Estimating Default with Discrete Duration and Structural Models*

Todd B. Walker[†]
University of Iowa

First Version: October 2005

Abstract

This paper estimates the conditional probability that a firm will default on its debt obligations and contributes to the corporate default literature in several ways. First, Bayesian estimation of discrete-time hazard models that include firm-specific effects is shown to outperform the traditional reduced-form methods. Second, a hybrid approach to default forecasting that reconciles the structural model of Merton (1974) with reduced-form models is introduced. The hybrid approach uses advanced option pricing techniques to improve the traditional estimate of expected default frequency. Finally, competing models are estimated and tested with over 36,000 firm-quarters of data from the U.S. industrial machinery and instruments sector. The new techniques are shown to significantly improve the forecasting of corporate default over existing methods.

Keywords: Default, Bankruptcy, Duration Analysis, Risk Neutral, Entropy
JEL Classification Numbers: C41, G33, G13

*I would like to thank Matt Billett, John Geweke, and Chuck Whiteman for their helpful comments and suggestions. All remaining errors are my own.

[†]E-mail: todd-walker@uiowa.edu, webpage: http://myweb.uiowa.edu/twaer

1 Introduction

This paper provides Bayesian estimates of corporate default probabilities by improving upon several well-established methods of default prediction. First, while it is common in the literature to assume that firms that are the same on a few measured dimensions are identical, I find that accommodating unobserved heterogeneity leads to models of firm survival that are empirically superior to the standard models. Second, I introduce a *hybrid* approach of default forecasting that enables me to use the best features of two diverse approaches to modeling: the widely-used structural model, first introduced by Merton (1974), and the increasingly popular reduced-form approach. The hybrid approach employs advanced option pricing techniques to improve default forecasting by effectively relaxing the unrealistic assumptions of the structural model, without compromising too much on the underlying theoretical structure. Finally, the default estimation methods introduced here are tested with over 36,000 firm-quarters of data from the U.S. industrial machinery and instrument sector. These methods are shown to improve forecasting of corporate default and bankruptcy.

I find evidence that a failure to account for firm heterogeneity in a discrete-time hazard model leads to significant *under-prediction* of default probabilities. This result is well known in other literatures (e.g., Cameron and Heckman (1998)) and has been discussed in the default setting, with respect to *industry* heterogeneity, by Chava and Jarrow (2004). Unobserved heterogeneity across firms could arise for several reasons; for example, differences in the cost of external financing (Davydenko (2005)), differences in management structure and quality, and changes in the laws that govern bankruptcy protection (Davydenko and Franks (2005)). A Markov chain Monte Carlo (MCMC) sampler is developed to estimate a model with firm heterogeneity. The Bayesian Deviance Information Criterion (DIC) is then calculated to assess the tradeoff between model fit and complexity. The DIC favors the model that accounts for firm heterogeneity over the model that ignores firm heterogeneity in spite of the nontrivial increase in model complexity. The two competing models are then tested out of sample, and again, the model accommodating firm heterogeneity is shown to outperform the traditional model.

Several studies of default prediction have included distance-to-default, as derived from the structural model of Merton (1974), as a time-varying covariate in reduced-form models (e.g., Duffie and Wang (2004), Bharath and Shumway (2004)). This paper also combines these two most popular models of default prediction, but in a different, yet intuitive, manner. In doing so, criticisms of the Merton model are taken into consideration. Specifically, I show how the output from a reduced-form model can be used to estimate a random default point for each firm. I then relax the assumption that the firm's value follows geometric Brownian motion. I assume that the distribution of firm value can be approximated by the distribution of the firm's equity and debt, which is assumed to follow a mixture-of-normals process. The risk-neutral probability of default is then derived using the maximum entropy technique described in Stutzer (1996).

Over 36,000 firm-quarters from the intersection of the Compustat and CRSP databases provided by the Wharton Research Database are used to show that the advances advocated here have more predictive power than traditional methods. The results and methods should be appealing to a large variety of practitioners especially given the expansion of credit derivatives markets and the implementation of the Basel II Accord, which requires all financial institutions to maintain some level of credit risk monitoring.

1.1 Related Literature

While there have been many variations of the structural model of corporate credit risk first introduced by Merton (1974), the empirical record of the structural model(s) for forecasting default and pricing corporate debt has been mixed. The most notable extensions have expanded upon the simplistic capital structure of the Merton model.[1] Merton assumed firms were capitalized with equity and a zero-coupon debt instrument. Geske (1977) assumed a more complex capital structure and showed that multiple default options for coupons, sinking funds, junior debt, safety covenants, or other payment obligations could be treated as compound options; Turnbull (1979) included corporate taxes and bankruptcy costs; Leland (1994) and Leland and Toft (1996) allowed the default barrier to be endogenously determined by stockholders. These improvements have made the Merton model more realistic, yet they have not translated into improved empirical performance. In testing five structural models of corporate bond pricing, Eom et al. (2004) found all models to have substantial prediction errors. Using a sample of 182 bond prices from 1986 to 1997, they found most models predicted yield spreads that were "often ridiculously small or incredible large." Nonetheless, the empirical implementation of the structural model has been championed by Moody's KMV (MKMV hereafter), a leading provider of corporate credit risk analysis. MKMV estimates a modified struc-

[1] See Elizalde (2005) for an exhaustive review.

tural model called the Vasicek-Kealhofer (VK) model (see Crosbie and Bohn (2001) and Vasicek (1984)), which uses an empirical distribution of distance-to-default to calculate commercially available *Expected Default Frequency*™(EDF). Recently, Bharath and Shumway (2004) constructed a simple alternative to the EDF of MKMV and showed that it outperforms the EDF as a predictor in hazard models and in out-of-sample forecasts.

In light of these empirical shortcomings, many have turned to reduced-form models as an alternative to the structural approach. Reduced-form models assume an unpredictable default time driven by a default intensity (see, Jarrow and Turnbull (1995), Jarrow et al. (1997), Duffie and Singelton (1999), and Hull and White (2001)). Duffie and Lando (2001) and Jarrow and Protter (2004) showed that the reduced-form approach is a special case of the structural model in which the value of the firm's assets are latent from the econometrician. Unlike the structural approach, reduced-form models assume an incomplete information set, implying the default time is appropriately modeled as a default intensity. This default intensity should then depend on *any* covariate that may reveal additional information about the firm's conditional probability of default. This affords flexibility in empirical implementation, and reduced-form models can be cast under the broad heading of duration analysis. Shumway (2001) demonstrated the superior performance of discrete-time hazard models over the traditional econometric models of Altman (1968) and Ohlson (1980) in estimating corporate default probabilities. Shumway showed that many of the accounting ratios used in Altman (1968) and Ohlson (1980) had little predictive power when market variables were included in the duration model.

More recently, Chava and Jarrow (2004) incorporated industry effects into a reduced-form model and argued that the likelihood of failure should differ across firms with otherwise similar financial conditions that operate in different industries. It seems plausible that this argument should extend to the firm level as well. In an empirical study of over 797 U.S. firms, Davydenko (2005) found that the market value of assets varied substantially across defaulting firms and concluded that "even if boundary-based models can be calibrated to predict the average probability of default, they are still likely to lack accuracy in the cross-section." He argued that while firm value and liquidity are significant predictors of default, their importance was mitigated or amplified by the ability of the firm to raise outside financing in times of need. Moreover, as recently emphasized by Hanson et al. (2005) and Pesaran et al. (2005), successful credit portfolio management depends critically on being able to account for firm-specific heterogeneity. Therefore accounting for firm heterogeneity, as is done here, may improve upon the existing methods of credit risk analysis.

While it seems natural to integrate both structural and reduced-form approaches so as to gain both the economic appeal of the structural approach and the tractability of the reduced-form model, there have been few attempts to do so. Most recently, Giesecke (2001) and Giesecke and Goldberg (2004) introduced an approach with a random default barrier based on incomplete information. The model assumed default is triggered when an observable firm value, that follows geometric Brownian motion, dropped below a random default barrier that is unobservable and independent of firm value. Giesecke and Goldberg (2004) documented the empirical consistency of their approach in predicting credit spreads relative to the Merton model. The hybrid approach introduced here is similar to the their model in that the default barrier is assumed to be a random variable, but my approach allows for a more flexible distribution for firm value.

2 Reduced-Form Model: Firm-Specific Effects

As recently emphasized by Shumway (2001) and Chava and Jarrow (2004), discrete duration analysis is a superior method of predicting corporate default relative to the static linear discriminant models of Altman (1968) and Zmijewski (1984). While most duration models are designed for a continuous-time approach (e.g., biomedical studies), economic data call for a discrete-time approach for two reasons. First, economic data arrive in weeks, months, or quarters and are therefore discrete by nature. Second, a basic assumption of continuous-time models is that failure times cannot tie. While banned in theory, a moderate number of tied failure times can be appropriately handled but a large number of ties, as is the case here, may lead to substantial bias in estimation (Fahrmeir and Tutz (1994)).

In discrete-time survival analysis, the timescale is partitioned into intervals $[a_0, a_1)$, $[a_1, a_2), ..., [a_{m-1}, a_m)$, $[a_m, \infty)$ where $a_0 = 0$ and a_m denotes the last observation. For the subsequent analysis, each interval will be defined as the duration between quarters for each firm. Identifying the discrete time index t with interval $[a_{t-1}, a_t)$, $t \in \{1, ...m + 1\}$, a discrete failure time T is considered, where $T = t$ denotes failure within interval t. In addition to duration T, we also observe a sequence of time-varying covariate vectors for firm i. Let $\mathbf{x}_{it}$ denote the $(q \times 1)$ vector of covariates for firm i at time t and let $\mathbf{X}_{it}$ denote the history of covariates up to quarter t for firm i. The discrete hazard function for firm i is defined as the conditional probability

$$\lambda_i(t|\mathbf{X}_{it}) = Pr(T = t|\, T \geq t, \mathbf{X}_{it}), \tag{2.1}$$

for failure in quarter t, given survival up to this quarter. The corresponding discrete survival function is given by

$$S(t|\mathbf{X}_{it}) = Pr(T > t|\mathbf{X}_{it}) = \prod_{j=1}^{t}[1 - \lambda(j|\mathbf{X}_{ij})], \tag{2.2}$$

which is the probability of surviving beyond a_t. The survival information for each firm i can be summarized by (T_i, δ_i), which records the observation time T_i and δ_i, an indicator of survival taking the value of 1 if the firm defaults, and 0 if the firm exits for reasons other than default (e.g., merger, acquisition, delisting, etc.).[2]

By defining the event indicators $y_{it}, t = 1, ..., T_i$ for firm i, where

$$y_{it} = \begin{cases} 0, & \text{if } t < T_i \\ \delta_i, & \text{if } t = T_i, \end{cases}$$

discrete-time models can be cast into the framework of binary regression models. This is because the hazard function (2.1) for firm i can be written as a binary response model

$$\lambda_i(t|\mathbf{X}_{it}) = Pr(y_{it} = 1|\mathbf{X}_{it}) = g(\eta_{it}) \tag{2.3}$$

where $g(\cdot)$ is a known cdf linking the probability of default with the predictor η_{it}, a function of the covariates. Common choices for discrete survival models of form (2.3) include the grouped Cox model, probit, and logit model.[3] I follow Shumway (2001), Hillegeist et al. (2004), and Chava and Jarrow (2004) in specifying $g(\cdot)$ as the logistic distribution. The conventional model is then

$$Pr(y_{it} = 1|\eta_{it}) = \frac{\exp(\eta_{it})}{1 + \exp(\eta_{it})} = \frac{1}{1 + \exp(-\eta_{it})}$$

[2]The data set includes several firms that exit for reasons other than default. Duffie and Wang (2004) model this process as a separate exit intensity.

[3]These models yield similar results when the intervals are relatively short (Thompson (1977)).

with linear predictor

$$\eta_{it} = \alpha_t + \boldsymbol{\beta}'\mathbf{x}_{it} + b_i. \tag{2.4}$$

The linear predictor consists of fixed effects $\beta_1, ..., \beta_q$ to capture the cross-sectional dependencies in the data for each covariate, a baseline hazard $\alpha_1, ...\alpha_{m+1}$ to capture temporal dependencies, and random effects $b_1, ..., b_N$ to account for unobserved individual heterogeneity in firms. Several studies have documented how not accounting for unobserved heterogeneity can lead to substantial bias in classical estimates, and it has been argued that it is safer to assume the presence of heterogeneity unless it is obvious that none exists (e.g., Heckman and Singer (1984), Cameron and Heckman (1998)).

Each period prior to T_i, firm i survives and contributes (2.2) to the likelihood; the firm then contributes (2.1) at default, and therefore firm i's contribution to the likelihood is the T_i-fold product,

$$\begin{aligned} p(y_{i1}, y_{i2}, ...y_{iT_i}|\mathbf{X}_{it}, \boldsymbol{\beta}, \boldsymbol{\alpha}, b_i) &= \prod_{t=1}^{T_i} [\lambda(t|\mathbf{X}_{it})]^{y_{it}} [1 - \lambda(t|\mathbf{X}_{it})]^{(1-y_{it})} \\ &= \prod_{t=1}^{T_i} \frac{\{\exp(\alpha_t + \boldsymbol{\beta}'\mathbf{x}_{it} + b_i)\}^{y_{it}}}{1 + \exp(\alpha_t + \boldsymbol{\beta}'\mathbf{x}_{it} + b_i)} \end{aligned} \tag{2.5}$$

Note that this is the exact likelihood obtained by treating the indicators y_{it} as independent Bernoulli trials across discrete time points t with $\mathbb{E}(y_{it}|\boldsymbol{X}_{it}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \lambda(t|\mathbf{X}_{it}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ (Allison (1982)). However, by including the random-effects parameters, this model cannot be estimated with a standard logistic regression model, and more advanced techniques are required (Biggeri et al. (2001)). I will take a Bayesian approach and estimate the parameters of the model via Markov chain Monte Carlo (MCMC) sampling.[4]

2.1 Bayesian Inference for Duration Model

The discrete-duration model is completed with specification of prior distributions for the free parameters. Combining the likelihood function with the prior distributions on $\boldsymbol{\beta}$, b_i for $i = 1, ..., N$ (where N is the total number of firms), α_k for $k = 1, ..., m+1$ and the variance components (σ^2, υ^2) I obtain, by Bayes rule, the joint posterior distribution for

[4]WinBugs software (Spiegelhalter et al. (2000)) was used to perform the MCMC analysis. WinBugs is freely available at http://www.mrc-bsu.cam.ac.uk/bugs/.

the hazard model:

$$p\left(\boldsymbol{\beta}, \alpha_1, ..., \alpha_{m+1}, b_1, ..., b_N, \sigma^2, \upsilon^2 \middle| \left[\mathbf{X}_{iT_i}, \delta_i, T_i\right]_{i=1}^{N}\right)$$
$$\propto \prod_{i=1}^{N} p(y_{i1}, y_{i2}, ...y_{iT_i}|\mathbf{X}_{it}, \boldsymbol{\beta}, b_i)p(\boldsymbol{\beta})p(b_i|\sigma^2)p(\sigma^2)p(\boldsymbol{\alpha}|\upsilon^2)p(\upsilon^2)$$

The five prior distributions are specified as follows: (i) *A priori*, the fixed effects parameters $\boldsymbol{\beta}$ are assumed to be distributed as multivariate normal with zero mean and covariance matrix ςI, where $\varsigma = 10^4$ and I is the $(q \times q)$ identity matrix. The large variance implies the the mass of the distribution is widely dispersed.

(ii) In order to avoid strong structural assumptions about the way the hazard depends on time, the baseline effects α_t are modelled as piecewise constants. Following Gamerman and West (1987) and Fahrmeir and Knorr-Held (1997), I allow for prior correlation of these parameters in the form of the random walk process:

$$\alpha_t = \alpha_{t-1} + \epsilon_t, \qquad t = 1, ..., m$$

where α_0=0, $\epsilon_t \sim N(0, \upsilon^2)$ and ϵ_t is independent α_{t-s} for all s.[5]

(iii) The random effect terms are assumed to be independent and identically distributed Gaussian random variables with mean zero and variance σ^2. By setting the mean equal to zero, the random effects are interpreted as deviations from the overall average.

(iv)– (v) For the variance components, I follow the usual practice of assigning inverted gamma prior distributions. That is, the prior pdf for σ^2 is given by

$$\sigma^2 \sim \mathcal{IG}(\kappa_\sigma, \omega_\sigma)$$

and the prior for υ^2 is of the same form but with shape and scale parameters κ_υ and ω_υ. The full model is summarized with the likelihood, $\mathcal{L}$, and prior distributions for the

[5] It is typical for the variance υ^2 to depend on the length of the interval, but in this study interval lengths are all the same.

parameters:

$$\mathcal{L} = \prod_{i=1}^{N} p(y_{i1}, y_{i2}, ...y_{iT_i} | \mathbf{X}_{it}, \boldsymbol{\alpha}, \boldsymbol{\beta}, b_i)$$

$$\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \varsigma I_{q \times q}),$$

$$b_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \quad \text{for } i = 1, ...N$$

$$\alpha_0 \sim \mathcal{N}(0, \upsilon^2), \qquad \alpha_t \sim N(\alpha_{t-1}, \upsilon^2) \qquad \text{for } t = 1, ..., m+1$$

$$\upsilon^2 \sim \mathcal{IG}(\kappa_\upsilon, \omega_\upsilon), \qquad \sigma^2 \sim \mathcal{IG}(\kappa_\sigma, \omega_\sigma).$$

Draws from the posterior distribution will be obtained via Gibbs sampling. Appendix B.1 contains the details of the sampling algorithm, which simulates a Markov chain in high dimensional space. After testing appropriate convergence diagnostics, draws from the posterior will be used to estimate default probabilities.

Since variations of (2.4) will be considered, an essential task of the analysis is the formal comparison of alternative models. One method for comparing models employed here is the Deviance Information Criterion (DIC) suggested by Spiegelhalter et al. (2001). DIC is especially useful for model comparison when complex hierarchical models are to be estimated. This is because Bayes factors, the usual method for carrying out Bayesian model comparison, are difficult to obtain, while DIC is easily computed given MCMC output.[6] Calculation of the Bayes factor for comparing any two models requires the marginal likelihoods and thus if the dimension of the parameter space is large, marginalization over each parameter vector can pose a formidable computational challenge. From a frequentist's point of view, model assessment is based on *deviance*, which is defined as the difference between the log-likelihoods of a fitted model and a "perfectly-fit" model; that is, a model with as many parameters as data points, yielding a perfect fit. From a Bayesian point of view, Dempster (1974) suggested the posterior distribution of deviance, given by

$$D(\theta) = -2 \ln L(y|\theta) + 2 \ln g(y),$$

as a measure of model assessment. This combines the likelihood, $L(y|\theta)$, with a fully specified standardizing term, $g(y)$, that is a function only of the data. DIC is based on

[6]In fact, one of the motivating examples used in Spiegelhalter et al. (2001) is the mixed effects (i.e., estimating both random and fixed effects) model of Laird and Ware (1982).

comparisons of posterior distributions of deviance, and is given by

$$DIC(\theta) = \overline{D(\theta)} + p_D \tag{2.6}$$
$$p_D = \overline{D(\theta)} - D(\bar{\theta}) = E_{\theta|y}[D(\theta)] - D(E_{\theta|y}[\theta]) = \mathbb{E}_{\theta|y}[-2\ln L(y|\ \theta)] + 2\ln L(y|\bar{\theta})$$

where $\overline{D(\theta)}$ is the posterior mean of the deviance ($\mathbb{E}_{\theta|y}[D]$), which summarizes the overall fit of the model. The better the model fits the data, the larger the values for the likelihood. Thus $D(\theta)$, defined as -2 times the log-likelihood, attains smaller values for increased model fit. Model complexity is measured by the effective number of parameters p_D, defined as the mean deviance $\overline{D(\theta)}$ minus the deviance of means $D(\bar{\theta}) = D(\mathbb{E}_{\theta|y}[\theta])$. The term $-2\ln L(y|\theta)$ can be interpreted as the residual information in the data y after conditioning on θ, and thus is a measure of uncertainty. Therefore, p_D can be regarded as the expected difference of the true over the estimated residual information in data y conditional on θ, and can be interpreted as a measure of expected reduction in uncertainty due to estimation. Note that the DIC can be rewritten as $D(\bar{\theta}) + 2p_D$, which can be interpreted as a more traditional measure of fit plus complexity. Spiegelhalter et al. (2001) give conditions for which the DIC is a generalization of the Akaike information criterion (AIC), and the DIC's relationship to the Bayesian (or Schwarz) information criterion (BIC).

3 Structural Model

The main drawback of the duration model just described is the lack of theoretical underpinning. Duration models do not consider the relationship between default and firm value in an explicit manner. Moreover, time-of-default is determined by an exogenously given jump process. In contrast, Merton (1974), using the methodology of Black and Scholes (1973), demonstrated how equity could be modeled as a call option on the assets of the firm with a strike price equal to the firm's liabilities. By assuming a simple capital structure, Merton was able to calculate the default probability via (now) well-known derivative pricing equations.

Specifically, assume the firm's assets are financed by equity issued at time t denoted by S_t, and zero-coupon debt issued at t (D_t) with a face value of F and maturity date J. The market value of the firm at any date t is given by the sum of the market value of debt and equity. Therefore, the accounting identity $V_t = S_t + D_t$, where V_t denotes firm value, holds for each period. Under these assumptions, the bondholders are entitled to a time-J cash flow of $\min[V_J, F]$ and since equity holders are the residual claimants, the value of equity at time J is given by $\max[V_J - F, 0]$. At any time $t < J$, the value of these derivative securities is

$$S_t = e^{-r(J-t)} \mathbb{E}_t^Q \big\{ \max[V_J - F, 0] \big\} \tag{3.1}$$
$$D_t = e^{-r(J-t)} \mathbb{E}_t^Q \big\{ \min[V_J, F] \big\} \tag{3.2}$$

where the expectation is taken with respect to the risk-neutral probability measure and the risk-free rate r is assumed to be constant over time.

Merton (1974) assumed firm value followed geometric Brownian motion

$$d \ln V_t = \left(\mu_V - \frac{\sigma_V^2}{2} \right) dt + \sigma_V dW_t$$

where μ_V and σ_V are, respectively, the expected return and volatility rates, and W_t is a Wiener process. Under this assumption, Black-Scholes derivative pricing equations imply that (3.1) and (3.2) become

$$S_t = V_t \Phi(d_t) - F e^{r(J-t)} \Phi(d_t - \sigma_V \sqrt{J-t}) \tag{3.3}$$
$$D_t = F e^{-r(J-t)} \left(\frac{V_t}{F e^{-r(J-t)}} \Phi(-d_t) + \Phi(d_t - \sigma_V \sqrt{J-t}) \right)$$

where $d_t = (\ln(V_t/F) - (r - \sigma_V^2/2)(J-t))/(\sigma_V \sqrt{J-t})$ and $\Phi(\cdot)$ is the standard normal distribution function. Given that the value of the firm's assets follows a geometric Brownian motion, the value of the assets at any future date J is given by:

$$\ln(V_{t+J}) = \ln(V_t) + (\mu_V - \sigma_V^2/2)J + \sigma_V \sqrt{J} \varepsilon_{t+J}$$
$$\varepsilon_{t+J} = \frac{W(t+J) - W(t)}{\sqrt{J}}, \qquad \varepsilon_{t+J} \sim N(0,1)$$
$$V_{t+J}/V_t \sim \mathcal{LN}((\mu_V - \sigma_V^2/2)J, \sigma_V^2 J) \tag{3.4}$$

where the last line is the well-know result that V_{t+J}/V_t has a log-normal distribution.

Therefore the probability of default is

$$\pi_t^D = Pr\left(\ln(V_t) - \ln(F) + \left(\mu - \frac{\sigma_V^2}{2}\right)J + \sigma_V\sqrt{J}\varepsilon_{t+J} \leq 0 \right)$$
$$= Pr\left(-\frac{\ln(V_t) - \ln(F) + (\mu_V - \sigma_V^2/2)J}{\sigma_V\sqrt{J}} \geq \varepsilon_{t+J} \right).$$

We can then define the distance to default as

$$DD_t = \frac{\ln(V_t/F) + (\mu_V - \sigma_V^2/2)J}{\sigma_V\sqrt{J}}$$

Default occurs when the ratio of firm value to debt (V_t/F) drops below unity or the log of the ratio is negative. The distance to default DD_t can be interpreted as a z-score, which gives the number of standard deviations the log of this ratio needs to deviate from its mean in order for default to occur. In other words, the probability of bankruptcy depends upon the distance between the current value of the firm's assets and the face value of its liabilities, adjusted for the expected growth in asset value relative to asset volatility. Figure 1 replicates Figure 8 of Crosbie and Bohn (2001), which is the graphical representation of DD_t.

In order to implement this approach empirically, we need reliable estimates of the mean and volatility of firm value. Of course, the value of the firm is not observable for all t, but the value of equity S_t is typically readily available. Notice (3.3) is derived under the assumption of risk-neutrality, and hence, only the risk-free rate enters into the call option equation. Therefore, given an initial guess for the volatility of firm value and knowledge of the strike price F, the nonlinear equation (3.3) can be used to solve for firm value at any date t. This provides a sample of firm values from which mean and volatility estimates can be used to approximate DD_t. The Moody's KMV (MKMV) estimation procedure takes advantage of this observation and is a simple two-step algorithm that is repeated until convergence.[7] To explain this algorithm, I adopt the notation of Duan et al. (2004); suppose we are able to observe a time series of equity values $\{S_0, S_h, S_{2h}, ..., S_{nh}\}$ where h is the length of time between two observations (measured in years) and $n+1$ is the length of the observed time series. Given an initial arbitrary starting value for asset volatility

[7] It should be noted that this procedure is not the exact method used by MKMV. In estimating asset volatility, MKMV incorporates an undisclosed Bayesian model that includes country, industry and size variables to produce a more accurate predictive estimate of the firm's asset volatility.

and an approximate value for F, the MKMV algorithm at iteration m is

1. Given a value for asset volatility ($\hat{\sigma}^{(m)}$) and a proxy for the default point F, compute the implied asset value time series $\{\hat{V}_0(\hat{\sigma}^{(m)}), \hat{V}_h(\hat{\sigma}^{(m)}), ..., \hat{V}_{nh}(\hat{\sigma}^{(m)})\}$ from $\{S_0, S_h, ..., S_{nh}\}$ using (3.3).
2. Compute the implied asset returns $\{\hat{R}_1^{(m)}, \hat{R}_2^{(m)}, ..., \hat{R}_n^{(m)}\}$ where

$$\hat{R}_i = \ln(\hat{V}_{ih}(\hat{\sigma}^{(m)})/\hat{V}_{(i-1)h}(\hat{\sigma}^{(m)}))$$

   Update the asset drift and volatility parameters

$$\bar{R}^{(m)} = \frac{1}{n}\sum_{k=1}^{n}\hat{R}_k^{(m)}$$

$$\left(\hat{\sigma}^{(m+1)}\right)^2 = \frac{1}{nh}\sum_{k=1}^{n}\left(\hat{R}_k^{(m)} - \bar{R}^{(m)}\right)^2$$

$$\hat{\mu}^{(m+1)} = \frac{1}{h}\bar{R}^{(m)} + \frac{1}{2}\left(\hat{\sigma}^{(m+1)}\right)^2$$

3. Repeat steps 1 and 2 until the absolute difference between $\hat{\sigma}^{(m)}$ and $\hat{\sigma}^{(m+1)}$ is less than 10^{-3}.

It is typical to achieve convergence within three iterations of the algorithm. In the analysis below, distance-to-default is computed on a monthly basis using daily equity data.

3.1 A Hybrid Approach

As noted in Section 1.1, the empirical performance of the structural model has been poor. Specifically, there are three main criticisms of the implementation of the Merton model just described. First, the Gaussian distribution (3.4) may not have sufficiently thick tails to capture swift changes in firm value. Figure 2 plots the empirical distribution of daily returns of a firm that eventually declared bankruptcy, along with the corresponding normal density with moments set equal to the sample mean and sample variance. Clearly, the assumption of normality is not a good one. Zhang et al. (2005) found superior performance by incorporating stochastic volatility and jumps in a structural model. Second, implementation of this approach must rely on a particular assumption about the level of the default boundary F, on which there has been little empirical guidance. Huang and Huang (2003) assumed that the boundary is 60 percent of the face value of debt; Leland (2004) used a model-implied level of 73 percent. The industry-standard MKMV model assumes that default happens if at maturity the value of assets is lower than short-term debt plus 50% of long-term debt, measured at book values (Crosbie and Bohn (2001)). Third, using daily equity data to calibrate the model could introduce substantial noise into the estimation procedure (Duan and Fulop (2005)).

Despite these criticisms, a nice property of the structural model is that it can be updated as often one wishes. It relies on equity data, which are available at a very high frequency. This is especially important to practitioners who must update, in many cases, on a weekly basis. In contrast, due to the dependence on accounting data, the discrete duration model can only be updated on a quarterly basis. Thus *forecasts* of default using the discrete duration model could only be updated quarterly.[8]

The hybrid model introduced here addresses each of the criticisms of the structural model, while at the same time, maintaining useful properties of the model. Suppose after estimating the discrete duration model, the *quarter* t posterior mean probability of default for firm i is given by 0.01. Suppose further that one has on hand a time series that proxies for *monthly* firm values $\{\hat{V}_{j-n}, ..., \hat{V}_{j-1}, \hat{V}_j\}$. Then assuming a distribution for firm value that better fits the data (e.g., a mixture of normals), we can treat the current default barrier F as a random variable. More specifically, given a predictive density for V_{j+J}, we can treat the random default barrier as a value-at-risk problem. That is, we seek an F such that

$$Pr(V_{j+J} \leq F) = Pr(\ln(V_{j+J}/V_j) \leq \ln(F/V_j)) = 0.01. \tag{3.5}$$

By simulating draws from the predictive distribution of V_{j+J}, we would be able to pin down the random default barrier F implied by the reduced-form model of Section 2. Figure 3 gives a graphical representation of the hybrid model. Instead of specifying a default barrier *a priori*, the hybrid model takes advantage of the output from the discrete

[8]Forecasting default with the discrete duration model requires a time-series model for the covariates x_{it}. This is a subject of ongoing research.

duration model to solve for an endogenous default barrier.

As mentioned above, the hybrid model addresses each of the criticisms of the structural model. First, the hybrid model relaxes the assumption that firm value follows a geometric Brownian motion process. Given a proxy for firm value, one is free to choose the corresponding distribution of firm value. Second, the hybrid model does not rely on daily equity data. The proxy for firm value could employ the monthly average of daily equity data, thus mitigating the presence of market noise. Finally, the hybrid model specifies an *endogenous* default barrier implied by the discrete duration model.[9] However, the hybrid model does assume that the probability of default is given by the discrete duration model. Thus, an important question at this point is; what "new" information does the hybrid model offer? The usefulness of the hybrid model comes from the ability to produce risk-neutral default probabilities. These probabilities serve as an upper bound on the probability of default implied by the discrete duration model. Most importantly, the hybrid model can be updated at a much higher frequency than the discrete duration model, and hence these upper bounds are available when practitioners need them. The next section describes how to obtain risk-neutral default probabilities using the technique of Stutzer (1996).

3.2 Maximum Entropy, Risk Neutral Probability of Default

Stutzer (1996) introduced a nonparametric approach to derivative pricing that, given a predictive density for the underlying asset, uses a maximum entropy principle to price the derivative security. Stutzer's method forces the empirical predictive density to satisfy a set of moment restrictions not used in constructing the predictive density. This transformed distribution is as close to the original density in the sense that the Kullback-Leibler Information Criterion (or relative entropy) is minimized.

Within the context of the structural model, implementing this technique amounts to using the predictive density of firm value to price the derivative securities D_t and S_t correctly. Given that the expectation is taken with respect to the risk-neutral probability distribution, one will obtain risk-neutral probabilities of default. As an illustrative example, suppose one had a sample of N weighted draws of firm value at time J, denoted by $\hat{V}_{iJ}$, $i = 1, ..., N$ with weights π_i, $i = 1, ..., N$ determined by the predictive density. A random sample from the predictive density would stipulate $\pi_i = 1/N$ for all i. Robertson

[9] Recall that this is actually the result of Duffie and Lando (2001) and can be proved formally. That is, if firm value is unobservable, the structural model "reduces to" the discrete duration model.

et al. (2005) define "new" information as a theoretically-justified moment restriction that does not hold given weights π_i and draws $\hat{V}_{iJ}$. For example, we know that in the present context, D_t and S_t should equal the expected (in risk-neutral measure) discounted (at the risk-free rate) present value of the firm at horizon J (see (3.1)). But the discrete duration model gives the objective predictive density, not the risk-neutral one. Stutzer's suggestion is to choose the risk-neutral probabilities as close as possible to the objective probabilities such that the pricing constraint (3.1) holds. In other words, we know that the following inequality

$$e^{-r(J-t)} \sum_{i=1}^{N} \pi_i \{\max[\hat{V}_{iJ} - F, 0]\} \neq S_t \tag{3.6}$$

holds because the π_i's are not the risk-neutral weights needed to price the derivative security. Thus, we need to reweight the draws with π_i^* for $i = 1, ..., N$ such that (3.6) is satisfied. Of course, there are an infinite number of ways to reweight the draws; however a straightforward, axiomatically rationalized and interpretable way to accomplish this task is to minimize the Kullback-Leibler Information Criterion (KLIC) divergence between the two probability measures.[10] This divergence is given by

$$K(\pi : \pi^*) = \sum_{i=1}^{N} \pi_i^* \log\left(\frac{\pi_i^*}{\pi_i}\right). \tag{3.7}$$

Therefore, the convex minimization problem is to minimize (3.7) subject to

$$\pi_i^* \geq 0, \qquad \sum_{i=1}^{N} \pi_i^* = 1, \qquad e^{-r(J-t)} \sum_{i=1}^{N} \pi_i^* \{\max[\hat{V}_{iJ} - F, 0]\} = S_t$$

Using the Lagrange multiplier method and taking the first-order condition yields

$$\pi_i^* = \frac{\pi \exp(\gamma \hat{V}_{iJ})}{\sum_{i=1}^{N} \pi_i \exp(\gamma \hat{V}_{iJ})}$$

where γ is the Lagrange multiplier on the moment restriction, which can be computed

[10]If it were the case that the draws came from a random sample, minimizing the KLIC divergence is equivalent to maximizing the Shannon entropy.

as the solution to the well-behaved convex minimization problem

$$\gamma = \arg\min \sum_{i=1}^{N} \pi_i \exp(\tilde{\gamma}[\max[\hat{V}_{iJ} - F, 0] - S_t]).$$

Therefore, given a predictive density for firm value, this procedure transforms the actual default probabilities implied by the discrete duration model into the risk-neutral default probabilities implied by the moment restriction (3.1).[11] As noted by Delianedis and Geske (1999), risk-neutral probabilities are more accurate and are just as valuable in pricing corporate debt.

4 Data

The subsequent empirical analysis focuses on publicly-traded firms that are members of the industrial machinery and instrument sector from 1962 to 2004.[12] I use quarterly accounting data for each individual firm obtained from COMPUSTAT, and daily equity data from the Center of Research in Security Prices (CRSP); therefore firms must have complete information on both the Compustat and CRSP databases. Duffie and Wang (2004) analyze a similar data set and note that focusing on one sector may improve inference by mitigating some of the "industry effects" documented by Chava and Jarrow (2004). Moreover, relative to other sectors, defaults in this sector occur more frequently and are distributed more uniformly across time.[13]

A firm is defined as bankrupt at the end of the current quarter if that firm files for Chapter 7 or Chapter 11 bankruptcy protection at any time during the current quarter.[14] I also consider a broader definition of default. Firms that are delisted from one of the major exchanges and declare bankruptcy within a five-year period, receive a D credit rating from Standard & Poor's Issuer Credit Rating (ICR), or default on any financial contract are categorized as being in default at the end of the current quarter.[15] This broader definition is relevant because firms that are delisted from one of the major ex-

[11] Stutzer (1996) describes this "risk-neutralization" procedure in more detail.

[12] More accurately, I use compustat's primary industrial classification numbers (DNUM) between 3500 and 3600.

[13] For example, consider the banking industry where a majority of the defaults in the last 60 years occurred in the decade between 1982 and 1992.

[14] This bankruptcy indicator is given primarily by annual COMPUSTAT data item AFTNT35, which documents reasons for withdraw from the database.

[15] Campbell et al. (2005) construct a similar default indicator.

changes are often in serious financial trouble and typically declare bankruptcy within a few years.

Since duration analysis specifies time-to-default as the dependent variable, the exact month of bankruptcy or default is recorded. This is to ensure that market and accounting variables are observable during the interval in which default occurs. Once a firm is classified as a default, it does not reenter the sample.[16] Compustat and CRSP data are supplemented with data obtained from SEC filings and the Directory of Obsolete Securities to identify the exact month of failure.

Following Shumway (2001), I use the firm's *trading* age as the dependent variable. This assumes some degree of homogeneity across firms that are newly listed on one of the three major exchanges, which is not unrealistic since firms are required to meet a specific set of requirements prior to listing. Therefore firms that have valid data on COMPUSTAT and CRSP but began trading prior to 1962 (the first date that COMPUSTAT begins quarterly coverage of publicly traded firms) are dropped from the sample. If the firm traded for less than one year, then the observation is also dropped from the sample. Missing quarterly data are replaced with annual data when available.[17]

The sample consists of 916 firms, 67 of which filed for bankruptcy and and 153 were classified as defaulting firms. Table 1 lists the number of active firms, bankrupt and defaulting firms since 1970. Prior to 1970 there were no recorded defaults in this industry and only 9 out of 1,283 recorded failures for the entire COMPUSTAT database occurred before 1971 (Duffie and Wang (2004)). Thus, the growth in the number of firms in the industry over time and the increase in the frequency of default implies that the subsequent analysis will be influenced largely by firms in the latter half of the time series.

Quarterly accounting data from COMPUSTAT along with monthly and daily equity daily from CRSP were used to construct explanatory variables at the individual firm level. This paper follows Campbell et al. (2005) in covariate construction.[18] The discrete duration model was fit with five firm-specific covariates, two accounting variables and three market variables, and one "extertnal" covariate. For the latter, I follow Duffie and Wang (2004) and use the percentage change in seasonally-adjusted personal income

[16] Very little information is lost given that only two firms that defaulted would have reappeared in the sample.

[17] By replacing the missing quarterly data with annual data, the number of firms included in the study increases from 871 to 916. This is mainly because quarterly data prior to 1971 are sparse.

[18] Several covariates employed in past studies were examined here, including the variables analyzed by Altman (1968) and Zmijewski(1984). The five covariates selected provided the best out-of-sample default prediction.

growth (INC) obtained from the National Economic Accounts of The Bureau of Economic Analysis as an external covariate to account for temporal dependencies not captured by the baseline parameter α_t. The next section describes the importance of this covariate in more detail.

The accounting ratios examined here include total liabilities relative to market total assets (TLMTA), which provides a standard measure of firm leverage. Following Campbell et al. (2005), I also find better explanatory power using market-valued total assets, defined as firm market equity plus total liabilities, instead of total assets. Net income to total assets (NITA) provides a measure of profitability. Unlike with TLMTA, I did not find better predictive power using the market value of total assets when constructing NITA. To the two accounting variables, I add three market variables: the monthly standard deviation of equity returns (SIGMA), the excess equity return relative to the return of the S&P 500 (EXRET), and relative size (RSIZE), measured as the log ratio of a firm's market capitalization to that of the S&P 500 index. To capture market data at a monthly frequency, I use a three-month moving average with increasing weights (1/6, 1/3, 1/2) for time periods $t-2$, $t-1$ and t, respectively in entering EXRET and SIGMA into the discrete hazard model for observation period t. The accounting variables are Winsorized at the 1st and 99st percentiles of their cross-sectional distributions to mitigate the influence of outliers. The variables TLMTA, NITA, and RSIZE exhibited substantial non-stationarity for a large majority of the firms in the sample. To induce stationarity, these covariates were first differenced. Appendix A describes the covariate construction in more detail.

Table 2 provides summary statistics for the five explanatory variables. As emphasized by Campbell et al. (2005), it is important to make a few observations regarding the distributions of the covariates. First, value-weighted statistics would look very different from those reported in Table 2. The differences between mean and median of most the variables suggests the distributions are skewed by under-performing firms. Second, the statistics represent a summary of the cross sectional *and* time series variation in the variables. The cross-sectional averages of the statistics have experienced several trends over the last 3 decades. Third, the statistics reported for the defaulted and bankrupt firms are only for the quarter in which the firm defaulted or declared bankruptcy. The preceding quarters had similar statistical properties and therefore one would expect that these covariates would have some predictive power.

5 Empirical Analysis

5.1 Duration Analysis

In analyzing the discrete-time duration model, in addition to (2.4) itself, I estimated three variations:

$$\mathcal{M}_1 : \alpha + \boldsymbol{\beta}'\mathbf{x}_{it}, \quad \mathcal{M}_2 : \alpha_t + \boldsymbol{\beta}'\mathbf{x}_{it}, \quad \mathcal{M}_3 : \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + b_i.$$

Model $\mathcal{M}_1$ examines only fixed-effects and does not allow for temporal dependency in the baseline (i.e., α does not contain a subscript t), nor does it account for firm-specific effects. The upside of this model is that it is very easy to estimate. As described in Shumway (2001), a simple logit program with modified standard errors can be employed for estimation. The downside is that the simplicity may not adequately capture the temporal and firm-specific nature of default data.

Model $\mathcal{M}_2$ allows for a baseline that changes with firm trading age. Figure 4 shows the empirical hazard estimate, defined as the number of firms defaulting in period t divided by the number of firms at risk of default in period t, for the first five years of trading. Figure 4 would suggest that no clear pattern for a time-varying baseline exists and thus would add little to default prediction. Moreover, the downside of estimating a baseline that changes with *every* quarter of duration is that at the end of the sample period, there are typically few observations and model fit is very poor. A solution is to introduce a series of dummy variables for different quarterly intervals (Baker and Rea (1998), Campolieti (2001)). I use dummy variables to control for a series of 15 intervals; each year for the first 10 years, years 11-15, 16-20, 21-25, 26-35, and 36-45. As discussed in Section 2.1, the baseline is estimated with a flexible random walk process. I follow the usual practice of assigning a noninformative prior to the variance component of the random walk process; that is, I assume that the precision $\tau_\nu = 1/\nu^2$ is distributed gamma with scale and shape parameters both set to 0.01. This implies that the precision has an *a priori* mean of 1^{-4} and therefore, the distribution of α_t is very diffuse about α_{t-1}.

Models $\mathcal{M}_3$ and $\mathcal{M}_4$ introduce firm-specific, random effects. These models account for unobserved heterogeneity, which could arise due to misspecification of the functional form of the model, and the omission of important but perhaps latent variables from the conditioning set (e.g., the cost of raising additional capital). One potential downside of mixed-effects models (i.e., models that include both fixed and random effects) is that convergence of the MCMC algorithm is considerably slower than the previous models. For example, the computational time required for a single draw of the mixed-effects model is, on average, 65% longer than a single draw from the fixed-effects model. In setting the prior distribution for the random-effect parameter b_i, I assume the precision $\tau_\sigma = 1/\sigma^2$ follows a gamma distribution with shape parameter of 3 and scale of 0.65.[19] This prior distribution seems plausible given that the model implies 95% of firms with identical covariates will have log(odds) of default occurring within the range of $2 \times 1.96/\sqrt{\tau_\sigma}$. The flexibility of the prior distribution allows for roughly one order of magnitude difference between the odds of default for firms with identical covariates. That is, 95% of firms are assumed to have log(odds) in a range of width $\log(10) \approx 2.3$, and therefore $(2 \times 1.96/2.3)^2 = 2.9$ seems reasonable for τ_σ. The prior also suggests low probability of a two-fold variation in default odds ($\log(100) = 4.6$, thus $\tau_\sigma = (2 \times 1.96/4.6)^2 = 0.73$). A gamma distribution with parameters (3, 0.65) has a mean of 3 and 89% probability of exceeding 0.73.

Each model was run with 42,000 draws from the posterior simulator with the first 2,000 draws discarded. Convergence diagnostics developed by Raftery and Lewis (1992) and Geweke (1992) suggested convergence of the posterior simulator.[20] The remaining 40,000 draws were used for posterior and predictive analysis. All of the coefficients are of the correct sign and are very similar to maximum likelihood estimates obtained by the author.

Table 3 reports the posterior mean, standard deviation, 95% credible interval and DIC for models $\mathcal{M}_1$ and $\mathcal{M}_3$. The posterior distributions for the parameters in all the models examined suggest the covariates are important predictors of default. The lower DIC for model $\mathcal{M}_3$ implies the data favor the model with random effects despite the nontrivial increase in model complexity (a ten-fold increase in P_D).[21] The prior specification assumed a one-fold variation in the log(odds) of default, implying a standard deviation for the random-effect parameter of roughly 0.587. The posterior mean and 95%

[19] Smith et al. (1995) discuss prior specification for mixed-effects models in more detail.

[20] Geweke's separated partial means statistic tests whether the mean from the first 20% of the MCMC sample and the last 50% are identical. A Z-test of the hypothesis of equality of these two means is carried out and the corresponding chi-squared marginal significance is calculated. This statistic had a low value of 0.5287 for the random effects variance σ^2.

[21] Formal comparison of the DIC statistic for each model continues to be a subject of on-going research. Since the DIC is computed with MCMC output it is subject to Monte Carlo sampling error; however, there is currently no easy and precise way to calculate standard errors for DIC values. Nonetheless, given the difference in DIC values across the two models, it would require a rather large standard error to render the difference insignificant.

credible interval for the random effect standard deviations are 0.689 (0.374, 0.981); thus the data suggest that firm-specific effects are also significant predictors of default and vary considerable across firms.

Figure 5 plots the cross-sectional industry-average default probability implied by the fixed-effects model and the random-effects model with random-effect parameter b_i of the 90th percentile firm (i.e., the firms were sorted according to the posterior mean of their respective random-effects parameters and the 90th percentile firm was chosen). The random-effects parameter more than doubles the estimated default probability of the industry average. Therefore, if there are many firms with similar covariates (or *forecasted* covariates) adding random effects to the model effectively sorts these firms.

Table 4 gives posterior estimates for model $\mathcal{M}_2$. The posterior mean baseline ($alpha_t$) varies from a low of -6.456 in trading-year 10 with standard deviation of 0.507, to a high of -4.225 in trading years 16-20 with standard deviation of 0.602 As shown in Table 3, both of these estimates are beyond the 95% credible interval $(-6.629, -5.643)$ obtained from the constant baseline, fixed-effects model $\mathcal{M}_1$. Therefore, as a firm trades beyond year 10, the probability of default increases. This counterintuitive result is probably due to the lack of data beyond trading year 20, which is reflected in the increased standard deviation and numerical standard error for the baseline towards the end of the sampling period. For example, there are only 108 firms that traded longer than 20 years and of these, 5 defaulted; an empirical hazard rate of 0.047. Moreover, the DIC statistic clearly favors models that do not include a time-varying baseline, suggesting the increase in model complexity over the fixed-effects model is not justified. This result suggests that external covariates that are correlated with the business cycle are a better way to capture temporal dependencies in the data. Indeed, when model $\mathcal{M}_3$ was estimated without personal income (INC) and the resulting DIC (1693.47) implied a relatively poor model fit.[22]

To compare how the different techniques respond to changes in a firm's economic environment, I follow Duffie and Wang (2004) and examine the default characteristics of General Binding Corporation (GBC). GBC is selected because it existed for a majority of sample period (1973-present), and is a relatively large firm with approximately 4,250 employees and a market capitalization of $188 million at the end of 2004. GBC is based

[22]Results of model $\mathcal{M}_4$ are similar to previous results and are therefore not reported here. The DIC of 1610.72 incorporates the improvement due to adding firm-specific effects (b_i), and the added complexity with no improvement in fit of the varying baseline (α_t).

in Skokie, Illinois and designs, manufactures, and distributes office equipment and related supplies.

Figure 6 plots the in-sample posterior mean probability of default for GBC from 1972:IV to 2004:IV for model $\mathcal{M}_1$. As a reference, the industry-average default probability is also plotted over the same time period. NITA began to turn negative for GBC in late 1999 (quarter 103). The average monthly stock price dropped from a high of \$39.27 in late 1998 to a low of \$6.67 in 2000. Any of the three models predict a 7-fold increase in default over the same time period. Figures 7 and 8 plot the posterior mean default probability for GBC for models $\mathcal{M}_2$ and $\mathcal{M}_3$. As the trading age of GBC increases, model $\mathcal{M}_2$ more than doubles the probability of default over model $\mathcal{M}_1$. The random-effects parameter b for GBC is one of the few centered about zero. Thus, the default probability suggested by $\mathcal{M}_3$ is almost identical to that of $\mathcal{M}_1$.

It is interesting to compare these in-sample default estimates obtained from the reduced-form model with those obtained from the MKMV algorithm. Figure 9 plots the default probability obtained from the MKMV algorithm using quarterly intervals for GBC. The MKMV estimate is much more sensitive to changes in the economic environment. The MKMV estimate changes from near zero (0.002) around the 100th quarter (1998:IV) to 0.94 in quarter 108 (2000:IV). Such sensitivity may be a useful property of a default estimator in that small changes in the economic environment are amplified, but it may also lead to spurious conclusions.[23] Obviously, a compromise between the structural model and reduced-form models might be preferred.

In order to calculate the risk-neutral default probabilities implied by the hybrid model, I must first specify a proxy for firm value. I use one-half of total liabilities and current liabilities (linearly interpolated from quarterly to monthly values) and add this to the monthly average of total market capitalization. This approximation to firm value is frequently used as an initial condition for the MKMV algorithm (Crosbie and Bohn (2001)). I assume that the monthly returns for firm value follow a two-state, univariate Markov mixture of normals (MMN). The Markov structure of the model permits serial correlation and persistence in higher moments, common characteristics of financial time series (see Geweke (2005)). The general form of the univariate linear model, given by

$$y_t = \boldsymbol{B}'\mathbf{x}_t + \varepsilon_t, \tag{5.1}$$

[23]This result is discussed in Eom et al. (2004).

is transformed to the MMN model by specifying that the $T \times 1$ vector of observables, y_t, occupies one of m discrete states, denoted by the latent variable $s_t \in 1, ..., m$. The discrete states are assumed to evolve according to a Markov process. Thus, conditional on $s_t = j$, $\varepsilon_t \sim \mathcal{N}(\mu_j, h_j^{-1})$ and the univariate MMN model can be described by

$$y_t | (\mathbf{x}_t, \mathbf{B}, \mu_j, h_j, s_t = j) \sim \mathcal{N}(\mathbf{B}'\mathbf{x}_t + \mu_j, h_i^{-1}) \quad j = 1, ..., m$$
$$s_t | s_{t-1} \sim Markov(P, \varpi_1)$$

where $P = [p_{ij}]$ is the one-step transition probability matrix of the chain, i.e., $p_{ij} = Pr(s_t = j | s_{t-1} = i)$ and ϖ_1 is the probability distribution at $t = 1$. I take a Bayesian approach to estimating the MMN model. Appendix B.2 describes the Bayesian estimation procedure and prior specification.

Figure 10 plots the estimated risk-neutral default probabilities of GBC assuming the random default barrier F is determined by the reduced-form model $\mathcal{M}_1$, as described in Section 3.1. The risk-neutral default probabilities serve as an upper bound on the default probability implied by model $\mathcal{M}_1$. The hybrid model was updated monthly, as opposed to quarterly, thus allowing the risk-neutral default probabilities to deviate substantially from the implied default probabilities of $\mathcal{M}_1$. The monthly risk-neutral estimates were then averaged to achieve quarterly estimates. As shown in Figure 10, by "forcing" the predictive density of firm value to satisfy the moment restriction implied by (3.1), the hybrid model responds aggressively to changes in GBC's total market capitalization, and yet effectively smooths the default probabilities implied by the MKMV algorithm.

5.2 Out-of-Sample Testing

Sobehart et al. (2001) developed several techniques for validating models of credit risk using out-of-sample testing.[24] I briefly describe and conduct two such tests. First, I estimate models $\mathcal{M}_1 - \mathcal{M}_3$ using 6 different subsets of the original data set. Specifically, I estimate the models using the starting date of 1962 with various stopping dates (1990, '92, '94, '96, '98, 2000). I then test the models' out-of-sample ability to sort firms according to posterior mean probability of default for the next 2 years. For example, I estimate models $\mathcal{M}_1 - \mathcal{M}_3$ from 1962 to 1990. I then use the *actual* covariates from years 1991 and 1992 to sort the firms based upon posterior mean probability of default

[24]Recently, the Basle Committee on Banking Supervision identified credit model validation as one of the most challenging issues facing quantitative credit model development.

implied by the model.

Sobehart et al. (2001) advocated the use of power curves to examine the tradeoff between Type I and Type II error associated with the one-year ahead and two-year ahead forecast.[25] Type I and Type II errors are the errors of misclassifying a bankrupt firm as healthy and misclassifying a healthy firm as bankrupt, respectively. Statistically, power curves represent the cumulative probability distribution of default events for different default probabilities. To plot power curves, companies are first ordered by model score, from riskiest to safest. For a given fraction of the total number of companies, x%, a power curve is constructed by calculating the percentage $f(x)$ of the defaulters whose risk score is equal to or lower than the one for fraction x. Figure 11 shows an example of a power curve. The curved line shows the performance of the model being evaluated. The horizontal axis (% of population excluded) depicts the probability of misclassifying a healthy firm as bankrupt (Type II error), while the vertical axis gives the probability of correctly classifying a bankrupt firm as bankrupt (1 - Type I error). Thus, a perfect model would be able to discriminate perfectly, catching all defaulting firms at the lowest model output. A model that randomly assigned default probabilities would follow the diagonal line indicated in Figure 11.

Sobehart et al. (2001) also advocated calculating the corresponding accuracy ratio (AR) associated with a given power curve.[26] This metric is obtained by comparing the power curve of a model with that of the perfect model and is defined to be twice the area between the power curve and the 45-degree line. Obviously, the closer the power curve is to the perfect power curve, the better the model performs.

Figure 12 gives the power curves for the hybrid model and models $\mathcal{M}_1$ and $\mathcal{M}_3$ and Table 5 gives the corresponding accuracy ratios. The power curves indicate that the hybrid model outperforms models $\mathcal{M}_1$ and $\mathcal{M}_3$ across the entire population. That is, the hybrid model sorted and ranked *all* of the defaulting firms better than did models $\mathcal{M}_1$ and $\mathcal{M}_3$. The power curves also show that model $\mathcal{M}_3$ slightly outperforms model $\mathcal{M}_1$. The random effects model better identified a handful of firms that were not obvious defaulters. This was mainly due to the differences in the parameter values across the two models. More specifically, these firms had more volatile returns, on average, which the random effects model weighted more heavily. It is also clear from Figure 12 that some

[25] These curves are also referred to as cumulative accuracy profiles, Lorenz diagrams and receiver-operator curves.

[26] Accuracy ratios are similar to Gini coefficients.

defaulting firms are very difficult to predict. Accounting ratios and market variables of a few defaulting firms did not display the typical characteristics of a defaulting firm.

6 Conclusion

The increasing popularity of credit derivatives and innovative corporate debt products has placed a premium on accurate forecasting of corporate default probabilities. Default rate probabilities are a necessary input to credit derivative and corporate debt pricing models. Therefore understanding and accurately modeling the factors that determine corporate default will continue to be a subject of future research.

This paper has addressed this need by introducing two new techniques for modeling corporate default probabilities. First, I show that accounting for firm-specific effects is an important factor when modeling corporate default probabilities. The empirical evidence suggests that failing to account for firm-specific effects leads to a significant under-prediction of default. Second, I introduce a hybrid model that combines the reduced-form model and the structural model of Merton. The hybrid model relaxes the unrealistic assumptions of the Merton model and provides upper-bound probabilities of default associated with the discrete duration model. I also document that a default prediction procedure that utilizes these two techniques markedly improves predictions over existing procedures that do not.

On going and future extensions to this research include: [i] Specifying a time-series distribution for the covariates of the discrete duration model and *forecasting* corporate default probabilities. This would provide a term structure of corporate default, and one could examine how the level and shape of the term structure of conditional future failure probabilities depends on a firm's accounting and market variables. [ii] Specifying a "private-firm model" that relies solely on accounting data to estimate corporate default. Many practitioners (e.g., banks) must estimate default probabilities for firms that are not publicly traded. [iii] Expanding the data set to further test the predictive power of the hybrid model relative to other model specifications. Estimating the models introduced here using a broad cross-section of credit default swap data would test how well the models price corporate credit risk.

Appendix A: Covariate Construction

The covariates were constructed using data from COMPUSTAT, CRSP and the Bureau of Economic Analysis National Economic Accounts database. COMPUSTAT quarterly data item numbers are in parenthesis.

$$NITA_{i,t} = \frac{NetIncome_{i,t}(Data69)}{TotalAssets_{i,t}(Data44)} - \frac{NetIncome_{i,t-1}}{TotalAssets_{i,t-1}}$$

$$TLMTA_{i,t} = \frac{TotalLiabilities_{i,t}(Data54)}{TotalLiabilities_{i,t}(Data54) + MarketEquity_{i,t}} - \frac{TL_{i,t-1}}{TL_{i,t-1} + ME_{i,t-1}}$$

$$RSIZE = \log\left(\frac{MarketEquity_{i,t}}{S\&P500MarketValue_t}\right) - \log\left(\frac{MarketEquity_{i,t-1}}{S\&P500MarketValue_{t-1}}\right)$$

$$INC_t = \log\left(\frac{PersonalIncome_t}{PersonalIncome_{t-1}}\right)$$

The monthly average excess return and monthly standard deviation in equity returns were changed to quarterly frequency using a three-month moving average with declining weights.

$$EXRET_{i,t} = \frac{1}{2}[\log(1 + R_{i,t}) - \log(1 + R_{S\&P500,t})]$$
$$+\frac{1}{3}[\log(1 + R_{i,t-1}) - \log(1 + R_{S\&P500,t-1})] + \frac{1}{6}[\log(1 + R_{i,t-2}) - \log(1 + R_{S\&P500,t-2})]$$
$$SIGMA_{i,t} = \frac{1}{2}STD(R_{i,t}) + \frac{1}{3}STD(R_{i,t-1}) + \frac{1}{6}STD(R_{i,t-2})$$

Appendix B: MCMC Calculations

Appendix B.1: Posterior and Conditional Posterior for the Duration Model

Recall that the joint posterior distribution of the reduced-form model is given by

$$p\left(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{b}, \sigma^2, \nu^2 \middle| \left[\mathbf{X}_{iT_i}, \delta_i, T_i\right]_{i=1}^{N}\right) \propto \prod_{i=1}^{N} p(y_{i1}, y_{i2}, ... y_{iT_i} | \mathbf{X}_{it}, \boldsymbol{\beta}, b_i) p(\boldsymbol{\beta})$$
$$\times p(b_i|\sigma^2) p(\sigma^2) p(\boldsymbol{\alpha}|\nu^2) p(\nu^2)$$

Thus, each of the posterior conditional distributions, at least up to a normalizing constant, may be found by combining the logistic likelihood and the appropriate prior distribution.

The univariate conditional posterior for a single fixed-effect parameter β_r is

$$p(\beta_r|\boldsymbol{\alpha}, \beta_{-r}, \boldsymbol{b}, \mathbf{y}_{it}, \mathbf{x}_{it}, \varsigma) \propto \left[\prod_{t=1}^{T}\prod_{i=1}^{N} \frac{\exp\{\alpha_t + \boldsymbol{\beta}'\mathbf{x}_{it} + b_i\}^{y_{it}}}{1 + \exp\{\alpha_t + \boldsymbol{\beta}'\mathbf{x}_{it} + b_i\}}\right] \times \exp[-\beta_r^2/2\varsigma]$$

This conditional distribution is log-concave in β_r and is therefore amenable to the adaptive-rejection sampling (ARS) technique of Gilks and Wilde (1992).

The joint prior density of the baseline hazards is given by

$$p(\boldsymbol{\alpha}|\nu^2) \propto (\nu^2)^{-m/2} \exp\left\{-\frac{1}{2\nu^2}\sum_{t=1}^{m}(\alpha_t - \alpha_{t-1})^2\right\}$$

and the prior for α_0 is drawn from $\mathcal{N}(0, \nu^2)$. The univariate conditional posterior for α_t is given by

$$p(\alpha_t|\boldsymbol{\beta}, \alpha_{t-1}, \nu^2, \mathbf{y}_{it}, \mathbf{x}_{it}) \propto \left[\prod_{i \in R_t} \frac{\exp\{\alpha_t + \boldsymbol{\beta}'\mathbf{x}_{it} + b_i\}^{y_{it}}}{1 + \exp\{\alpha_t + \boldsymbol{\beta}'\mathbf{x}_{it} + b_i\}}\right]$$
$$\times \left[\exp\left\{-\frac{(\alpha_t - (\alpha_{t+1} + \alpha_{t-1})/2)^2}{2\nu^2}\right\}\right] \qquad t = 1, ..., m$$

where R_t is the set of all firms at risk of defaulting in quarter t. For t=1, $\alpha_{t-1} = \alpha_{t+1}$ and for $t = m$, $\alpha_{t+1} = \alpha_{t-1}$. This conditional distribution is also log-concave and is sampled using ARS.

The univariate conditional posterior distribution for the random effect parameter

$$p(b_i|\boldsymbol{\beta}, \boldsymbol{\alpha}, b_{-i}, \sigma^2, \mathbf{x}_{it}, y_{it}) \propto \left[\prod_{t=1}^{T_i} \frac{\exp\{\alpha_t + \boldsymbol{\beta}'\mathbf{x}_{it} + b_i\}^{y_{it}}}{1 + \exp\{\alpha_t + \boldsymbol{\beta}'\mathbf{x}_{it} + b_i\}} \right] \exp\left\{ - \frac{b_i^2}{2\sigma^2} \right\}$$

is log-concave and sampled via ARS.

The precision $\tau_\nu = \nu^{-2}$ of the baseline has the full conditional posterior

$$p(\tau_\nu|\boldsymbol{\alpha}) \propto \tau_\nu^{(k_\nu + m/2)-1} \exp\left[- (\omega_\nu + \sum \left(\alpha_t - \alpha_{t-1})^2/2\right)\tau_\nu \right]$$

which is a gamma distribution and can be sampled directly.
The univariate conditional posterior distribution for the precision of the random effects is given by

$$p(\tau_\sigma|\mathbf{b}) \propto \left(\tau_\sigma\right)^{N/2+k_\sigma-1} \exp\left\{ - \left(\omega_\sigma + \frac{1}{2}\sum_{n=1}^{N} b_i^2\right)\tau_\sigma \right\}$$

which is sampled by slice sampling (Neal (1997)).

Appendix B.2: Bayesian Estimation of MMN Model

I assume that the proxy for firm value follows a univariate Markov mixture of normals (MMN). The general form of the univariate MMN model is given by

$$\begin{gathered} y_t|(\theta, s_t = i) \sim N(\mu_i, h_i^{-1}) \quad i = 1, ..., m \\ P[s_t = j|s_{t-1} = i, s_{t-u}(u > 1)] = p_{ij} \quad (i = 1, .., m;\ j = 1, ..., m) \\ s_t|s_{t-1} \sim Markov(P, \pi_1) \end{gathered}$$

where $\theta = (\mu_j, h_j)$ and $P = \{p_{ij}\}$ is the one-step transition probability matrix of the Markov chain. Denote $S_t = (s_1, ..., s_t)$, $S^{t+1} = (s_{t+1}, ..., s_n)$, $Y_t = (y_1, ..., y_t)$, and $Y^{t+1} = (y_{t+1}, ...y_n)$. The joint density $p(s_1, ..., s_n|Y_n, \theta)$ can be written as

$$p(S_n|Y_n, \theta) = p(s_n|Y_n, \theta) \cdots p(s_t|Y_n, S^{t+1}, \theta) \cdots p(s_1|Y_n, S^2, \theta)$$

By Bayes theorem,

$$\begin{aligned} p(s_t|Y_n, S^{t+1}, \theta) &\propto p(s_t|Y_t, \theta)p(Y_{t+1}, S^{t+1}|Y_t, s_t, \theta) \\ &\propto p(s_t|Y_t, \theta)p(s_{t+1}|s_t, \theta)p(Y^{t+1}, S^{t+2}|Y_t, s_t, s_{t+1}, \theta) \\ &\propto p(s_t|Y_t, \theta)p(s_{t+1}|s_t, \theta) \end{aligned}$$

where the last equation follows due to independence of s_t and $p(Y^{t+1}, S^{t+2}|Y_t, s_t, s_{t+1}, \theta)$. To calculate $p(s_t|Y_t, \theta)$, assuming $p(s_{t-1}|Y_{t-1}, \theta)$ is available, the following steps are repeated

1. Prediction Step: By the law of total probability,

$$p(s_t|Y_{t-1}, \theta) = \sum_{k=1}^{m} p(s_t|s_{t-1} = k, \theta)p(s_{t-1} = k|Y_{t-1}, \theta)$$

2. Update Step: By Bayes theorem, the mass function of the state given information up to time t is now

$$p(s_t|Y_t, \theta) \propto p(s_t|Y_{t-1}, \theta)p(y_t|Y_{t-1}, \theta_{s_t})$$

Prior specification is based on the set of conditionally conjugate priors for the means, precisions, and transition matrix.

$$\begin{aligned} p(\mu_i) &\sim N(\underline{\mu}_i, \underline{h}_i^{-1}) \\ p(h_i) &\sim \mathcal{G}(\underline{\alpha}_i, \underline{\beta}_i) \\ p(P_i) &\sim \mathcal{D}(\underline{c}_{i1}, ..., \underline{c}_{im}) \end{aligned}$$

where $\mathcal{G}$ denotes the gamma distribution and $\mathcal{D}$ denotes the Dirichlet. In estimating the MMN model, I assumed two states ($m = 2$) with $\mu_i = 0$, $h_i = 10^6$, $\alpha_i = 2$, $\beta_i = 3/2$,$\underline{c}_{i1} = 10$, and $\underline{c}_{i2} = 2$ (see Geweke (2005) for an interpretation of this prior for the MMN model).

Given the above priors and the likelihood, the posterior of θ can be expressed as

$$p(\theta|Y_t) \propto \prod_{i=1}^{m} \exp\left[\frac{-(\mu_i - \underline{\mu}_i)^2}{2\underline{h}_i^{-1}}\right] h_i^{\underline{\alpha}_i - 1} \exp(-\underline{\beta}_i h_i) \prod_{i=1}^{m}\prod_{j=1}^{m} p_{ij}^{\underline{c}_{ij}-1} \prod_{t=1}^{T} p(y_t|Y_{t-1}, \theta)$$

Conditioning on the states $\{s_t\}$, the posterior becomes

$$p(\theta|Y_t) \propto \prod_{i=1}^{m} \exp\left[\frac{-(\mu_i - \underline{\mu}_i)^2}{2\underline{h}_i^{-1}}\right] h_i^{\underline{\alpha}_i - 1} \exp(-\underline{\beta}_i h_i) \prod_{i=1}^{m}\prod_{j=1}^{m} p_{ij}^{\underline{c}_{ij}-1} \prod_{i=1}^{m} h_i^{n_2/2}$$
$$\times \exp\left[\frac{-\sum_{k=1}^{n_i}(y_{i,k} - \mu_i)^2}{2h_i^{-1}}\right]$$

where $y_{i,k=1,..n_i}$ are the observations assigned to state s_i. The Gibbs sampler is given by running the following simulations successively:

$$s_t \sim s_1, s_2, ..., s_T|Y_t, \theta \qquad t = 1, ..., T$$
$$\mu_i \sim N\left[\frac{\underline{h}_i\underline{\mu}_i + n_i h_i \overline{y}_{n_i}}{\underline{h}_i + n_i h_i}, (\underline{h}_i + n_i h_i)^{-1}\right] \qquad i = 1, ..., m$$
$$h_i \sim \mathcal{G}\left[\underline{\alpha}_i + n_i/2, \underline{\beta} + \frac{\sum_{k=1}^{n_i}(y_{i,k} - \mu_i)^2}{2}\right] \qquad i = 1, ..., m.$$
$$p_i \sim \mathcal{D}(\underline{c}_{i1} + n_{i1} + \cdots + \underline{c}_{im} + n_{im}) \qquad i = 1, ..., m$$

References

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. In Leinhardt, S., editor, *Sociological Methodology*. Jossey-Bass, San Francisco.

Altman, E. I. (1968). Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance*, 23:589–609.

Baker, M. and Rea, S. (1998). Employment spells and unemployment insurance eligibility requirements. *Review of Economics and Statistics*, 80:80–94.

Bharath, S. and Shumway, T. (2004). Forecasting default with the kmv-merton model. Working Paper, University of Michigan.

Biggeri, L., Bini, M., and Grilli, L. (2001). The transition from university to work: A multilevel approach to the analysis of time to obtain the first job. *Journal of the Royal Statistical Society*, A(164):293–305.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654.

Cameron, S. V. and Heckman, J. J. (1998). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political Economy*, 106:79–97.

Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2005). In search of distress risk. Working Paper. Havard University.

Campolieti, M. (2001). Bayesian semiparametric estimation of discrete duration models: An application of the dirichlet process prior. *Journal of Applied Econometrics*, 16:1–22.

Chava, S. and Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. Working Paper. Cornell University.

Crosbie, P. J. and Bohn, J. R. (2001). Modeling default risk. KMV LLC.

Davydenko, S. A. (2005). When do firms default? a study of the default boundary. Working Paper. London Business School.

Davydenko, S. A. and Franks, J. R. (2005). Do bankruptcy codes matter? a study of defaults in france, germany and the uk. Working Paper. London Business School.

Delianedis, G. and Geske, R. (1999). Credit risk and risk neutral default probabilities: Information about rating migrations and defaults. Working Paper, UCLA.

Dempster, A. P. (1974). The direct use of likelihood for significance testing. In *Proceedings of Conference on Foundational Questions in Statistical Inference*. University of Aarhus.

Duan, J.-C. and Fulop, A. (2005). Estimating the structural credit risk model when equity prices are contaminated by trading noises. Working Paper. University of Toronto.

Duffie, D. and Lando, D. (2001). Term structures of credit spreads with incomplete accounting information. *Econometrica*, 69(3):633–664.

Duffie, D. and Singelton, K. J. (1999). Modeling the term structure of defaultable bonds. *Review of Financial Studies*, 12:687–720.

Duffie, D. and Wang, K. (2004). Multi-period corporate failure prediction with stochastic covariates. Working Paper. Stanford University.

Elizalde, A. (2005). Credit risk models ii: Structural models. Working Paper. Universidad Pública de Navarra.

Eom, Y. H., Helwege, J., and Huang, J.-Z. (2004). Structural models of corporate bond pricing: An empirical analysis. *Review of Financial Studies*, 17:499–544.

Fahrmeir, L. and Knorr-Held, L. (1997). Dynamic discrete-time duration models: Estimation via markov chain monte carlo. *Sociological Methodology*, 27:417–452.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag, New York.

Gamerman, D. and West, M. (1987). An application of dynamic survival model in unemployment studies. *Statistician*, 36:269–274.

Geske, R. (1977). The valuation of corporate liabilities as compound options. *Journal of Financial and Quantitative Analysis*, 12:541–552.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In Bernando, J., Berger, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics 4*. Clarendon Press, Oxford, UK.

Geweke, J. F. (2005). *Contemporary Bayesian Econometrics and Statistics*. Wiley-Interscience, New York.

Giesecke, K. (2001). Default and information. Working Paper, Cornell University.

Giesecke, K. and Goldberg, L. (2004). Forecasting default in the face of uncertainty. *Journal of Derivatives*, 12(1):1–15.

Hanson, S., Pesaran, M. H., and Schuermann, T. (2005). Firm heterogeneity and credit risk diversification. Working Paper, Wharton Financial Institutions Center.

Heckman, J. J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52:271–320.

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

Huang, J. and Huang, M. (2003). How much of the corporate-treasury yield spread is due to credit risk. Working Paper, Stanford University.

Hull, J. and White, A. (2001). Valuing credit default swaps ii: Modeling default correlations. *Journal of Derivatives*, 8:12–22.

Jarrow, R. A., Lando, D., and Turnbull, S. M. (1997). A markov model of the term structure of credit risk spreads. *Review of Financial Studies*, 10:481–523.

Jarrow, R. A. and Protter, P. (2004). Structural versus reduced form models: A new information based perspective. *Journal of Investment Management*, (2):1–10.

Jarrow, R. A. and Turnbull, S. M. (1995). Pricing derivatives on fiancial securities subject to credit risk. *Journal of Finance*, 50:53–86.

Leland, H. E. (1994). Risky debt, bond covenants and optimal capital structure. *Journal of Finance*, 49:1213–1252.

Leland, H. E. (2004). Predictions of default probabilities in structural models of debt. Working Paper. University of California, Berkeley.

Leland, H. E. and Toft, K. B. (1996). Optimal capital structure, endogenous bankruptcy and the term structure of credit spreads. *Journal of Finance*, 50:789–819.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29:449–470.

Neal, R. (1997). Markov chain monte carlo methods based on slicing the density function. Technical Report, Department of Statistics, University of Toronto.

Ohlson, J. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 19:109–131.

Pesaran, M. H., Schuermann, T., Treutler, B.-J., and Weiner, S. M. (2005). Macroeconomic dynamics and credit risk. Working Paper, Wharton Financial Institutions Center.

Raftery, A. L. and Lewis, S. (1992). One long run with diagnostics: Implementation strategies for markov chain monte carlo. *Statistical Science*, (7):493–497.

Robertson, J. C., Tallman, E. W., and Whiteman, C. H. (2005). Forecasting using relative entropy. *Journal of Money, Credit, and Banking*, 37(3):383–401.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business*, 74:104–124.

Smith, T., Spiegelhalter, D. J., and Thomas, A. (1995). Bayesian graphical modelling applied to random effects meta-analysis. *Statistics in Medicine*, (14):2685–2699.

Sobehart, J., Keenan, S., and Stein, R. (2001). Risk models: A validation methodology. *Algo Research Quarterly*, 4:57–72.

Spiegelhalter, D., Best, N. G., Carlin, B. P., and van der Linde, A. (2001). Bayesian measures of model complexity and fit. *Journal of Royal Statistical Societ*, B(64):583–639.

Spiegelhalter, D., Thomas, A., and Best, N. G. (2000). Winbugs version 1.3: User manual. MRC Biostatistics Unit, Cambridge, U.K.

Stutzer, M. (1996). A simple nonparametric approach to derivative security valuation. *Journal of Finance*, 51:1633–1652.

Thompson, W. A. (1977). On the treatment of grouped observations in life studies. *Biometrics*, 39:141–147.

Turnbull, S. M. (1979). Debt capacity. *Journal of Finance*, 34:931–941.

Vasicek, O. (1984). Credit valuation. *White Paper, Moody's KMV.*

Zhang, B. Y., Zhou, H., and Zhu, H. (2005). Explaining credit default swap spreads with equity volatility and jump risks of individual firms. BIS Working Paper No. 181.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Acccounting Research*, 22:59–82.

Figure 1: Distance to Default

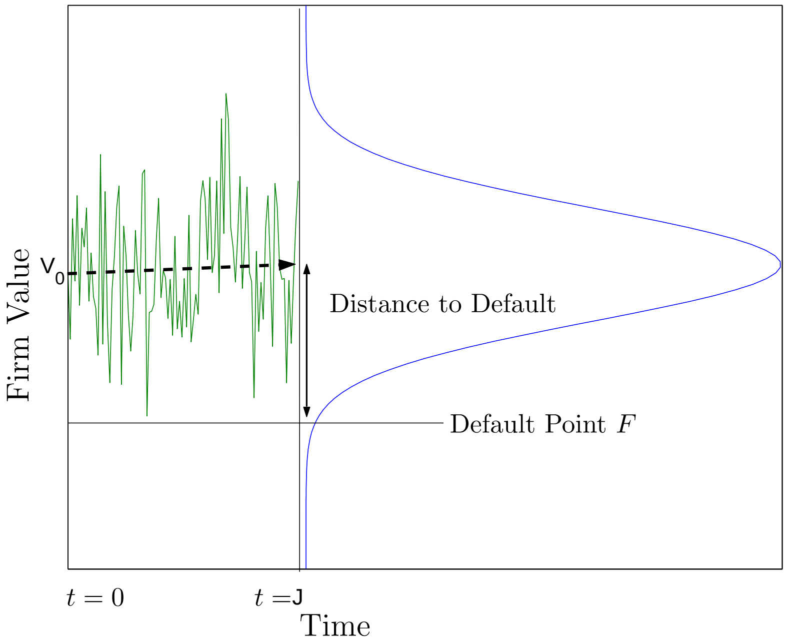

Figure 1: This figure replicates Figure 8 of Crosbie and Bohn (2001). Given some initial value V_0, an assumed process for the predictive density at time J, and an assumed default boundary F, one can calculate distance to default.

Figure 2: Empirical Kernel Density (solid) and Gaussian Density for Bankrupt Firm

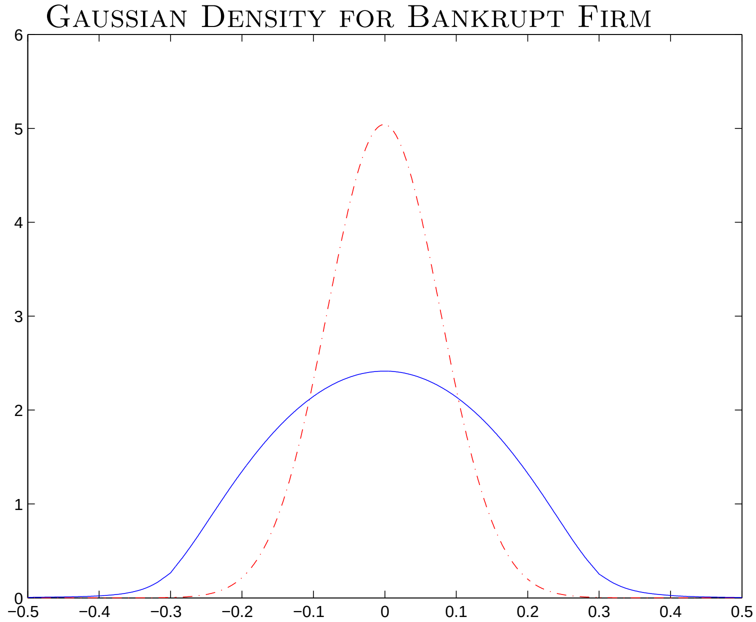

Figure 2: This figure plots the empirical kernel density (solid) using an Epanechnikov kernel with 0.3 bandwidth. The normal density is also plotted with mean and variance equal to the sample moments.

Figure 3: Random Default Barrier

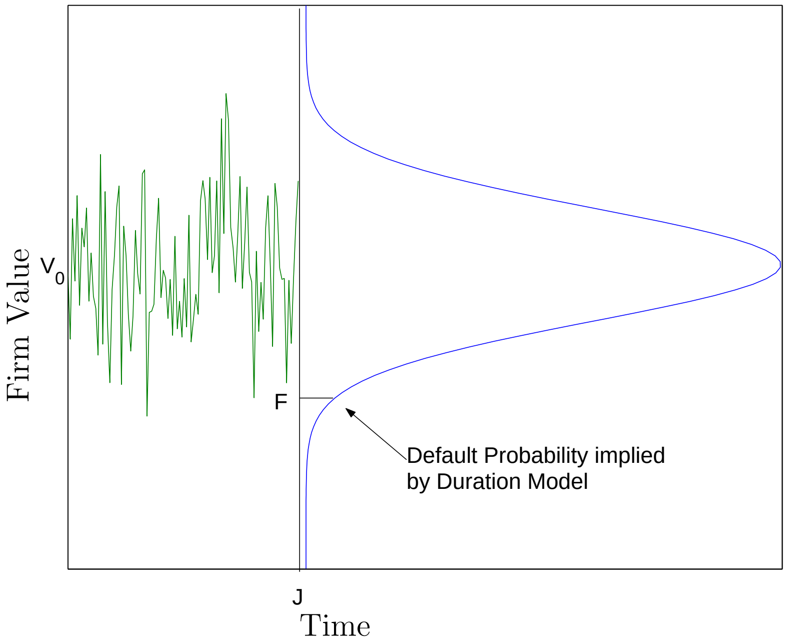

Figure 3: This figure shows that given a default probability implied by the discrete duration model, finding a corresponding default barrier is a value-at-risk problem.

Figure 4: Empirical Hazard Function

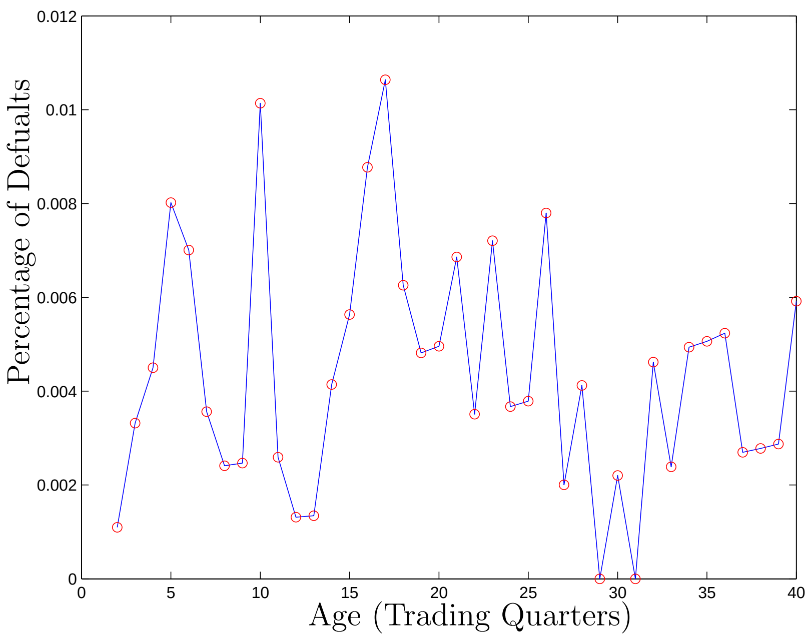

Figure 4: This figure plots the empirical hazard function for the first 40 trading quarters. The empirical hazard function is defined as the number of defaulting firms in quarter t divided by the number of firms that are at risk of defaulting at time t.

Figure 5: Posterior Mean Probability of Default for Industry Average: $\mathcal{M}_1$ (solid) and $\mathcal{M}_3$

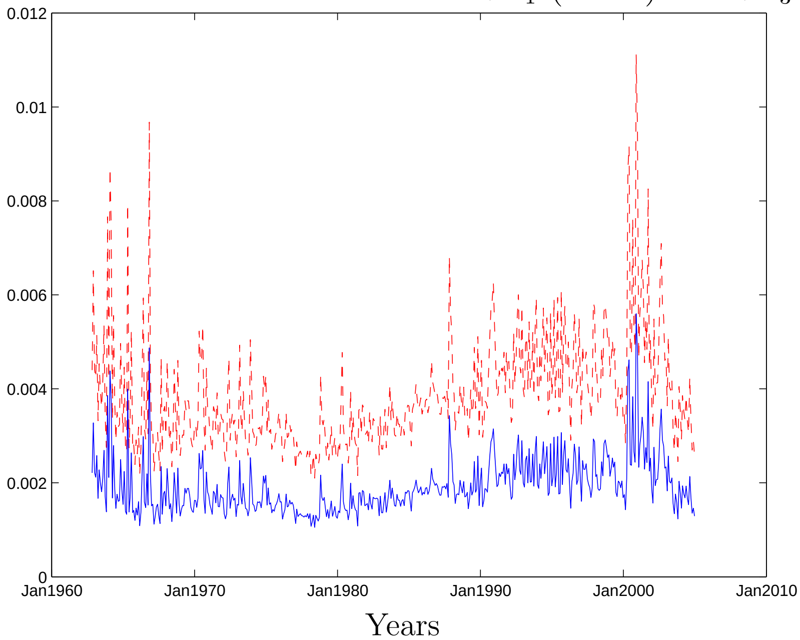

Figure 5: This figure plots the posterior mean probability of default for the cross-sectional industry average implied by models $\mathcal{M}_1$ (solid) and $\mathcal{M}_3$ with b_i set equal to the 90th percentile.

Figure 6: Posterior Mean Probability of Default GBC (Solid) v. Industry Avg.: 1972:IV to 2004:IV

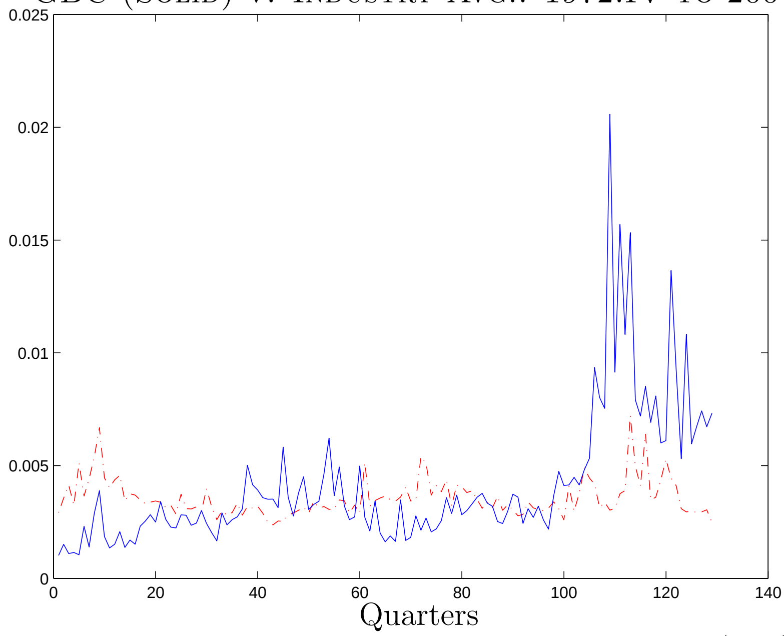

Figure 6: This figure plots the posterior mean probability of default for GBC (solid) and the industry average implied by $\mathcal{M}_1$ for 1972:IV to 2004:IV.

FIGURE 7: GBC POSTERIOR MEAN PROBABILITY OF DEFAULT

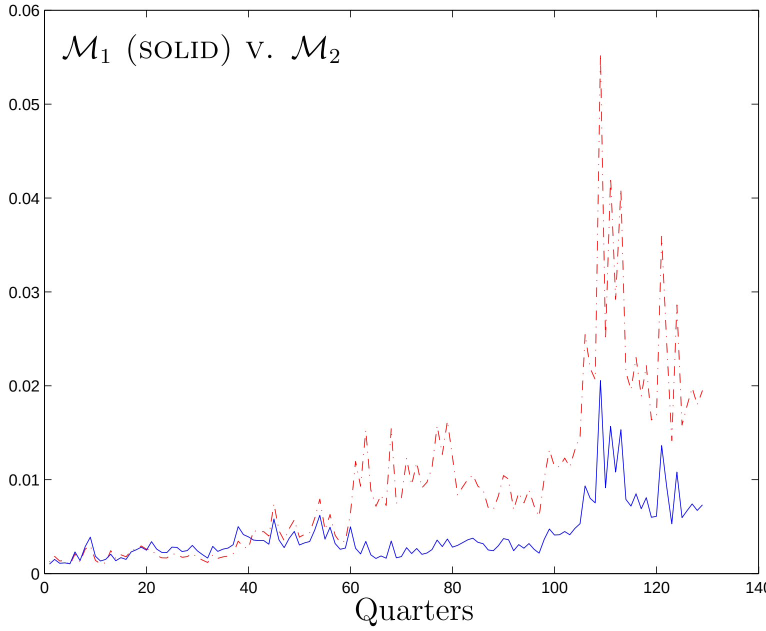

Figure 7: This figure plots the posterior mean probability of default for GBC (solid) implied by models $\mathcal{M}_1$ and $\mathcal{M}_2$.

FIGURE 8: GBC POSTERIOR MEAN PROBABILITY OF DEFAULT

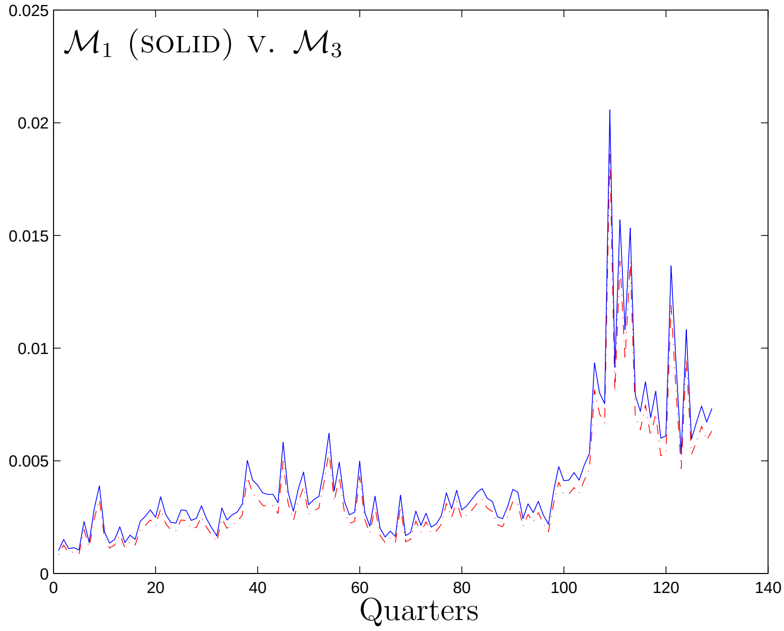

Figure 8: This figure plots the posterior mean probability of default for GBC (solid) implied by models $\mathcal{M}_1$ and $\mathcal{M}_3$.

Figure 9: MKMV Probability of Default for GBC

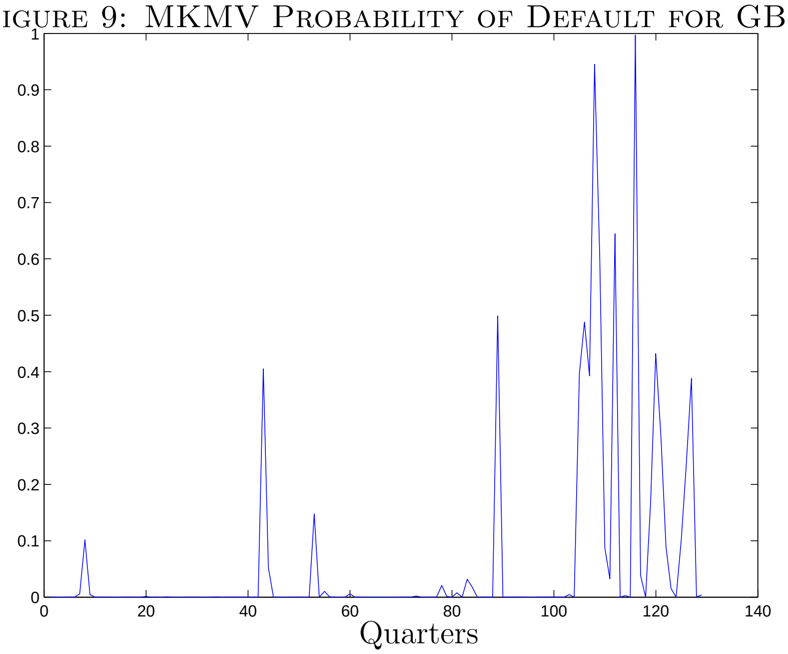

Figure 9: This figure plots the default probabilities implied by the MKMV algorithm for GBC from 1972:IV to 2004:IV. Monthly default estimates were averaged to obtain quarterly estimates.

Figure 10: Risk-Neutral Probability of Default for GBC

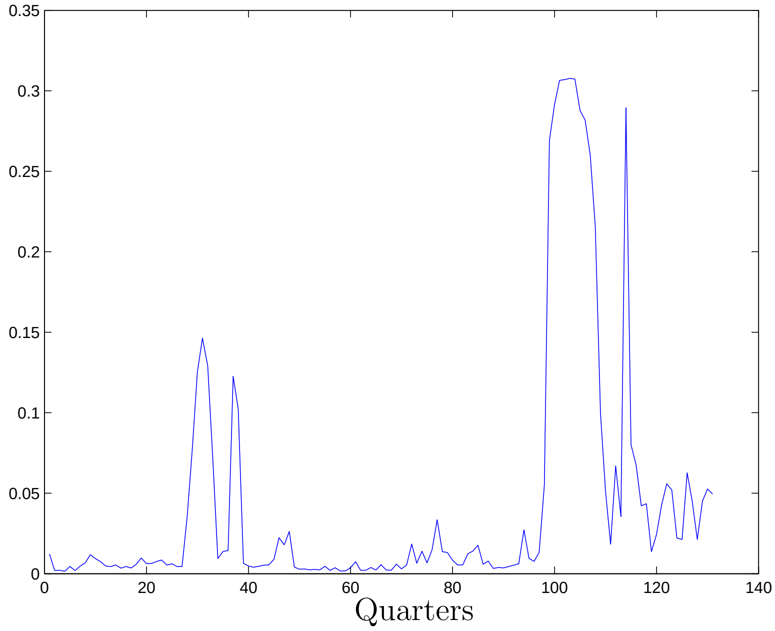

Figure 10: This figure plots the risk-neutral default probability for GBC implied by the hybrid model. Total market capitalization, averaged monthly, along with one-half total liabilities plus current liabilities, interpolated from quarterly to monthly values, was used to proxy for firm value. This proxy was assumed to follow a Markov mixture of normals process. The posterior mean default probability implied by model $\mathcal{M}_1$ was used to find the random default barrier.

Figure 11: Power Curve

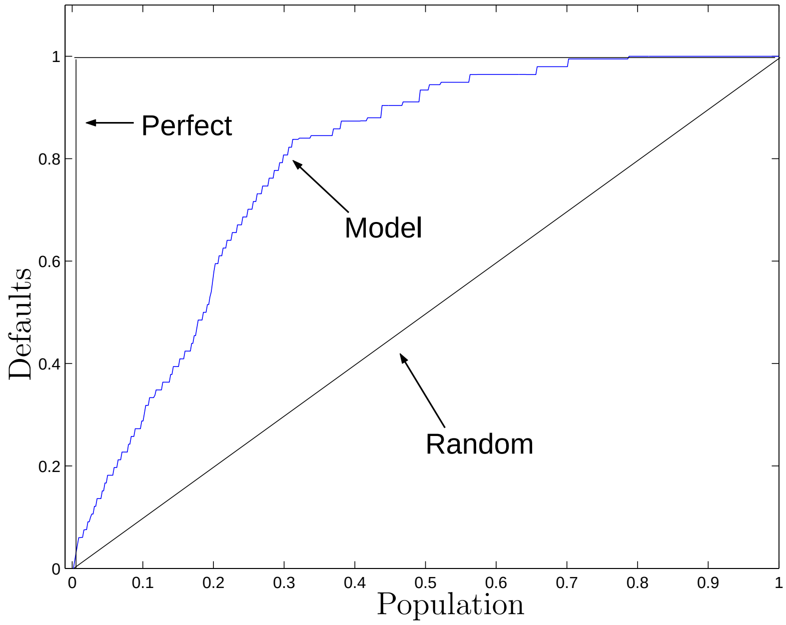

Figure 11: An example of a power curve with a perfect and random model. The horizontal axis (% of population excluded) depicts the probability of misclassifying a healthy firm as bankrupt (Type II error), while the vertical axis gives the probability of correctly classifying a bankrupt firm as bankrupt (1 - Type I error).

Figure 12: Power Curves for $\mathcal{M}_1$, $\mathcal{M}_3$, and Hybrid

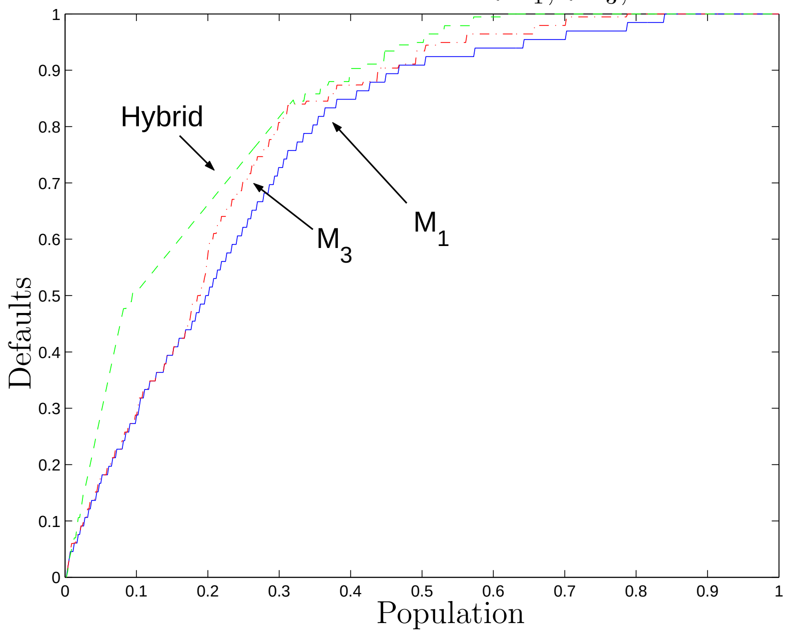

Figure 12: Power curve of model $\mathcal{M}_1$, $\mathcal{M}_3$ and the hybrid model.

Figure 13: Power Curves for MKMV and Hybrid Model

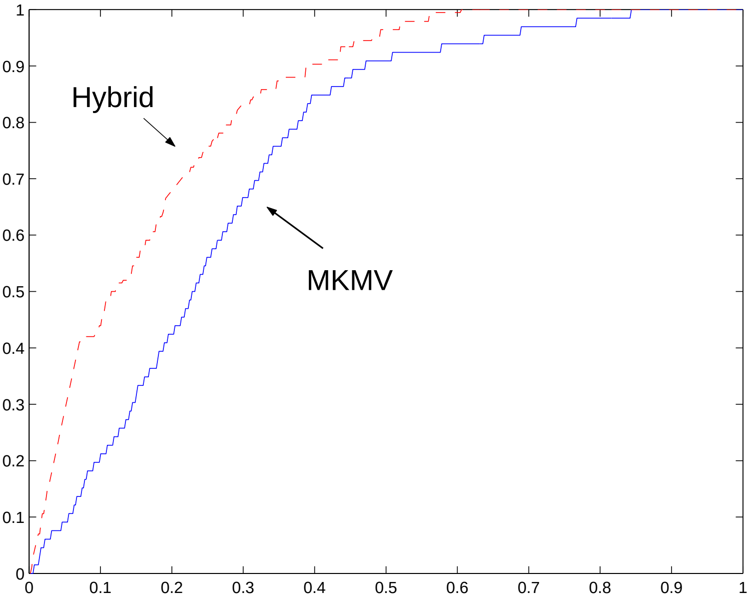

Figure 13: This figure plots the power curves for the MKMV model and the hybrid model.

Table 1: Number of Bankruptcies and Defaults per Year

This table lists the number of active firms, bankruptcies and defaults for 1970-2004. There are 916 total firms, 67 total bankruptcies and 153 total defaults.

Year	Active Firms	Bankruptcies	Failures
1970	68	0	0
1971	77	0	0
1972	142	0	0
1973	182	0	0
1974	190	1	1
1975	193	0	0
1976	201	0	0
1977	201	0	0
1978	199	0	0
1979	194	0	0
1980	205	0	1
1981	229	0	1
1982	236	2	2
1983	281	0	2
1984	296	5	12
1985	298	2	8
1986	317	1	4
1987	324	2	6
1988	327	1	6
1989	314	8	10
1990	301	7	10
1991	295	8	12
1992	303	4	9
1993	328	4	6
1994	356	4	8
1995	381	2	3
1996	409	1	6
1997	423	2	5
1998	402	6	17
1999	377	3	5
2000	315	0	5
2001	315	3	9
2002	209	0	2
2003	270	0	2
2004	244	0	0

Table 2: Summary Statistics of Covariates

This table lists the summary statistics for the winsorized explanatory variables. The number of observations is total firm-quarters. The second and third panels give summary statistics for the firm-quarters of default and bankruptcy.

Entire Data ($N = 37,750$)

	Mean	**Median**	**Std. Dev**	**Min**	**Max**
NITA	-0.005	0.012	0.086	-0.471	0.153
TLMTA	0.353	0.309	0.244	0.012	0.993
EXRET	-0.000	-0.000	0.005	-0.109	0.1237
RSIZE	-10.872	-10.952	1.815	-16.824	-2.582
SIGMA	0.038	0.032	0.027	0.002	0.681
INC	0.016	0.016	0.008	-0.159	0.0466

$N = 36,835$

ΔNITA	-0.000	-0.000	0.162	-0.462	0.471
ΔTLMTA	0.001	0.002	0.319	-0.980	0.980
ΔRSIZE	-0.000	-0.000	0.010	-0.228	0.207

Default Data ($N = 153$)

	Mean	**Median**	**Std. Dev**	**Min**	**Max**
NITA	-0.332	-0.157	0.541	-0.471	0.091
ΔNITA	-0.318	-0.166	0.551	-0.461	0.381
TLMTA	0.574	0.613	0.265	0.012	0.783
ΔTLMTA	0.239	0.244	0.328	-0.509	0.905
EXRET	-0.008	-0.005	0.014	-0.074	0.030
RSIZE	-13.514	-13.529	1.292	-15.952	-9.760
ΔRSIZE	-0.011	-0.008	0.026	-0.127	0.085
SIGMA	0.093	0.083	0.056	0.011	0.376

Bankruptcy Data ($N = 67$)

	Mean	**Median**	**Std. Dev**	**Min**	**Max**
NITA	-0.355	-0.193	0.185	-0.471	0.081
ΔNITA	-0.347	-0.180	0.572	-0.462	0.240
TLMTA	0.642	0.676	0.0.241	0.017	0.977
ΔTLMTA	0.353	0.334	0.244	0.012	0.993
EXRET	-0.011	-0.010	0.015	-0.075	0.017
RSIZE	-13.383	-13.472	1.356	-15.375	-9.761
ΔRSIZE	-0.015	-0.009	0.028	-0.127	0.033
SIGMA	0.095	0.077	0.055	0.011	0.284

Table 3: Posterior Estimates for Fixed and Random Effects

This table reports the posterior mean, standard deviation, 95% credible interval and numerical standard errors (i.e., the standard error of the Monte Carlo approximation assuming a 4% autocovariance tapered estimate) in parenthesis for the fixed-effects model ($\mathcal{M}_1$) and the random-effects model ($\mathcal{M}_3$).

	Default Data					
	Fixed Effects ($\mathcal{M}_1$)			Random Effects ($\mathcal{M}_3$)		
	Mean	Std Dev	95% CI	Mean	Std Dev	95% CI
ΔRSIZE	-9.204	9.086	(-25.791, 9.782)	-10.219	9.299	(-27.430, 8.669)
	(0.009)			(0.010)		
SIGMA	10.891	1.208	(8.478, 13.264)	13.841	1.889	(10.361, 17.783)
	(0.015)			(0.028)		
EXRET	-64.890	14.023	(-93.084, -38.372)	-61.808	14.311	(-90.721, -34.389)
	(0.095)			(0.118)		
ΔNITA	-2.543	0.245	(-3.025, -2.056)	-2.753	0.296	(-3.344, -2.183)
	(0.002)			(0.011)		
ΔTLMTA	1.714	0.271	(1.216, 2.275)	1.770	0.284	(1.215, 2.329)
	(0.002)			(0.003)		
INC	-9.858	12.006	(-33.251, 13.592)	-7.191	12.250	(-31.171, 16.539)
	(0.027)			(0.098)		
Constant	-6.128	0.251	(-6.629, -5.645)	-6.553	0.321	(-7.225, -5.967)
	(0.006)			(0.021)		
$\mathbf{P_D}$	6.961			64.015		
DIC	1611.72			1591.42		

Table 4: Posterior Estimates for Fixed Effects

This table reports the posterior mean, standard deviation, 95% credible interval and numerical standard errors (i.e., the standard error of the Monte Carlo approximation assuming a 4% autocovariance tapered estimate) in parenthesis for the fixed-effects model ($\mathcal{M}_2$).

Fixed Effects ($\mathcal{M}_2$)

	Mean	Std Dev	95% CI	NSE
ΔRSIZE	-8.752	9.222	(-24.900, 9.546)	0.010
SIGMA	10.661	1.264	(8.270, 13.020)	0.017
EXRET	-68.450	14.628	(-96.565, -41.170)	0.108
ΔNITA	-2.630	0.256	(-3.115, -2.155)	0.002
ΔTLMTA	1.700	0.270	(1.191, 2.209)	0.003
ΔINC	-16.578	11.983	(-38.725, 5.774)	0.075
$\boldsymbol{\alpha_1(1)}$	-6.138	0.357	(-6.858, -5.458)	0.006
$\boldsymbol{\alpha_2(2)}$	-6.177	0.345	(-6.875, -5.552)	0.006
$\boldsymbol{\alpha_3(3)}$	-6.334	0.374	(-7.095, -5.623)	0.006
$\boldsymbol{\alpha_4(4)}$	-5.796	0.341	(-6.486, -5.148)	0.006
$\boldsymbol{\alpha_5(5)}$	-5.582	0.350	(-6.584, -5.218)	0.007
$\boldsymbol{\alpha_6(6)}$	-6.301	0.402	(-7.119, -5.543)	0.007
$\boldsymbol{\alpha_7(7)}$	-6.277	0.411	(-7.111, -5.509)	0.007
$\boldsymbol{\alpha_8(8)}$	-6.201	0.418	(-7.054, -5.433)	0.006
$\boldsymbol{\alpha_9(9)}$	-6.283	0.448	(-7.211, -5.454)	0.006
$\boldsymbol{\alpha_{10}(10)}$	-6.456	0.507	(-7.507, -5.521)	0.008
$\boldsymbol{\alpha_{11}(11-15)}$	-5.663	0.305	(-6.282, -5.074)	0.007
$\boldsymbol{\alpha_{12}(16-20)}$	-4.225	0.602	(-5.499, -3.133)	0.007
$\boldsymbol{\alpha_{13}(21-25)}$	-4.635	1.874	(-8.653, -1.299)	0.013
$\boldsymbol{\alpha_{14}(26-35)}$	-4.624	2.786	(-10.391, 0.685)	0.019
$\boldsymbol{\alpha_{15}(36-45)}$	-4.649	3.480	(-11.805, 2.096)	0.027
$\mathbf{P_D}$	17.29			
DIC	1615.38			

Table 5: Accuracy Ratios for Models $\mathcal{M}_1$, $\mathcal{M}_3$, MKMV and Hybrid Models

Model	AR
$\mathcal{M}_1$	0.77
$\mathcal{M}_2$	0.80
MKMV	0.73
Hybrid	0.83