



BondScore 3.0: A Credit Risk Model for Corporate Debt Issuers

September 2004

Susan K. Lewis
slewis@creditsights.com

Copyright © 2004 CreditSights, Inc. All rights reserved. Reproduction of this report, even for internal distribution, is strictly prohibited. The information in this report has been obtained from sources believed to be reliable. However, neither its accuracy and completeness, nor the opinions based thereon are guaranteed. If you have any questions regarding the contents of this report, contact CreditSights, Inc. at (212) 340-3840.

CreditSights Limited is regulated by The Financial Services Authority. This product is not intended for use in the UK by Private Customers, as defined by the Financial Services Authority

Summary

BondScore is CreditSights' set of quantitative tools for predicting the credit risk of publicly-held corporate debt issuers. This report describes the model that forms the core of BondScore, as well as enhancements in version 3.0. The BondScore 3.0 model estimates the probability that a US, non-financial issuer will default on public debt within a year, as a function of empirically-verified predictors including financial ratios and equity market variables. The model is structured to balance timeliness and stability, and to allow integration with company- or industry-level qualitative analysis.

The model is a statistically derived one (described in more detail below and in the appendix), which predicts the risk of default within a year as a function of seven ratios, derived both from companies' financial statements and from equity market data. We show that the original version of the model, based on data up through 2000, was quite effective at predicting defaults during the 2001-2003 default wave. The newly updated model is equally accurate in contemporaneous out-of-sample tests. As expected, greater margins, asset turnover, and liquidity reduce the risk of subsequent default. In contrast, higher leverage, long-term earnings volatility, and short-term equity volatility increase that risk.

The basic output of the model is a one-year forward estimate of default probability, known as a Credit Risk Estimate (CRE). CreditSights uses the BondScore 3.0 model to calculate CREs weekly for publicly-traded companies with assets over \$250 million. We also convert these scores to implied categorical ratings and spreads for ease of comparison with agency ratings and corporate bond market spreads. The model-derived scores and the ratings and spreads based on them can complement qualitative judgement by rating firms too small to merit ongoing qualitative coverage, by giving frequent updates of a company's status in between qualitative evaluations, and by providing a systematic algorithm for scoring risk that can be used in the context of an integrated analysis as a check and jumping-off point for in-depth discussion.

Summary of changes from the original version of the BondScore model

- We evaluated how successfully the original BondScore model predicted defaults during the wave of failures in 2001-2003, and found it performed well (see pages 11-13 for details.)
- The model has been reestimated to take advantage of the new data for 2001-2003. All factors in the original model remain significant predictors of default, and have effects roughly comparable to those in the original model, save for asset size (pp. 13-15)
- The equity volatility measure has changed to a one-year measure based on daily data, to enhance the model's immediacy and stability (pp. 10-11).
- We have changed the method of mapping of CREs to alphanumeric ratings from a static method to a dynamic one, to make BondScore ratings more directly and consistently comparable with agency ratings at a given point in time. Thus, BondScore 3.0 ratings are now designed for comparing companies to each other or to agency ratings at a given time, while CREs themselves are the best index of change in credit risk over time (pp. 16-17).
- We have released a new tool that uses CREs to calculate implied corporate bond spreads, or I-Spreads (described briefly on pages 17-18, and in more detail in a separate white paper).

BondScore is CreditSights' set of quantitative tools for predicting corporate bond issuers' defaults. This report describes the model that forms the core of BondScore, as well as enhancements in version 3.0. The BondScore 3.0 model estimates the probability that a US, non-financial issuer will default on public debt within a year, as a function of company financial ratios and equity market variables. It has several advantages over other tools for evaluating default risk: it is timely, stable, specific, and accurate.

- 1) **Timely.** Many contend that ratings agencies cannot respond quickly to changes in a firm's financial outlook (see, for example, "Ratings Firms—First to Know, Last to Say," Bloomberg News Service, Dec. 4, 2000). Credit scoring models, including the BondScore model, can be more current because they can be updated in a matter of hours or days when inputs change, while ratings can lag those changes. More timely credit quality evaluations are more accurate on average, all else being equal.
- 2) **Stable.** Timeliness is desirable, but not at the cost of excessive volatility. Some default risk predictions are based almost entirely on equity market movements, which can be more volatile than merited by changes in fundamentals (more on this below). The BondScore model, in contrast, takes advantage of the immediacy of equity market data but tempers it with direct measurement of fundamentals as well.
- 3) **Specific.** The BondScore model was created using rated corporate bond issuers—the large issuers which make up large concentrations of risk in many portfolios. Many credit scoring models are developed using a broad universe, including bank loan obligors and small, unrated companies. In such models, those types of obligors make up the vast majority of cases of default and therefore drive model results. Those obligors differ from rated public debt issuers in size, financial profile, and default risk, such that models driven by them are likely to be less accurate predictors for the larger names which default less often but which represent greater potential losses.
- 4) **Accurate.** Other credit scoring models have been shown to be better default predictors than are ratings (Kealhofer, Kwok, and Weng 1999; Miller 1998), and this is true for the BondScore model as well. The BondScore model also predicts defaults better than some published models (including Altman 1968) and at least as well as others (including Shumway 2001).

The following sections discuss existing default risk models, the procedure we used to develop and test ours, and the resulting BondScore 3.0 model. We also compare its ability to predict out-of-sample defaults to that of other methods.

Existing Default Risk Models

The best-known quantitative credit scoring models are Altman's Z-Score and KMV Corporation's implementation of a Merton-type model. Each exemplifies one of the two main approaches to modeling default risk. Altman's (1968) Z-Score was the first of the two to be used for default prediction. Models like the Z-Score are commonly called statistical models because they are based on empirical study of quantifiable correlates of default. In essence their strategy is to try to quantify the sort of examination of fundamental financial ratios that analysts have long used to gauge a company's health. Thus Altman focused on such ratios as default predictors: for example, liquidity measures like the current ratio, and turnover measures like the ratio of sales to total assets.

In contrast, default models that embody a theory of economic structure are known as structural models. The Merton (1974) model is the most widely used of these, often in the form of KMV's implementation (Vasicek 1984). Merton argued that equity

represents a call option on the assets of the firm, with a strike price that depends on the firm's debt level. Therefore, the equity holders will place the firm in default (decline to exercise the call) when the firm's asset value falls below a point determined by its debt level. This gap between the asset value and the default point, together with the volatility of the firm's asset value, determines the firm's default probability. The asset value and volatility can be derived from the firm's equity value and volatility.

Each of these two strategies has advantages and disadvantages. The Merton model is based on a compelling and intuitive theory, and because it uses equity market data it incorporates the most current information available about a firm's financial health. On the other hand, it produces unrealistically low expected default probabilities (Falkenstein 2000), such that KMV's adaptation must adjust them by mapping them to observed historical default data. And such models assume equity markets are efficient, accurately reflecting true changes in firms' conditions. But two decades of academic research convincingly demonstrate that this assumption is problematic: stock prices are more volatile than changes in fundamentals merit (Shiller 1981; Daniel, Hirshleifer, and Subrahmanyam 1998; Daniel and Titman 1999).

Statistical models like Altman's are based on direct measurement of fundamentals historically used to predict defaults. Because of this, their predictions naturally reflect historical experience more closely than those of the Merton model. This is not a claim about whether either strategy predicts better than the other going forward. Rather, the point is simply that statistical models require no post-hoc mapping to adjust their outputs to empirically observed default rates—instead, their mapping to observed default rates is integral to their estimation. Reliance on historical data is not costless, though: it produces model parameters that depend on the data used to estimate them, which would not be true of purely structural models.

No models in practical application, though, are purely structural. All are one of three types of hybrid of structural and statistical strategies. KMV's implementation of the Merton model is an implicit hybrid. Because it uses historical data to adjust its outputs, it produces default risk projections that depend on the data used to calibrate it, just as do statistical models. Moody's proprietary public firm default model (described in Sobehart and Stein [2000], but discontinued since Moody's purchase of KMV) is an explicit hybrid: what we will call a structurally focused statistical model. It centers on a Merton-like distance to default measure, but adds financial ratios. Finally, Shumway's (2001) published default model is also an explicit hybrid: what we will call a structurally informed statistical model. It mainly uses financial ratios to predict default risk, but also adds measures taken from equity markets, particularly equity volatility.

The BondScore model is of the third type: a structurally informed hybrid. This strategy reflects two judgments on our part: (1) equity markets and financial statements each provide information not completely captured in the other; and (2) the nature of the relationship between equity volatility and default risk is an empirical question, best determined by the data itself rather than specified a priori as in structurally-driven models.¹ The BondScore model uses financial ratios as well as equity prices and equity volatility to predict default. The next sections describe in more detail the

¹ The first judgment is based on the substantial literature, cited above, demonstrating excessive market volatility relative to fundamentals. We base our second judgment on the belief that while theory should guide models, systematic evaluation of the data must be the final arbiter when the theory's predictions are internally inconsistent or inconsistent with empirical observation (as the Merton model is—see Sobehart and Stein's [2000] discussion of the findings of Kim, Ramaswamy, and Sunderasan [1993] and Wei and Gou [1997]).

statistical form of BondScore and other models, and the specific measures they use to predict default.

Statistical form

The Merton model is not statistically estimated. It is a theoretically specified mathematical statement of relationships among the market value of a firm's equity, the value of its liabilities, and the volatility of its equity value. In contrast, statistical models specify only their component variables and the family of forms to which the relationships among them must belong, with the data being used to estimate the exact nature of those relationships. Statistical default models estimate those relationships using variants of regression-based techniques.²

The earliest, Altman's Z-score (1968), used discriminant analysis, a close analog of simple linear regression. A latent variable, L (the discriminant function), is a linear function of discriminating variables x_j and coefficients b_j that maximize the distance between the means of the discriminating variables. That is, $L = b_0 + b_1x_1 + b_2x_2 + \dots + b_jx_j$. The coefficients are usually estimated via ordinary least squares. The discriminating variables can be standardized along with the latent variable, then termed Z , as in Altman's model. Discriminant analysis shares with linear regression several assumptions (and adds assumptions of its own): predictors with multivariate normal distributions, normally distributed errors, linear relationships between independent and dependent variables, and a dependent variable with roughly similar-size groups. Violating those assumptions, as default data and many other kinds of data often do, can produce biased results.

In the decades since Altman published his Z-score model, logistic regression has replaced discriminant analysis as the preferred tool for dichotomous outcomes, largely because logistic regression has assumptions that are less restrictive, and so less frequently violated. The natural log of the odds of an event occurring conditional on independent variables x_j is a function of those independent variables and coefficients b_j that maximize the likelihood function of the data.³ That is, $\text{Log}_e(p / (1 - p)) = b_0 + b_1x_1 + b_2x_2 + \dots + b_jx_j$. This model allows for predictors that need not be normally distributed and that are nonlinear in their effects on the probability of an event, and for dependent variables with different sized groups. For more details, see the appendix.

Modern statistical default models generally use variants of logistic regression. In particular, they use a subset of a class of techniques variously called survival analysis, event history models, and hazard models—we will use the last name, and note that although names and emphasis differ, the underlying models are the same. Hazard models predict the probability of an event occurring, where that probability is a function of time. Subtypes of the broader class differ in how they parameterize time and how they account for change over time in covariates.

The particular subset of hazard models now often used to predict defaults is the discrete-time or piecewise exponential hazard model.⁴ It is mathematically identical to, and can be estimated as, a logistic regression applied to firm-year data. That is, for each firm, the independent variables are measured at the end of each fiscal year,

² Another possible technique for predicting default risk based on analysis of historical data is the use of neural networks. Neural nets have been intermittently fashionable as tools for financial prediction, but their adoption has been limited by their computational complexity and the fact that their applications intrinsically lack theoretical or even conceptual justification. Indeed, since neural nets are based purely on data-sifting rather than specification of rules or relationships, it can be difficult to interpret their results in a substantively sensible fashion.

³ For a discussion of maximum likelihood estimation, see Arminger (1995).

⁴ See Blossfeld, Hamerle, and Mayer (1989); or Yamaguchi (1991).

and used to predict whether the firm will default within the following fiscal year. The data contain a record for each firm for every year up to the one in which it defaults. If it never defaults, the firm contributes an annual record to the data for each year until it disappears (by purchase, for example) or until the data end.

Discrete-time hazard models have important advantages that make them ideal for the purpose of estimating default risk. They allow for differences in time to default. They are not biased by the fact that some firms that would eventually default are lost to observation before they do so (as long as the causes of censoring are unrelated to the causes of the event), whereas censoring would bias conventional regression models. And they allow for independent variables that change over time. It is these advantages that led Shumway (2001) and Sobehart & Stein (2000), and that lead us, to use this type of model to predict default risk.⁵

Predictors

Credit scoring models may seem to differ more in their predictors of default than in form, but this appearance is deceptive. While the variables used in these models are many, all tap three simple concepts: debt, returns, and risk. Models based on Merton's use a distance-to-default measure that combines these three phenomena in a pre-specified way. Distance to default, in their formulation, is a function of how many standard deviations (*risk*) the firm's equity value (*return*) is from its default point, where its equity value is the difference between the value of its assets and its liabilities (*debt*). Some models (e.g., Sobehart & Stein 2000) supplement that combined measure with financial ratios also intended to tap a firm's debt, return, and risk. Others disaggregate the three phenomena, measuring each separately, and specifying their relationship empirically according to the data rather than *a priori*.

Debt

The simplest of the three concepts is debt. Highly leveraged firms should be more likely to default than otherwise similar firms with less leverage. Measures of leverage include the ratio total liabilities to assets, or, more narrowly, total debt to assets. The debt-only construction of the numerator is predicated on the idea that not all liabilities are sufficiently legally binding to force a firm into default, and only those that are should be considered relevant predictors of default. Because leases are among those legally binding liabilities, some include capitalized lease costs, estimated as eight times annual rent costs, in the numerator of leverage ratios. And the denominator of a leverage ratio, instead of using the book value of the firm's assets, often uses the market value: that is, the sum of the binding obligations in the numerator and the firm's current equity market capitalization. Finally, in addition to leverage measures, some models include related measures of a firm's ability to service its debt in the short term, such as the interest coverage ratio (EBIT / interest) or the fixed charge coverage ratio: $(\text{EBIT} + \text{Lease payments}) / (\text{Interest} + \text{Lease payments} + [\text{Sinking fund payments} / \{1 - \text{Tax rate}\}])$.

Returns

Firms with higher returns should be less likely to default than otherwise similar firms with lower returns. "Returns," however, covers several related but not identical concepts. Measures of returns can be taken from equity markets, from financial statements, or from a combination of the two. Measures taken from financial statement alone are backward looking: returns over the past accounting period are captured by ratios like EBIT to assets or net income to assets. The latter can be decomposed instructively into its components, margins (net income to sales) and

⁵ Sobehart and Stein state that theirs is a nested logit formulation. The term "nested logit" refers to a set of variants of logistic regression, used for modeling processes that involve discrete choices among multiple outcomes in several steps.

asset turnover (sales to assets). Liquidity measures like the quick ratio are also related to past returns, albeit over a longer period, insofar as firms' current cash position is a product of past returns' accumulation over time. In contrast, measures based on current market valuations tend to give forward-looking assessments of returns. That is, they capture the market's best guess about future returns. Such measures include the market to book ratio and the ratio of enterprise value (market value of equity + total debt – cash) to assets. Leverage ratios that compare firms' debt load to their market value can also be seen as indexing the level of the firm's debt relative to the market's estimate of its likely future returns.

Risk

The riskier—that is, the more variable—firms' returns, the more often they will fall below the level needed to meet their obligations. So all else being equal, firms with more variable returns will be more likely to default. Just as returns can be measured by equity markets or financial statements, so can variation in those returns. Equity volatility can be measured as the standard deviation of monthly or daily stock returns over a given time (usually the past 12 or 24 months), or as the firm's residual volatility in stock returns (Shumway 2001), σ . The latter reflects equity volatility that is uncorrelated with broader market movements, and is calculated as the standard deviation of the residual of regressing the firm's monthly or daily return on the market's return. Measures of variability in returns based on financial statements may provide a more stable evaluation of long-term risk, given the above-cited evidence that stock returns are excessively volatile in relation to fundamentals. Altman et al.'s (1977) Zeta model includes the standard deviation of the past ten years' EBITDA to assets ratio. On the other hand, because volatility in stock returns can capture information not yet present in long-term earnings variation, both sorts of risk measures have value.

Ways of assessing model fit

In addition to being based on substantively sensible measures, any event prediction model must meet two quantitative criteria: (1) it must fit the data used to estimate it; and (2) it must predict the event with reasonably accuracy. The first criterion is a formal one. In the case of logistic regression based models the likelihood ratio statistic is the most widely used fit test: the difference between the log-likelihood function of the estimated model and that of a no-effects model, multiplied by -2 , yields a test, with a χ^2 distribution, of whether the model fits the data significantly better than chance guessing (Godfrey 1988).

The second criterion is quantitative but less formal: in what proportion of cases does the model correctly predict whether or not a firm defaults? What we really want to know is how well a model developed using the data available up to today will predict events in the future. That kind of prediction is impossible to evaluate prospectively. For example, only now can we know with certainty how well the original BondScore model (developed using data up to November 2000) did at predicting defaults during the years 2001 to 2003. We will discuss that specific example later, since those years turned out to be particularly interesting ones, but even certain knowledge about that period does not help us evaluate how well the original model would do in the next three years from now. Nor can it tell how well the new model will do in the future.

Given the logical impossibility of truly prospective testing, three strategies are commonly used to instead test models' contemporaneous out-of-sample performance—the next best thing. The simplest, and usually most conservative, out-of-sample testing method is to divide the available data a priori into estimation and holdout samples. The first sample is used to determine the variables that will enter the model and calculate the coefficients of those variables. The second sample is then used to evaluate the performance of the model developed in the first step. In the

case of the BondScore model, we divided the data randomly into a development sample of 2/3 of the data and a holdout sample of 1/3. Unless otherwise specified, parameters reported below for the model itself (and used in applications of the model) are from the 2/3 estimation sample, while tests are on the 1/3 holdout sample.

A variant of this method, a K-fold test, is sometimes used instead. This method starts by calculating the model on all available data. The K-fold test then divides the data into K equal subsets by random assignment, recalculates the model coefficients for the selected variables using to the first of those subsets, then uses those model coefficients to score the cases in other K-1 subsets. These steps are repeated for each of the K subsets. In each case, the scores are based on model coefficients derived only from data that excluded that case. Finally, results of all predictions are aggregated and evaluated. In the K-fold test employed below we divide the sample into three equal groups.

The logic of the walk-forward test is similar, with the exception that the data are divided chronologically rather than randomly. In the walk-forward test below, for example, we begin with the first six years of data, calculate model coefficients, score the remainder of the data using those coefficients, then repeat the process, adding a year's worth of data with each successive iteration.

All of these testing strategies ensure that the predictions being evaluated are based at least on coefficients not derived from the same cases being predicted. The first method is yet more conservative than the other two, in that it ensures even the initial variable selection does not rely on data that is also used in testing. Predicting defaults well in a holdout sample does not guarantee that a model will perform well on all new data to which it is applied. But evaluating a model's performance on a holdout sample at least identifies and allows users to reject models that fit only the data on which they are developed.

Whichever of the three methods is used to create a set of out-of-sample predictions, the next step is to evaluate how well the model does at those out-of-sample predictions. Models of the type used here (and in most other recent default studies) produce a probability. So to identify firms as likely defaulters requires setting a probability cutoff, above which firms are classified as likely future defaulters and below which they are not. How often the model correctly identifies firms that subsequently default, and how often it identifies as defaulters firms that actually do not (false positives), depends on where that cutoff is set.

The goal is maximizing sensitivity (the proportion of failures accurately predicted) while maintaining specificity (the proportion of non-failures accurately predicted). There is a tradeoff between sensitivity and specificity. The higher the probability cutoff is set, the more false negatives and the fewer false positives. The lower it is set, the reverse. The model with the best predictive accuracy overall is one that minimizes this tradeoff. In other words, a good model correctly identifies most defaults without incorrectly identifying many non-defaults.

The receiver operating characteristic (ROC) curve provides a useful and well-accepted way to evaluate models on this criterion. The ROC curve plots the proportion of failures correctly predicted against the proportion of non-failures incorrectly predicted, as the probability cutoff is varied. When the number of failures is small relative to the number of cases, this is essentially identical to what is described elsewhere (see Falkenstein 2000) as the Cumulative Accuracy Profile (CAP) or power curve. The CAP curve plots the percentage of defaults that would be excluded (on the vertical axis) by ranking the scores by estimated default risk and excluding a given percentage of the sample (on the horizontal axis).

Figure 1A: Example power curve, random-guess model

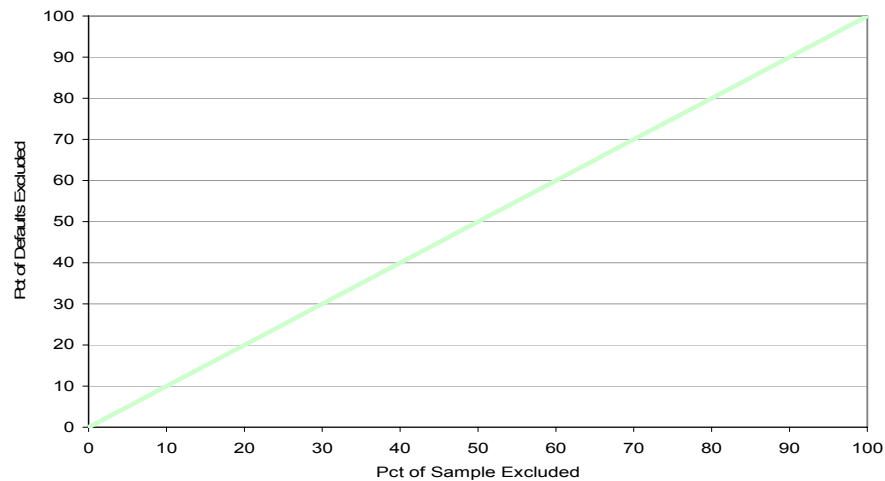


Figure 1B: Example power curve, model with moderate fit

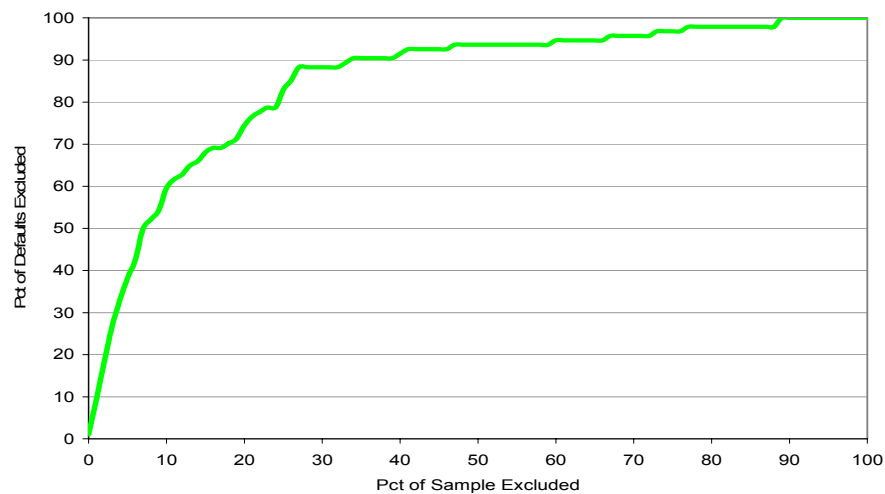
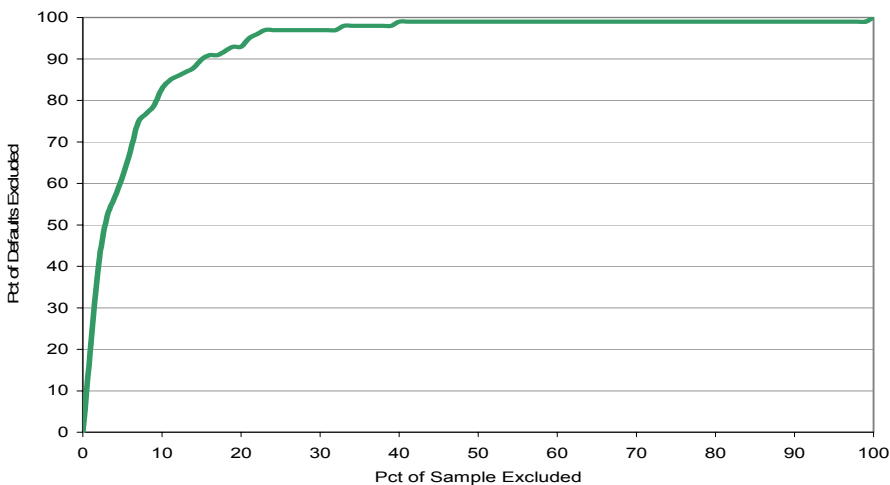


Figure 1C: Example power curve, model with good fit



Whichever of these two methods is used to plot the curve, for models with greater predictive accuracy the plotted line has a steeper curve and greater area below it. Figure 1A, with half the area below the power curve, illustrates a model equivalent to random guessing. Figure 1B, with 86% below the curve, is a better model. Figure 1C, with 95% below the curve, has the greatest overall predictive accuracy of the three. The area below the power curve can serve as an index of how successfully a model predicts defaults while simultaneously holding false positives to a minimum.

The BondScore 3.0 Model

Predictors

The variables in the BondScore model are intended to capture the three constructs of debt, returns, and risk. The goal is not just to choose the model that best fits the data (although it does, in fact, fit the data well), but to design a substantively sensible model of the forces that push a firm toward or away from default. Firms default when they are unable to meet their obligations: a state triggered by the size of those obligations relative to the level of their resources available to meet them, and how much those resources vary.

The model measures leverage by the ratio of a firm's debt to the market value of its assets. Our measure does not include all liabilities because not all liabilities are sufficiently legally binding to force a firm into default. Rather, we limit the numerator to the sum of long term debt, short term debt, and capitalized leases (estimated as eight times the most recent year's rent). The denominator of this ratio is the sum of the firm's equity market capitalization and the liabilities in the numerator.

The resources available to meet those obligations include expected future returns, recent returns and cash accumulated as a result of past returns. The debt to market value ratio just described captures expected future returns in the denominator. The BondScore model measures recent returns using information from financial statements: the firm's margins, or EBITDA / sales; and its asset turnover, or sales / assets. The quick ratio indexes liquidity, the accumulated results of past returns on the firm's current cash position. Additionally, we include a measure of the firm's relative size (the ratio of the firm's assets to the average assets of all firms) since there has historically been evidence that smaller firms are more likely to default, all else equal, probably because larger firms have better access to short term funding that can help them weather temporary liquidity crises.

BondScore 3.0 includes two measures of risk, one based long-term variation in earnings and the second based on relatively short-term equity movements. Our first measure of business risk, emphasizing the long term, is the standard deviation of the firm's EBITDA to assets ratio over the past ten years.⁶ The second measure, σ , is measured in BondScore 3.0 as the residual volatility in the firm's daily stock returns (as described above) over the past year. In our original version of the BondScore model, we likewise used a measure of residual stock return volatility, differing only in that it reflected monthly returns over two years rather than daily returns over one year. The switch to a shorter-duration daily measure allows both greater immediacy and greater stability in capturing the equity markets' information about firm risk.

Users of the original version of BondScore should note that switching to a shorter-duration daily measure of risk has three implications that are worthwhile to understand. First, changing the term and periodicity of the residual equity volatility measure changes its scale, so that any given company's equity volatility on the old measure is not interchangeable with its equity volatility on the new measure. Second,

⁶ If available. If fewer years are available, we use as few as five, or, lacking five years, the industry mean.

a company's rank on the new measure may differ from its rank on the old measure. For example, a company whose equity returns have been very stable in the most recent year but were quite volatile in the year before that would be judged relatively less volatile under the new measure than under the old one.

And third, in times of declining market volatility, the average credit risk estimates produced by the new model can be lower than those produced by the old model, because the new equity volatility emphasizes the more recent, less volatile history. That pattern holds, for example, at the time of this writing (August 2004), when the most recent year's equity market volatility is lower than the preceding year's. Conversely, in times of increasing volatility, average CREs can be higher under the new model than under the old. In short, on average CREs will respond more quickly to sustained trends in equity market volatility under the new model than the old.

Data

Estimating a model with the variables just listed requires data on firm's annual financial statements, monthly returns and corresponding index returns, and defaults of rated issuers, all for a long enough period to cover multiple business cycles. Financial statements are available in Compustat. CRSP provides daily returns for most Compustat firms, along with corresponding index returns. Our default records cover all rated firm known to have defaulted on public debt, and are derived from multiple sources. Each fiscal year's financial and returns data are used to predict defaults occurring within the one-year period following the fiscal year end.

In creating the original BondScore model, we used data for 1975 up through November 2000, imposing the lower time limit because default records are sparse before that date. For BondScore 3.0, we have reestimated the model, updating its coefficients using data up to December 2003. The BondScore model is designed for, and thus the data limited to, US-incorporated, non-financial firms that carry a senior debt rating. 224 such firms with complete financial and returns data defaulted by November 2000, and 268 by December 2003.

Like most financial variables, the financial statement inputs to the model from Compustat and the stock returns-based inputs from CRSP can have highly nonnormal distributions, with a few cases having quite extreme values on some variables. Uncorrected, these outliers would dominate the model, producing estimates that fit best for those firms but are less accurate for the large majority of firms. To minimize the effect of outliers we transform the variables used in the BondScore model to their percentiles.

How well did the original model do?

Before we turn to describing and evaluating the fit of the new model, we must evaluate how well the original model predicted defaults after it was created, particularly since the years immediately following its release, 2001-2003, saw an unusually large number of defaults. Some, like Enron, Worldcom, and others, were characterized by accounting fraud and correspondingly distorted equity market signals. Other companies defaulting in these years, even in the absence of accounting fraud, were atypically large compared with earlier defaulters. So this, perhaps more than any other recent time, should challenge the ability of quantitative models based on accounting ratios and/or equity market signals to successfully distinguish companies that were likely to default.

The original BondScore, like the updated model, was created using companies with agency-issued debt ratings, in order to optimize the model's fit to companies that issue the largest pools of liquid debt (rather than small companies that would otherwise be the majority of defaults and so dominate the model). We have also provided subscribers with BondScore Credit Risk Estimates (CREs) for

a broader universe that includes both rated and unrated companies with assets above \$250 million. This lets subscribers evaluate potential debt issuers on the same basis as already-rated ones, but it also means it is necessary to judge how well the model predicted defaults in this broader universe, not just for rated issuers.

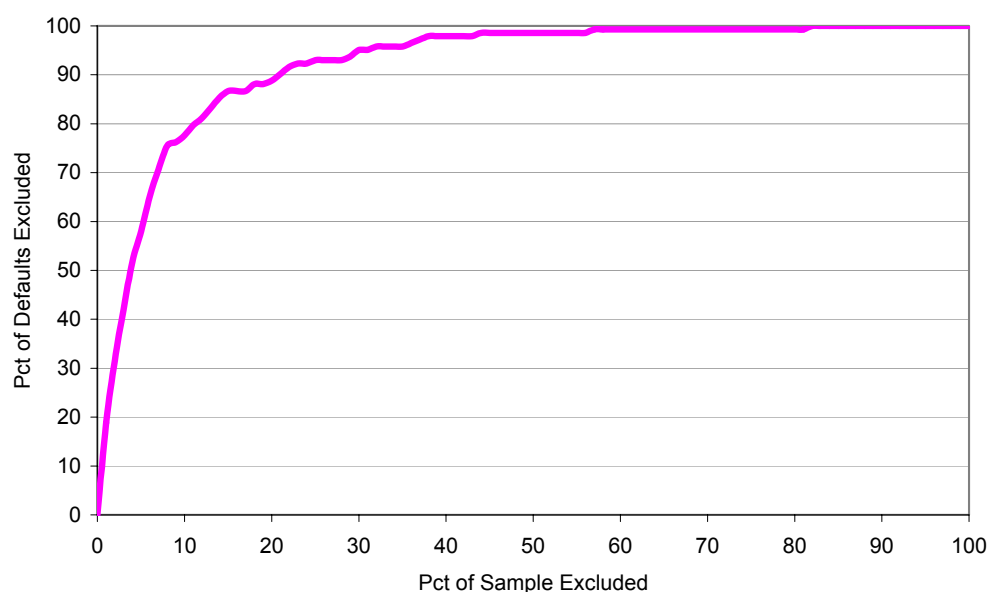
We identified 143 companies that defaulted between 2001 and the end of 2003, within a year of having been scored in the data published for BondScore subscribers.⁷ If customers used the published BondScore CREs to classify any company with a CRE worse than 0.5% as probable to default, they would have correctly identified 95% of those that did go on to default within a year (but would also have incorrectly identified 29% of those that did not default in a year). At a more reasonable threshold of a 1% CRE, the corresponding numbers would have been 92% and 22%. Table 1, at right, lists the original model's sensitivity (D, or the percent of defaults correctly identified) at varying CRE cutoffs, along with the proportion of companies that did not default within a year misidentified at each cutoff.

Table 1: Predictive accuracy of original BondScore model in 2001-03

Percent of defaults correctly identified (D) and non-defaults misidentified (M) at varying cutoffs.

<i>CRE > X%</i>	<i>D</i>	<i>M</i>
0.5	95	29
1.0	92	22
1.5	89	19
2.0	88	17
2.5	87	14
3.0	87	13
3.5	83	12

Figure 2: Power curve, original BondScore model on 2001-2003 data



⁷ We made every effort to ensure this is as complete a listing as possible, using several sources to identify defaults. Completeness cannot be guaranteed because some companies may not have been scored, some whose defaults were reported may have been impossible to match to corresponding records in our data, or some that defaulted may not have been reported in any of our sources. But this list should be a fairly comprehensive set of those companies with assets over \$250 million that were scored and that defaulted during this period.

To create a power curve as described above, we sorted all the companies scored during 2001 to 2003 by their scores at the beginning of each year, and determined what percentage of the 143 defaults would have been excluded by excluding specified percentages of the cases. For example, excluding only the worst-scoring ten percent of cases—that is, those with CREs worse than about 5%—excluded 78% of the defaults. The power curve in Figure 2 summarizes this balance over the range of possible thresholds. The area below the power curve, 93%, indicates that the original BondScore model was quite successful at identifying defaults without sacrificing an unreasonably large amount of specificity in the form of false positives.

Results from the updated model

Table 2 gives the median and interquartile range for the ratios in the data used to update the BondScore model. Converting the ratios to percentiles and re-estimating the model as described above yields the results reported in Table 3. Each variable significantly influences the chance of default (with one exception, discussed below), and does so in the expected fashion, as the z-statistics in Table 3 show (z's of 1.86 or greater [absolute value] indicate statistical significance at conventional levels). Greater leverage increases the default rate, higher margins decrease it, and greater asset turnover decreases it. Firms with more liquidity are less likely to default. Firms with more volatile earnings and more volatile stock returns default more often. The model provides a good fit to the data used to estimate it, as the log likelihood shows.

Table 2: Descriptive statistics for BondScore 3.0 variables

	<u>25th %ile</u>	<u>Median</u>	<u>75th %ile</u>
Debt / market value	.29	.45	.63
EBITDA / sales	.08	.14	.21
Sales / assets	.59	1.00	1.42
Quick ratio	.73	1.02	1.41
Log of relative asset size	-2.08	-1.04	.02
Volatility of EBITDA / assets	.021	.034	.055
σ	1.50	2.04	2.92
Defaults		268	
Firm-years		14,215	

The one exception to the rule that the inputs to the model are all significant predictors of default remains in the BondScore 3.0 model for purposes of comparison with the original BondScore model, and because it makes an instructive, albeit unsurprising, point. In the data up to November 2000 that we used to create the original BondScore model, larger companies (measured by the log of their assets relative to average assets of other companies) were less likely to default. When we add the data for the next three years, that effect disappears—size has no significant effect on default risk. Considering that a notable proportion of the wave of defaults in those years were large companies (Worldcom, US Air, etc.), this makes intuitive sense. A statistical model systematically ties prediction to the guidance of history: when historically established patterns change, so must the model.

How well does the updated BondScore 3.0 model do at predicting defaults while not incorrectly rejecting companies that will not ultimately default? Within the 2/3 of the available data used for estimation purposes—that is, for the 9,581 company-years and 187 defaults used to derive the model coefficients—the model does a good job of distinguishing defaulters from non-defaulters. The area under the in-sample power curve (the pink line in Figure 3) is 94%. The model's predictive ability on the one-third

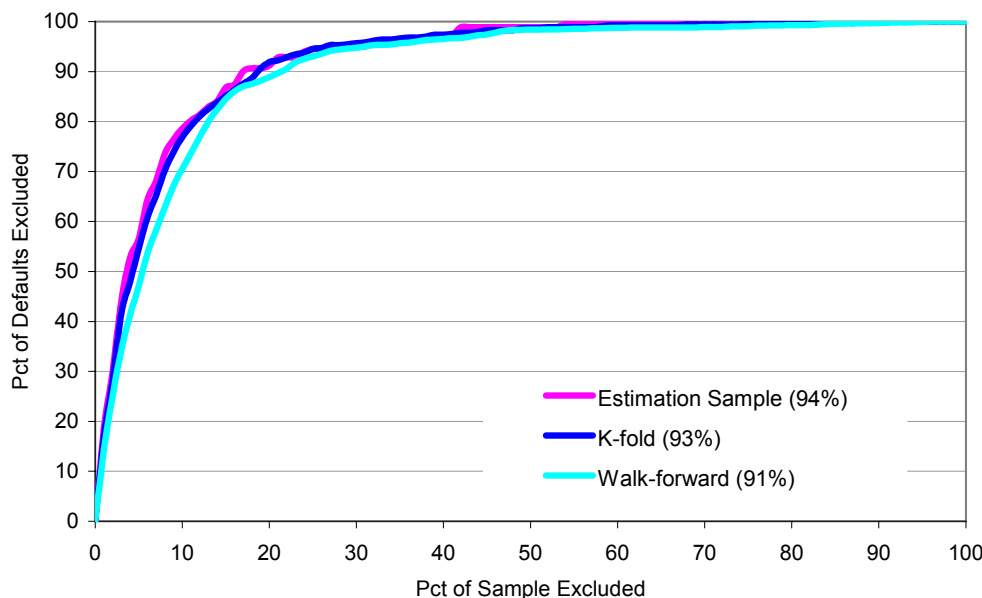
holdout sample is nearly the same, with the area under that power curve at 93% (not shown). Similarly, the K-fold and walk-forward tests whose power curves are shown on Figure 3 indicate that the model has good out-of-sample predictive power. The area under the K-fold test's power curve is 93%, and for the walk-forward test it is 91%. So over a range of possible probability cutoffs for marking a firm as a likely defaulter, the model does a good job of correctly identifying both firms that default and those that do not.

If all firms with more than a 1% chance of defaulting are marked as likely defaults, the model identifies 95% of future defaults in the holdout sample but misidentifies 23% of non-defaults. If the cutoff is moved to 1.5% the model identifies 91% of defaults but misidentifies 19% of non-defaults. The model retains fairly good sensitivity over the range of substantively sensible cutoffs tested, at a 3% cutoff accurately predicting 86% of events and mistakenly predicting events in 13% of non-default cases.

Table 3: The BondScore 3.0 model

	Z
Debt / market value	9.21
EBITDA / sales	-2.51
Sales / assets	-4.68
Quick ratio	-2.86
Log of relative asset size	0.25
Volatility of EBITDA / assets	3.48
σ	6.60
Constant	-12.96
Log-likelihood	622.44
Defaults	268 (total) 187 (estimation sample)
Firm-years	9,581

Figure 3: Power curves for BondScore 3.0 model



Comparison with other models

We evaluated the ability of the BondScore 3.0 model to predict defaults going forward one year in the 1/3 holdout sample relative to several other tools: ratings alone, an early and well known statistical model (Altman 1968), and a recent model published in the academic literature (Shumway 2001). Using the originally published

coefficients for the other two quantitative models (Altman's and Shumway's) gave fairly poor predictive power in our sample. Therefore, we reestimated these two models using the current estimation sample and used those reestimated coefficients to allow at least a variant of these models to have the best possible opportunity to fit the current testing sample. Table 4 reports the results of those comparisons. The table contains the percent of defaults correctly identified and non-defaults misidentified by each model, at each of several substantively sensible probability cutoffs for classifying a firm a likely default. The overall predictive accuracy of each model as given by the area under the power curve is reported in parentheses. (For purposes of this table, ratings are S&P ratings).

Table 4: Predictive accuracy of BondScore 3.0 versus other models

Percent of defaults correctly identified (D) and non-defaults misidentified (M) in holdout sample, at varying cutoffs. Area under power curve in parentheses.

<i>CRE ></i>	<i>BondScore (93%)</i>		<i>Ratings (88%)</i>		<i>Z-Score (88%)</i>		<i>Shumway (91%)</i>	
<i>X%</i>	<i>D</i>	<i>M</i>	<i>D</i>	<i>M</i>	<i>D</i>	<i>M</i>	<i>D</i>	<i>M</i>
0.5	96	29	98	48	92	54	93	35
1.0	95	23	98	42	89	41	91	25
1.5	91	19	94	36	89	32	90	20
2.0	90	17	91	26	87	26	86	16
2.5	88	15	91	26	84	21	84	10
3.0	86	13	73	17	81	16	85	14
3.5	85	12	73	17	79	13	83	12

Both recent quantitative models (BondScore 3.0 and Shumway) are more accurate default predictors in the holdout sample than Altman's (1968) Z-Score—compare the areas under their respective ROC curves, 93% and 91%, to that of the earlier model, 88%. This is expected, since the newer models have access to 30 years' more data and modern statistical techniques in the variable selection stage. Those three models all also surpass ratings at predicting one-year forward default rates.⁸ That may be in part due to the fact that ratings can become less accurate over time as firms' credit quality changes faster than agencies' ability to keep up with those changes.

Extensions of the BondScore Model

The fundamental output of the BondScore model is an estimate of the probability that a firm will default on any debt obligation in the following year. For brevity, in applications of the model we refer to this as a Credit Risk Estimate or CRE. CreditSights generates and distributes current CREs from the BondScore model for approximately 2,400 companies as of this writing. The CREs are generated by applying the model described above to normalized ratios based on the last twelve months' financial data for the companies⁹, along with the long-term earnings volatility and one-year (up to the date of measurement) equity volatility described above. Companies' current CREs, the history of those CREs, and the inputs that determine them are published on the BondScore website. The website also offers a calculator

⁸ For this model, we represented detailed S&P rating as a 21-category ordinal variable in a logistic regression on default rates. Of the several parameterizations of ratings we tried, this one predicted default best. The percent of defaults predicted and non-defaults mispredicted remain constant over a range of cutoffs, as shown in the table, because ratings are step-functional rather than continuous.

⁹ LTM income statement items are the sum of the last four quarters. LTM balance sheet items average the last four quarters, except for long-term debt. Only the most recent quarter's long-term debt is used, because by its nature long-term debt is "sticky"—that is, current levels of long-term debt contain more information about the future than do past levels.

for scenario analysis that lets users estimate the effect on credit risk of, for example, an acquisition, debt increase, off balance sheet liability, etc. Finally, information derived from the model is distributed in a desktop tool—BondScore Reports—that permits maintenance of portfolios and creation of risk reports based on user-specified criteria. These tools are described in more detail elsewhere.

Ratings

While the Credit Risk Estimate is the fundamental output of the BondScore model, for users' convenience we also provide an alphanumeric rating derived from the CRE. The step of mapping a default risk estimate to a rating is not intrinsic to a default risk model, and is grounded in heuristic guidelines rather than precise statistical tests, but it is useful because it gives subscribers a way to compare model-estimated default risk to a familiar metric, agency ratings. In the original version of our BondScore model applications, we designed the method of mapping BondScore CREs to BondScore ratings to ensure that each BondScore rating reflected a constant range of risk levels over time. We did this by determining the median CRE of companies with specified agency ratings in the past, selecting fixed cutoffs between those medians, and assigning current CREs to ratings according to those cutoffs.¹⁰

However, agency ratings are not likewise fixed to specific, unchanging risk levels. Companies with a given agency rating have had varying default rates over time, and during cyclical downturns, default rates can rise before the distribution of agency ratings shifts. This has the effect of creating distributions of BondScore ratings that can differ from distributions of agency ratings over time. So although a set of constant-risk ratings has advantages for some purposes, it has the disadvantage of making it harder to compare companies' BondScore ratings to their agency ratings over time, as the default risk reflected by given agency ratings has varied. Therefore, at the request of users, we have changed the design of our BondScore ratings to ensure more consistent comparability between agency and BondScore ratings, even if the underlying risk reflected within those ratings categories changes.

In BondScore 3.0, CREs are still the fundamental model output and ratings are still derived from them. But the mapping from CREs to ratings is now recalibrated weekly and separately by broad industry groupings we call ratings sectors (eleven groups, in categories like industrials, consumer discretionary, and so on). The first step in this mapping process is determining, for each ratings sector, the distribution of agency ratings. We calculated this from monthly records of S&P ratings available in Compustat for the ten-year period from mid-1993 to mid-2003 (chosen because it is relatively recent, yet long enough to provide a sufficient frequency of even fairly uncommon ratings and to span a period of generally high ratings and of lower ratings). The distribution is calculated on a firm-month basis. That is, a firm rated for the entire period would contribute 120 observations; one rated BBB+ for the first year and A- thereafter would contribute 12 observations to the frequency of BBB+'s in the ratings distribution and 108 observations to the frequency of A-'s.

The second step is to apply this distribution to the current week's CRE's for agency-rated companies. This is done by sorting the scored, agency-rated companies by CRE, then assigning ratings proportional to the ratings distributions derived as above. For example, if 2% of company-years in the ten-year ratings distribution for a given

¹⁰ This was the general guideline, but it was necessary to depart from it for deriving CRE cutoffs between some ratings categories, because for some adjacent agency ratings categories the better agency rating has a worse average CRE. This is equally true for actual historical default frequencies: counterintuitively, some agency ratings categories actually have greater one-year historical default frequencies than the adjacent worse ratings. That is, while default risk generally increases with agency rating, that increase is not consistently monotonic.

ratings sector were AAA's, the agency-rated companies in that sector with CREs in the best 2% in a given week are assigned a BondScore rating equivalent of AAA.

Finally, the BondScore CREs and ratings assigned to scored, agency-rated companies are used to derive BondScore ratings for the scored companies that are not rated by the agencies. This is done by setting cutoffs at the midpoints between the CREs of the nearest adjacent agency-rated companies assigned different BondScore ratings. For example, if the worst-scoring agency-rated company assigned a BondScore rating of AAA has a CRE of 0.003% while the best-scoring agency-rated company assigned a BondScore rating of AA1 has a CRE of 0.005%, the cutoff between AAA and AA1 for companies without agency ratings will be set at 0.004%. Note that the BondScore ratings distribution for companies without agency ratings can differ from the BondScore ratings distribution for agency-rated companies, or from the agency ratings distribution itself.

A weekly calibration of CREs to the long-term distribution of agency ratings ensures that for agency-rated names, the distribution of BondScore ratings will always be similar to the distribution of agency ratings in the same rating sector over the medium to long term. It also means that BondScore ratings now reflect moving risk bands in a way that mirrors agency ratings, rather than constant risk bands. Thus, the new BondScore ratings are optimized for direct comparison to agency ratings, or for comparing companies at a point in time. CREs, in contrast, are the best metric for examining the change in a company's risk over time.

Users of the original version of BondScore, with its static ratings mappings, should be aware of the implications of this new dynamic ratings mapping method. The broadest implication is that the meaning of BondScore ratings has changed from a reflection of absolute risk to a reflection of relative risk. That broad change gives rise to more specific phenomena: under the new dynamic mappings, companies with similar CREs may be assigned different BondScore ratings, if they are in different ratings sectors. A company's rating may stay the same when its CRE worsens, if the scores of other companies in the sector also worsen. These phenomena happen by design, and are consistent with similar phenomena in agency ratings: the level of default risk represented by a rating may vary by sector, and over time. While such patterns provide an advantage for comparability between BondScore ratings and agency ratings, the original static method of mapping provides advantages for other types of analysis. For subscribers who wish to continue using that static method to instead map CREs to fixed ratings that reflect constant risk bands, we will provide a set of static breakpoints on request that can be used to create a static mapping.

I-Spread

While the BondScore CRE can be used to estimate the default risk of corporate bonds and the BondScore rating gives a shorthand comparison tool, portfolio managers, traders, and other credit market practitioners also need a risk measure that translates CRE change into portfolio value and potential P&L. Such a measure helps users quantify risk and easily perform better rich/cheap relative value analysis. CreditSights' new I-Spread model, released with BondScore 3.0, meets this need by translating a company's CRE to an implied spread of an individual bond.

The model uses a hybrid approach to the corporate bond valuation problem (further detailed in a forthcoming white paper to be released soon on the BondScore website). It uses a structural model as the valuation framework to give the model a sound theoretical basis for equating default risk with pricing. The structural model uses the option theory based methodology pioneered by Merton. However we extend the model so that it does not take the default probability from the structural model. Instead it takes as input the BondScore CRE as the default probability. This approach

combines the best of structural and statistical models by fully using all available information to derive empirical default risk estimates then taking advantage of the mathematical relationships specified in structural theory to convert those default risk estimates to spreads. The I-Spread model also implements a simple recovery rate model that uses Moody's historical bond recovery rate statistics from 1982-2003. The modeled recovery rate depends on the sector, quality and seniority of the bond.

The I-Spread model replaces our existing Fair Value Curves model. There are some common aspects and major differences between these two models. They both translate CRE to bond spread, using market spreads from Merrill indices to calibrate model parameters. However, the I-Spread model leverages a well-understood theoretical model, where the FVC model was purely empirical. Because of this, the I-Spread model can give spreads for a company at different maturity points, that is, it can give the term structure of the spread curve of a company. Thus, the I-Spread model allows the user to compare the model-derived spread with the market spread of a specific bond. The FVC, in contrast, only gave an average issuer-level spread for one maturity point. The added capabilities of the I-Spread model are significant improvements over the FVC and make it a better tool for BondScore users to integrate CREs into risk management and bond valuation analysis.

Closing Comments

Because the BondScore 3.0 default model does a good job of classifying firms in cases not used to develop it, predicting the large majority of defaults while maintaining a comparatively low rate of false positives, we believe it is a useful tool for evaluating credit quality going forward. Analysts use the BondScore CREs generated by this model as a screen to flag potential problems for more in-depth examination; they use the scenario analysis it facilitates to quantify the likely credit risk impact of events like acquisitions and new debt issues; they compare market spreads to the spreads implied by companies' model-estimated credit risk to seek trading opportunities. This integration of scoring model and qualitative analysis allows us to offer our subscribers the breadth of an algorithmic approach, within the context of intelligent case-by-case analyses that interpret its results and consider the effects of anomalies and events that may not systematically appear in model inputs.

Appendix A: The Statistical Model

We apply a statistical model to historical data on companies' characteristics and later defaults. The particular form of statistical model we use is known as a discrete-time event history model. Event history models (or hazard models, as they are also known) are designed to predict the risk of an event occurring, as a function of specified variables measured before the event occurs. There are several forms of event history models, but the one most appropriate for our purposes—and most akin to the familiar linear regression model—is the discrete time model. It predicts the risk of an event occurring within a year (or some other specified, fixed time). It is mathematically equivalent to, and is estimated via, applying a logistic regression to firm-year data (that is, data containing a separate record for each year for each firm at risk of the event, up to the year in which the event occurs).

The logistic regression takes the following form:

$$\log\left(\frac{p}{1-p}\right) = \sum_{k=1}^K \beta_k x_k$$

where p is the probability of the event occurring, and the K independent variables, x , are each weighted by a coefficient, β . This differs from the conventional linear regression model,

$$y = \sum_{k=1}^K \beta_k x_k$$

in that it predicts not the value of the dependent variable, y , but rather the natural logarithm of the ratio of the probability of the event occurring to the probability of it not occurring (what is known as the logit function, which gives this form of regression its name). The logit function is preferable to the linear regression function because it limits p , the probability of the event, to be between 0 and 1 (or zero and 100%), where applying the conventional linear regression to dichotomous outcomes would instead allow nonsensical results like probabilities greater than 1 or less than zero.

As a simple mathematical transformation of the first equation above, the logistic model gives the probability of the event as:

$$p = \frac{e^{\sum_{k=1}^K \beta_k x_k}}{1 + e^{\sum_{k=1}^K \beta_k x_k}}$$

where e is Euler's constant, the base of the natural logarithm. This difference in structure between this model and more familiar linear regression models gives rise to differences in how the results of the model can be interpreted. Conventional linear regression models have a simple, linear, additive form. Changing one independent variable by a fixed amount changes the level of the dependent variable by an amount that is identical, no matter what the levels of the other independent variables. For example, if we used a linear regression to predict the risk of a company failing, whether its current risk was 2% or 10%, decreasing its ratio of net income to assets by a fixed quantity would increase its risk of failure by the same amount—let's say 1%, for the sake of argument—to, respectively 3% or 11%. In contrast, in a logistic regression, a change in one factor changes the risk by an amount that is proportional to the level of the other factors. If the other factors produce a failure probability of 2%,

decreasing the ratio of EBITDA to assets might double the risk to 4%, whereas if the initial risk is 10% the same decrease in EBITDA to assets would double the risk to 20%. This is both statistically more appropriate to the data than a linear additive model, and substantively more sensible: the amount of additional risk a company incurs with a fixed change in any given input should depend on the company's other characteristics.¹¹

As with any regression model, an estimating algorithm is used to find the coefficients (β 's) that best satisfy the relationship expressed in the regression equation for the estimation data sample. The technique used to find those coefficients for logistic regression, as for most nonlinear regression models, is known as maximum likelihood estimation. Because the details of maximum likelihood estimation are routine and unlikely to be of interest to most readers, we leave them to standard statistical texts (for example, see Wonnacott and Wonnacott 1979). Suffice it to say that "the principle of maximum likelihood is to choose as parameter estimates those values that maximize the likelihood (probability) of observing the data that have actually been observed," (Allison 1984: 67). In other words, an algorithm iteratively tries coefficients until it finds the set that maximizes the value of a mathematical function that gives the joint probability of observing the given data.

That function, L , the likelihood function, also forms the basis of a statistical test of how well the model fits the observed data:

$$L^2 = 2(\log L_T - \log L_B)$$

where L_T is the likelihood function of the tested model and L_B is the likelihood function of a baseline model containing only an intercept term. L^2 is known as the likelihood ratio chi-square test, and this statistic can be compared with standard chi-square tables to determine whether the tested model fits significantly better than a baseline model. The same logic can be used to compare two models, one of which includes a smaller subset of the variables in the other, to establish whether the added variables significantly improve the fit to the data, or whether conversely the smaller subset is equally sufficient.

¹¹ This also means questions that are simple to ask in the context of the conventional linear regression model are not sensible in the context of this kind of logistic model. For example, users of linear regression models frequently calculate the percent change in the predicted value of the dependent variable due to one unit change in one of the independent variables. In a logistic regression, that value depends both on the current level of that variable and on the levels of the other variables. It changes as they change, such that there are no general answers to this question—only answers that are specific to the hypothetical situation of only that variable changing while all else remains constant. While this sacrifices the convenient simplicity of the linear model, in doing so it more accurately reflects the complexity of the real world.

References

- Altman, E.I. (1968), "Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy," *Journal of Finance* 23, pp. 589-609.
- Altman, E.I., R. Haldeman, and P. Narayanan (1977), "ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations," *Journal of Banking and Finance* 10, pp. 29-54.
- Arminger, G. (1995), "Specification and estimation of mean structures," in eds. G. Arminger, C.C. Clogg, and M.E. Sobel, *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, New York: Plenum, pp. 77-183.
- Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998), "Investor Psychology and Security Market Under- and Over-Reactions," *Journal of Finance* 53, pp. 1839-86.
- Daniel, K. and S. Titman (1999), "Market Efficiency in an Irrational World," *Financial Analysts Journal* 55 (6), pp. 28-40.
- Godfrey, L.G. (1988), *Misspecification Tests in Econometrics*, New York: Cambridge University Press.
- Kealhofer, S., S. Kwok, and W. Weng (1998), Uses and Abuses of Bond Default Rates," *CreditMetrics Monitor* (First quarter) pp. 37-55.
- Kim, J. K. Ramaswamy, and S. Sunderasan (1993), "Does Default Risk in Coupons Affect the Valuation of Corporate Bonds? A Contingent Claims Model," *Financial Management* 117-131.
- Merton, R.C. (1974), "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *Journal of Finance* 29 pp. 449-70.
- Miller, R. (1998), "Refining Ratings." *RISK*, August.
- Shiller, R.J. (1981), "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review* 71, pp. 421-98.
- Shumway, T. (forthcoming 2001), "Forecasting Bankruptcy More Accurately: A Simple Hazard Model," *Journal of Business*.
- Sobehart, J. and R. Stein (2000), "Moody's Public Firm Risk Model: A Hybrid Approach to Modeling Short Term Default Risk," Moody's Investors Service, March.
- Vasicek, O.A. (1984), "Credit valuation," KMV Corporation.
- Wei, D.G. and D. Gou (1997), "Pricing Risky Debt: An Empirical Comparison of the Longstaff and Schwartz, and Merton Models," *The Journal of Fixed Income* 7, pp. 8-28.