings categories even for public firms. RiskCalc shows that lower ranked firms within the B class tended to default more frequently than higher ranked firms over the subsequent 1 through 5 years. We estimated the model through 12/92, so that no default information or financial statements subsequent to 9/92 were used in the construction of RiskCalc for this test (the transformations and the coefficients excluded this information). This doesn't imply that ratings are flawed, just that even within small risk bands RiskCalc can add granularity to risk groupings.

Parameter Stability and Input Correlation

Description	#Defaults	CIER	AR	Model
Compustat, 5 year, Z-score	5413	0.0485 0.0328	0.3574 0.2972	Z-score(4) Z-score(5)
Compustat, 1 year, Z-score	835	0.1183 0.0756	0.6248 0.5132	Z-score(4) Z-score(5)
CRD, 5 year, Z-score	4888	0.0338 0.0231	0.3250 0.2525	Z-score(4) Z-score(5)
CRD, 5 year, Z-score	603	0.0554 0.0309	0.4554 0.3314	Z-score(4) Z-score(5)

Z-score (4 variable): $6.56 * \text{Working Capital/Assets} + 3.26 * \text{Retained Earnings/Assets} + 6.72 * \text{EBIT/Assets} + 1.05 * \text{Net Worth/Liabilities}$
Z-score (5 variable): $.717 * \text{Working Capital/Assets} + .847 * \text{Retained Earnings/Assets} + 3.107 * \text{EBIT/Assets} + .420 * \text{Net Worth/Liabilities} + 0.998 * \text{Sales/Assets}$

The 4-variable Z-score outperformed the 5-variable Z-score on both datasets on both horizons

Description	#Defaults	CIER	AR	Model
Compustat, 1 year, simple models	834	0.1453	0.6873	NI/A - L/A
		0.1432	0.6801	Shumway
		0.1183	0.6251	Z-score(4)
		0.1195	0.6194	Liab/Assets
		0.1188	0.6145	NI/A
Compustat, 5 year, simple models	5399	0.0674	0.4255	Shumway
		0.0660	0.4254	NI/A - L/A
		0.0500	0.3690	Liab/Assets
		0.0482	0.3636	NI/A
		0.0487	0.3587	Z-score(4)
CRD, 1 year, simple models	603	0.0668	0.4782	Shumway
		0.0598	0.4662	NI/A - L/A
		0.0554	0.4554	Z-score(4)
		0.0592	0.4477	NI/A
		0.0414	0.3986	Liab/Assets
CRD, 5 year, simple models	4888	0.0359	0.3378	NI/A - L/A
		0.0366	0.3358	Shumway
		0.0338	0.3250	Z-score(4)
		0.0303	0.3093	Liab/Assets
		0.0279	0.2762	NI/A
<i>Shumway=-6.307*Net Income/Assets+4.068*Liabilities/Assets-0.158*Current Assets/Current Liabilities</i>				
<i>For both public and private firms, at the 1 and 5 year horizon, Shumway's model and the simple improper linear model Net Income/Assets - Liabilities/Assets dominated the univariate ratios and Z-score. Z-score's performance is not significantly different than that of the univariate ratios.</i>				
Description	#Defaults	CIER	AR	Model
Compustat, 1 year, alternative estimations	465	0.2149	0.7853	Percentiles and their squares
		0.2007	0.7678	Percentiles
		0.1912	0.7648	RiskCalc*
		0.1356	0.6585	Ratio Level
		0.1271	0.6477	NI/A - L/A
Compustat, 5 year, alternative estimations	2403	0.0931	0.5107	Percentiles and their squares
		0.0893	0.5024	RiskCalc*
		0.0719	0.4586	Percentiles
		0.0649	0.4342	Ratio Level
		0.0536	0.3931	NI/A - L/A
<i>Other than NI/A - L/A, all models used the same 10 input ratios estimated within a probit model. Percentiles used the percentile rank of the various ratios, percentiles and squares used the percentile rank and the square of this input, and ratio level used the ratio levels truncated at the 2% and 98% points. RiskCalc* used the univariate transformation process described in the document. All models were re-estimated annually, up to the year prior to the target forecast year (i.e., they were all estimated on a different sample than what they were forecast upon), and started in 12/31/89.</i>				
Description	#Defaults	CIER	AR	Model
CRD, 1 year, alternative estimations	603	0.0839	0.5412	RiskCalc*
		0.0662	0.4937	Percentiles
		0.0634	0.4934	Percentiles and squares
		0.0584	0.4710	Ratio Levels
		0.0598	0.4662	NI/A - L/A
CRD, 5 year, alternative estimations	4888	0.0430	0.3689	RiskCalc*
		0.0431	0.3657	Percentiles
		0.0422	0.3616	Ratio levels
		0.0376	0.3470	Percentiles and squares
		0.0359	0.3378	NI/A - L/A
<i>Other than NI/A - L/A, all models used the same 10 input ratios estimated within a probit model. Percentiles used the percentile rank of the various ratios, percentiles and squares used the percentile rank and the square of this input, and ratio level used the ratio levels truncated at the 2% and 98% points. RiskCalc* used the univariate transformation process described in the document. All models were estimated prior to 1995 on Compustat, which makes these out-of-sample tests.</i>				
Description	#Defaults	CIER	AR	Model
CRD, 5 year, in-sample comparison	5346	0.0620	0.4212	RiskCalc (in-sample)
		0.0413	0.3599	RiskCalc*
CRD, 1 year, in-sample comparison	686	0.0949	0.5554	RiskCalc (in-sample)
		0.0819	0.5279	RiskCalc*
<i>The final in-sample estimation provides for an even greater performance lift, which is suggested, but probably exaggerated, by the in-sample performance of the ultimate RiskCalc model</i>				

The correlation of the inputs is interesting to many users of the model. Our transformation approach in general increases the correlations of the inputs, but to modest levels that still allow for robust estimation. The average correlation between the transformed inputs is only 0.2. For the raw inputs, the correlation is

$$IER_A = \frac{IE_0 - IE_A}{IE_A}$$

$$IE_A = -\frac{1}{n} \sum_{i=1}^n p_d^{iA} \ln(p_d^{iA}) + (1-p_d^{iA}) \ln(1-p_d^{iA})$$

$$IE_0 = -\frac{1}{n} \sum_{i=1}^n \bar{p}_d \ln(\bar{p}_d) + (1-\bar{p}_d) \ln(1-\bar{p}_d)$$

p_d^{iA} = probability of default for bucket i using model A

\bar{p}_d = average sample default rate

p_d^{iA} is derived by ranking the set of firms by the model in question, then grouping these into 50 percentile buckets, and examining the year-ahead default rate. The division by the sample IE is to normalize each number to make it comparable across groupings (higher default rates have higher IE irrespective of model power). As IEA is concave in p_d^{iA} , firms with greater p_d^{iA} variance will generate lower IE scores, and thus higher information entropy ratios. This ratio and its application is described more fully in Keenan and Sobehart (1999)

an almost insignificant 0.02. The transformations, therefore, increase the correlations. This makes sense, since the transformations are to their univariate probabilities of default (smoothed), and so a weak firm will generally have a low NI/A ratio and a high L/A ratio. Both ratios will be mapped into a 'high' transform, making them positively correlated, while the levels will be negatively correlated. In this example, the L/A and NI/A are negatively correlated at -0.36 in levels, while their transforms are positively correlated, 0.29.

Correlation Over Time

The relation of ratios by Moody's grade suggests that ratios and risk are consistent over time. Higher leverage, lower profitability, and smaller firm size all imply higher risk today just as yesterday. Subtle changes in these inputs suggest one should be monitoring and adjusting these relationships over time. Yet, defaults experience is cyclical, and in the past 20 years, the credit cycle bottomed-out approximately in 1991.

What we would like to see is that the model is relatively consistent over time in the face of new information, as significant changes over the cycle would imply that the model is less robust than we would like. We examine the changing parameters of the model over time in Exhibit 8.12. The transforms of the various inputs make these comparisons a little less straightforward, but they are still informative. The forecasts from the rolling Compustat forecasts using the RiskCalc algorithm are all above 0.9, and we see that over time the coefficients are generally similar, with a few significant changes in magnitude though none of sign. Note that for one of the Compustat estimations (5-year, 1980-90) the sign on NI/A is 'wrong', and this is because in this sample it is very highly correlated with EBIT/interest.⁵²

Industry Performance

RiskCalc does not treat industries differently, yet this does not mean that RiskCalc is not robust across industries. In fact, the differences in default rates and industry ratios is not a coincidence, but instead information that is directly related. By estimating upon the entire dataset-excluding finance, insurance and real estate-the model is better estimated than if we separately estimated models upon each individual sector. Nonetheless, going forward as we gather more defaults one of our priorities is to modify the model for various industry classifications.

To see how the model performs across industries, we generated a CAP plot for four major groupings: trade (wholesale and retail), manufacturing, services (excluding hospitals), and other (which includes everything else). The different default rates for the various groupings makes strict comparison of the model between groups not straightforward (i.e., these are by definition not intersecting samples), yet the results are informative nonetheless. We can see from Exhibit 8.13 that RiskCalc performs best in the manufacturing and 'other' classifications, as compared to trade and services. The difference, however, is slight.

Appendix 8A: Accuracy Ratios And Conditional Entropy Ratios

Presented below are the accuracy ratios and conditional entropy ratios for selected groupings of models, utilizing different time horizons and different datasets. 'CRD' refers to the Moody's proprietary private firm database, which includes information on 30,000 firms, of which 1,502 are recorded as having defaulted over the years 1988-99, although 90% of the observations are subsequent to 1996. 'Compustat' refers to the public firms database, which covers 1,400 defaulting firms over the years 1980 - 99. All the sets used identical records for the comparison (i.e., statistics are only valid for comparison within each subgrouping). CIER refers to the Conditional Information Entropy Ratio, and is defined in Appendix 8B. AR refers to Accuracy Ratios, which are defined in section 7. RiskCalc* refers to the out-of-sample performance of the RiskCalc* algorithm, which is not identical to the ultimate RiskCalc algorithm used in our final model.

Appendix 8B : Information Entropy Ratios

The formulas for the information entropy ratios (IER) are presented below. In this case, IEA is the information entropy for model A, and IE_0 is the information entropy for the entire sample.

Section IX: Conclusion

Default estimation is directly and essentially related to such issues as loan decisioning, pricing, investor and regulatory transparency, provisioning, securitization, and incentive compensation.

Are private firm loans similar to Ba- or to B-rated firms? This is not a simple question. But over a five-year horizon, these rating grades have an annualized difference of 150 basis points in expected loss - a distinction with a difference. In such a case, inefficiencies arise as various parties take either too much or too little risk.

Clearly, each loan that is mispriced, or mistakenly granted or declined, represents a lost opportunity. It is difficult in this context to underestimate the importance of better commercial loan default estimation.

Validated credit scoring models are useful for all lenders, even those with well-established credit cultures. Like the character who was surprised to find he had always spoken prose, many lenders are surprised to know they are always using a model; the question going forward is whether the model should be implicit or explicit. Unarticulated models, especially in environments that do not warehouse data, are difficult to criticize. The downside to this cozy state of affairs is significant in areas where quantitative tools really work. The main problem facing a credit analyst is how to integrate all the financial statement information given to her. There are hundreds of potentially relevant inputs and thousands of permutations of this data. Endless 'elevator analysis' can explain why certain ratios went up as others went down, but for many the primary purpose of analysis is to predict, not simply explain. To that end, a succinct and informative summary of this information is desired. This takes risk management out of the role of audit, where one explains why certain portfolios underperformed, to the line where decisions affecting the future revenues of the company are made.

RiskCalc helps directly answer questions such as whether the current spread on loans is sufficient to add shareholder value and plays an essential role in determining whether bank loans should be kept on or off balance sheet and an appropriate capital attribution.

It has been said that a good business strategy is to build those components that will give you a competitive advantage and buy everything else. In this case, internal development of default models is disadvantaged relative to RiskCalc because of Moody's large dataset of private and public firm defaults, as well as the significant amount of resources we can allocate to this problem. The latest upgrade in RiskCalc is sufficiently meaningful that nonadopters are potentially at a competitive disadvantage.

RiskCalc efficiently and accurately summarizes financial statement information. Other information is useful, and the final judgement is best left to an old-fashioned biological neural net. That said, the distillation of the balance sheet and income statement as they pertain to default, is best suited for a statistical model, which has the added benefit of producing a truly comparable measure of commercial credit quality between institutions.

References

- Antonov, Ivo, 2000, Crafting a Market Landscape, *The Journal of Lending & Credit Risk Management*, February.
- Alpert, W. and Raiffa, H. 1969, A Progress Report on the Training of Probability Assessors, Unpublished manuscript.
- Altman, E. I., 1968, Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy, *Journal of Finance* 23.
- Altman, E. I., R. Haldeman, and P. Narayanan, 1977, ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporation, *Journal of Banking and Finance*, 29-55.
- Altman, E.I., 1993, *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting and Avoiding Distress*, (Wiley and Sons, New York).
- Altman, Edward I., and Saunders, Anthony, *Credit Risk Measurement: Developments over the Last 20 Years*.
- Altman, N.S., 1992, An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression, *The American Statistician*, 46, 175-85.
- Baestaens, Dirk-Emma, 1999, Credit Risk Strategies: The Road to Serfdom? *International Journal of Intelligent Systems in Accounting, Finance and Management* 8, 225-235.
- Banz, Rolf, 1981, The Relationship Between Return and Market Value of Common Stocks, *Journal of Financial Economics*, 9, 3-18.
- Barber, Brad, and Terrance Odean, 1999, The Courage of Misguided Convictions, *Financial Analysts Journal*, November/December, 1999.
- Barniv and Raveh, 1989, Identifying Financial Distress: A New Nonparametric Approach, *Journal of Business Finance and Accounting*, 361-383.
- Beaver, W., 1966, Financial Ratios as Predictors of Failure, *Journal of Accounting Research, Supplement on Empirical Research in Accounting*, pp. 71-111.
- Begley, J., J. Ming, and S. Watts, 1996, Bankruptcy Classification Errors in the 1980's: An Empirical Analysis of Altman's and Ohlson's Models, *Review of Accounting Studies* 1, 267-284.
- Benishay, H, 1973, Discussion of A Prediction of Business Failure Using Accounting Data, *Journal of Accounting Research, Supplement on Empirical Research in Accounting*, 180-182.
- Berger, Allen, R.J. Herring, and G.P. Szego, 1995, The Role of Capital in Financial Institutions, *Journal of Banking and Finance* June, 19:3-4.
- Birkes, David, and Yadolah Dodge, 1993, *Alternative Methods of Regression* (John Wiley & Sons, Inc. New York).
- Blum, M, 1974, Failing Company Discriminant Analysis, *Journal of Accounting Research*, Spring.
- Blume, Marshall, Felix Lim, and Craig MacKinlay, 1998, The Declining Credit Quality of U.S. Corporate Debt: Myth or Reality? *Journal of Finance* August, 1998.
- Boothe and Hutchinson, 1989, Distinguishing Between Failing and Growing Firms: A Note on the Use of Decomposition Measure Analysis, *Journal of Business Finance and Accounting*, 267-271.
- Caouette, John B, Edward Altman, and Paul Narayanan, 1998, *Managing credit risk: the next great financial challenge* (Wiley).
- Casey and Bartczak, 1984, Cash Flow - It's Not the Bottom Line, *Harvard Business Review*, July-August, 61-66.
- Chen, K., and T. Shimerda, 1981, An Empirical Analysis of Useful Financial Ratios, *Financial Management* Spring.
- Claudy, J. G., 1972, A Comparison of Five Variable Weighting Procedures, *Education and Psychological Measurement*, 32, 311-322.
- Dambolena and Khoury, 1980, Ratio Stability and Corporate Failure, *Journal of Finance*, 1017-1026.
- Dawes, Robyn, and B. Corrigan, 1974, Linear Models in Decision Making, *Psychological Bulletin*, 81, 95-106.

- Dawes, Robyn M, 1979, The Robust Beauty of Improper Linear Models in Decision Making, *American Psychologist*, 34, 571-582.
- Deakin, E, 1972, A Discriminant Analysis of Predictors of Business Failure, *Journal of Accounting Research* Spring, 167-179.
- Dev, S., 1974, Ratio Analysis and the Prediction of Company Failure in Ebits, Credits, Finance and Profits, ed. H.C. Edy and B.S. Yamey, Sweet and Maxwell: London, 61-74.
- Diamond, H., 1976, Pattern Recognition and the Detection of Corporate Failure, Ph. D. dissertation, New York University.
- Dichev, I., 1998, Is the Risk of Bankruptcy a Systematic Risk? *Journal of Finance*, forthcoming.
- Edmister, 1997, An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction, *Journal of Financial and Quantitative Analysis*, 1477-1493.
- Eisenbeis, R., Pitfalls in the Application of Discriminant Analysis in Business and Economics, *The Journal of Finance*.
- El Hennawy and Morris, 1983, The Significance of Base Year in Developing Failure Prediction Models, *Journal of Business Finance and Accounting*, 209-223.
- Fama, Eugene and Kenneth French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance* 47:2, June.
- Fitzpatrick, 1932, A Comparison of Ratios of Successful Industrial Enterprises with those of Failed Firms, *Certified Public Accountant*, 598-605, 656-662, 727-731.
- Foulke, R.A., 1968, *Practical Financial Statement Analysis* (McGraw-Hill, Hew York).
- Gombola, M.J., and J.E. Ketz, 1983. A Note on Cash Flow and Classification Patterns of Financial Ratios, *Accounting Review*, 105-114.
- Gupta, Y. P., R.P. Rao, and P.K. Bagchi, 1990, Linear Goal Programming as an Alternative to Multivariate Discriminant Analysis: A Note, *Journal of Business Finance and Accounting*.
- Hamer, M., 1983, Failure prediction: Sensitivity of classification Accuracy to Alternative Statistical Methods and Variable Sets, *Journal of Accounting and Public Policy*, 289-307.
- Hardle, W., 1990, *Applied Nonparametric Regression*. (Cambridge University Press, Cambridge, U.K.).
- Herrity, J., Keenan, S.C., Sobehart, J.R., Carty, L.V., Falkenstein, E., 1999, *Measuring Private Firm Default Risk*, Moody's Investors Service Special Comment, June.
- Hodrick, Robert, and Edward Prescott, 1997, Post-War U.S. Business Cycles: An Empirical Investigation, *Journal of Money, Credit and Banking* 29, 1-16.
- Kahneman, Daniel, Paul Slovic, and Amos Tversky, 1982, *Judgement Under Uncertainty* (University of Cambridge, New York).
- Katz, S., S. Linien, and B. Nelson, 1985, Stock Market Behavior Around Bankruptcy Model Distress and Recovery Predictions, *Financial Analysts Journal*, Jan-Feb, 70-73.
- Keasey and McGuinness, 1990, The Failure of UK Industrial Firms for the Period 1976-84, Logistic Analysis and Entropy Measures, *Journal of Business Finance and Accounting*, 119-135.
- Keasey, K., P. McGuinness and H. Short, 1990, Multilogit Approach to Predicting Corporate Failure - Further Analysis and the Issue of Signal Consistency, *Omega*, 85-94.
- Keenan, Sean, and Jorge Sobehart, 1999, *Performance Measures for Credit Risk Models*, Moody's Research Report 1-10-10-99.
- Keenan, Sean, Igor Shtogrin, and Jorge Sobehart, January 1999, *Historical Default rates of corporate bond issuers, 1920-98*, Moody's Special Comment.
- Lau, A.H.L., 1987, A Five-State Financial Distress Prediction Model, *Journal of Accounting Research* 18, 109-131.
- Lennox, Clive, 1999, Identifying Failing Companies: A Re-evaluation of the Logit, Probit and MDA approaches, *Journal of Economics and Business*, Vol. 51, No.4: 347-364.

- Lennox, Clive, 1999, The Accuracy and Incremental Information Content of Audit Reports in Predicting Bankruptcy, *Journal of Business, Finance and Accounting*, Vol. 26, Nos. 5 & 6: 757-778.
- Lev, 1971, *Financial Statement Analysis: A New Approach* (Prentice-Hall, Englewood Cliffs, N.J.).
- Libby, R, 1975, Accounting Ratios and the Prediction of Failure: some behavioral evidence, *Journal of Accounting Research*, 150-161.
- Lovie, A.D., and P Lovie, 1986. The Flat Maximum Effect and Linear Scoring Models for Prediction, *Journal of Forecasting* 5,159-168.
- Marais, M., J. Patell, and M. Wolfson, 1984. The Experimental Design of Classification Models: An Application of Recursive Partitioning and Bootstrapping to Commercial Bank Loan Classifications, *Journal of Accounting Research, Supplement on Current Econometric Issues in Accounting Research*, 87-118.
- McNees, Stephen, 1995, An Assessment of the Official Economic Forecasts, *New England Economic Review* Jul-August, 17-32.
- Meehl, P.E., 1954, *Clinical Versus Statistic Prediction: A Theoretical Analysis and a Review of the Evidence* (University of Minnesota Press, Minneapolis).
- Merton, R. C., 1973, Theory of Rational Option Pricing, *Bell Journal of Economic and Management Science* 4, 141-83.
- Meyer, P., and H. Pfifer, 1970, Prediction of Bank Failures, *Journal of Finance*, September.
- Miller, Ross, 1998, Refining Risk, *Risk Magazine*, August.
- Moody's Investor Service. Global Credit Analysis. 1991. Edited by Dave Stimpson. Moody's Investors Service Inc.: New York.
- Morris, Richard, 1997, *Early Warning Indicators of Corporate Failure* (Ashgate, Brookfield, US).
- Moyer, 1984, *Forecasting Financial Failure: A Re-examination*, *Financial Management*, 11-15.
- Narula, S.C., 1978, Orthogonal Polynomial Regression for Unequal spacing and Frequencies, *Journal of Quality Technology*, 9, 170-179.
- Nisbet, Richard, David Krantz, Christopher Jepson, and Geoffrey Fong. 1982, *Improving Inductive Inference*. In *Judgement Under Uncertainty* (Kahneman et al, 1982).
- Ohlson, J. S., 1980, Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research*.
- Peel and Peel, 1988, A Multilogit Approach to Predicting Corporate Failure-Some Evidence for the UK Corporate Sector, *Omega*, 309-18.
- Pinches, G., K. Mingo, J. Caruthers, 1973, The Stability of Financial Patterns in Industrial Organizations, *Journal of Finance*, May.
- Pridgen, Tracy, and Mario F. Verna, 1995, CBO/CLO Rating Criteria, Fitch Research, Fitch Investors Service, L.P., March.
- Ryan, Thomas P. 1997, John Wiley & Sons. *Modern Regression Methods*. New York.
- Santomero, A., and J. Vinso, 1977, Estimating the Probability of Failure for Firms in the Banking System, *Journal of Banking and Finance*, 185-215.
- Sawyer, J., 1966, Measurement and Prediction, Clinical and Statistical. *Psychological Bulletin*, 66, 178-200.
- Schmidt, F. L., 1971, The Relative Efficiency of Regression and Simple Unit Predictor Weights in Applied Differential Psychology, *Educational and Psychological Measurement*, 31, 669-714.
- Scott, J., 1981, The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models, *Journal of Banking and Finance*, 317-344.
- Sherdan, William A., 1999, *The Fortune Sellers: The Big Business of Buying and Selling Predictions*, (John Wiley & Sons, New York).
- Shimko, David, 1999, *Credit Risk, Models and Management* (Risk Books, London)
- Shumway, Tyler, 1999, Forecasting Bankruptcy More Accurately. *A Simple Hazard Model*, University of Michigan, Working Paper.

- Silverman, B.W., 1985, Some Aspects of the Spline Smoothing Approach to Nonparametric Regression Curve Fitting, *Journal of the Royal Statistical Society, Series B*, 47, 1-21.
- Sims, Christopher, 1982, Policy Analysis with Econometric Models, *Brookings Papers on Economic Activity* 1, 107-52.
- Sobehart, Jorge, and Sean Keenan, 1999, *An Introduction to Market-Based Credit Analysis*, Moody's Research Report, 2-11/9/99.
- Sobehart, J.R., S.C. Keenan, and R. Stein, 2000a, *Benchmarking Quantitative Default Risk Models: A Validation Methodology*, Risk Management Services Rating Methodology March.
- Sobehart, Jorge, and Roger Stein, 2000b, *Moody's Public Firm Risk Model: A Hybrid Approach to Modeling Short Term Default Risk* March.
- Society of Actuaries, 1996, 1986-1992 *Credit Risk Loss Experience Study: Private Placement Bonds*, (Society of Actuaries, Schaumburg, IL).
- Spanos, Aris, 1986, *Statistical foundations of econometric modeling* (Cambridge University Press, Cambridge, U.K.).
- Stumpp, Pamela, 1999, *Credit Considerations in Assigning a Caa1*, Moody's Investors Service.
- Taffler, 1984, Empirical Models for the Monitoring of U.K. Corporations, *Journal of Banking and Finance*, 199-227.
- Taffler and H.J. Tisshaw, 1977, Going, Going, Gone-Four Factors which Predict, *Accountancy*, March, 50-54.
- Vinso, J.D., 1979, A Determination of the Risk of Ruin, *Journal of Financial and Quantitative Analysis*, 77-100.
- Wilcox, 1977, September, A Gambler's Ruin Prediction of Business Failure Using Accounting Data, *Sloan Management Review*, 12
- Wilcox, 1971, A Simple Theory of Financial Ratios as Predictors of Failure, *Journal of Accounting Research*, 389-395.
- Wilcox, 1973, A Prediction of Business Failure Using Accounting Data, *Journal of Accounting Research, Supplement on Empirical Research in Accounting* 163-190.
- Wilks, S. S., 1938, Weighting Systems for Linear Functions of Correlated Variables when there is no Dependent Variable. *Psychometrika*, 8, 23-40.
- Zarnowitz, Victor, 1979, An Analysis of Annual and Multiperiod Quarterly Forecast of Aggregate Income, Output, and Price Level. *Journal of Business* 52 (1), 1-32.
- Zavgren, C. 1983, The Prediction of Corporate Failure: The State of the Art, *Journal of Accounting Literature* 2, 1-37.
- Zavgren, C., 1982., An Analysis of the Relationship between Failure Likelihood and Certain Financial Variables for American Industrial Firms, working paper, Krannert Graduate School of Management, Purdue University.
- Zavgren, C.Y., 1985, Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis, *Journal of Business Finance and Accounting*.
- Zmijewski, M. E., 1984, Methodological Issues Related to the Estimation of Financial Distress Prediction Models, *Journal of Accounting Research, Supplement on Current Econometric Issues in Accounting Research*, 59-82.

RiskCalc™ For Private Companies: Moody's Default Model

Rating Methodology

*To order reprints of this report (100 copies minimum), please call 800.811.6980 toll free in the USA.
Outside the US, please call 1.212.553.1658.*

Report Number: 56402