Forecasting Default with the Merton Distance to Default Model

Sreedhar T. Bharath
Ross School of Business, University of Michigan

Tyler Shumway
Ross School of Business, University of Michigan

We examine the accuracy and contribution of the Merton distance to default (DD) model, which is based on Merton's (1974) bond pricing model. We compare the model to a "naïve" alternative, which uses the functional form suggested by the Merton model but does not solve the model for an implied probability of default. We find that the naïve predictor performs slightly better in hazard models and in out-of-sample forecasts than both the Merton DD model and a reduced-form model that uses the same inputs. Several other forecasting variables are also important predictors, and fitted values from an expanded hazard model outperform Merton DD default probabilities out of sample. Implied default probabilities from credit default swaps and corporate bond yield spreads are only weakly correlated with Merton DD probabilities after adjusting for agency ratings and bond characteristics. We conclude that while the Merton DD model does not produce a sufficient statistic for the probability of default, its functional form is useful for forecasting defaults. (*JEL* G12, G13, G33)

1. Introduction

Due to the advent of innovative corporate debt products and credit derivatives, academics and practitioners have recently shown renewed interest in models that forecast corporate defaults. One innovative forecasting model, which has been widely applied in both academic research[1] and practice, is a particular application of Merton (1974) that was developed by the proprietors of the KMV corporation.[2] We refer to this model as the Merton distance to default model,

---

Previously circulated with the title "Forecasting Default with the KMV-Merton Model." This research was supported by the NTT fellowship of the Mitsui Life Center at the Ross School of Business, University of Michigan. We thank seminar participants at Michigan, Boston College, Indian School of Business, Moody's 2006 Credit Risk Conference, and Stanford. We also thank Bill Beaver, Jeff Bohn, Darrell Duffie, Eric Falkenstein, Wayne Ferson, Ravi Jagannathan, Kyle Lundstedt, Ken Singleton, and Jorge Sobehart for their comments. All errors are ours. Address correspondence to either Sreedhar T. Bharath and Tyler Shumway, Ross School of Business, University of Michigan, 701 Tappan Street, E7606, Ann Arbor, MI 48109; telephone: (734) 763-0485; e-mail: sbharath@umich.edu or Tyler Shumway, Ross School of Business, University of Michigan, Michigan Business School, 701 Tappan Street, Ann Arbor, MI 48109-1234; telephone: (734) 763-4129; e-mail: shumway@umich.edu.

[1] The model is discussed in Duffie and Singleton (2003) and Saunders and Allen (2002). It is applied by Vassalou and Xing (2004); Duffie, Saita, and Wang (2007); and Campbell, Hilscher, and Szilagyi (2007), among others.

[2] We do not intend to imply that we are using exactly the same algorithm that Moody's KMV (which acquired KMV in 2002) uses to calculate distance to default. Differences between our method and that of Moody's KMV are discussed in Section 2.2 and in Table 2.

doi:10.1093/rfs/hhn044 Advance Access publication May 19, 2008

or the Merton DD model. This paper assesses the accuracy and the contribution of the Merton DD model.

The Merton DD model applies the framework of Merton (1974), in which the equity of the firm is a call option on the underlying value of the firm with a strike price equal to the face value of the firm's debt. The model recognizes that neither the underlying value of the firm nor its volatility is directly observable. Under the model's assumptions both can be inferred from the value of equity, the volatility of equity, and several other observable variables by using an iterative procedure to solve a system of nonlinear equations. After inferring these values, the model specifies that the probability of default is the normal cumulative density function of a z-score depending on the firm's underlying value, the firm's volatility, and the face value of the firm's debt.

The Merton DD model is a clever application of classic finance theory, but how well it performs in forecasting depends on how realistic its assumptions are. The model is a somewhat stylized structural model that requires a number of assumptions. Among other things, the model assumes that the underlying value of each firm follows geometric Brownian motion and that each firm has issued just one zero-coupon bond. If the model's strong assumptions are violated, it should be possible to construct a reduced-form model with more accuracy.

We examine three hypotheses in this paper. First, we ask whether the probability of default given by the Merton DD model is a sufficient statistic for forecasting bankruptcy. If the Merton model is literally true, it should be impossible to improve on the model's implied probability for forecasting. If it is possible to construct a reduced-form model with better predictive properties, we can conclude that the probability implied by the Merton DD model (π_{Merton}) is not a sufficient statistic for forecasting default.

The second hypothesis we test is that the z-score functional form used by the Merton DD model is an important construct for predicting default. We hypothesize that the probability calculated with the z-score functional form cannot be completely replaced in a forecasting model by a linear combination of simple variables, including the variables used to calculate the probability. In other words, we test whether a sufficient statistic for default probability can be calculated without considering the Merton DD model's functional form.

Our third hypothesis is that the solution of the Merton model is important for forecasting default. As discussed above, the Merton DD probability is calculated by solving the classic Merton model for the total value of the firm and the firm's volatility (given the value and volatility of the firm's equity) and substituting these values into the z-score functional form. By testing this hypothesis, we ask whether the forecasting ability of the Merton DD model is sensitive to the manner in which total firm value and firm volatility are calculated. Again, we effectively test whether a sufficient statistic for default probability can be constructed without considering the Merton DD algorithm for calculating these values.

We test these three hypotheses in five ways. First, we incorporate π_{Merton} into a hazard model that forecasts defaults from 1980 through 2003. With the hazard model, we compare π_{Merton} to a naïve alternative ($\pi_{\text{naïve}}$) which is much simpler to calculate, but mimics the functional form of π_{Merton}. This helps us to separate the relative importance of the functional form from the solution procedure. We also compare it to several other default forecasting variables. Second, we compare the short-term, out-of-sample forecasting ability of π_{Merton} to that of $\pi_{\text{naïve}}$. Third, we examine the forecasting ability of several alternative predictors, each of which calculates Merton DD model probabilities in a slightly different way. Fourth, we examine the ability of the Merton DD model to explain the probability of default implied by credit default swaps; and fifth, we regress corporate bond yield spreads on π_{Merton}, $\pi_{\text{naïve}}$ and other variables.

Assessing the Merton DD model's value is important for two reasons. Perhaps the most salient reason is that many researchers and practitioners are applying the model and yet we do not know very much about its statistical properties. For example, Vassalou and Xing (2004) use π_{Merton} to examine whether default risk is priced in equity returns. As a second example, the Basel Committee on Banking Supervision (1999) considers exploiting the Merton DD model (and its commercial implementations) a viable practice currently employed by numerous banks. To have confidence in both the risk management of the banking sector and the accuracy of academic research, the power of the Merton DD model must be examined.

A second reason to assess the Merton DD model is to test the classic Merton (1974) model in a new way. If the classic Merton model is literally true, π_{Merton} should be the best default predictor available. The classic Merton model has been rejected previously for failing to fit observed bond yield spreads.[3] Comparing the model to reduced-form alternatives gives us a fresh perspective about how realistic the model's assumptions are.

Over the past several years, a number of researchers have examined the contribution of the Merton DD model. The first authors to examine the model carefully were practitioners employed by either KMV or Moody's. Some papers, including Stein (2000); Sobehart and Stein (2000); Sobehart and Keenan (1999, 2002a, 2002b); and Falkenstein and Boral (2001), argue that Merton DD models can easily be improved upon. Other papers, including Kealhofer and Kurbat (2001), argue that the Merton DD-like model originally developed by the KMV corporation captures all of the information in traditional agency ratings and well-known accounting variables. An academic literature has also recently developed that critically assesses the model. Both Hillegeist et al. (2004) and Du and Suo (2004) examine the model's predictive power in ways that are similar to some of our analyses. Duffie, Saita, and Wang (2007) show that Merton DD probabilities have significant predictive power in a model of default probabilties over time, which can generate a term structure of default

[3] See Jones et al. (1984).

probabilities. Campbell, Hilscher, and Szilagyi (2007) estimate hazard models that incorporate both π_{Merton} and other variables for bankruptcy, finding that π_{Merton} seems to have relatively little forecasting power after conditioning on other variables. While our findings are consistent with the findings of all of these papers, we analyze the performance of π_{Merton} in several novel ways. In particular, we introduce and assess our naïve predictor and we examine the ability of π_{Merton} to explain credit default swap premiums and bond yield spreads.

We find that it is fairly easy to reject hypothesis one, that π_{Merton} is not a sufficient statistic for default probability. It is possible to improve upon π_{Merton} by conditioning on other default prediction variables. However, we find some support for hypothesis two. We find that using the *z*-score functional form implied by the Merton model in $\pi_{\text{naïve}}$ improves our forecasting ability. In hazard models that include both $\pi_{\text{naïve}}$ and the variables that are used to construct $\pi_{\text{naïve}}$, the naive probability remains statistically significant. More impressive, the out-of-sample forecasting performance of $\pi_{\text{naïve}}$ by itself is slightly better than that of a reduced-form hazard model that incorporates the variables used to calculate $\pi_{\text{naïve}}$. Finally, we find some evidence against hypothesis three. The contribution of π_{Merton} to a well-specified reduced-form model that includes $\pi_{\text{naïve}}$, or the variables used to construct $\pi_{\text{naïve}}$, is fairly low. We conclude that while π_{Merton} has some predictive power for default, most of the marginal benefit of π_{Merton} comes from its functional form rather than from the solution of the Merton model.

The paper proceeds as follows. The next section details the Merton DD model, our naïve alternative default probability, and the hazard models that we use to build reduced-form models. Section 2 also lists several ways in which our Merton DD model differs from the model that Moody's KMV actually sells. Section 3 discusses the data that we use for our tests and Section 4 outlines our results. We conclude in Section 5.

2. Default Forecasting Models

As discussed above, we examine our hypotheses by examining the statistical and economic significance of the Merton DD model default probabilities (π_{Merton}) and a simple, naïve alternative ($\pi_{\text{naïve}}$). Before examining the empirical value of these variables, we need to describe them carefully.

2.1 The Merton DD model

The Merton DD model produces a probability of default for each firm in the sample at any given point in time. To calculate the probability, the model subtracts the face value of the firm's debt from an estimate of the market value of the firm and then divides this difference by an estimate of the volatility of the firm (scaled to reflect the horizon of the forecast). The resulting *z*-score, which is sometimes referred to as the distance to default, is then substituted

into a cumulative density function to calculate the probability that the value of the firm will be less than the face value of debt at the forecasting horizon. The market value of the firm is simply the sum of the market values of the firm's debt and the value of its equity. If both these quantities were readily observable, calculating default probabilities would be simple. While equity values are readily available, reliable data on the market value of firm debt is generally unavailable.

The Merton DD model estimates the market value of debt by applying the classic Merton (1974) bond pricing model. The Merton model makes two particularly important assumptions. The first is that the total value of a firm follows geometric Brownian motion,

$$dV = \mu V dt + \sigma_V V dW, \tag{1}$$

where V is the total value of the firm, μ is the expected continuously compounded return on V, σ_V is the volatility of firm value and dW is a standard Wiener process. The second critical assumption of the Merton model is that the firm has issued just one discount bond maturing in T periods. Under these assumptions, the equity of the firm is a call option on the underlying value of the firm with a strike price equal to the face value of the firm's debt and a time-to-maturity of T. Moreover, the value of equity as a function of the total value of the firm can be described by the Black-Scholes-Merton Formula. By put-call parity, the value of the firm's debt is equal to the value of a risk-free discount bond minus the value of a put option written on the firm, again with a strike price equal to the face value of debt and a time-to-maturity of T.

Symbolically, the Merton model stipulates that the equity value of a firm satisfies

$$E = V\mathcal{N}(d_1) - e^{-rT}F\mathcal{N}(d_2), \tag{2}$$

where E is the market value of the firm's equity, F is the face value of the firm's debt, r is the instantaneous risk-free rate, $\mathcal{N}(\cdot)$ is the cumulative standard normal distribution function, d_1 is given by

$$d_1 = \frac{\ln(V/F) + \left(r + 0.5\sigma_V^2\right)T}{\sigma_V\sqrt{T}}, \tag{3}$$

and d_2 is just $d_1 - \sigma_V\sqrt{T}$.

The Merton DD model makes use of two important equations. The first is the Black-Scholes-Merton Equation (2), expressing the value of a firm's equity as a function of the value of the firm. The second relates the volatility of the firm's value to the volatility of its equity. Under Merton's assumptions the value of equity is a function of the value of the firm and time, so it follows directly from

Ito's lemma that

$$\sigma_E = \left(\frac{V}{E}\right)\frac{\partial E}{\partial V}\sigma_V. \tag{4}$$

In the Black-Scholes-Merton model, it can be shown that $\frac{\partial E}{\partial V} = \mathcal{N}(d_1)$, so that under the Merton model's assumptions, the volatilities of the firm and its equity are related by

$$\sigma_E = \left(\frac{V}{E}\right)\mathcal{N}(d_1)\sigma_V, \tag{5}$$

where d_1 is defined in Equation (3).

The Merton DD model basically uses these two nonlinear equations, (2) and (5), to translate the value and volatility of a firm's equity into an implied probability of default. In most applications, the Black-Scholes-Merton model describes the unobserved value of an option as a function of four variables that are easily observed (strike price, time-to-maturity, underlying asset price, and the risk-free rate) and one variable that can be estimated (volatility).[4] In the Merton DD model, however, the value of the option is observed as the total value of the firm's equity, while the value of the underlying asset (the total value of the firm) is not directly observable. Thus, while V must be inferred, E is easy to observe in the marketplace by multiplying the firm's shares outstanding by its current stock price. Similarly, in the Merton DD model, the volatility of equity, σ_E, can be estimated but the volatility of the underlying firm, σ_V, must be inferred.

The first step in implementing the Merton DD model is to estimate σ_E from either historical stock returns data or from option-implied volatility data. The second step is to choose a forecasting horizon and a measure of the face value of the firm's debt. For example, it is common to use historical returns data to estimate σ_E, assume a forecasting horizon of 1 year ($T = 1$), and take the book value of the firm's total liabilities to be the face value of the firm's debt. The third step is to collect values of the risk-free rate and the market equity of the firm. After performing these three steps, we have values for each of the variables in Equations (2) and (5) except for V and σ_V, the total value of the firm and the volatility of firm value, respectively.

The fourth, and perhaps most significant, step in implementing the model is to solve Equation (2) numerically for values of V and σ_V. Once this numerical solution is obtained, the distance to default can be calculated as

$$DD = \frac{\ln(V/F) + \left(\mu - 0.5\sigma_V^2\right)T}{\sigma_V\sqrt{T}}, \tag{6}$$

[4] Of course, it is common to infer an implied volatility from an observed option price.

where μ is an estimate of the expected annual return of the firm's assets. The corresponding implied probability of default, sometimes called the expected default frequency (or EDF), is

$$\pi_{\text{Merton}} = \mathcal{N}\left(-\left(\frac{\ln(V/F) + \left(\mu - 0.5\sigma_V^2\right)T}{\sigma_V\sqrt{T}}\right)\right) = \mathcal{N}(-DD). \qquad (7)$$

If the assumptions of the Merton model really hold, the Merton DD model should give very accurate default forecasts. In fact, if the Merton model holds completely, the implied probability of default defined above, π_{Merton}, should be a sufficient statistic for default forecasts. Testing this hypothesis is one of the central tasks of this paper.

Simultaneously solving Equations (2) and (5) is reasonably straightforward. However, Crosbie and Bohn (2003) explain that "In practice the market leverage moves around far too much for [Equation (5)] to provide reasonable results." To resolve this problem, we follow Crosbie and Bohn (2003) and Vassalou and Xing (2004) by implementing a complicated iterative procedure. First, we propose an initial value of $\sigma_V = \sigma_E[E/(E+F)]$ and we use this value of σ_V and Equation (2) to infer the market value of each firm's assets every day for the previous year. We then calculate the implied log return on assets each day and use that returns series to generate new estimates of σ_V and μ. We iterate on σ_V in this manner until it converges (so the absolute difference in adjacent σ_Vs is less than 10^{-3}). Unless specified otherwise, in the rest of the paper values of π_{Merton} are calculated by following this iterative procedure and calculating the corresponding implied default probability using Equation (7).

The Merton DD model is a rather unusual forecasting model. Most forecasting models constructed by econometricians involve posing a model and then estimating the model with method of moments or maximum-likelihood techniques. The Merton DD model actually involves very little estimation. Instead, it replaces estimation with something more like calibration–solving for implied parameter values. Since the Merton DD model is not a typical econometric model, it is not clear how the model can be extended to consider additional default predictors or how its parameters might be estimated with alternative techniques. It is also unclear how standard errors for forecasts can be calculated for the Merton DD model. While we propose a few simple alternative ways to calculate Merton DD probabilities below, applying more standard econometric methods to the model seems like a promising topic for future research.

Before describing alternative models, it is useful to interpret the Merton DD model a little. The most critical inputs to the model are clearly the market value of equity, the face value of debt, and the volatility of equity. As the market value of equity declines, the probability of default increases. This is both a strength and weakness of the model. For the model to work well, both the Merton model assumptions must be met and markets must be efficient and well informed.

Moody's KMV markets a model that is similar to our Merton DD model in a number of ways. In its promotional material, Moody's KMV points to the case of Enron as an example of how their method is superior to that of traditional ratings agencies. When it became clear that Enron had serious accounting problems, Enron's stock price began to fall and its distance to default immediately decreased. The ratings agencies took several days to downgrade Enron's debt. Clearly, using equity values to infer default probabilities allows the Merton DD model to reflect information faster than traditional agency ratings. However, before Enron's accounting problems were well known, when Enron's stock price was arguably unsustainably high, the expected default frequency (calculated by both Moody's and Merton's methods) for Enron was actually significantly lower than the default probability assigned to Enron by standard ratings. If markets are not perfectly efficient, then conditioning on information not captured by π_{Merton} probably makes sense.

2.2 Our method versus Moody's KMV

We should point out that there are a number of things which differentiate the Merton DD model, which we test from that actually employed by Moody's KMV. One important difference is that we use Merton's (1974) model while Moody's KMV uses a proprietary model that they call the KV model. Apparently the KV model is a generalization of the Merton model that allows for various classes and maturities of debt. Another difference is that we use the cumulative normal distribution to convert distances to default into default probabilities. Moody's KMV uses its large historical database to estimate the empirical distribution of changes in distances to default and it calculates default probabilities based on that distribution. The distribution of distances to default is an important input to default probabilities, but it is not required for ranking firms by their relative probability. Therefore, several of our results will emphasize the model's ability to rank firms by default risk rather than its ability to calculate accurate probabilities.[5] Finally, Moody's KMV may also make proprietary adjustments to the accounting information that they use to calculate the face value of debt. We cannot perfectly replicate the methods of Moody's KMV because several of the modeling choices made by Moody's KMV are proprietary information.

While our method does not match that of Moody's KMV exactly, it is the same method employed by Vassalou and Xing (2004); Duffie, Saita, and Wang (2007); Campbell, Hilscher, and Szilagyi (2007), and other researchers. Our results can be considered relevant for a "feasible" Moody's KMV-like model, which can be estimated and implemented by academic researchers. In order to compare our method with that of Moody's KMV, in Section 4 we compare our estimates with the estimates produced by Moody's KMV for a sample of large firms in the US.

[5] If the model ranks firms accurately then using historical data to map relative rankings into accurate probabilities is a straightforward task.

2.3 A naïve alternative

To test whether π_{Merton} adds value to reduced-form models, we construct a simple alternative "probability" that does not require solving Equations (2) and (5) by implementing the iterative procedure described above. We construct our naïve predictor with two objectives. First, we want our naïve predictor to have a reasonable chance of performing as well as the Merton DD predictor, so we want it to capture the same information the Merton DD predictor uses. We also want our naïve probability to approximate the functional form of the Merton DD probability. Second, we want our naïve probability to be simple, so we avoid solving any equations or estimating any difficult quantities in its construction. We wrote down the form for our naïve probability after studying the Merton DD model for a little while. None of the numerical choices below is the result of any type of estimation or optimization.

To begin constructing our naïve probability, we approximate the market value of each firm's debt with the face value of its debt,

$$\text{naïve } D = F, \tag{8}$$

Since firms that are close to default have very risky debt, and the risk of their debt is correlated with their equity risk, we approximate the volatility of each firm's debt as

$$\text{naïve } \sigma_D = 0.05 + 0.25 * \sigma_E. \tag{9}$$

We include the five percentage points in this term to represent term structure volatility, and we include the 25% times equity volatility to allow for volatility associated with default risk. This gives us an approximation to the total volatility of the firm of

$$\begin{aligned} \text{naïve } \sigma_V &= \frac{E}{E + \text{ naïve } D}\sigma_E + \frac{\text{naïve } D}{E + \text{naïve } D}\ \text{naïve } \sigma_D \\ &= \frac{E}{E+F}\sigma_E + \frac{F}{E+F}(0.05 + 0.25 * \sigma_E). \end{aligned} \tag{10}$$

Next, we set the expected return on the firm's assets equal to the firm's stock return over the previous year,

$$\text{naïve } \mu = r_{it-1}. \tag{11}$$

This allows us to capture some of the same information that is captured by the Merton DD iterative procedure described above. The iterative procedure is able to condition on an entire year of equity return data. By allowing our naïve estimate of μ to depend on past returns, we incorporate the same information. The naïve distance to default is then

$$\text{naïve } DD = \frac{\ln[(E+F)/F] + \left(r_{it-1} - 0.5\ \text{naïve } \sigma_V^2\right)T}{\text{naïve } \sigma_V \sqrt{T}}. \tag{12}$$

This naïve alternative model is easy to compute; however, it retains the structure of the Merton DD distance to default and expected default frequency. It also captures approximately the same quantity of information as the Merton DD probability. Thus, examining the forecasting ability of this quantity helps us separate the value of solving the Merton model from the value of the functional form of π_{Merton}. We define our naïve probability estimate as

$$\pi_{\text{naïve}} = \mathcal{N}(-\text{naïve } DD). \tag{13}$$

It is fairly easy to criticize our naïve probability. Our choices for modeling firm volatility are not particularly well motivated and our decision to use past returns for μ is arbitrary at best. However, to quibble with our naïve probability is to miss the point of our exercise. We have constructed a predictor that is extremely easy to calculate, and it may have significant predictive power. If the predictive power of our naïve probability is comparable to that of π_{Merton}, then presumably a more carefully constructed probability that captures the same information should have superior power.

2.4 Alternative predictors

One purpose of our paper is to examine the relative importance of several of the components of the Merton DD calculation. Comparing the predictive performance of our naïve probability to that of π_{Merton} is one way to accomplish this. Another way we accomplish this purpose is by examining the predictive performance of several alternative predictors, or predictors that calculate Merton DD probabilities in alternative, somewhat simpler ways.

One predictor, $\pi^{\mu=r}_{\text{Merton}}$, is calculated in exactly the same manner as π_{Merton}, except that the expected return on assets used for π_{Merton} is replaced by the risk-free rate, r. Considering this predictor helps us gauge the importance of estimating the expected return on assets for the distance to default. A second alternative predictor, $\pi^{simul}_{\text{Merton}}$, is calculated by simultaneously solving Equations (2) and (5). This predictor avoids the iterative procedure in the text, estimating equity volatility with 1 year of historical returns data and using r as the expected return on assets. The third alternative predictor, $\pi^{imp\sigma}_{\text{Merton}}$, uses the option-implied volatility of firm equity (implied σ_E) to simultaneously solve Equations (2) and (5).

2.5 Hazard models

In order to assess the Merton DD model's accuracy, we need a method to compare π_{Merton} to alternative predictor variables. We employ a Cox proportional hazard model to test our three hypotheses. Hazard models have recently been applied by a number of authors and probably represent the state of the art in default forecasting with reduced-form models.[6] Proportional hazard models

[6] Shumway (2001) and Chava and Jarrow (2004) argue that hazard models are superior to other types of models.

make the assumption that the hazard rate $\lambda(t)$ or the probability of default at time t conditional on survival (lack of default) until time t is

$$\lambda(t) = \phi(t)[\exp(x(t)'\beta)], \tag{14}$$

where $\phi(t)$ is referred to as the "baseline" hazard rate and the term $\exp(x(t)'\beta)$ allows the expected time to default to vary across firms according to their covariates, $x(t)$. The baseline hazard rate is common to all firms. Note that in this model the covariates may vary with time. Most of our default predictors, including π_{Merton}, vary with time. The Cox proportional hazard model does not impose any structure on the baseline hazard $\phi(t)$. Cox's partial likelihood estimator provides a way of estimating β without requiring estimates of $\phi(t)$. It can also handle censoring of observations, which is one of the features of the data. Details about estimating the proportional hazard model can be found in many sources, including Cox and Oakes (1984).

Our first hypothesis, that π_{Merton} is a sufficient statistic for forecasting default, implies that no other variable in a hazard model should be a statistically significant covariate. Our second hypothesis, that the functional form of the Merton DD model is useful, implies that either $\pi_{\text{naïve}}$ or π_{Merton} should remain statistically significant in a hazard model that includes all the variables used to calculate these quantities. Our third hypothesis, that the algorithm specified by the Merton DD model to calculate the inputs to π_{Merton} is important, implies that π_{Merton} should remain a statistically significant default predictor in our hazard models regardless of the other variables that we include in the models. As a robustness check, we will also examine the out-of-sample performance of our models.

2.6 Implied probabilities of default

In addition to examining the default prediction ability of the Merton DD model, we examine its ability to explain the variation in two market-based default probability variables. We regress both the implied probability of default from credit default swaps (CDS) and the yield spread on corporate bonds on π_{Merton} and $\pi_{\text{naïve}}$. While there is a large literature on explaining bond yield spreads, using CDS data to assess default probabilities is relatively new. Other recent papers that use CDS data include Longstaff, Mithay, and Neis (2005) and Berndt et al. (2004) (referred to as BDDFS hereafter).

Credit default swaps are one example of credit derivatives, and credit derivative markets have experienced explosive growth in recent years. According to the British Bankers' Association, the total notional principal for outstanding credit derivatives increased from $180 billion in 1997 to more than $2 trillion by the end of 2002 and it is expected to have reached $8.4 trillion by the end of 2006. Popular credit derivatives such as the credit default swap allow market participants to trade credit risks with each other. We use the information in

CDS to extract a direct measure of default probabilities and compare it with the estimates obtained from our methods.

In a credit default swap, the party buying credit protection pays the seller a fixed premium until either default occurs or the swap contract matures. In the event of a default, because these payments are made in arrears, a final accrual payment by the buyer is required. In return, if the underlying firm (the reference entity) defaults on its debt, the protection seller is obligated to buy back from the buyer the defaulted bond at its par value. The payoff from a credit default swap is simply one minus the recovery rate, which is the loss given default for every dollar of notional principal. Thus a CDS is similar to an insurance contract compensating the buyer for losses arising from a default.

Let s be the CDS spread, which is the amount paid per year as a percentage of the notional principal. Most CDS contracts have a maturity of 5 years. Let T determine the life of the CDS contract. Further, assume that the probability of a reference entity defaulting during a year conditional on no earlier default is π_{CDS}. For simplicity, we assume that defaults always happen halfway through the year, and the payments on the CDS are made once a year, at the end of each year. Thus the final accrual payment will be made halfway through the year and will be equal to $0.5s$. We also assume that the risk-free (LIBOR) rate is r with continuous compounding and the recovery is δ.

Thus the expected present value of the payments made on the CDS (assuming a notional principal of \$1) is given by

$$\sum_{t=1}^{t=T}(1-\pi_{\text{CDS}})^{t}e^{-rt}s+(1-\pi_{\text{CDS}})^{t-1}\pi_{\text{CDS}}e^{-r(t-0.5)}0.5s. \tag{15}$$

The first term represents the discounted present value of the expected payments made at the end of each year provided the reference entity survives until period t and the second term represents the present value of the accrual payments made in the case of a default assuming default happens midway through the year.

Similarly, the expected present value of the payoff is given by

$$\sum_{t=1}^{t=T}(1-\pi_{\text{CDS}})^{t-1}\pi_{\text{CDS}}(1-\delta)e^{-r(t-0.5)}. \tag{16}$$

We need an implied estimate of recovery rate δ in order to value the payoff. The same recovery rate is typically used to (a) estimate implied default probabilities and (b) value the CDS. Hull, Predescu, and White (2004) argue that the net result of this is that the value of a CDS (or the estimate of a CDS spread) is not very sensitive to the recovery rate. This is because implied probabilities of default are approximately proportional to $1/(1–\delta)$ and the payoffs from a CDS are proportional to $(1–\delta)$, so that the expected payoff is almost independent of δ.

We use a risk-neutral loss given default rate of 75% as suggested by BDDFS (2004) in our analysis.

Setting the present value of the expected payments equal to the expected payoffs, we can solve for π_{CDS} from the resulting nonlinear equation. Since this calculation assumes that there is no risk-premium associated with default, the resulting implied probability should be considered a risk-neutral default probability. As described above, we solve for π_{CDS} for a sample of CDS spreads. We then calculate the distribution of the ratio of π_{CDS}, the CDS risk-neutral default probability to π_{Merton} or $\pi_{naïve}$. We also follow BDDFS (2004) and regress log of the CDS spread against the log of π_{Merton} or $\pi_{naïve}$. The results of our analysis are described in Section 4.5 and Table 6.

3. Data

We begin by examining all firms in the intersection of the Compustat Industrial file—Quarterly data and CRSP daily stock return for NYSE, AMEX, and NASDAQ stocks between 1980 and 2003. We exclude financial firms (SIC codes 6021, 6022, 6029, 6035, 6036) from the sample.

We obtain default data for the period 1980–2000 from the database of firm default maintained by Edward Altman (The Altman default database). We supplement this information for 2001 through 2003 by using the list of defaults published by Moody's at their website www.moodys.com. In all, we obtain a total of 1449 firm defaults covering the period 1980–2003.

The inputs to the Merton DD model include σ_E, the volatility of stock returns F, the face value of debt r, the risk-free rate, and the time period T. σ_E is the annualized percent standard deviation of returns and is estimated from the prior year stock return data for each month. For r, the risk-free rate, we use the 1-year Treasury Constant Maturity Rate obtained from the Board of Governors of the Federal Reserve system.[7] E, the market value of each firm's equity (in millions of dollars), is calculated from the CRSP database as the product of share price at the end of the month and the number of shares outstanding. Following Vassalou and Xing (2004), we take F, the face value of debt, to be debt in current liabilities (COMPUSTAT data item 45) plus one-half of long-term debt (COMPUSTAT data item 51). In addition to the above variables, following Shumway (2001), we measure each firm's past excess return in year t as the return of the firm in year $t-1$ minus the value-weighted CRSP NYSE/AMEX index return in year $t-1$ ($r_{it-1} - r_{mt-1}$). Each firm's annual returns are calculated by cumulating monthly returns. We also collect each firm's ratio of net income to total assets. These variables, though not required for the Merton DD model, will augment the information set for the alternative models we consider later in the paper.

[7] Available at http://research.stlouisfed.org/fred/data/irates/gs1 (H.15 Release).

Table 1
Summary statistics

Panel A: Means, standard deviations, and quantiles

Variable	Mean	Std. dev.	Quantiles Min	0.25	Median	0.75	Max
E	808.80	2453.15	1.21	18.56	76.52	394.83	17534.72
F	229.92	729.66	0.02	2.67	15.56	96.65	5175.50
r (%)	6.46	2.82	1.01	4.85	5.85	7.89	16.72
$r_{it-1} - r_{mt-1}$ (%)	−8.69	63.02	−99.89	−46.79	−14.21	16.94	272.00
NI/TA	−1.08	6.99	−41.13	−0.94	0.73	1.85	7.83
V	1072.33	3228.60	1.52	26.43	105.24	530.12	22949.32
σ_V (%)	56.00	36.83	10.03	30.41	46.32	70.61	230.19
μ (%)	3.25	57.17	−253.58	−21.72	4.36	29.34	210.37
π_{Merton} (%)	10.95	23.32	0.00	0.00	0.01	6.41	100.00
naïve σ_V (%)	50.67	30.97	10.48	28.17	42.29	64.70	162.70
r_{it-1} (%)	13.75	82.07	−85.45	−27.01	2.27	34.13	294.94
$\pi_{\text{naïve}}$ (%)	8.95	20.57	0.00	0.00	0.00	3.46	100.00

Panel B: Correlations

Corr (σ_V, naïve σ_V) = 0.8748
Corr (π_{Merton},$\pi_{\text{naïve}}$) = 0.8642

This table reports summary statistics for all the variables used in the Merton DD model and the hazard models. E is the market value of equity in millions of dollars and is taken from CRSP as the product of share price at the end of the month and the number of shares outstanding. F is the face value of debt in millions of dollars (computed as Compustat item 45 + 0.5 * Compustat item 51). r is the risk-free rate measured as the 3-month Treasury-bill rate. The past returns variable, $r_{it-1} - r_{mt-1}$, is the difference between the prior year return of the firm and the return on the CRSP value weighted index during the same period, and NI/TA is the firm's ratio of net income to total assets. V is the market value of firm assets in millions of dollars, σ_V is the asset volatility measured in percentage per annum, and μ is the expected return on the firm's assets. All three of these variables are generated as the result of solving the Merton DD model for each firm-month in the sample using the iterative procedure described in the text. π_{Merton} is the expected default frequency in percent and is given by Equation (7). naïve σ_V is calculated by Equation (10), and the firm's equity return from the previous year, r_{it-1}, is used as a proxy for the firm's expected asset return to calculate the naïve probability of default, $\pi_{\text{naïve}}$. Our naïve probability, $\pi_{\text{naïve}}$, is calculated according to Equation (13). All variables except the default probabilities are winsorized at the first and 99th percentiles. Our sample spans 1980 through 2003, containing 1,016,552 firm-months with complete data.

There are a number of extreme values among the observations of each variable constructed from raw COMPUSTAT data. To ensure that statistical results are not heavily influenced by outliers, we set all observations higher than the 99th percentile of each variable to that value. All values lower than the first percentile of each variable are winsorized in the same manner. The minimum and maximum numbers reported in Table 1 are calculated after winsorization.[8] Table 1 provides summary statistics for all the variables described above.

It may seem slightly odd that the summary statistics in Table 1 show that the average firm's past excess return is −8.7%. This value is negative because of the winsorization of the upper tail extreme values at the 99th percentile level. More significantly, the distribution of the expected default frequency obtained from the Merton DD model, π_{Merton}, is very similar to the naïve alternative $\pi_{\text{naïve}}$.

[8] We do not winsorize the expected default frequency measures from the Merton DD model and the naïve alternative, since these are naturally bounded between 0 and 1.

Our point estimate of 10.95% for the mean value of π_{Merton} in 1980–2003 is a bit higher than the estimate of 4.21% for the period 1971–1999 reported in Vassalou and Xing (2004). The correlation between the naïve and Merton DD model expected default frequencies is very high at 86%, and it is significant at the 1% level. The similarity in distributions is also evident between the naïve and Merton DD model estimates of asset volatility. The correlation between the two asset volatilities is 87%, and it is also significant at the 1% level.

Given that the naïve counterparts ($\pi_{\text{naïve}}$ and naïve σ_V) of the output from the Merton DD model (π_{Merton} and σ_V) are quite similar, what is the incremental value of solving the Merton DD model? The next section addresses this question.

4. Results

We present a number of empirical results, including correlations of our probability estimates with those published by Moody's KMV, estimates of hazard models for time to default, out-of-sample forecast assessments, CDS implied default probability regressions, and bond yield spread regressions. We discuss each type of result in turn.

4.1 Comparing Moody's KMV probabilities to ours

As mentioned above, our method for calculating π_{Merton} and that employed by Moody's KMV differ in several potentially important respects. In order to gauge how close our methods are, we would like to compare our probability estimates to those calculated by Moody's KMV. It would be desirable to acquire data directly from Moody's KMV for this purpose, but Moody's KMV data are prohibitively expensive for us. Fortunately, in November of 2003, Ronald Fink of Moody's KMV published an article in *CFO Magazine* titled "Ranking America's top debt issuers by Moody's KMV Expected Default Frequency." This magazine article included a table with Moody's KMV EDF data for 100 firms. We are able to calculate default probabilities for 80 of the firms listed in the article. We include a comparison of our probability estimates and those of Moody's KMV in Table 2. Each default probability is computed as of August 2000.

Among the 80 firms for which we have data, the rank correlation between our calculated π_{Merton} and that calculated by Moody's KMV is 79%. The rank correlation between our naïve probability and the Moody's KMV probability is also 79%. These high correlations indicate that both of our probability measures do a good job of capturing the information in the probability estimates published by Moody's KMV. Table 2 also shows that the rank correlation between our (iterated) estimate of firm volatility and that of Moody's KMV is only 57%, while the rank correlation of our naïve estimate of firm volatility and the firm volatility published by Moody's KMV is much higher, at 85estimate, this high correlation is remarkable. Again, this demonstrates that we are able to capture much of the information in Moody's KMV estimates with our measures.

Table 2
Comparison with Moody's KMV EDF

Correlation	Estimate
Rank Corr(Moody's π_{KMV}, Our π_{Merton})	0.788
Rank Corr(Moody's π_{KMV}, Our $\pi_{\text{naïve}}$)	0.786
Rank Corr(Moody's σ_V, Our σ_V)	0.574
Rank Corr(Moody's σ_V, Our naïve σ_V)	0.853

This table compares the expected default frequency computed by Moody's KMV corporation and the methods used in this paper. We obtain the data for Moody's KMV EDF and asset volatility for 80 firms for August 2000 from the article "Ranking America's Top Debt Issuers by Moody's KMV Expected Default Frequency" by Ronald Fink, in *CFO Magazine*, November 2003. The second column of the table provides the rank correlations between the various measures listed in the first column. All correlations are significant at the 0.1% level or lower.

4.2 Hazard model results

Table 3 contains the results of estimating several Cox proportional hazard models. Models 1 and 2 are univariate hazard models, which explain time-to-default as a function of the Merton DD probability and the naïve probability. While these are relatively simple univariate models, the fact that their explanatory variables vary with time means they are more complicated than they might at first appear. Models 1 and 2 confirm that the Merton DD probability and the naïve probability are both extremely significant default predictors. Interestingly, both probabilities, which have similar magnitudes, also have similar coefficients and standard errors. Unreported models that use either the log of market equity or the log of the Merton DD distance to default, rather than the Merton DD probability, perform uniformly worse than the results reported.

Model 3 in Table 3 combines the Merton DD and the naïve probability in one hazard model. Both covariates are very statistically significant, allowing us to conclude that the Merton DD is not a sufficient statistic for default probability, or allowing us to reject our first hypothesis. While the coefficients have similar magnitudes and similar statistical significance, their significance and magnitude are much smaller in Model 3 than in Models 1 and 2. This reflects the fact that the Merton DD and naïve probabilities are highly correlated. In fact, in our sample, their correlation coefficient is 0.86.[9]

[9] Given this high correlation, we investigate the issue of multicollinearity in some depth. As Menard (2002, p. 76.) notes, "Because the concern is with the relationship among the independent variables, the functional form of the model for the dependent variable is irrelevant to the estimation of collinearity." We, therefore, use the standard diagnostic tools for detecting multicollinearity. We compute the variance inflation factors (VIFs) for the independent variables used in the hazard model estimates. The lowest VIF is for $r_{it-1} - r_{mt-1}$ at 1.13 and the highest VIF is for π_{Naive} at 4.27. All the VIFs are much less than 10, a commonly used guideline for detecting multicollinearity. We also compute the condition number, which is the ratio of the square root of the largest to the smallest eigenvalue for our set of variables. We obtained a condition number of 4.38, much less than 15, a commonly used guideline for detecting multicollinearity. Most of our individual coefficient estimates are significantly different from zero, while a classic symptom of multicollinearity is insignificant coefficients. Furthermore, our coefficient estimates are relatively stable when we randomly divide our sample into two groups and reestimate the model. As we add variables in models one through seven, we do not obtain any changes in the signs of coefficients that seem theoretically questionable. Finally, there are no big changes in the estimated coefficients as we move through the specifications. We conclude that multicollinearity is not a significant concern in the data.

Table 3
Hazard model estimates

	Dependent Variable: Time to Default						
Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
π_{Merton}	3.635***		1.697***			0.230	
	(0.068)		(0.142)			(0.164)	
$\pi_{\text{naïve}}$		4.011***	2.472***		1.675***	1.366***	1.526***
		(0.067)	(0.147)		(0.149)	(0.178)	(0.138)
$\ln(E)$				−0.499***	−0.308***	−0.247***	−0.255***
				(0.019)	(0.027)	(0.024)	(0.023)
$\ln(F)$				0.423***	0.238***	0.263***	0.269***
				(0.013)	(0.022)	(0.020)	(0.020)
$1/\sigma_E$				−0.858***	−0.587***	−0.506***	−0.518***
				(0.046)	(0.049)	(0.047)	(0.046)
$r_{it-1} - r_{mt-1}$				−1.426***	−1.081***	−0.819***	−0.834***
				(0.078)	(0.086)	(0.081)	(0.080)
NI/TA						−0.044***	−0.044***
						(0.002)	(0.002)

This table reports on the estimates of several Cox proportional hazard models with time-varying covariates. There are 15,018 firms and 1449 defaults in the sample. π_{Merton} is the Merton DD probability, $\pi_{\text{naïve}}$ is our naïve alternative, $\ln(E)$ and $\ln(F)$ are the natural logarithms of market equity and face value of debt, respectively. $1/\sigma_E$ is the inverse of equity volatility, measured with daily data over the previous year, $r_{it-1} - r_{mt-1}$ is the stock's return over the previous year minus the market's return over the same period, and NI/TA is the firm's ratio of net income to total assets. A positive coefficient on a particular variable implies that the hazard rate is increasing in that variable, or that the expected time to default is decreasing in that variable. Standard errors are in parentheses (*** Significant at 1% level, ** significant at 5% level, and * Significant at 10% level).

Model 4 is a simplereduced-form model, which uses the same inputs as the Merton DD model: the log of the firm's equity value, its returns over the past year, the log of the firm's debt, and the inverse of the firm's equity volatility. Each of these covariates is strongly statistically significant. Model 5 includes all the covariates of Model 4, and it also includes $\pi_{\text{naïve}}$. Comparing the estimates of Models 4 and 5, we see that $\pi_{\text{naïve}}$ is a significant predictor even when all the quantities used to calculate $\pi_{\text{naïve}}$ are included in the hazard model. This implies that the functional form of $\pi_{\text{naïve}}$ is a useful construct for default forecasting, providing evidence in favor of our second hypothesis.

Model 6 adds two other covariates to Model 5: the Merton DD probability and the firm's ratio of net income to total assets. From the estimates of Model 6, we infer that each of the predictors other than the Merton DD probability is statistically significant, making our rejection of our first hypothesis quite robust. Interestingly, the ratio of net income to total assets is quite significant, even though the Merton DD model has no obvious way to incorporate this kind of accounting information. With all of the predictors of Model 6 included in the hazard model, the Merton DD probability is no longer statistically significant, but the naïve probability continues to be significant. The magnitude of the Merton DD coefficient is much smaller in Model 6 than it is in Model 3, while its standard error is quite similar. Thus, the insignificance of π_{Merton} is not driven by difficulty in measuring its coefficient accurately. This evidence suggests that we can reject our third hypothesis, and that the algorithm specified

by the Merton DD model for calculating total firm value and firm volatility is not important.

The naïve probability retains its statistical significance in Model 6 even though its coefficient drops by approximately one half. Furthermore, in an unreported hazard model that includes all the covariates of Model 6 except the naïve probability, π_{Merton} is again statistically significant. Like the results described previously, these results suggest that the value of the Merton DD model lies in its functional form rather than in its solution of the Merton model. Since the Merton DD probability is insignificant in Model 6, we estimate one more hazard model (Model 7) without this variable. In Model 7, we continue to find that the naïve probability is significant, even in the presence of all the other covariates.

Overall, Table 3 shows that the Merton DD probability is not a sufficient statistic for forecasting default. It also shows that the naïve default probability measure is at least as important as the Merton DD measure for forecasting default, and the contribution of π_{Merton} to a reduced-form model that includes $\pi_{\text{naïve}}$ is quite small. This implies that the functional form suggested by the Merton model is more important than the solution of the Merton model for the inputs to the Merton DD model.

4.3 Out-of-sample results

Table 4 contains our assessment of the out-of-sample predictive ability of several variables. To create the table, firms are sorted into deciles each quarter based on a particular forecasting variable. Then the number of defaults that occur in each of the decile groups is tabulated, with the percentage of defaults in the highest probability deciles reported in the table. One advantage of this approach is that the rankings of firms into default probability deciles can be done without estimating actual default probabilities. If our model for translating distances into default probabilities is slightly misspecified (in particular, if the normal CDF is not the most appropriate choice), our out-of-sample results will remain unaffected.[10]

Panel A compares the predictions of the Merton DD model to the naïve model, market equity, and past returns. While the Merton DD model probability is able to classify 64.9% of defaulting firms in the highest probability decile at the beginning of the quarter in which they default, the naïve model is able to classify 65.8% of defaulting firms in the top decile. Fully 80.0% of defaults occur in the highest π_{Merton} quintile, while 80.1% occur in the highest $\pi_{\text{naïve}}$ quintile. It is remarkable that the out-of-sample performance of π_{Merton} is worse than that of $\pi_{\text{naïve}}$.

[10] A rough calibration of probabilities associated with distance to default rankings can be inferred from the data in Table 4. For example, the probability that firms in the top decile of π_{Merton} will default in the next quarter is equal to the number of defaults occurring in the top decile (1449 * 0.649) divided by one-tenth of the number of firm-quarter observations used to create the table (350,662 * 0.1), giving a probability of 2.7%.

Table 4
Out-of-Sample Forecasts

Panel A: 1980–2003

	350,662 firm-quarters, 1449 defaults				
Decile	π_{Merton}	$\pi_{\text{naïve}}$	E	$r_{it-1} - r_{mt-1}$	NI/TA
1	64.9	65.8	35.7	44.4	46.8
2	15.1	14.3	17.5	25.1	23.8
3	6.0	6.7	14.3	9.2	10.6
4	4.6	4.1	9.1	5.4	5.9
5	2.9	2.4	6.1	2.9	4.2
6–10	6.5	6.7	17.3	13.0	8.7

Panel B: 1991–2003

	226,604 firm-quarters, 847 defaults					
Decile	π_{Merton}	$\pi_{\text{naïve}}$	Model 4	Model 5	Model 6	Model 7
1	69.4	70.0	68.8	70.4	75.6	75.8
2	14.9	12.4	12.8	12.9	11.0	11.0
3	5.2	6.9	5.3	4.7	4.4	4.5
4	2.8	3.2	2.8	2.6	1.8	1.9
5	1.7	1.8	1.8	1.2	1.2	0.8
6–10	6.0	5.7	8.5	8.2	6.0	6.0

This table reports on the success of various forecasting quantities by sorting firms each quarter by forecast and counting the fraction of defaults that correspond with each decile of the forecast variable. Panel A examines accuracy over the entire period for which we have data, from 1980 to 2003. There are 350,662 firm-quarters in our sample, with 1449 defaults. π_{Merton} is the Merton DD probability, $\pi_{\text{naïve}}$ is our naïve alternative, E is market equity, $r_{it-1} - r_{mt-1}$ is the stock's return over the previous year minus the market's return over the same period, and NI/TA is the firm's ratio of net income to total assets. Panel B only considers defaults from 1991 to 2003, and it includes the fitted values of three hazard models (models 5, 6, and 7 from Table 3) as predictors. These models are estimated each quarter using all available data in each quarter, and the resulting coefficients are used to form the predictors assessed in columns 4 and 5 of Panel B.

The out-of-sample performance of both π_{Merton} and $\pi_{\text{naïve}}$ is quite a bit better than simply sorting firms on their market equity. This is consistent with the results of Vassalou and Xing (2004) and indicates that the success of π_{Merton} is not simply reflecting the predictive value of market equity. Apparently, it is quite useful to form a probability measure, by creating a z-score and using a cumulative distribution to calculate the corresponding probability. Given that π_{Merton} does not perform better than $\pi_{\text{naïve}}$ in either hazard models or out-of-sample forecasts, the functional form of the probability measure suggested by the Merton DD model appears to be a more valuable innovation than the solution of Equations (2) and (5).

Simply sorting firms on their excess equity return over the last year has surprisingly good forecasting power, as does sorting firms by their value of net income over total assets. This is consistent with the economic and statistical significance of both of these variables in the hazard model results reported in Table 3. Since the Merton DD model has no simple way to capture innovations in past returns or income, it is difficult to believe that π_{Merton} can be a sufficient statistic for default. Any reasonable default prediction model probably needs to include some measure of past returns and net income.

Panel B reports similar forecasting assessments for a shorter time period, from 1991 to 2003. Looking at defaults in this shorter period allows us to examine the out-of-sample performance of the hazard models reported in Table 3. We estimate Models 4 through 7 from Table 3 each quarter, using data available in that quarter, to define our decile groups. For example, we sort firms in the second quarter of 1995 based on the fitted values of a hazard model estimated with data from 1980 through the first quarter of 1995. The forecasting success of hazard Models 4 through 7 appears in Columns 3 through 6 of the table.

Remarkably, the naïve variable and the Merton DD probability, both of which exploit the z-score functional form suggested by theory, do better at default prediction than a reduced-form model that uses all the inputs to the naïve variable (Model 4). As in Panel A, the naïve variable performs slightly better than the Merton DD probability out of sample. Both models perform quite well in classifying low-risk firms. The misclassification of risk firms into the lower risk deciles (deciles 6–10) is the lowest for $\pi_{\text{naïve}}$. Using the functional form suggested by theory (as in $\pi_{\text{naïve}}$) produces fewer low-risk misclassifications than any of the other models considered. While the simple reduced-form model that uses all of the inputs of the naïve model (Model 4) has a misclassification rate of 8.5%, $\pi_{\text{naïve}}$ produces a misclassification rate of only 5.7%. All these results again confirm that we should accept our second hypothesis and reject our third hypothesis. The functional form specified by the Merton DD model adds significantly to our predictive power, but solving the model for implied parameter values is not useful.

The hazard models that incorporate our income variable (Models 6 and 7) clearly outperform both π_{Merton} and $\pi_{\text{naïve}}$. This is not surprising, given that they employ more information in making their forecasts. Model 7 performs substantially better than π_{Merton}, categorizing 75.8% of defaults in the highest hazard decile when they default versus 69.4% for π_{Merton}. Model 7 also performs substantially better than $\pi_{\text{naïve}}$. It appears that combining the functional form suggested by theory along with other covariates, including accounting measures, delivers the best performance.

4.4 Alternative predictors

In order to further assess the importance of various calculations required to generate the Merton DD probabilities, we examine the forecasting performance of three alternative probabilities in Table 5. The first of our three measures is a Merton DD probability that is calculated under the assumption that the expected return of each firm is the risk-free rate. We examine this predictor, which we denote by $\pi^{\mu=r}_{\text{Merton}}$, to determine how important the calculation of μ is for the distance to default in Equation (6). Our second predictor, which we denote by $\pi^{simul}_{\text{Merton}}$, is a Merton DD probability that is calculated by simultaneously solving Equations (2) and (5) rather than following the more complicated iterative procedure described in the text. Our third alternative predictor is a Merton DD probability that is calculated with option-implied volatility rather

than historical-equity volatility. The implied volatility predictor, which we denote by $\pi^{imp\sigma}_{\text{Merton}}$, simultaneously solves Equations (2) and (5) instead of using the iterative procedure. Each of our three alternative predictors can be thought of as Merton DD probabilities that are calculated with some strong simplifying assumptions. If these predictors perform as well as π_{Merton} then we can conclude that the simplifying assumptions are valid.

It is important to point out that while our sample for $\pi^{\mu=r}_{\text{Merton}}$ and $\pi^{simul}_{\text{Merton}}$ is the same as the samples described in the rest of the paper, our sample for $\pi^{imp\sigma}_{\text{Merton}}$ is much smaller, spanning 1996 through 2003 and containing 101,201 firm-months with complete data. We obtain the implied volatility of 30-day at-the-money call options from the IVY Database of Optionmetrics LLC. IVY is a comprehensive database of historical price, implied volatility and sensitivity information for the entire US listed index and equity options market and contains historical data beginning in January 1996. The implied volatilities are calculated by Optionmetrics in accordance with the standard conventions used by participants in the equity option market, using a Cox-Ross-Rubinstein binomial tree model, which is iterated until convergence of the model price to the market price of the option.

Table 5 reports summary statistics for each variable, correlations between each of the variables and π_{Merton}, and measures of out-of-sample prediction accuracy that correspond to those in Table 4. Looking at the correlations in Panel B, each of our alternative predictors is highly correlated with π_{Merton} and with the other predictors in the table. Interestingly, the simultaneously solved $\pi^{simul}_{\text{Merton}}$ is more correlated with $\pi_{\text{naïve}}$ than π_{Merton}. The probability calculated with implied volatility is less correlated with π_{Merton} than most of the other probabilities.

The out-of-sample forecast accuracy in Panel C allows us to gauge the relative importance of the iterative procedure, the estimation of μ, and the estimation of equity volatility. As in Table 4, Panel C sorts all firm-quarters by each predictor and then counts the number of defaults that occur among firms in each decile of the predictor. In Panel C, the results for $\pi^{\mu=r}_{\text{Merton}}$ and $\pi^{simul}_{\text{Merton}}$ in the second and third columns are directly comparable to the results in Panel A of Table 4. However, because of the different sample size, the result for $\pi^{imp\sigma}_{\text{Merton}}$ are not comparable to any results in Table 4. To provide a performance benchmark for the $\pi^{imp\sigma}_{\text{Merton}}$ results, the fifth and sixth columns of Table 5 report on the success of π_{Merton} and $\pi_{\text{naïve}}$ using the subset of firms for which implied volatility is available.

The estimation of μ for distance to default [Equation (6)] is apparently quite important. The probability that sets μ equal to the risk-free rate performs substantially worse than π_{Merton} out of sample, classifying only 60% of defaulting firms in the highest probability decile at the beginning of the quarter in which the firms default. π_{Merton} is able to classify almost 65% of defaulting firms in the highest decile. Calculating Merton DD probabilities with the iterative procedure described in the text is apparently less important. The simultaneously

Table 5
Alternative predictors

Panel A: Summary statistics (%)

			Quantiles				
Variable	Mean	Std. Dev.	Min	0.25	Mdn	0.75	Max
$\pi^{\mu=r}_{\text{Merton}}$	7.71	17.97	0.00	0.00	0.01	3.77	100.00
$\pi^{simul}_{\text{Merton}}$	8.13	20.76	0.00	0.00	0.00	1.48	100.00
Implied σ_E	58.47	26.54	4.01	39.25	52.48	72.27	500.00
$\pi^{imp\sigma}_{\text{Merton}}$	4.11	14.70	0.00	0.00	0.00	0.10	100.00

Panel B: Correlation matrix

	π_{Merton}	$\pi_{\text{naïve}}$	$\pi^{\mu=r}_{\text{Merton}}$	$\pi^{simul}_{\text{Merton}}$
$\pi_{\text{naïve}}$	0.8642			
$\pi^{\mu=r}_{\text{Merton}}$	0.8575	0.7486		
$\pi^{simul}_{\text{Merton}}$	0.8338	0.9755	0.7220	
$\pi^{imp\sigma}_{\text{Merton}}$	0.6858	0.9624	0.4102	0.6259

Panel C: Out-of-sample forecasts

Decile	$\pi^{\mu=r}_{\text{Merton}}$	$\pi^{simul}_{\text{Merton}}$	$\pi^{imp\sigma}_{\text{Merton}}$	π_{Merton}	$\pi_{\text{naïve}}$
1	60.0	65.1	84.1	80.7	83.0
2	17.7	15.0	8.0	9.1	9.1
3	8.0	7.7	4.6	3.4	5.7
4	4.1	3.4	0.0	5.7	1.1
5	3.4	3.2	1.1	0.0	0.0
6–10	6.8	5.6	2.2	1.1	1.1
Defaults	1,449	1,449	88	88	88
Firm-Quarters	350,662	350,662	36,274	36,274	36,274

This table reports on the success of three alternative predictors, or predictors that calculate Merton DD probabilities in alternative ways. One predictor, $\pi^{\mu=r}_{\text{Merton}}$, is calculated in exactly the same manner as π_{Merton}, except that the expected return on assets used for π_{Merton} is replaced by the risk-free rate, r. A second alternative predictor, $\pi^{simul}_{\text{Merton}}$, is calculated by simultaneously solving Equations (2) and (5). This predictor avoids the iterative procedure in the text, estimating equity volatility with 1 year of historical returns data and using r as the expected return on assets. The third alternative predictor, $\pi^{imp\sigma}_{\text{Merton}}$, uses the option-implied volatility of firm equity (implied σ_E) to simultaneously solve Equations (2) and (5). Our sample for $\pi^{\mu=r}_{\text{Merton}}$ and $\pi^{simul}_{\text{Merton}}$ is the same as the samples described in Table 1, including 1,016,552 firm-months from 1980 through 2003. Our sample for $\pi^{imp\sigma}_{\text{Merton}}$ spans 1996 through 2003, containing 101,201 firm-months with complete data. Panel A reports summary statistics on our alternative predictors, Panel B reports correlations between all of our predictors, and Panel C reports on the out-of-sample predictive success of our alternatives. As in Table 4, Panel C sorts all firm-quarters by each predictor and then counts the number of defaults that occur among firms in each decile of the predictor. In Panel C, the results for $\pi^{\mu=r}_{\text{Merton}}$ and $\pi^{simul}_{\text{Merton}}$ in the second and third columns are directly comparable to the results in Panel A of Table 4. However, because of the different sample sizes, the results for $\pi^{imp\sigma}_{\text{Merton}}$ are not comparable to any results in Table 4. To provide a performance benchmark for the $\pi^{imp\sigma}_{\text{Merton}}$ results, the fifth and sixth columns of Table 5 report on the success of π_{Merton} and $\pi_{\text{naïve}}$ using the subset of firms for which implied volatility is available.

solved $\pi^{simul}_{\text{Merton}}$ actually has better out-of-sample predictive performance than the iteratively solved π_{Merton}, though its performance still does not dominate that of $\pi_{\text{naïve}}$. This is consistent with the relative success of our naïve probability in Tables 3 and 4. Finally, using implied equity volatility rather than

estimated equity volatility in our probability improves out-of-sample performance substantially. However, given that there are only 88 defaults to forecast in our sample of firms with corresponding options contracts, it is difficult to apply this finding to the broader sample of firms.

4.5 CDS spread regressions

Our previous results demonstrate that while π_{Merton} appears to be a useful quantity for forecasting defaults, it is not a sufficient statistic for the purpose of forecasting. Our next two sets of results examine whether π_{Merton} and $\pi_{\text{naïve}}$ are important explanatory variables for pricing credit-sensitive securities. First, we regress log spreads and implied default probabilities of credit default swaps on our probability measures. Bond yield spread regressions are our final set of results.

CDS default probability results are reported in Table 6. We obtain the data on CDS spreads from www.credittrade.com for the period December 1998 to July 2003. From this source, we are able to collect 3833 firm-months of CDS spread observations. We calculate the probability that a firm defaults in the next year, π_{CDS}, according to the algorithm described in Section 2.5. Following BDDFS, we also winsorize π_{Merton} and $\pi_{\text{naïve}}$ at an upper limit of 20% and a lower limit of 0.20% to facilitate comparison of our results with theirs.

Panel A of Table 6 provides the summary statistics on the CDS spreads and the default probabilities. We note that the difference between the average default probabilities π_{Merton} and $\pi_{\text{naïve}}$ is comparable to the overall sample difference in Table 1. In both tables, the mean of π_{Merton} is greater than that of $\pi_{\text{naïve}}$, and in both tables the difference in means is statistically significant. While the average value of π_{CDS} is smaller than those of π_{Merton} and $\pi_{\text{naïve}}$, the median value is larger. It seems likely that the means of π_{Merton} and $\pi_{\text{naïve}}$ are influenced by a few large influential observations. Looking at the correlations between π_{CDS} and our probability measures in Panel B, we see that $\pi_{\text{naïve}}$ is much more correlated with π_{CDS} (at 51%) than π_{Merton} (at 38%).

If π_{Merton} is a well calibrated and accurate probability of default, then the ratio of π_{CDS} to π_{Merton} for any particular firm should be greater than or equal to one. If there is no risk-premium for default risk and π_{Merton} captures all the information that π_{CDS} contains, then the ratio should be exactly 1. If there is a risk-premium for default risk, then the risk-neutral default probability (π_{CDS}) will generally be larger than the physical probability, and the ratio of the two probabilities will be greater than 1. If the ratio is less than 1, there must be some information in π_{CDS} that is absent from π_{Merton}. Since π_{CDS} is determined by market participants, and therefore is conditional on all information available at the time it is determined, it seems likely that π_{CDS} contains some information that is not captured by the relatively simple π_{Merton}.

We report summary statistics of the ratio of π_{CDS} to the other probabilities for the same firm in Panel C. The means and medians for both ratios ($\frac{\pi_{\text{CDS}}}{\pi_{\text{Merton}}}$ and $\frac{\pi_{\text{CDS}}}{\pi_{\text{naïve}}}$) are greater than 1, as predicted by theory. Further, the distributions of

Table 6
CDS spread regressions

Panel A: Summary statistics (%)

	Quantiles						
Variable	Mean	Std. Dev.	Min	0.25	Mdn	0.75	Max
CDS Spread	165.89	170.49	9.50	60.83	100.00	204.30	1650.00
π_{CDS}	2.13	2.14	0.13	0.79	1.31	2.65	19.64
π_{Merton}	3.82	6.88	0.20	0.20	0.20	2.75	20.00
$\pi_{naïve}$	2.73	5.84	0.20	0.20	0.20	0.73	20.00

Panel B: Correlation matrix

Corr (π_{CDS}, π_{Merton})	0.3762***
Corr (π_{CDS}, $\pi_{naïve}$)	0.5199***

Panel C: Distribution of the ratios of risk-neutral to actual probabilities

	Quantiles						
Ratio	Mean	Std. Dev.	Min	0.25	Mdn	0.75	Max
π_{CDS}/π_{Merton}	5.09	5.70	0.02	0.80	3.73	6.71	28.86
$\pi_{CDS}/\pi_{naïve}$	5.54	5.61	0.05	1.62	4.14	7.29	28.89

Panel D: Regressions

	Dependent variable: log(CDS)				Dependent variable: π_{CDS}		
Variable	Model 1	Model 2	Model 3	Variable	Model 4	Model 5	Model 6
Const.	4.0047***	4.2404***	4.2605***	Const.	1.6842***	1.6109***	1.6027***
	(0.3950)	(0.1854)	(0.1689)		(0.0283)	(0.0266)	(0.0278)
$\log(\pi_{Merton})$	0.1737***		−0.0194	π_{Merton}	0.1172***		0.0053
	(0.0083)		(0.0105)		(0.0074)		(0.0062)
$\log(\pi_{naïve})$		0.2771***	0.2931***	$\pi_{naïve}$		0.1908***	0.1864***
		(0.0076)	(0.0118)			(0.0090)	(0.0100)
Obs.	3833	3833	3833	Obs.	3833	3833	3833
R^2	0.26	0.38	0.38	R^2	0.14	0.27	0.27

This table reports on a comparison of the probability of default implied by credit default swap (CDS) spreads with π_{Merton} and $\pi_{naïve}$. We obtain the data on CDS spreads from www.credittrade.com for the period December 1998 to July 2003. CDS spread is the credit default swap spread in basis points. π_{CDS} is the probability of default backed out from the CDS spread. All other measures are described in Table 1. The total number of firm-month observations is 3833. Panel A reports summary statistics, Panel B reports correlations, and Panel C reports the distribution of the ratios of the CDS default probability to π_{Merton} and $\pi_{naïve}$. Panel D reports the results of regressing the log of CDS spread and π_{CDS} on π_{Merton}, $\pi_{naïve}$ and time dummies. In Panel D, standard errors are shown in parentheses (*** Significant at 1% level, ** Significant at 5% level).

the estimates are comparable to the estimates produced by BDDFS. However, the values in the lowest quartile of the distribution of both the ratios are less than 1, counter to theory. This suggests that the procedures for calculating π_{Merton} and $\pi_{naïve}$ might produce useful estimates for rank ordering firms by their default risk, but the procedures may not always produce estimates that are reliable for pricing credit-sensitive instruments. The proprietary mapping that Moody's KMV uses to translate distances to default into default probabilities

may be important for pricing purposes. Alternatively, since π_{CDS} is a conditional probability, it is likely that π_{Merton} is a noisy estimate of π_{CDS}.

In the first three columns of Panel D, we replicate the spread-KMV EDF regressions of BDDFS with our data. The dependent variable in these regressions is the log CDS spread. While BDDFS use 5-year EDFs as measured by Moody's KMV in their regressions, we use our 1-year default probabilities π_{Merton} and $\pi_{naïve}$. BDDFS note that the 1-year EDF is highly correlated with the 5-year EDF, and the log specification of this regression makes the intercept reflect the average level of the probability. Our coefficients on the default probabilities are 0.17 and 0.28, lower, but of the same order of magnitude as their estimate of 0.76. Our R^2 values of 0.26 and 0.38 are also lower than their estimate of 0.74. These results once again imply that the mapping of distance to default into default probabilities may be important for pricing applications.

Interestingly, adding both π_{Merton} and $\pi_{naïve}$ in the same regression (Model 3) shows that the statistical significance of π_{Merton} is driven out by $\pi_{naïve}$. This result is similar to the hazard model results in Table 3, again confirming the importance of the functional form of the probability of default over the solution procedure. The last three columns in Panel D regress π_{CDS}, the risk-neutral default probability, on π_{Merton} and $\pi_{naïve}$. The conclusions are similar to the CDS spread regressions. Further, the coefficients on π_{Merton} and $\pi_{naïve}$ in these regressions are less than 1, counter to the theory. Given that one-quarter of firms have risk-neutral probabilities less than their physical probabilities, the low coefficients may be driven by a few influential observations. Overall, the results in Table 6 show that the naïve probability estimate performs at least as well as the Merton DD probability.

4.6 Bond spread results

Our final set of results are regressions of bond yields on our default probabilities. Before discussing our regression results, we describe the sample used to estimate the regressions. Summary statistics for the bond yield sample appear in Panel A of Table 7.

Our bond data are extracted from the Lehman Brothers Fixed Income Database distributed by Warga (1998). This database contains monthly price, accrued interest, and return data on all corporate and government bonds from 1971–1997. We use the data from the 1980–1997 period to be consistent with the default prediction sample. This is the same database used by Elton et al. (2001) to explain the rate spread on corporate bonds. In addition, the database contains descriptive data on bonds, including coupons, ratings, and callability. A subset of the data in the Warga database is used in this study. First, all bonds that were matrix priced rather than trader priced are eliminated from the sample.[11] Employing matrix prices might mean that all our analysis uncovers

[11] For actively traded bonds, dealers quote a price based on recent trades of the bond. Bonds for which a dealer did not supply a price have prices determined by a rule of thumb relating the characteristics of the bond to dealer-priced bonds. These rules of thumb tend to change very slowly over time and do not respond to changes in market conditions.

Table 7
Bond yield spread regressions

Panel A: Summary statistics

Variable	Mean	Std. Dev.	Min	Quantiles 0.25	Mdn	0.75	Max
Spread (bp)	108.09	69.43	28.11	67.53	90.31	121.75	605.91
σ_E (%)	27.67	8.73	6.60	22.34	26.36	31.06	250.38
Maturity	10.32	8.20	1.00	4.59	7.79	12.63	39.25
Amount	190,000	120,000	15,700	100,000	150,000	250,000	750,000
r (%)	5.50	1.39	3.18	4.94	5.54	5.87	16.72
Coupon (%)	8.36	1.46	4.50	7.25	8.38	9.38	14.25
π_{Merton} (%)	9.42	27.49	0.00	0.00	0.00	0.00	100.00
$\pi_{\text{naïve}}$ (%)	2.02	10.71	0.00	0.00	0.00	0.00	99.49

Panel B: Regressions

Variable	Dependent variable: Bond yield spread Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Const.	125.77***	118.27***	126.2***	153.11***	147.28***	153.8***
	(16.55)	(16.59)	(16.59)	(17.43)	(17.36)	(17.46)
σ_E	.86***	.95***	.86***	.79***	.87***	.75***
	(.05)	(.05)	(.05)	(.06)	(.06)	(.06)
Maturity	1.49***	1.5***	1.49***	1.5***	1.51***	1.5***
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Ln(amount)	−5.78***	−5.92***	−5.81***	−8.13***	−8.27***	−8.01***
	(.38)	(.39)	(.39)	(.49)	(.49)	(.49)
r	−.46	.23	−.46	−.61*	.07	−.60
	(.34)	(.34)	(.34)	(.37)	(.37)	(.37)
Coupon	3.31***	3.31***	3.31***	3.49***	3.53***	3.51***
	(.2)	(.21)	(.2)	(.22)	(.22)	(.22)
Coverage < 5				−1.4*	−3.1***	−2.12**
				(.85)	(.91)	(.89)
5 <= Coverage < 10				−6.37***	−7.37***	−6.65***
				(.71)	(.73)	(.73)
10 <= Coverage < 20				−1.79***	−2.13***	−1.87***
				(.68)	(.69)	(.68)
Operating income to sales				−36.11***	−40.93***	−38.78***
				(4.46)	(4.87)	(4.85)
Long-term debt to assets				12.11***	20.08***	16.29***
				(2.57)	(2.83)	(2.78)
Total debt to capitalization				8.88***	5.82***	4.26***
				(1.62)	(1.59)	(1.57)
$\pi_{\text{naïve}}$	.5***		.49***	.63***		.6***
	(.03)		(.03)	(.04)		(.04)
π_{Merton}		.08***	.007		.19***	.1***
		(.008)	(.007)		(.02)	(.01)
Rating dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	61776	61776	61776	51831	51831	51831
R^2	0.70	0.70	0.70	0.72	0.71	0.72

is the rule used to matrix-price bonds rather than the economic influences at work in the market. Eliminating matrix-priced bonds leaves us with a set of prices based on dealer quotes. This is the same type of data as that contained in the standard academic source of government bond data: the CRSP government

Table 7
(Continued)

Panel C: Using Bond Spreads to predict bankruptcy

	Dependent variable: Time to default		
Variable	Model 1	Model 2	Model 3
π_{Merton}	0.994		−0.064
	(0.665)		(0.784)
$\pi_{\text{naïve}}$		2.284***	2.316***
		(0.679)	(0.783)
Spread	0.00109***	0.00098***	0.00097***
	(0.0004)	(0.0003)	(0.0003)
$\ln(E)$	−0.213	−0.120	−0.125
	(0.153)	(0.141)	(0.156)
$\ln(F)$	0.173	0.105	0.111
	(0.132)	(0.116)	(0.134)
$1/\sigma_E$	−0.769***	−0.629***	−0.633***
	(0.198)	(0.195)	(0.200)
$r_{it-1} - r_{mt-1}$	−0.800**	−0.351	−0.356
	(0.379)	(0.410)	(0.415)
NI/TA	−0.025	−0.021	−0.021
	0.015	(0.016)	(0.016)

This table reports the results of bond yield spread regressions. Spread is the difference between the yield to maturity on the bond and the yield of the closest maturity treasury in basis points. σ_E is the standard deviation of equity returns. Maturity is the remaining time to maturity in years of the bonds. Coupon is the coupon rate on the bond issue. r is the risk-free rate measured as the 3-month T-bill rate. Amount is the dollar amount of the bond issue. π_{Merton} is the expected default frequency in percent, given by Equation (7), and $\pi_{\text{naïve}}$ is the corresponding naïve default probability measure given by Equation (13). All observations except the default frequency measures are winsorized at the first and 99th percentiles. The data span 1980 through 1997 and there are 61,776 bond-months with complete data in our sample. Panel A reports summary statistics for the sample used in the regressions and Panel B reports the regression results. In addition to the variables described in Panel A, all regressions have year, rating, and one-digit sic code dummies. Heteroscedasticity consistent standard errors are shown in parentheses. Panel C reports the estimates of several Cox proportional hazard models with time-varying covariates with the bond spread included as an additional explanatory variable. With this additional restriction, there are 571 firms and 58 defaults in the sample. Standard errors are in parentheses. (*** Significant at 1% level, ** significant at 5% level , * significant at 10% level).

bond file. Next, we eliminate all bonds with special features that would result in their being priced differently. This means we eliminate all bonds with options (e.g., callable bonds or bonds with a sinking fund), all corporate floating rate debt, bonds with an odd frequency of coupon payments, and inflation-indexed bonds. In addition, we eliminate all bonds not included in the Lehman Brothers bond indexes, because researchers in charge of the database at Lehman Brothers indicate that the care in preparing the data was much less for bonds not included in their indexes. This also results in eliminating data for all bonds with a maturity of less than 1 year. This exclusion is also consistent with our estimates of π_{Merton} and $\pi_{\text{naïve}}$, which are based on a 1-year forecasting horizon. Finally, we also remove AAA (Moody's rating Aaa) bonds because the data for these bonds appear problematic. Both Elton et al. (2001) and Campbell and Taksler (2003) exclude AAA bonds from their analysis for this reason. We are finally left with 61,776 bond-months with complete data in our sample.

There are a number of extreme values among the observations of each variable constructed from the Warga data. To ensure that statistical results are not heavily influenced by outliers, we set all observations higher than the 99th percentile of each variable to that value. All values lower than the first percentile of each variable are winsorized in the same manner. The minimum and maximum numbers reported in Panel A for the bond are calculated after winsorization.[12]

We compute the spread on the corporate bond as the difference between the yield to maturity on a corporate bond in that particular month and the yield to maturity on a government bond of the closest maturity in the same month. For the benchmark treasuries, we use the CRSP fixed-term indices, which provide monthly yield data on notes and bonds of 1, 2, 5, 6, 10, 20, and 30 target years to maturity. We assume that each quoted price in the Warga data is at the end of the month when the CRSP indices are published, but this should have little impact on the measured spreads. As can be seen from Panel A, the average spread is about 108 basis points over this sample period (1980–1997), similar in magnitude to spreads reported in the other studies. We find that the magnitudes of π_{Merton} and $\pi_{\text{naïve}}$ are smaller than the values reported in Table 1, suggesting that firms that have issuances in the bond market are better credit risks. The average maturity outstanding for the bonds in our sample is about 10 years and the Coupon rate is around 8.3%.

In Panel B of Table 7, we report the results of regressing bond yield spreads on a number of explanatory variables. Looking at the results, it appears that both π_{Merton} and $\pi_{\text{naïve}}$ are significantly related to bond yield spreads. However, looking again at the results, it quickly becomes clear that while spreads are correlated with both π_{Merton} and $\pi_{\text{naïve}}$, the coefficients on these default probabilities are too low. For example, given the coefficient of 0.5 for $\pi_{\text{naïve}}$ in Model 1, if $\pi_{\text{naïve}}$ increased from 0% to 5%, the expected bond yield would increase by just 2.5 basis points. The magnitude of the coefficients can be explained by the fact that bond rating dummies are included in these regressions, and bond ratings capture a large fraction of the variation in spreads. The regression coefficients must be interpreted as capturing the explanatory power of our probability measures conditional on being in a particular ratings class.

In univariate regressions, the magnitude and statistical significance of π_{Merton} is much smaller than that of $\pi_{\text{naïve}}$. Combining both π_{Merton} and $\pi_{\text{naïve}}$ in one spread regression makes the coefficient of π_{Merton} become insignificant. When other explanatory variables are included in the regression, the coefficient on $\pi_{\text{naïve}}$ loses some of its significance but remains statistically distinguishable from zero. π_{Merton} is less significant, both statistically and economically.

[12] As in the summary statistics in Table 1, we do not winsorize the expected default frequency measures from the Merton DD model and the naïve alternative, since these are naturally bounded between 0 and 1.

Overall, the regressions indicate that π_{Merton} is not strongly related to bond yield spreads after conditioning on bond ratings. This is consistent with the hazard model and out-of-sample results discussed previously.

Finally, we incorporate the spread information into a hazard model for defaults, to see whether it helps in forecasting. We are particularly interested to examine if using credit spreads as an input to the bankruptcy forecasting model drives out all of the other forecasting variables in Table 3. The difficulty with this test is the availability of data. Introducing the spread variable into the hazard model reduces the number of firms from 15,018 to 571 and the number of defaults from 1449 to 58. Thus, the results obtained from this restricted sample may not be generalizable. With this caveat in mind, we replicate Models 5 through 7 in Table 3 with the spread as an additional explanatory variable. The results are reported in Panel C of Table 7.

As Panel C indicates, spread is a significant variable in all the models and seems to contain useful information in predicting defaults over and above the other explanatory variables. It seems to drive out the significance of most of the other explanatory variables, though it is not clear how much the reduced sample size contributes to this result. However, $\pi_{\text{naïve}}$ continues to be significant even in this small sample, and this reinforces the importance of mimicking the functional form suggested by theory over the actual solution procedure employed in obtaining π_{Merton}. Using spreads to predict bankruptcies appears to be an interesting area for further research.

5. Conclusion

We examine the accuracy and the contribution of the Merton distance to default (DD) model. Looking at hazard models that forecast default, the Merton DD model does not appear to produce a sufficient statistic for default. It appears to be possible to construct an accurate default forecasting model without considering the iterated Merton DD probability. The naïve probability that we propose, which captures both the functional form and the same basic inputs of the Merton DD probability, performs surprisingly well. Looking at the out-of-sample forecasting ability, it is fairly simple to construct a model that outperforms the Merton DD model without using the Merton DD probability as an explanatory variable. However, hazard models that use the Merton DD probability with other covariates have slightly better out-of-sample performance than models that omit the Merton DD probability. Looking at CDS spread regressions and bond yield spread regressions, the Merton DD probability does not appear to be a significant predictor of either quantity when our naïve probability, agency ratings, and other explanatory variables are accounted for.

We conclude that the Merton DD probability is a useful variable for forecasting default, but it is not a sufficient statistic for default. The usefulness of the Merton DD probability is due to the functional form suggested by the Merton model. The iterative procedure used to solve the Merton model for default

probability does not appear to be useful. Our results indicate that structural models like the Merton model provide useful guidance for building default forecasting models.

References

Basel Committee on Banking Supervision. 1999. *Credit Risk Modelling: Current Practices and Applications*. Available at www.bis.org.

Berndt, A., R. Douglas, D. Duffie, M. Ferguson, and D. Schranz. 2005. Measuring Default Risk Premia from Default Swap Rates and EDFs. Working Paper, Stanford University.

British Bankers' Association. 2002. *BBA Credit Derivatives Report 2001/2002*.

Campbell, J. Y., J. Hilscher, and J. Szilagyi. 2007. In Search of Distress Risk. *Journal of Finance*. http://www.afajof.org/afa/forthcoming/3185.pdf.

Campbell, J. Y., and G. B. Taksler. 2003. Equity Volatility and Corporate Bond Yields. *Journal of Finance* 58:2321–49.

Chava, S., and R. Jarrow. 2004. Bankruptcy Prediction with Industry Effects. *Review of Finance* 8:537–69.

Cox, D. R., and D. Oakes. 1984. *Analysis of Survival Data*. New York: Chapman and Hall.

Crosbie, P. J., and J. R. Bohn. 2003. Modeling Default Risk. http://www.ma.hw.ac.uk/∼mcneil/F79CR/Crosbie_Bohn.pdf (accessed May 2008).

Du, Y., and W. Suo. Assessing Credit Quality from Equity Markets: Is a Structural Approach a Better Approach? *Canadian Journal of Administrative Studies* (forthcoming).

Duffie, D., and K. J. Singleton. 2003. *Credit Risk: Pricing, Measurement, and Management*. Princeton, NJ: Princeton University Press.

Duffie, D., L. Saita, and K. Wang. 2007. Multi-Period Corporate Failure Prediction with Stochastic Covariates. *Journal of Financial Economics* 83:635–65.

Elton, E. J., M. J. Gruber, D. Agrawal, and C. Mann. 2001. Explaining the Rate Spread on Corporate Bonds. *Journal of Finance* 56:247–77.

Falkenstein, E., and A. Boral. 2001. Some Empirical Results on the Merton Model. *Risk Professional*. April.

Hillegeist, S. A., E. K. Keating, D. P. Cram, and K. G. Lundstedt. 2004. Assessing the Probability of Bankruptcy. *Review of Accounting Studies* 9(1):5–34.

Hull, J., M. Predescu, and A. White. 2004. The Relationship between Credit Default Swap Spreads, Bond Yields and Credit Rating Announcements. *Journal of Banking and Finance* 28:2789–2811.

Jones, E., S. Mason, and E. Rosenfeld. 1984. Contingent Claims Analysis of Corporate Capital Structures: an Empirical Investigation. *Journal of Finance* 39:611–25.

Kealhofer, S., and M. Kurbat. 2001. *The Default Prediction Power of the Merton Approach., Relative to Debt Ratings and Accounting Variables* (KMV LLC).

Longstaff, F. A., S. Mithal, and E. Neis. 2005. Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit-Default Swap Market. *Journal of Finance* 60:2213–53.

Menard, S. 2002. *Applied Logistic Regression Analysis*. Thousand Oaks, CA: Sage Publications.

Merton, R. C. 1974. On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance* 29:449–70.

Saunders, A., and L. Allen. 2002. *Credit Risk Measurement: New Approaches to Value at Risk and Other Paradigms*. New York: John Wiley.

Shumway, T. 2001. Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business* 74:101–24.

Sobehart, J. R., and S. C. Keenan. 1999. *An Introduction to Market-Based Credit Analysis*. New York: Moody's Investors Services.

Sobehart, J. R., and S. C. Keenan. 2002. Hybrid Contingent Claims Models: a Practical Approach to Modelling Default Risk, in Michael Ong, (ed.), *Credit Rating: Methodologies, Rationale, and Default Risk* 125–45. New York: Risk Books.

Sobehart, J. R., and S. C. Keenan. 2002. The Need for Hybrid Models. *Risk* February, 73–7.

Sobehart, J. R., and R. M. Stein. 2000. *Moody's Public firm Risk Model: A Hybrid Approach to Modeling Short Term Default Risk*. New York: Moody's Investors Services.

Stein, R. M. 2000. *Evidence on the Incompleteness of Merton-type Structural Models for Default Prediction*. New York: Moody's Investors Services.

Vassalou, M., and Y. Xing. 2004. Default Risk in Equity Returns. *Journal of Finance* 59:831–68.

Warga, A. 1998. Fixed Income Data Base. Working Paper, University of Houston.