DefProb™
A Corporate Probability of Default Model

Eric Falkenstein

This version
7/15/08

Abstract

DefProb™ is a tool for estimating corporate default risk. Unique among models, it uses Agency ratings such as Moody's and S&P's within the model. The output is calibrated to default probabilities and the simple estimation procedure is both powerful and robust.
See www.defprob.com

While much of the general ideas of my approach come from my prior work at Moody's, all data in the creation and testing of this model are from public sources available to any researcher.

Section 1

Introduction

DefProb™ is a model that estimates corporate default probabilities based on firm characteristics. The model is calibrated to default rates and implied agency ratings. It is transparent, as a downloadable Excel spreadsheet allows one to see exactly how the inputs are used to create the output with only minor redaction of the true parameters. The key advantages of this model over alternatives are the following:

1) It uses implied volatilities when available
2) It uses agency rating information when available
3) It is not overfit

Modeling credit risk is a deceptively simple problem made complicated by perverse incentives. On one hand, academics are interested in presenting elegant, often complicated, nuances, because academics are generally interested in the parochial problem of predicting default only to demonstrate a general technical innovation or as part of a larger issue.[1] On the other hand, vendors also have an inordinate fondness of innovation because they want to appear to add value, and the most obvious way to do that is to present a complicated nuance to a simple framework, and then state that within their proprietary dataset, the nuance generates huge lift in statistical power. These approaches tend to understate the power of straightforward models and probably sacrifices true out-of-sample power.

The technical problem of predicting default is not subtle. Unlike in equity markets where it is difficult to find a metric of risk that consistently relates to future returns, firms with high default risk have theoretically straightforward characteristics that strongly correlate with future default rates: profitability, leverage, liquidity, growth, and volatility. Default data are a stationary target – and similar to explaining the post WW2 recessions – there is no 'out of sample' dataset. Thus the temptation to overfit and ability to rationalize is great. One of the key attributes of DefProb is that it is modest.

Yet the model does contain one novel and powerful – though simple – twist. DefProb includes all information – financial statements, traded equity information and agency ratings – necessary to generate a powerful and consistent output. It is ironic that the major rating agencies' quantitative modeling divisions do not use their own agency ratings in their quantitative models. We are all aware of the agency's imperfections: ratings are not optimal, they are not sufficient, but they are clearly powerful in that investment grade issuers default at a predictably lower rate than non-investment grade companies, and often this assessment is not obvious from the financial statements or

[1] Shumway (2001) introduced a hazard model to predict default, while Campbell, Hilscher and Szilagyi (2007) generate a model to assess how distress risk affects equity returns.

stock price. On the other hand, fitting financial statements or stock market data to agency ratings and applying any such model to non-rated companies results in a significant bias.

Agency ratings are not as powerful as most quantitative models and integrating agency ratings with models highlights this unfortunate fact, which leaves currently vended quantitative models to ignore valuable information (agency ratings) where credit information is most needed: large agency-rated companies.

Modelers of credit risk can learn a lot from the development of other less sophisticated, but useful, products. Toothpaste is a useful product that is pretty straightforward. Nearly all toothpaste sold in the United States has about 1000 parts per million fluoride ion, an ingredient shown to be quite effective at reducing tooth decay. It took civilization several thousands years to develop toothpaste, but the secret is out, and companies offer many redundant versions.

In a similar way, models that integrate quantitative financial and market information to generate an evaluation of a company's credit worthiness are like toothpaste. It is nontrivial to discover what makes a near-optimal credit model (it took us well into the 20th century), but it's not an incredibly subtle or difficult problem, such as solving a cube root in your head. The seminal model of Altman (1968) contained only 33 defaults, whereas currently some vendors have claimed to use 7,900 defaults (Fitch). While the benefit to more defaults is clear, there is also a clear diminishment in information as we go from 10, to 100, to 1000, to 10,000, especially because reaching for defaults tends to impart a selection bias to these observations. The main ingredients of a credit default model are well known, and the weightings of these main ingredients do not have large effects – credit models exhibit a 'flat maximum'. Vendors assert their models show subtle nuances that generate large benefits in model power. As they have proprietary data and proprietary models, one cannot disprove the claim that these models are significant improvements over the state-of-the-art; however, given the flat maximum in models and the clear self-interest of vendors, these claims are dubious.

The model presented here is simple in that it uses equal weights on its inputs, which are themselves selected from straightforward nonparametric univariate correlations with future default rates. Simple is not simplistic, such as not transforming the inputs (which are often nonlinearly distributed), or assuming the sample default rate is the population default rate. The idea that ‘less is more’ is a paradox worth appreciating here (see Cohen (1990)). Desirable simplicity is the product of sophisticated design, an appeal to practicality as opposed to insecurity – if you want a model to impress people with abstruse refinements, this is not for you.

Section 2

Database Description and Applicability

The modern era of commercial default modeling began with the works of Beaver and Altman in the late 1960's, but remained an academic curiosity until KMV's popularization of the Merton Model in the 1990's (see Crosbie (2003)). The keys to the practicality of credit modeling are the data and computing resources needed to calibrate and test a model. For example, consumer modeling used to be a qualitative process where loan officers interviewed neighbors, asked about job status and personal finances and applied intuition plus a few rules of thumb. Today a bureau score captures much of the risk inherent in a consumer loan, can be purchased for a few dollars and often makes possible getting a $5000 credit line in minutes. This was accomplished because of hundreds of thousands of bad credit card debts, along with the information in the applications and subsequent performance that allowed quants to detect patterns, the majority of which are rather straightforward.

Financial statement databases are standard, but identifying a set of 'bad credits' within this dataset is non-trivial, and it is the number of 'bads' that defines the degrees of freedom in this model. For example, if there are 101 bad firms out of 1 million firm-year observations, then 101 imperfectly correlated firm characteristics would allow one to perfectly 'explain' these 101 firm failures. Thus the practice of reporting time-firm data points is very misleading: while one may have daily observations on 5000 companies, this does not imply 10 million observations over a 10-year horizon because the number of firm 'bads' is the only relevant number.

Corporate default data come from three sources. First, CRSP reports delisting information, including bankruptcy. Moody's reports the names of defaulted corporate bond issuers in their annual default studies, but these are generally confined to rated companies. BankruptcyData.com reports bankruptcy data on many, but not all, public companies. Only when a score was available 3 months prior to the default date was it used in the creation and testing of this model.

CRSP provides a delisting code for a variety of circumstances: some are good news (takeover), some indifferent (moving to a new exchange) and some clearly bad (bankruptcy). We interpret delisting codes 500 and 505 to 588 as being performance-related (i.e., a negative change for the firm) and exclude delistings due to mergers, moving to different exchanges, etc (see Shumway (2001)).

Firms were only used for observations if they had market capitalization greater than USD $50 million. This is because a publicly traded firm with significantly lower market capitalization tends to have several problems, including late and incomplete financial statements. As a median non-financial firm has a market capitalization of around $250mm in this sample, we exclude about 33% of the nonfinancial firms in the database.

While our dataset is not as large as other datasets, we avoid some of the problems that come from trying to create a very large credit database. As studies prior to 1990 tended to have at most 100 defaults, one must assume that subsequent studies that included defaults prior to 1990 had to retroactively identify observations. This process is biased by the size and conspicuousness of failed companies. I recall, while working for a bank in the 1990's, seeing that an informal collection of "bad companies" data was biased by this fact. If you ask a credit analyst for bad companies from 5 years ago, he or she invariably brings forth companies large enough to be memorable or worth keeping, generally the larger companies. This bias also shows up in some proprietary databases.

The first step in creating a model is to collect as many 'bads' as possible, while avoiding these kinds of selection biases. Moody's RiskCalc[tm], which is applied to private companies, does a good job in this area because it uses a broad set of data. Moody's has such a lead on competitors that they do not have an incentive to overstate their database.

Table 2.1

Number of Defaulted Companies	953
Number of Delisted Companies	1,924
Number of Companies	5,780
Number of Agency Rated Companies	2,421
Number of Agency Rated Defaulted Companies	953
Number of Firm/month Observations	321,996

Data November 1995 to December 2006

The data only include nonfinancial firms as financial firms have higher leverage and lower transparency in their financial statements. For example, a loan to a risky borrower, as long as it is current (what US regulators call noncriticized), is not distinguished from a loan to a much safer unrated borrower: the sea of non-agency rated credits is indistinguishable from a regulatory and GAAP perspective. Further, the annuity-nature of assets generates cash flows that can be accelerated or extended at a firm's discretion, making earnings generally more easily managed and less meaningful than for nonfinancial companies. Lastly, very few banks default on their debts. Generally, failed financial institutions are taken over or otherwise rescued. For example, there were zero bank failures in the US during 2005 and 2006, even though there were over 8,000 banks. Clearly, this is a different industry from a 'default' perspective, than non-financial public companies. Consequently, this model is not applicable to financial firms, trusts, or ETFs.

Further, given the unique nature of extremely high-grade companies (>A rated) and their low number of defaults, a model is poorly suited towards distinguishing between AAA and A rated credits (which, not coincidentally, is where most financial companies exist, providing another reason not to apply this model to financial companies). There simply is not enough data to distinguish between a 0.01% and 0.07% annualized default rate with a model. There are currently about 180 A- and higher rated nonfinancial companies in the US, implying one needs several decades to generate a 0.0X% point estimate.

Section 3

Rated vs. Unrated Firms

Rated companies are systematically different from unrated companies, and among unrated companies, publicly traded companies are different from privately held companies. Estimating a correlation between firm characteristic and future default rate from one universe to another is unwise because of these systematic differences. So we ignore rated firms in the estimation and testing of the unrated model.

There is a sample bias in that listing requirements for public exchanges are primarily based on a size threshold (market capitalization). Fama and French (2003) report that the average newly listed company is near the median market capitalization for its exchange (Nasdaq new listing being lower than NYSE/AMEX), which is around $250mm over the past 10 years. Public firms with market capitalizations of less than $50mm are already failures: they have most likely lost market value and are therefore underachieving original revenue forecasts. The cost structure of firms at this level of market cap, including pre-existing debt, by itself suggests smaller public firms are a higher risk. Therefore, there is good reason to believe these smaller *publicly* traded firms will fail at a higher frequency. But an unlisted firm of similar size is not such an underachiever, and indeed loss rates experienced by commercial loan portfolios at banks suggests unlisted firms in the $50-$500mm size range have comparable default risk to rated firms.

Just as there are systematic differences between rated and unrated firms, there are systematic differences between public and private firms. Because this model uses equity information as a key factor, it is not appropriate for private firms.

Another element of bias in size metrics is that larger companies will get a greater benefit from the fixed cost of an agency rating, so the fact that most AAA companies are larger than B rated companies is largely explained by the sample selection bias of larger firms seeking ratings when they are of high quality, as opposed to their intrinsic size characteristic. A model that fits firm data to agency ratings would conclude that the 'size' characteristic explains much of a firm's credit risk because it is strongly positively correlated with higher ratings. In fact, this is purely a function of the benefits of an investment-grade rating, as opposed to this factor causing the investment-grade rating.

Liquidity ratios, including the ratio of short-term to total debt, are systematically different for investment-grade firms because of their investment grade status, as opposed to their investment grade status stemming from their liquidity and short-term debt ratios. For example, commercial paper is generally only available to investment grade companies.[2] Access to commercial paper facilities reduces the need for cash on the balance sheet. The

[2] Ratings of A1/P1 or A2/P2 are generally required for active CP issuers, corresponding to the upper tier of investment-grade companies.

fact that the highest rated companies have lower levels of cash on their balance sheets would lead to a model that suggests having less cash is better, because it is correlated with higher ratings, when in fact the causality arrow goes the other way. We can see this in figure 3.1, which depicts the relation between the quick ratio ((current assets – inventories)/current liabilities) and rating category. Note that higher (better) ratings correlate with lower quick ratios. A model fit to ratings would assume that lower quick ratios imply less risk. But figure 3.2 shows that even among rated firms, lower quick ratios have higher default rates, so less liquidity is a signal of higher credit risk. This highlights the problems of fitting firm characteristics to agency ratings, as one might do with a multinomial logit model (see S&P's CreditModel, or Resti (2002)).

Figure 3.1

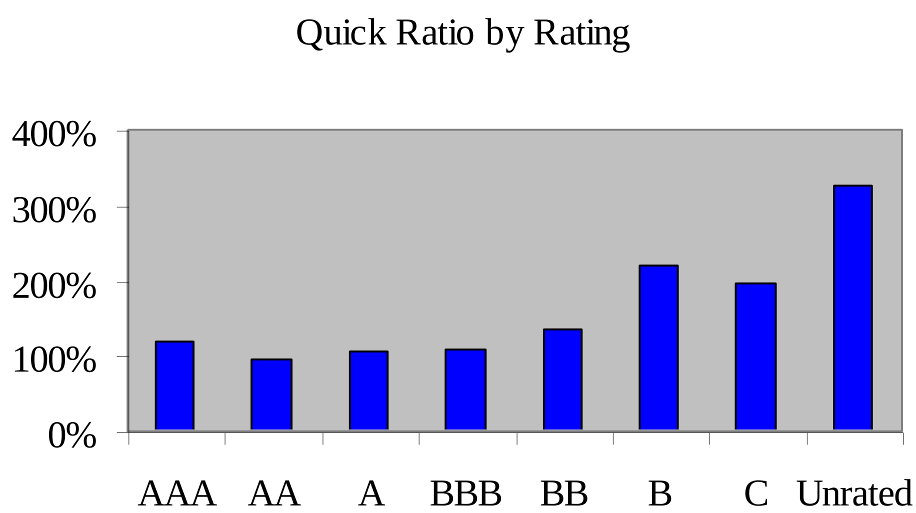
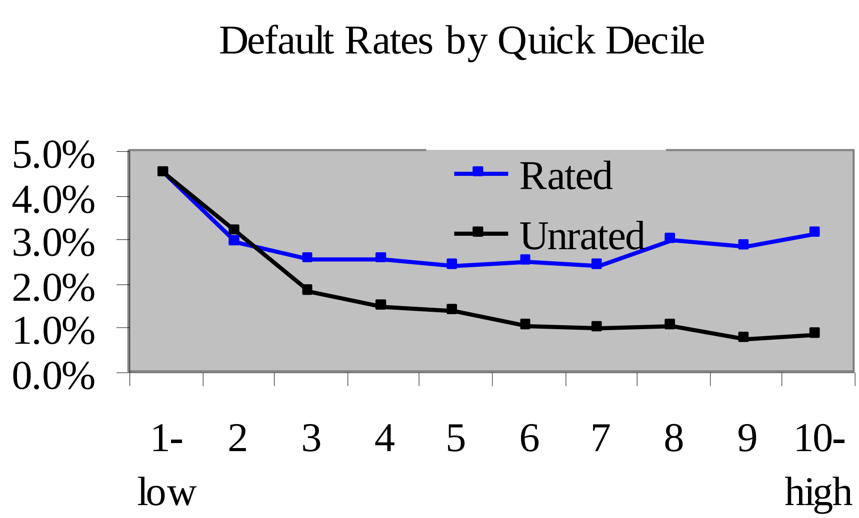


Figure 3.2



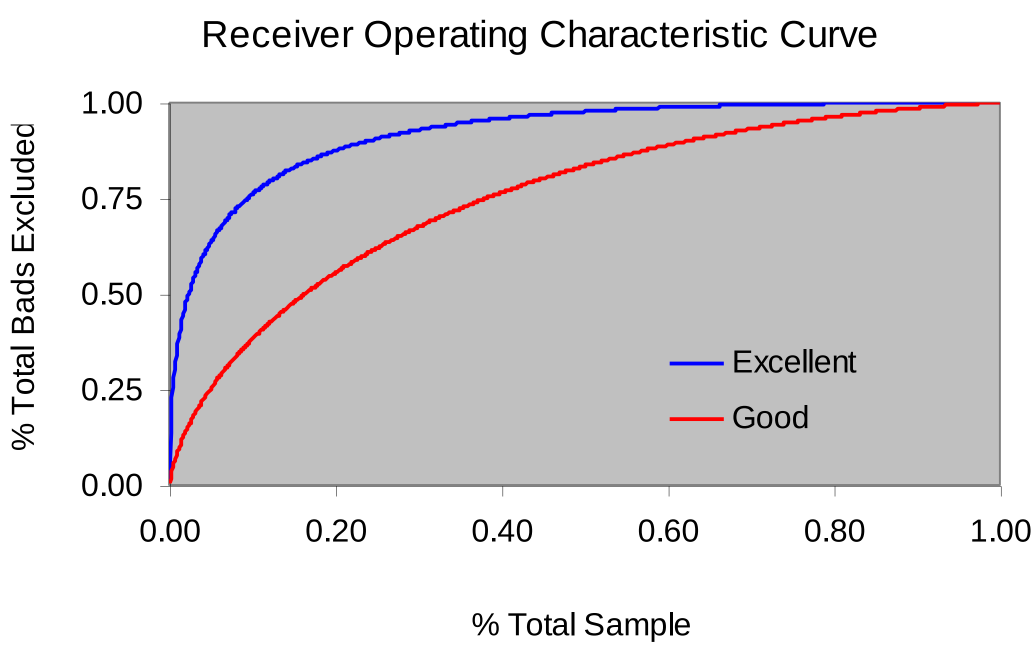
A clear implication is that companies with published agency ratings will have systemically differing financial ratios for a given level of risk, because the agency rating in essence implies an off-balance sheet asset – namely, the ability to borrow quickly and cheaply – that enhances a firm's viability and probably signals some intangible value. This creates a major problem if a naive modeler takes a model intended for unrated companies and calibrates it to rated credits, or vice versa: typical BB financial ratios are indicative of much greater risk for a company without an agency rating; a rated company with typical financial ratios for an unrated firm will have poor survival rate.

Section 4

Measures of Statistical Power

The primary metric for evaluating default models are Receiver Operating Characteristic (ROC) curves, which are known by a variety of other names: Power Curves, Cumulative Accuracy Profiles, Accuracy Ratios, Lorenz curves and Gini curves (see Sobehart (2000) or Bohn (2005)).

Figure 4.1



From figure 4.1, we see that by avoiding exposure to the lowest-ranked 25% of the population – according to the “Excellent” ranking system – we excluded 90% of the defaulters, while the “Good” ranking system avoided just 60% of the defaulters at this level. The ability to exclude more bad firms for any cutoff implies this is better, because it implies lower default and thus loss rates. As the better model lies above (and more northwesterly), it has a larger area under the curve, and so the Area Under Curve (AUC) is a straightforward metric of model power: higher is better. This is more general than a discrete approach that measures false positives, true positives, false negative and finally true negatives, because that approach assumes a particular cut-off, and different users will use different cutoffs. The AUC will typically lie between 0.5 and 1, because a totally non-informative model will exclude just as many defaulters and companies, putting a practical bound of 0.5 on models.

It is important to note that power metrics such as AUC are heavily dependent on the model and the data. For example, assume one could identify only 50% of the *actual* defaults for a particular sample. In that case, *any* model would be imperfect because a really good model would identify ‘bads’ that were not flagged as ‘bad’ in the data, making the test results suggest the model has less power than it actually offers. A model applied to ‘noisier’ data will exhibit less power than on cleaner data. This is especially relevant to default modeling, where the construction of ‘bads’ is very idiosyncratic as opposed to, say, equity returns, where data issues are small. This idiosyncrasy is sensitive to the definition of ‘bads’, the cut-off for inclusion of firms (>$50mm, >100mm, etc.), the cut-off for timing (use report data immediately prior to default versus prior to default by 180 days) and sample period (recessionary periods generally provide less power than nonrecessionary periods). Therefore it is generally not possible to compare models by using separate tests on different datasets.

Section 5

Popular Models

DefProb™ uses a linearly separable model. That is, firm-level financial characteristics (inputs), after a transformation, are additive. There are no combination terms, such as income times leverage, or the quick ratio divided by a Merton variable. Further, a version of the Merton model is one of the inputs. Thus, it is useful to review the origin of these two ideas, the Merton model and the linearly separable model.

The Merton Model

Equity can be considered an option on the assets of a firm, where the strike price is the value of debt. In Merton's original formulation (Merton (1974)), debt is assumed to have an unambiguous maturity, and the option value is computed using this singular date. If the market value of a firm's assets is below the value of its liabilities on the maturity date, owners will exercise their option to walk away and the firm will default; if the market value of assets exceeds the face value of debt, the equity owners keep the residual value and debt-holders are covered in full (no default).

Models based on this approach have a certain indisputable logic in that as a firm nears bankruptcy its equity value goes to zero. Under the Merton model, the asset value of the firm in the future has a probability distribution characterized by an expected value and standard deviation. The area under the distribution below liabilities of the firm is the probability of default. The greater the distance between a firm's total asset value (equity + debt) and debt (the default barrier) and the smaller the standard deviation of total asset value, the lower the probability of default. The following outlines how this default probability is calculated. 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:

$$E = MVA \cdot N\left(d_1\right) - DP \cdot e^{-rT} N\left(d_2\right) \quad , \qquad (0.1)$$

where

$$d_1 = \frac{\ln\left(\frac{MVA}{DP}\right) + \left(r + \sigma_A^2/2\right)T}{\sigma_A\sqrt{T}}$$

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

.

Here, r is the interest rate, MVA the market value of assets, E the market value of equity, σ_A the volatility of MVA, DP the default point (total debt in the simple model), N the cumulative Normal distribution and T is the time horizon.

The problem is that we have two unknowns: MVA and σ_A, so we need another equation. Luckily we have one, in that the relation between the volatility of equity and the volatility of assets is proportional to the value of assets over equity. Specifically, the volatility of equity can be found using the function:

$$\sigma_E = \frac{N(d_1)MVA}{E} \quad . \qquad (0.2)$$

Given equations 5.1 and 5.2 we can solve for MVA and σ_A using standard iterative techniques for solving two equations and two unknowns. The final step is to calculate a default metric from this information. Given the MVA and its volatility σ_A, along with the default point, and assuming that a default occurs if MVA< DP at time T, the number of standard deviations to default is expressed as:

$$\text{Distance to Default} = \frac{E(MVA) - DP}{MVA \cdot \sigma_A \cdot \sqrt{T}} \quad .$$

We can observe the market value of equity (E), the volatility of equity (σ_E), the book value of liabilities, (L) and the time horizon (T). A solution can be found using standard numerical methods since both functions are monotonic in the variables of interest (no closed form exists).

From this information we can determine the number of standard deviations the asset value lies from the default point. In this case, the 'distance to default' (DD) is simply the distance in standard deviations between the expected value of assets and the value of liabilities at time T. It is the conversion of the problem to a Z-statistic, similar to that of dividing a normally distributed variable by its standard deviation; one can use the standard normal table in the back of most statistics textbooks. If DD = 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, income, country and size differences should all be captured in these simple inputs.

Most people do not apply a strict Merton model as described in his initial paper. The most obvious adjustment is the default barrier, which he calls Total Debt. First, because debt maturities are staggered, the simple Merton default point is ambiguous in practice. Second, a firm can remain current on its debt even though technically insolvent (Default Point<Debt) forestalling, and with luck, avoiding bankruptcy. Lastly, the long maturity of some corporate debt (greater than 10 years) implies too small a sample size for calibrating and validating the model. Thus, as applied by most practitioners, the barrier is

not total debt, but a function of the maturity of the various maturities, where longer maturities are weighted less because they are less pressing. This makes sense because the model implies that, empirically, we should see the highest recovery rates lying near 100% and declining exponentially to zero. But in practice, the modal 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 liabilities, the most probable recovery rate would be in the highest range towards 99%; it is not, which suggests the trigger lies well below the amount of total debt. DefProb™ assume the default point is Current Liabilities + ½(Long Term Liabilities), or 30% of Total Assets, whichever is greater.

Moody's/KMV, CreditGrades (a joint work of RiskMetrics, Deutsche Bank, Goldman Sachs and JP Morgan), and Fitch all use a perpetual barrier option, so that at any time if the barrier is reached, default is assumed. Moody's/KMV adds other nuances, not just debt and not just current and long-term liabilities, but rather five classes of liabilities (short and long term, preferred and convertible stock). It also assumes coupons and dividends leak from the firm value. CreditGrades uses the volatility skew from option markets in creating volatility curves for the equity. CSFB uses a jump-diffusion process, as opposed to Brownian motion.

The refinements add realism, but the gist of the Merton model is mainly contained in the ratio Market Equity/Liability and the volatility. Higher volatility is bad, and lower market value/liability is bad, and these effects compound each other. Bharath and Shumway (2007) show that a naïve implementation of these two variables generates the same power as the Merton model, and Campbell, Hilscher and Szilagyi (2007) show that a reduced form model works better than a Merton model. It would seem improbable that the further extensions – perpetual barriers, stochastic barriers, stochastic jumps – truly add power to the model, just as the basic Black-Scholes model applied to a volatility surface does as well as more refined option models. I have examined the above extensions and found them to be inferior distractions.

Practitioners who generate a distance-to-default generally map it into an empirically determined default probability, as opposed to thinking that the normal CDF is going to give a true probability. While important, this does not affect ordinal ranking, and thus model power, or accuracy. Thus, practitioners generally look at the empirical frequency of defaults associated with various distance-to-defaults and generate a nonparametric transformation that maps a distance-to-default into a default probability. The adjustments indicate that the Merton model is a heuristic as opposed to anything as precise as Newton's law of gravitation.

Just as an altimeter predicts plane crashes with certainty at some level, market equity values go to zero prior to firm bankruptcy. Excepting cases of fraud, defaulting firms have low distance-to-default immediately prior to bankruptcy or default, so this measure has a near zero Type 2 error rate (false negatives), but the Type 1 error (false positives) is relevant. The key is, over an intermediate horizon of say 6 months to 2 years, do these models have greater power at separating defaulters from non-defaulters or those with rating upgrades?

Nonstructural models

The most famous quantitative credit scoring model is Altman's Z-score.

$$Z_{1968} = 1.2 \cdot \frac{WC}{TA} + 1.4 \cdot \frac{RE}{TA} + 3.3 \cdot \frac{EBIT}{TA} + 0.6 \cdot \frac{MVE}{TL} + 1.0 \cdot \frac{S}{TA} \qquad (0.3)$$

$$Z_{1977} = 6.56 \cdot \frac{WC}{TA} + 3.26 \cdot \frac{RE}{TA} + 6.72 \cdot \frac{EBIT}{TA} + 1.05 \cdot \frac{BE}{TL} \qquad (0.4)$$

The big idea in this approach is that a model that uses a combination of inputs, as opposed to merely single ratios (Beaver) is better. This is a pretty straightforward idea, in that rather than look at just one attribute, we can look at several and weight them proportionately to their explanatory power. Thus, although DefProb™ uses the Altman approach, in that it looks at several financial ratios, the connection is a fairly weak.

Altman's seminal 1968 model, which used Discriminate Analysis on a sample of 33 defaulting companies, is a familiar citation. As the first multivariate approach to the problem of credit analysis, it clearly has precedence, but one should recognize it is an anachronism in the true sense of the word. DA was preferred to more intensive techniques such as probit or logit back in 1968 because it is much less computationally intensive to invert a matrix than to solve a maximum likelihood function. But computer power has made that a non-issue. The ratios chosen by Altman include some perennials, such as EBIT/assets, but also some mistakes, such as Sales/Assets (which is biased by industry variation). The demonstration of power is only impressive in a vacuum, and a similar issue occurs when looking at agency ratings in that virtually any model will generate statistical power at predicting default: profitability, leverage, etc., all show very straightforward correlations with future default rates. The real question is, 'power relative to what?' My choice would utilize simple, straightforward alternatives. In the case of Altman, (Net Income - Liabilities)/Assets dominates; in the case of rating agencies, the Merton model dominates.

Altman's model is ubiquitous because as the first multivariate model it is an obvious reference in the introduction to a new model, and further, it has the attractive qualities of having some power, but not too much power (great when presenting a new model). The real key to any successful model is, at the very least, truncating the inputs, so that extremums due to denominators near zero do not generate outliers that overwhelm an estimation procedure. This is often accomplished by "Winsorizing" the data (e.g., applying the 98 percentile level to anything greater than the 98th percentile), or using sigmoidal transforms that map a variable into a 0-1 range, as in an arctangent function.

Section 6

Inputs

While almost all popular ratios are correlated with default risk to some degree, most are redundant. There are generally several categories of characteristics that relate to credit risk:

- Leverage: debt/assets, market cap/liabilities, total liabilities/assets
- Coverage: EBIT/interest expense, cash flow/assets
- Profitability: net income/assets, EBITDA/market cap, income accruals vs. cash flow (earnings quality)
- Liquidity: current assets/current liabilities, quick ratio, short term debt/total debt, cash/assets, inventories/sales
- Volatility: equity volatility, volatility of cash flows or earnings
- Growth: asset growth, sales growth, net income growth, stock return

Virtually all potential explanatory variables are ratios in that they are normalized for size (e.g., EBIT/assets, equity/assets and EBIT/interest expense). Further, by using ratios one can avoid differences from time variation in the value of money or differences in currency. I am excluding size as a ratio, which is often listed as a key firm credit characteristic because 1) its impact should be reflected by volatility that is directly measured, and 2) the biases to size in default collection are impossible to eliminate (see Shumway (1997) and Shumway and Warther (1999)).

Table 6.1
AUC (power metric) for various default predictors

	AUC	**Name**
1	84.39%	Merton European Option
2	83.94%	Merton American Option
3	81.64%	Market Cap/Liabilities
4	81.42%	Merton w/ Dividend
5	81.07%	Shumway
6	79.24%	Altman's Z-score (5 variable)
7	77.84%	Volatility
8	77.56%	Net Income Minus Ex Items/Assets
9	77.05%	Income Before Ex Items/Assets
10	74.68%	Cash Flow/Assets
11	74.41%	Mkt Cap
12	74.37%	Debt/Assets
13	74.26%	(Liabilities-pref stock)/Assets
14	73.50%	Liabilities/Assets
15	73.50%	Book Equity/Liabilities
16	73.30%	Retained Earnings/Assets
17	72.57%	(Liab - Cash)/Assets
18	72.41%	Altman's Z-score (4 variable)

19	71.81%	Ebit/Assets
20	70.04%	Book/MktCap
21	66.99%	Debt Change
22	66.48%	Asset Growth
23	65.03%	Sales Growth
24	63.42%	Equity Change
25	63.35%	Quick Ratio
26	62.04%	Current Ratio
27	61.57%	PreferredDebt/Assets
28	60.34%	Dividend/MktCap
29	60.20%	Cash/Assets
30	59.68%	Accrual
31	57.50%	Assets
32	56.22%	Beta
33	55.99%	Sales/Assets
34	54.79%	Inventories/Sales

Altman's Z-score (5 variable)	1.2*WC/A+1.4*RE/A+3.3*EBIT/A+0.6*MVE/lt+1*S/TA
Altman's Z-score (4 variable)	6.56*WC/A+3.26*RE/A+6.72*EBIT/A+1.05*BE/lt
Shumway	(2*NI-L)/A

All inputs were first transformed via univariate mappings to default, so that non-monotonicities would be correctly captured. For example, Asset Growth has a U-shaped relation to default, but after first transforming to default, then ranking, we get a monotonic (ie strictly non-decreasing) relation.

Looking at Table 6.1, we see that Merton models are at the top of the univariate ratio list. As they include market information, which is generally more informative than financial statements alone (efficient markets), all ratios with market capitalization or equity volatility tend to cluster at the top of the list. But note that the Merton model variants – an American Option, or subtracting dividends from the growth rate, or a barrier option – were all inferior to the simple Black Formulation listed in Section 5. While there are good theoretical arguments as to why these adjustments are useful, refinements such as these tend to create more noise than orthogonal information. Of course, I have no vested interest in them, so perhaps I did not torture the refinements until they behaved as expected. The deleterious effects of more information are discussed more fully in the next section, which deals with variable selection.

Interestingly, the Altman Z-score (the 5 variable version that includes market cap) is inferior to Shumway's simple (2*Net Income - Liabilities)/Assets model. This again highlights the power of simplicity and the flat maximum. The top variables are Merton, profitability, and leverage. Ratios related to growth or liquidity are generally lower in power.

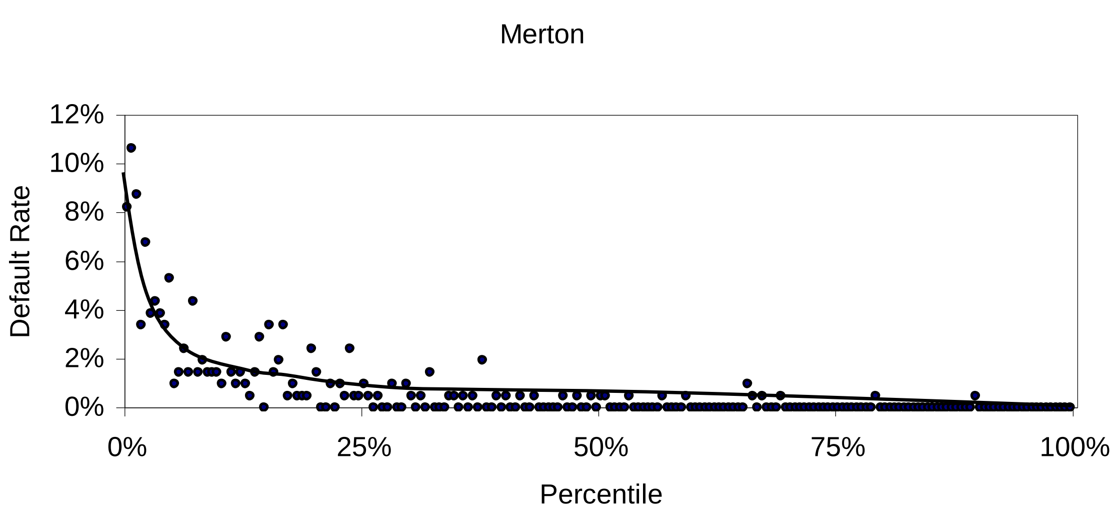
Section 7

Model Creation

Fundamentally, DefProb™ is a simple sum of univariate predictors mapped to a default rate. That is, we look at a handful of known correlates with historical default rates, and correlate their nonlinear relationship with defaults. We then take the most powerful univariate predictors and add them together. We then take this summation and empirically map it to sample default rates. Finally, we adjust the sample default rate to the population default rate based on data from aggregate historical default rate studies. To recap, the procedure is as follows:

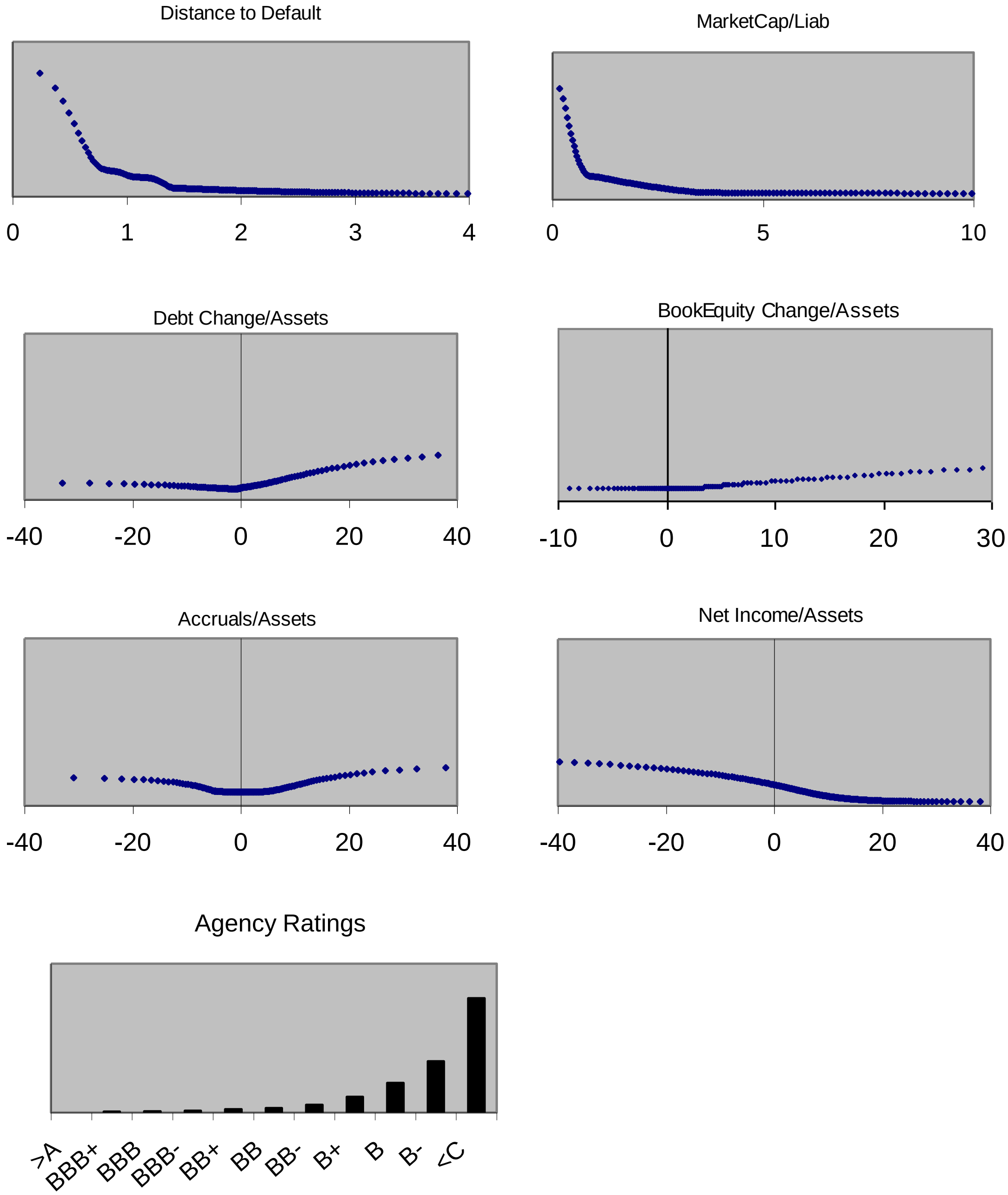
1) **Transform Univariate Inputs**: Estimate the two-year cumulative univariate default frequency as a function of each input's level. Use this univariate default frequency function, T(x) = Prob(default|x), to transform the raw input ratios. This first step allows the model to capture nonlinearities in the firm characteristic/default rate nexus. Most modelers who propose non-traditional models do not transform the inputs so that highly nonlinear relations are suboptimally captured in standard models. But this is either careless or calculated (calculated because it makes for an easy-to-beat benchmark). Any good default model transforms the inputs. The basic process is the following. First, bucket firm-month observations into percentiles, say 100 groups. Second, find the default rate over the next 2 years for each of these buckets. This gives you a set of points. You then fit a line through these points, and you have a 'nonparametric univariate default model' that serves as an input to the final algorithm. Figure 7.1 below illustrates.

Figure 7.1



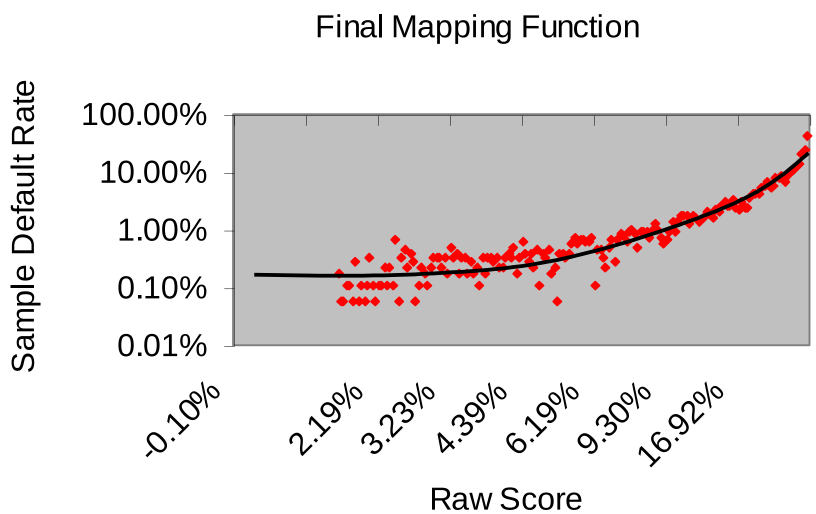
The transformations for the inputs are as below in figure 7.2. All the y-axes are the same, so we see that the inputs with the greatest impact on the final output are the Distance-to-Default (Merton Model) and Market Cap/Liabilities, and Agency Ratings.

Figure 7.2



2) **Sum Transforms, Map Final Output to Default**: Given the above transformations, simply sum, and map the output to default frequency by looking at the sample default rate. This is really just applying the same algorithm as above, only this time to the sum of the previously transformed inputs.

Figure 7.3



3) **Adjust final default rates to population means**. Compare sample default rates to 'through the cycle' default rates suggested by broader data sources.

The result is a simple linear model of the transformed (nonlinear) inputs. The inputs all have equal weights, making it 'simple' (see Dawes (1979)). Note that certain inputs, such as the Merton model's distance-to-default, are more predictive than others, in that there is greater disparity in the transformed value along the input domain. Thus the increased power of some inputs speaks through the greater variance of the range of the transformation, as opposed to a coefficient on that transformation.

DefProb™ is an amalgam of two models depending on which inputs are available. Thus, the optimal scenario has financial statements, stock returns, and agency ratings. But if no agency rating exists, the model is separately calibrated for that. Clearly, the more information available, the more powerful model, and agency ratings should be used because this generates a more powerful score than if you do not have this information.

Table 7.1
Inputs used in DefProb™

	Rated	Unrated

DebtChange		**X**
Accruals/Assets		**X**
EquityChange		**X**
Merton Model	**X**	**X**
MktEquity/Liab	**X**	**X**
Net Income/Assets	**X**	**X**
AgencyRating	**X**	

DebtChange is the year-over-year change in debt divided by total assets
EquityChange is the year-over-year change in book equity divided by total assets
Accruals is the change in working capital divided by total assets
Merton Model is the Distance to Default
MktEquity/Liab is the market capitalization divided by total liabilities.
Net Income/Assets is (Net Income minus Extraordinary Items) divided by Total Assets
Agency Rating is the composite senior unsecured rating based on an average of Standard and Poor's, Moody's and Fitch IBCA, depending on what is available.

DefProb™ is compared to other models across identical sample sets – for companies with agency ratings, market information, or only financial statement – because the universe of agency rated companies is different from the universe of public companies that are unrated, and from the universe of companies without volatility information. As mentioned above, it is essential to have the same sample sets when evaluating a model, which unfortunately makes for a messy presentation.

Section 8

Why a Simple Model is Better

The Flat Maximum

In the words of Dawes and Corrigan (1979), 'the whole trick is to know what variables to look at and then know how to add.' Once you have the most important variables, and the independent variables are standardized so they have comparable distributions, and if those independent variables are positively correlated, the problem develops a somewhat 'flat maximum': deviations from optimal coefficient weights produce almost the same result as the optimal coefficient weights. This occurs especially when the best cues are reasonably predictive, somewhat redundant and the model is not extremely powerful. Most of our inputs are correlated. As per the model being not sufficiently powerful, note that C-rated issuers tend to 'only' default at a 20% annual rate, so even the riskiest firms tend to *not* default, generating false negative errors for these really 'bad' issuers. Though a 20% default rate is economically large, from a pure information entropy perspective, statistically the model has relatively weak power. This does not mean it is suboptimal, merely that the problem is inherently difficult. Thus the costs of misspecification are low, in that coefficients can vary by orders of magnitude and still generate statistically similar power.

The empirical section shows that the weightings on the parameters matter very little.

Overfitting

There is no standard database for corporate bond defaults as there is for stock prices or financial statements, which implies that, in general, corporate credit modelers are each referencing a different particular dataset. Therefore it is impossible to prove whether any vendor's refinements are marketing hype or truly valuable. If refinements worked, one would expect to see more of a cluster around particular refinements, because things that work tend to be emulated. Among corporate credit models, however, vendors tend to emphasize different idiosyncratic nuances, suggesting that these primarily have marketing, as opposed to statistical power. Examples include stochastic default boundaries, modifying the liability structure based on various long-term debt maturities, the inclusion of dividends, off-balance sheet leases, interaction terms, or industry dummies.

Industry Dummies

Industry variation is a good example of the dangers of overfitting because with hindsight certain industries tend to do much worse in any particular recession. In 1970, the industry with the highest number of defaults was railroads; in 1990 it was hotels, and in 2001 it was telecom (which didn't even exist as a category previously). While it is tempting to add a 'plus factor' for these industries, the problem is that poorly performing industries do not repeat their anomalous performance, so this kind of reasoning merely adds to error as opposed to adding power. Given most industry participants react to their experiences, there are good reasons to understand this effect as *endogenous*: people don't make the same mistakes every business cycle, they make different ones. For example, the commercial real estate crisis of 1990 created a new benchmark for lenders: going forward, all loans, and CDO's became robust to this level of recession, which was historically quite extreme. Commercial mortgage CDOs have shown much lower-than-average default rates since 1991, suggesting that the underwriters overreacted. I imagine the recent revisions in residential mortgages will make this area better than average for residential mortgage loans *originated* today.

This industry pattern of non-repeating shows up not only in defaults but in stock returns: the oil industry did the worst in the 1981 bear market, but was one of the best industries in the 1990 and 1974-5 bear markets; or retail stores, which were among the worst stock performers in the 1973-4 contraction became one of the best performers in the 1980-81 and 2001 contractions. Any model with industry dummies, especially given the bias in our default sample towards the very latest recession, will assume the worst performers from the last recession will also fair worst in the next, which the record simply does not bear out for understandable reasons.

Macroeconomic Conditions

Instead of predicting which loans in a basket will default, wouldn't it be better to know when the economy is going bad? That is, if we could just take our loans off the table when there was a recession, this would be better than predicting, cross-sectionally, which loans have higher probabilities of default. The key is that *an aggregate default forecast is generally unrelated to the question of cross sectional risk,* and so conflating a credit default model with both a cross-sectional and time-series objective obfuscates both. Is it a 1.5% default rate because the model foresees a recession? Do I care about the model's macroeconomic forecast? No one can predict recessions well, including experts who have spent their lives studying these issues (see Stock and Watson (1992)). In hindsight, each recession looks predictable, but, alas, they are unique in ways that make them hard to foresee. Oil shocks, commercial real estate, the dot-com bubble, these will probably not be the center of the next recession.

So when one changes behind the scenes – to accommodate a business cycle forecast – a default model where the power is primarily cross-sectional, it is hard to evaluate the output. As we learn in the serenity prayer, it is essential to know what one can forecast and what one can't. I would note that next year's default rate is typically the historical average plus 0.7 times the deviation of last year's default rate from the average.[3] But my model ignores this. Default probabilities are mapped to 'thru the cycle' default rates. Shade them to your idiosyncratic business cycle views.

[3] e.g., DF(t+1)=1.5%+0.7(DF(t)-1.5%)

Global Variation

Generally speaking, different countries employ different accounting standards, have different bankruptcy laws, different government policies (that might affect a company's ability to downsize in bad times), and different attitudes about getting bail-out loans from government-owned banks. Nonetheless, financial statements and equity valuations allow investors to make apples-to-apples comparisons. The US is very like Canada, and next like Anglo-countries (including Singapore), next like European countries, and lastly like Japan. And then there are the developing markets.

The selection biases in non-US databases are much greater because it is difficult to see the historical guarantors of companies that propped them up. For example, in Japan, there were 'zombie' companies kept alive by the banks that feared that if they allowed them to fail, the true extent of Japan's bad loan problem would be revealed. The problem is similar to estimating a model for banks that will be bailed out by the government. A model fit to that data would suggest they will not default in the future, which is questionable.

DefProb™ as applied to Japan or other developed countries is exactly the same as that used for the US. It is estimated and calibrated against US firms. I suspect the US-derived model is still highly valuable. But this is merely an educated guess.

Changes v. Levels

Most on-line models offer snapshots of the 'big movers', kind of like 'off-label' usage of prescription drugs.[4] Many investors think that looking at recent changes in default probabilities are more important than the levels themselves in managing a portfolio. This is wrong. In predicting default, changes in inputs – such as profitability, leverage, or the Merton variable – are all dominated by the overall level (see Falkenstein, Carty, and Boral (2000)). This too, holds for the final output of the model. This is not surprising, in that if changes in a model dominated a model's level in predicting default, one would have a manifestly suboptimal model that could be bettered by incorporating changes in inputs directly, so a modeler suggesting you look at changes, is saying that he could improve his model by adding more trend or growth terms within the model. If you want to apply the model to the cutting edge, it is best to group some credits by agency rating, or a spread bucket (e.g., 300-500 basis points) and find those with the highest and lowest Default Probability, not those with the biggest *change* in default probability.

Nuances

As mentioned above, vendors are eager to differentiate themselves from competitors. They are also eager to build barriers to entry. A complicated, vague, modeling technique (e.g., support vector machines) is a way to do this at the expense of transparency and

[4] **Off-label use** is the practice of prescribing drugs for a purpose outside the scope of the drug's approved label, most often concerning the drug's indication. For example, Botox has become the basis of a popular cosmetic anti-wrinkle procedure, which is much more economically important than the approved use.

credibility. Generally, these nuances add significant power when used on proprietary datasets, and we are supposed to be impressed that they are using the 'anisotropic Gaussian kernel' function within a 'Pool Adjacent Violator' algorithm, because only 10 people really understand what this means (allusions to authority are common in this field – how often is "Nobel Prize winning" used when describing the Merton model?). Further, any nuance adds degrees of freedom to the model. As all modelers know the data very well, they can parameterize these nuances to fit the data better, but only in sample. They will all strenuously assert they did vigorous out-of-sample testing, but this out-of-sample testing is invariably in sample, in that it was not done once, but many times, with many slight variations, until the final 'out-of-sample' results are sufficiently strong. It would be more accurate to call out-of-sample testing 'subperiod stability' tests, but even that is inaccurate because it assumes it was done once, which it never is.

Another issue is the counterintuitive finding that adding information that is theoretically useful often decreases model power. For example, in the 1960's and 1970's there was a big drive to create large models of the macro economy, with small models containing 60 equations, larger ones 200 equations. As they grew, they reflected the pressing issues of the day: inflation in the 60's, oil shocks in the 70's, etc. Chris Sims (1980) showed that a simple 6 variable Vector Auto Regression (VAR) performed as well as these more complicated models. The macromodeling industry has been dying of neglect ever since. The problem with default models is that they generally include many correlated inputs. Because the researcher knows the right answer – and a clever researcher can rationalize any relationship (e.g., More cash higher default: a sign of desperation. Less cash higher default: a sign of no resources.) – in-sample correlations can cause the parameter estimate standard errors to explode, worsening out-of-sample power. Alas, too many modelers cannot accept that their neat, intuitive, more complicated model was over-fit.

If a subtle, complicated, and elegant refinement added as much power to these models as presented by the various proprietary modelers, it would be unprecedented in empirical finance. Anyone who has worked as a risk manager knows that the models that get used tend to be rather simple, and that mainly academics or consultants champion complicated approaches. The Bharath and Shumway (2004) paper is instructive here because the authors – two researchers without a vested interest – found that a much simpler implementation was as powerful as the Merton model. The Black-Scholes option model has spawned hundreds of refinements – jumps, stochastic volatility and finite difference methods, to name a few – but the basic Black-Scholes model works pretty well when accompanied by an empirically derived volatility curve, as is usually done in practice. Elegant, complicated nuances are like putting fruit in your beer.

While every modeler pays lip service to overfitting, it is just too tempting to avoid. Which is why *not* overfitting a modeling is a virtue – everyone is aware of it, but it is so hard to resist. The desire to differentiate and build a barrier to emulation, to create a transcendent innovation that will outlive oneself, are often overwhelming. Thus most models are overwrought.

Section 9

Test Results

We describe below four models built *off the same set of transformed inputs*:

1. DefProb™ (i.e., an equal-weighted sum of nonparametric univariate estimators)
2. Probit
3. OLS
4. A 'Superfit Linear Model'

Credit modeling is generally viewed as a binary choice problem, where the dependent variable is 1 if 'bad' (delist or default, as the case may be), and 0 if not bad. This makes OLS biased and inefficient relative to Probit, but I include it as an example merely to highlight the flat maximum property of this problem.

As Probit minimizes the difference between the implied probability of the model and either 1 or 0, it will not necessarily maximize the AUC metric that is generally used to measure power. So I designed a simple hill-climbing algorithm that performs a grid search and adjusts the input weights to maximize the AUC. That is, starting at unit weights for each input, I then find the marginal numerical derivative for each input and then move the weights by 0.1 in the direction of the greatest increase in the AUC. This is repeated until the AUC cannot go any higher. I do this only on the entire sample for comparison purposes. This is the 'Superfit Linear Model', referred to below as Optimized.

I also show the AUC for Agency Ratings and the Merton model in this sample for comparison.

I test these approaches three ways.

1. By estimating the model, including the transformations that underlie the inputs, on 'delisting' as defined by CRSP. Delistings are a weaker definition of bad, in that they include only those companies that were delisted for performance reasons, and these include companies that never defaulted, but merely had accounting issues, or their equity price approached zero. I then tested the model on the defaults, which is a subset of these delistings.
2. Estimating the model, including the transformations that underlie the inputs, on one half of the firms, and then applying it to the other half, and vice versa, and averaging the AUC. This is the 'out-of-sample' column.
3. Estimating the model on the same set of defaults I tested it upon, in the third column labeled 'defaults'.

Table 9.1 shows the power results on the rated universe of public companies.

Table 9.1
AUC for various models and datasets

Rated

	delist	out-of-sample	default
DefProb™	90.78%	89.73%	90.80%
Probit	89.43%	88.19%	89.22%
OLS	90.82%	89.68%	90.77%
Agency	N/A	N/A	84.96%
Merton	N/A	N/A	87.93%
Optimized	N/A	N/A	90.89%

UnRated

	delist	out-of-sample	default
DefProb™	89.50%	89.01%	89.52%
Probit	88.96%	87.87%	88.50%
OLS	89.73%	88.97%	89.76%
Merton	N/A	N/A	84.50%
Optimized	N/A	N/A	89.79%

Delistings vs. Defaults for estimating models

First, note that all three estimated models (OLS, Probit, and DefProb™) show relatively little change in power on the default sample, whether estimated on the defaults and applied to the defaults, or estimated on the delistings and applied to the defaults. The weakness of delisting is clearly mitigated by the greater number of these observations, and that ‘delisting’, while not necessarily a default, is still bad enough to be informative.

Agency Ratings

The rated sample shows the Agency rating were indeed powerful—84.96%--but they were weaker than the Merton model. The strength of the Agencies comes from their ability to validate their rating via a long history, but it should be noted that their performance is usually presented without a quantitative benchmark. Further, Moody’s own subsidiary that runs a Merton model states that models outperform the Moody’s rating, so there is little debate here.

Out-of-Sample Power

Out-of-sample estimation does generate a worse performance, but DefProb™ does relatively better than OLS and Probit on the Unrated sample, with about the same diminution of power on the rated sample.

Marginal difference between DefProb™, Probit, OLS and Optimized models

As mentioned, I tried several method of weighting the inputs, and these generated significantly different input weights. The weightings were as follow:

Table 9.2
Parameters of Various Models

Rated

	Simple	OLS	Probit	Optimized
nixa	1.00	0.39	16.26	0.90
ddblack	1.00	0.51	1.17	1.90
rat	1.00	0.41	1.98	0.90
mel	1.00	0.30	3.30	1.00

Unrated

	Simple	OLS	Probit	Optimized
nixa	1.00	0.28	13.75	0.70
accat	1.00	0.21	6.88	0.70
eqchlag	1.00	0.28	5.18	0.30
dachlag	1.00	0.30	9.36	0.80
mel	1.00	0.49	4.46	1.20
ddblack	1.00	0.42	1.69	1.10

As is clear from Table 9.2, Probit put much more weight on net income/assets. Also, the weightings were quite variable across the models. Nevertheless, the DefProb™ model performed better on the out-of-sample test, and best on the unrated sample. However, it was slightly suboptimal on the delisted/unrated sample. Further, the optimized in-sample fit generated only a marginal lift in power. Thus, there appears to be little cost to – and considerable *a priori* justification for – a simple model of unit weights. The transformations are already implicitly weighted because the range of each input is scaled to its correlation with default on the estimation sample so that the relative power of the Merton and Market Equity/Liabilities inputs is implicit in their transformations. Thus, further weighting is not necessary to reflect this. It's as if you had two inputs, say GPA and SAT scores, and instead of adding them together, you first normalized them by demeaning and dividing by their standard deviation.

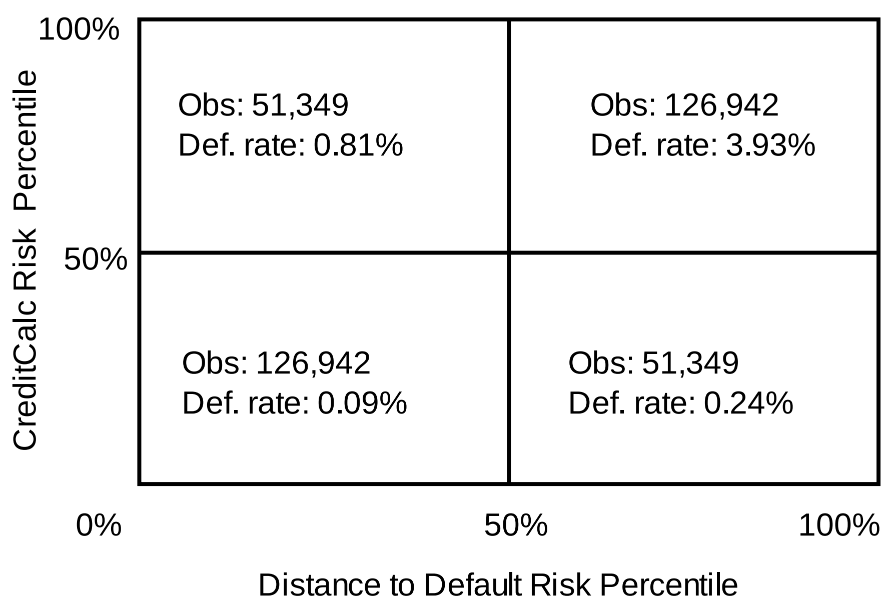
Sundry

The out-of-sample performance for DefProb™ (or any of the alternative methods) dominates both the Merton model and Agency Ratings. The relative performance is similar on the Unrated, and Rated universes.

Comparisons to Agency Ratings and the Merton Model

As the Merton approach is a popular model, I compare it to DefProb™ in more detail here. Figure 9.1 shows the data sorted by the DefProb™ model score and the Merton model. Data were separated into 4 quadrants based on the median scores from these two models and default rates were calculated for each sample. Each diagonal quadrant therefore has an identical number of observations, because high-high and low-low are equally likely given the positive correlation of these risk scores, while high-low and low-high are equally unlikely. Note that the upper right quadrant has a higher default rate than the lower left quadrant, which makes sense, as the upper right quadrant has a higher default score from both models, while the lower left quadrant has lower risk scores for both models. But when the DefProb™ implies above average risk AND the Merton model implies below average risk, the default rate is 0.81%, vs. only 0.24% for the reverse. This highlights the greater power of DefProb™ in that future default rates are higher when DefProb™ is above average and Merton is not, compared to the reverse.

Figure 9.1



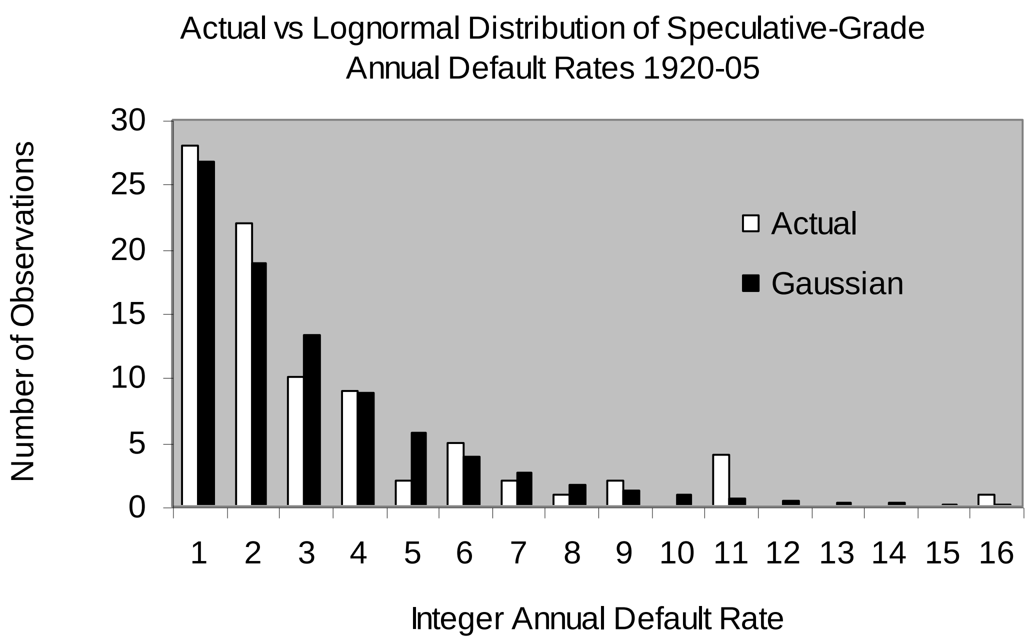
The approach taken here – a simple model with unit weights – does as well as methods that theoretically fit the data better because of the overfitting issues common to this environment. This implies there is little cost to applying the simple model. The upside is that one is less susceptible to overfitting. The model dominates the Merton Model and Agency Ratings [see table 9.1] because it uses both, and these inputs are not perfectly correlated.

Section 10

Mapping to Default Rates and Ratings

The final part of the modeling process, mapping to actual default probabilities, is similar to how we transforms the inputs. One could take the output of the model and find the best fitting function that maps the output into the sample default probability. But this final transformation is complicated for two reasons. First, defaults tend to be correlated over time, so that 80% of the defaults are smushed into 40% of the years over the 1920-2005 period. This implies sample default rates are rarely good estimates of population default rates, especially with only 10 years of data. Thus many years will have near zero defaults in good times, then much above-average defaults in bad times. A lognormal distribution for default rates is a relatively decent approximation, as figure 10.1 shows. We can observe an exponentially decreasing default intensity distribution for the 1920-2005 sample period.

Figure 10.1



In a paper written with Richard Cantor in 2001, we note that default rates exhibit significant time and industry variation, and these common shocks drive the sample standard errors more than the number of firms. Thus, even a sample period of 10 years, with over 1,000 defaults, should be considered a 'sample estimate' as opposed to a population estimate.

Further, there have been changes in the debt markets that make data prior to the 1980's less relevant. Prior to Michael Milken's popularization of 'high yield' debt, speculative grade firms were almost all 'fallen angels', whereas subsequent to the mid-1980's firms issued debt in this range. In the old days, B and Caa rated firms had below-average default rates (relative to today). This was consistent with Milken's assertion that high yield bonds generated significant above-average returns relative to their default rates.

Highlighting the ambiguity of forward-looking default rates is Moody's concept of "idealized default rates." As the Collateralized Debt Market grew, the importance of ratings consistency demanded that each rating grade imply a specific loss rate. Otherwise a "B" rating is ambiguous – a function of both country and markets (e.g., European Bank "B" vs. a US Industrials "B"). It is in the agency's self-interest to have consistent ratings in order to be a meaningful benchmark. But if you want "B" rated collateral to have tranches with "B" ratings, you need a good estimate of the B rating's default (and loss) rate in order to numerically peg the appropriate degree of cushion ahead of the B-rated tranches in the CDO waterfall of cashflows (see Cifuentes and Wilcox, 1996). With CDOs, and the regular reporting of default rates by Moody's across rating and horizon, a cardinal interpretation of default rates is now unavoidable.

So what are default rates by rating category? Looking at Table 10.1, we see that over the last few credit cycles they vary considerably.

Table 10.1
Average US Annual Default Rates

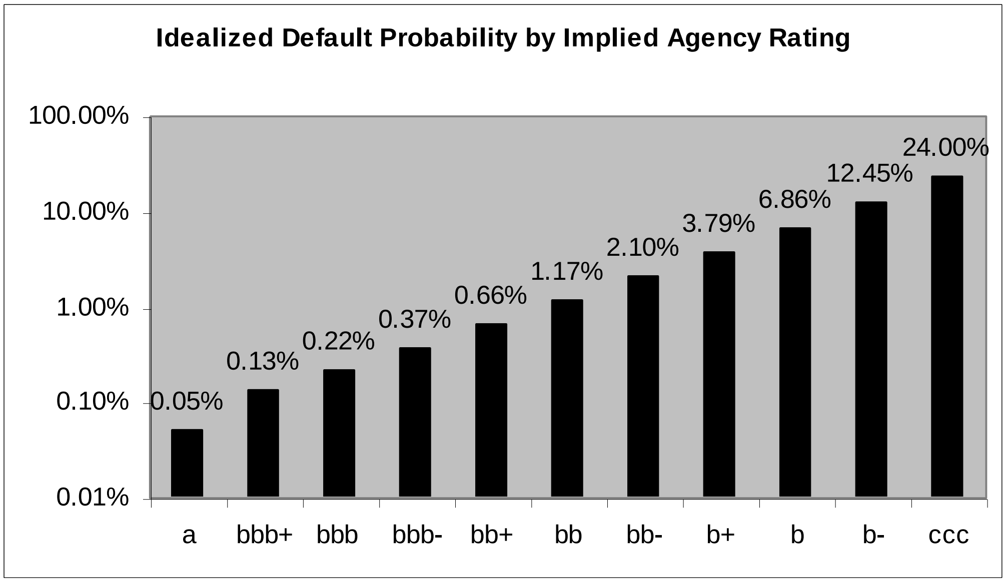
	1920-1969	1970-1988	1989-1999	2000-06
Aaa	0.00	0.00	0.00	0.00
Aa	0.09	0.03	0.05	0.00
A	0.14	0.01	0.00	0.05
Baa	0.34	0.20	0.10	0.32
Ba	1.02	1.21	1.39	0.85
B	1.90	6.15	7.00	4.06
Caa-C	6.12	24.87	23.47	21.22
Inv.Gr.	0.20	0.08	0.04	0.13
Spec.Gr.	1.87	2.83	4.79	5.78
All	0.98	0.85	1.59	2.05

Source: Moody's

One could argue for standard errors of 0.15% on any Baa estimate, or 2.0% on any B estimate. Moreover, merely omitting one year, or emphasizing the relevance of a particular credit cycle, will always have some considerable logic. It would be impossible to prove or refute that the expected default rate on B rated bonds is 5%, or even 7%, for the foreseeable future.

I assumed the following 'idealized default rates' for agency ratings and this underlies the mapping of the DefProb™ output to a model-implied agency rating. (Note the log scale.)

Figure 10.3



Section 11

Conclusion

DefProb™ is both simple and powerful. While human judgment will always be important, it is best left to qualitative or idiosyncratic considerations: the cumulative information from quantitative inputs via a model is hard to beat. The advantage of this model is that it uses multiple sources of information – equity market, financial statements, and agency ratings – in a very modest way that minimizes overfitting errors and is carefully mapped to a default probability that allows-apples-to-apples comparison.

You are invited to download an Excel Spreadsheet of DefProb™ outputs in order to see how it performs on actual listed companies (available at www.defprob.com).

References

Altman, Edward (1968), Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, Journal of Finance, 589-609, 23

Altman, Edward (1993), Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting and Avoiding Distress

Basel Committee (2005), Validation of low-default portfolios in the Basel II Framework, September, Bank for International Settlement

Bharath, Sreedhar and Tyler Shumway (2004), Forecasting Default with the KMV-Merton Model, AFA 2006 Boston Meetings Paper

Bharath, Sreedhar, and Tyler Shumway (2007), Forecasting Default with the Merton Distance to Default Model, Review of Financial Studies, 2007.

Bohn, Jeffrey, Navneet Arora, and Irina Korablex (2005), Power and Level Validation of the EDF Credit Measure in the US Market, MKMV White Paper,

Campbell, John, Jens Hilscher and Jan Szilagyi (2007), In Search of Distress Risk, *Journal of Finance*, forthcoming.

Chava, Sudheer and Robert Jarrow (2004), Bankruptcy prediction with industry effects, Review of Finance, 537--569, 8

Cifuentes, A and C Wilcox (1998), "The Double Binomial Method and Its Application to a Special Case of CBO Structures". Moody's Investors Service Structured Finance Special Report, March.

Cohen, Jacob (1990), Things I have learned (so far), American Psychologist 45, 1304-–1312, 45

Crosbie, Peter J. and Jeffrey R. Bohn (2003), Modeling Default Risk, KMV LLC, 0

CSFP, CUSP® 2007, http://www.csfb.com/institutional/research/cusp.shtml

Dawes, Robyn (1979), The Robust Beauty of Improper Linear Models in Decision Making, American Psychologist, 571--582, 34

Dawes, Robyn, and B. Corrigan (1974), Linear Models in Decision Making, Psychological Bulletin, 95--106, 81

Falkenstein, Eric (2002), Credit Scoring for Public Companies, Credit Ratings,

Falkenstein, Eric, Lea Carty and Andrew Boral (2000), RiskCalc for Private Companies: Moody's Default Model, Moody's Investors Service Special Comment

FitchRatings (2007), Fitch Equity Implied Rating and Probability of Default Model. Liu, Bo, Ahmet Kocagil, and Greg Gupton,

Lovie, A. and P. Lovie (1986), The Flat maximum Effect and Linear Scoring Models for Prediction, Journal of Forecasting, 159--168, 5

Merton, Robert (1974), On the pricing of corporate debt: the risk structure of interest rates, Journal of Finance, 449--470, 29

RiskMetrics (2002), Credit Grades Technical Document. Chris Finger et al. http://www.creditgrades.com/resources/pdf/CGtechdoc.pdf

Shumway, Tyler (2001), Forecasting bankruptcy more accurately: a simple hazard model, Journal of Business, 101--124, 74

Shumway, Tyler, and Vince Warther (1999), The Delisting Bias in CRSP's Nasdaq Data and its Implications for the Size Effect, Journal of Finance, December 1999, 2361-2379.

Sims, C. A. (1980). "Macroeconomics and Reality" Econometrica, 48 (10), pp.1-48.

Shumway, Tyler (1997), The Delisting Bias in CRSP Data, Journal of Finance, March, 327-340.
Sobehart, Jorge, Sean Keenan and Roger M. Stein (2000), Benchmarking quantitative default risk models: A validation methodology, Moody's Investors Service
Stock, James H. and Watson, Mark W. (1992), "A Procedure for Predicting Recessions With Leading Indicators: Econometric Issues and Recent Experience" (March 1992). NBER Working Paper Series, Vol. w4014, pp. -, Available at SSRN: http://ssrn.com/abstract=226874