

Frailty Correlated Default*

Preliminary and Incomplete Version

Darrell Duffie, Andreas Eckner,
Guillaume Horel and Leandro Saita

First Version: April, 2006
Current Version: May 12, 2006

Abstract

We analyze portfolio credit risk in light of dynamic “frailty,” in the form of incompletely observed default. Common dependence by firms on unobservable time-varying default covariates is estimated to cause large changes in conditional mean default rates above and beyond those predicted by observable factors, and large increases in the fatness of the tails of the distributions of portfolio default losses for U.S. corporates. We also allow for unobserved heterogeneity across firms.

Keywords: correlated default, doubly stochastic, frailty, latent factor.
JEL classification: C11, C15, C41, E44, G33

*We are grateful for financial support from Moody’s Corporation and Morgan Stanley, and for research assistance from Sabri Oncu and Vineet Bhagwat. We are also grateful for remarks from Stav Gaon and Richard Cantor. We are thankful to Moodys and to Ed Altman for data assistance. Duffie and Saita are at The Graduate School of Business, Stanford University. Eckner and Horel are at the Department of Statistics, Stanford University.

1 Introduction

This paper introduces and estimates a new model of frailty-correlated defaults, according to which firms have an unobservable common source of “frailty,” a default risk factor that changes randomly over time. The posterior distribution of this frailty factor, conditional on past observable covariates and past defaults, represents a significant additional source of uncertainty to creditors. Our model is estimated for U.S. non-financial public firms for the period 1979-2004. The results show that the frailty factor induces a large estimated increase in default clustering, and significant additional fluctuation over time in the conditional expected level of default losses, above and beyond that predicted by our observable default covariates, including leverage, volatility, and interest rates.

The usual duration-based model of default probabilities is based on the doubly-stochastic assumption, by which firms’ default times are conditionally independent given the paths of observable factors influencing their credit qualities. Under this assumption, different firms’ default times are correlated only to the extent implied by the correlation of observable factors determining their default intensities. For example, Couderc and Renault (2004), Shumway (2001), and Duffie, Saita, and Wang (2006) use this property to compute the likelihood function, which is to be maximized when estimating the coefficients of a default intensity model, as the product across firms of the covariate-conditional likelihoods of each firm’s default or survival. This significantly reduces the computational complexity of the estimation.

The doubly-stochastic assumption is violated in the presence of “frailty,” meaning unobservable explanatory variables that may be correlated across firms. For example, the defaults of Enron in 2001 and WorldCom in 2002 may have revealed faulty accounting practices that could have been used at other firms, and thus may have had an impact on the conditional default probabilities of other firms. Even if all relevant covariates are observable in principle, some will inevitably be ignored in practice. The impacts of missing and unobservable covariates are essentially equivalent from the viewpoint of estimating default probabilities or portfolio credit risk.

Our primary objective is to measure the degree of frailty that has been present for U.S. corporate defaults, and then to examine its empirical implications. We are particularly interested in the implications of common unobserved covariates on aggregate default rates and on default correlation. We find strong evidence of persistent unobserved covariates. For example,

even after controlling for the “usual-suspect” covariates, both firm-specific and macroeconomic, we find that defaults were persistently higher than expected during lengthy periods of time, for example 1986-1991, and persistently lower in others, for example during the mid-nineties. From trough to peak, the impact of frailty on the average default rates of U.S. corporations is roughly a factor of 2. This is quite distinct from the effect of time fixed effects (time dummy variables, or baseline hazard functions), because of the discipline placed on the behavior of the unobservable covariate through its transition probabilities. Deterministic time effects eliminate an important potential channel for default correlation, namely uncertainty regarding the current level of the time effect and its future evolution.

Incorporating unobserved covariates also has an impact on the relative default probabilities of individual issuers because it changes the relative weights placed on different observable covariates, although this effect is not especially large for our data because of the dominant role of a single covariate, the distance to default, which is a volatility corrected measure of leverage.

We anticipate several types of applications for our work. Understanding how corporate defaults are correlated is particularly important for the risk management of portfolios of corporate debt. For example, as backing for the performance of their loan portfolios, banks retain capital at levels designed to withstand default clustering at extremely high confidence levels, such as 99.9%. Some banks do so on the basis of models in which default correlation is assumed to be captured by common risk factors determining conditional default probabilities, as in Gordy (2003) and Vasicek (1987). If, however, defaults are more heavily clustered in time than currently captured in these default-risk models then significantly greater capital might be required in order to survive default losses with high confidence levels. An understanding of the sources and degree of default clustering is also crucial for the rating and risk analysis of structured credit products that are exposed to correlated default, such as collateralized debt obligations (CDOs) and options on portfolios of default swaps. The Bank of International Settlements (BIS) has cited reports¹ that cash CDO volumes reached \$163 billion in 2004, while synthetic CDO volumes reached \$673 billion. While we do not address the pricing of credit risk in this paper, a “risk-neutral” frailty effect could play a useful role in the valuation of relatively senior tranches of CDOs.

The remainder of the paper is organized as follows. The rest of this

¹Data are provided in the 75th BIS Annual Report, June 2005.

introductory section gives an overview of related literature and describes our dataset. Section 2 specifies the basic probabilistic model for the joint distribution of default times. Section 3 shows how we estimate the model parameters using a combination of the Monte Carlo EM algorithm and the Gibbs sampler. Section 4 summarizes some of the properties of the fitted model and of the posterior distribution of the frailty variable, given the entire sample. Section 5 is concerned with the posterior of the frailty variable at any point in time, given only past history, which determines the dynamics of conditional default risk from the viewpoint of investors in corporate debt. Section 6 addresses the term-structure of default probabilities implied by the model. Section 7 examines the impact of the frailty variable on default correlation, and on the tail risk of a U.S. corporate debt portfolio. Section 8 examines the out-of-sample performance of our model, while Section 9 concludes and suggests some areas for future research. Appendices outline the Gibbs sampling methodology for maximum likelihood estimation.

1.1 Related Literature

A standard structural model of default timing assumes that a corporation defaults when its assets drop to a sufficiently low level relative to its liabilities. For example, the models of Black and Scholes (1973), Merton (1974), Fisher, Heinkel, and Zechner (1989), and Leland (1994) take the asset process to be a geometric Brownian motion. In these models, a firm's conditional default probability is completely determined by its distance to default, which is the number of standard deviations of annual asset growth by which the asset level (or expected asset level at a given time horizon) exceeds the firm's liabilities. This default covariate, using market equity data and accounting data for liabilities, has been adopted in industry practice by Moody's KMV, a leading provider of estimates of default probabilities for essentially all publicly traded firms (see Crosbie and Bohn (2002) and Kealhofer (2003)). Based on this theoretical foundation, it seems natural to include distance to default as a covariate.

In the context of a standard structural default model of this type, however, Duffie and Lando (2001) show that if distance to default cannot be accurately measured, then a filtering problem arises, and the resulting default intensity depends on the measured distance to default and other covariates that may reveal additional information about the firm's condition. More generally, a firm's financial health may have multiple influences over time.

For example, firm-specific, sector-wide, and macroeconomic state variables may all influence the evolution of corporate earnings and leverage. Given the usual benefits of parsimony, the model of default probabilities estimated in this paper adopts a relatively small set of firm-specific and macroeconomic covariates.

Altman (1968) and Beaver (1968) were among the first to estimate statistical models of the likelihood of default of a firm within one accounting period, using accounting data. Early in the empirical literature on default time distributions is the work of Lane, Looney, and Wansley (1986) on bank default prediction, using time-independent covariates. Lee and Urrutia (1996) used a duration model based on a Weibull distribution of default times. Duration models based on time-varying covariates include those of McDonald and Van de Gucht (1999), in a model of the timing of high-yield bond defaults and call exercises. Related duration analysis by Shumway (2001), Kavvathas (2001), Chava and Jarrow (2004), and Hillegeist, Keating, Cram, and Lundstedt (2004) predict bankruptcy. Shumway (2001) uses a discrete duration model with time-dependent covariates. Duffie, Saita, and Wang (2006) provide maximum likelihood estimates of term structures of default probabilities by using a joint model for default intensities and the dynamics of the underlying time-varying covariates. These papers make the doubly-stochastic assumption, and therefore do not account for possibly unobservable or missing covariates affecting default probabilities. Das, Duffie, Kapadia, and Saita (2006), using roughly the same data studied here and by Duffie, Saita, and Wang (2006), provide strong evidence that defaults are significantly more correlated than would be suggested by the doubly stochastic assumption and the assumption that default intensities are explained by the observable covariates.

Empirical studies such as those of Collin-Dufresne, Goldstein, and Helwege (2003) and Zhang (2004) find that major credit events are associated with significant increases in the credit spreads of other firms, consistent with the existence of a frailty effect for actual or risk-neutral default probabilities. Collin-Dufresne, Goldstein, and Hugonnier (2004), Giesecke (2004), and Schönbucher (2003) explore learning-from-default interpretations, based on the statistical modeling of frailty, under which default intensities include the expected effect of unobservable covariates. In a frailty setting, the arrival of a default causes, via Bayes' Rule, a jump in the conditional distribution of hidden covariates, and therefore a jump in the conditional default probabilities of any other firms whose default intensities depend on the same

unobservable covariates. For example, the collapses of Enron and WorldCom could have caused a sudden reduction in the perceived precision of accounting leverage measures of other firms. Delloy, Fermanian, and Sbai (2005), in independent research that appeared shortly before ours, estimate default probabilities using a dynamic frailty model of rating transitions. They suppose that the intensities of downgrades from one rating to the next lower rating depend on a common unobservable factor, which gives rise to fatter tails for the distribution of aggregate portfolio losses.

Yu (2005) finds empirical evidence that, other things equal, a reduction in the measured precision of accounting variables is associated with a widening of credit spreads.

1.2 Data

Our dataset, drawing from Bloomberg, Compustat, CRSP and Moody's, is almost the same as that used in Duffie, Saita, and Wang (2006) and Das, Duffie, Kapadia, and Saita (2006). We have slightly improved the data by using The Directory of Obsolete Securities and the SDC database to identify additional mergers, defaults, or failures. The few additional defaults and mergers identified through these sources do not change significantly the results in Duffie, Saita, and Wang (2006). Our dataset contains 402,434 firm-months of data between January 1979 and March 2004. Because of the manner in which we define defaults, it is appropriate to use data only up to December 2003. For the total of 2,793 companies in this improved dataset, Table 1 shows the number of firms in each exit category. Of the total of 496 defaults, 176 first occurred as bankruptcies, although many of the other defaults eventually led to bankruptcy. In particular, many of the "other defaults" led subsequently to bankruptcy. We refer the interested reader to Section 3.1 in Duffie, Saita, and Wang (2006) for an in-depth description of the construction of the dataset and an exact definition of these event types.

Figure 1 shows the total number of defaults (bankruptcies and other defaults) in each year. Moody's 13th annual corporate bond default study² provides a detailed exposition of historical default rates for various categories of firms since 1920.

The model of default intensities estimated in this paper adopts a parsimonious

²Moody's Investor Service, "Historical Default Rates of Corporate Bond Issuers, 1920-1999."

Exit type	Number
bankruptcy	176
other default	320
merger-acquisition	1,047
other	671

Table 1: Number of firm exits of each type.

monious set of observable firm-specific and macroeconomic covariates:

- Distance to default, a volatility-adjusted measure of leverage. Our method of construction, based on market equity data and Compustat book liability data, is along the lines of that used by Vassalou and Xing (2004), Crosbie and Bohn (2002), and Hillegeist, Keating, Cram, and Lundstedt (2004).
- The firm’s trailing 1-year stock return.
- The 3-month Treasury bill rate.
- The trailing 1-year return on the S&P500 index.

Duffie, Saita, and Wang (2006) give a detailed description of these covariates and discuss their relative importance in modeling corporate default intensities. Other macroeconomic variables, such as GDP growth rates, industrial production growth, and the industry average distance to default, were also considered but found to be at best marginally significant after controlling for distance to default, trailing returns of the firm and the S&P 500, and the 3-month treasury-bill rate. We also considered a firm size covariate, measured as the logarithm of the model-implied assets. Size may be associated with market power, management strategies, or borrowing ability, all of which may affect the risk of failure. For example, it might be easier for a big firm to re-negotiate with its creditors to postpone the payment of debt, or to raise new funds to pay old debt. In a “too-big-to-fail” sense, firm size may also negatively influence failure intensity. The statistical significance of size as a determinant of failure risk has been documented by Shumway (2001). For our data and our measure of firm size, this covariate did not play a statistically significant role.

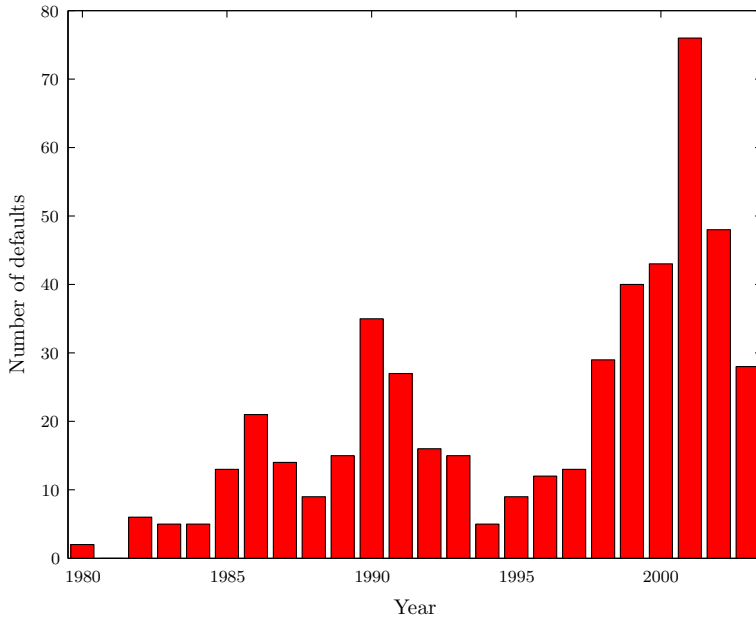


Figure 1: The number of defaults in our dataset for each year between 1980 and 2003.

2 The Model

We fix a probability space and a filtration of information³ for the purpose of an introduction to the default timing model, which will be made precise as we proceed. For a given borrower whose default time is τ , we will say that a non-negative progressively measurable process λ is the default intensity of the borrower if a martingale is defined by $1_{\tau \leq t} - \int_0^t \lambda_s 1_{\tau > s} ds$. This means that for a firm that is currently active the default intensity is the conditional mean arrival rate of default, measured in events per year (or per month, in some of our results).

Our model is based on a Markov state vector X_t of firm-specific and macroeconomic covariates, that may be only partially observable. If all of these covariates were observable, the default intensity of firm i at time t would be $\lambda_{it} = \Lambda(S_i(X_t), \theta)$, where θ is a parameter vector to be estimated and $S_i(X_t)$ is the component of the state vector relevant to the default in-

³For precise mathematical definitions not offered here, see Protter (2004).

tensity of firm i . The doubly-stochastic assumption, as stated for example in Chapter 11 of Duffie (2001) or Duffie, Saita, and Wang (2006), is that, conditional on the path of the underlying state process X determining default and other exit intensities, exit times are the first event times of independent Poisson processes. In particular, this means that, given the path of the state-vector process, the merger and failure times of different firms are conditionally independent.

A major advantage of the doubly-stochastic formulation is that it allows decoupled maximum-likelihood estimation of the parameter vector γ determining the time-series dynamics of the covariate process X as well as the parameter vector θ determining the default intensities. The two parameter vectors γ and θ can then be combined to obtain the maximum-likelihood estimator of, for example, a multi-year portfolio loss probability.

Coupled with the model of default intensities that we adopt here, the doubly-stochastic assumption is overly restrictive for US corporations during 1979-2004, according to tests developed in Das, Duffie, Kapadia, and Saita (2006). There are several channels by which the excessive default correlation shown in Das, Duffie, Kapadia, and Saita (2006) data could arise. With “contagion,” for example, default by one firm could have a direct influence on the default likelihood of another firm. This would be anticipated to some degree if one firm plays a relatively large role in the marketplace of another. The influence of the bankruptcy of autoparts manufacturer Delphi in November 2005 on the survival prospects of General Motors’ illustrates how failure by one firm could weaken another, above and beyond the default correlation induced by common and correlated covariates.

In this paper, we will examine instead the implications of a “frailty” effect, by which many firms could be jointly exposed to one or more unobservable risk factors. We restrict attention for simplicity to a single common frailty factor and to firm-by-firm idiosyncratic frailty factors, although a richer model and sufficient data could allow for the estimation of additional frailty factors, for example at the sectoral level.

The mathematical model that we adopt is actually doubly stochastic once the information available to the econometrician is artificially enriched to include the frailty factors. That is, conditional on the future paths of both the observable and unobservable components of the state vector X , firms are assumed to default independently. Thus, there are two channels for default correlation: (i) future co-movement of the observable and unobservable factors determining intensities, and (ii) uncertainty in the current conditional

distribution of the frailty factors, given past observations of the observable covariates and past defaults.

We let U_{it} be a firm-specific vector of covariates that are observable for firm i until its exit time T_i , let V_t denote a vector of macro-economic variables that are observable at all times, and let Y_t be an vector of unobservable frailty variables. The complete state vector is then $X_t = (U_{1t}, \dots, U_{mt}, V_t, Y_t)$, where m is the total number of firms in the dataset.

We let $W_{it} = (1, U_{it}, V_t)$ be the vector of observed covariates for company i (including a constant). Since we observe these covariates on a monthly basis but measure default times continuously, we take $W_{it} = W_{i,k(t)}$, where $k(t)$ is the time of the most recent month end. We let t_i and T_i be the first and last points in time, respectively, at which company i is observed.

The information filtration $(\mathcal{H}_t)_{0 \leq t \leq T}$ generated by firm-specific covariates is defined by

$$\mathcal{H}_t = \sigma(\{U_{i,s} : 1 \leq i \leq m, t_i \leq s \leq t \wedge T_i\}).$$

The default-time filtration $(\mathcal{U}_t)_{0 \leq t \leq T}$ is given by

$$\mathcal{U}_t = \sigma(\{D_{is} : 1 \leq i \leq m, t_i \leq s \leq t \wedge T_i\}),$$

where D_i is the default indicator process of company i (0 before default, 1 afterwards). The econometrician's information filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$ is defined by the join,

$$\mathcal{F}_t = \sigma(\mathcal{H}_t \cup \mathcal{U}_t \cup \{V_s : 0 \leq s \leq t\}).$$

The complete-information filtration $(\mathcal{G}_t)_{0 \leq t \leq T}$ is the yet larger join

$$\mathcal{G}_t = \sigma(\{Y_s : 0 \leq s \leq t\}) \vee \mathcal{F}_t.$$

With respect to the complete information filtration (\mathcal{G}_t) , defaults are assumed to be doubly stochastic, with the default intensity of firm i given by $\lambda_{it} = \Lambda(S_i(X_t); \theta)$, where $S_i(X_t) = (W_{it}, Y_t)$. We take the proportional-hazards form

$$\Lambda((w, y); \theta) = e^{\beta_1 w_1 + \dots + \beta_n w_n + \eta y} \tag{1}$$

for a parameter vector $\theta = (\beta, \eta)$ common to all firms. We can write

$$\lambda_{it} = e^{\beta \cdot W_{it}} e^{\eta Y_t} \equiv \tilde{\lambda}_{it} e^{\eta Y_t},$$

so that $\tilde{\lambda}_{it}$ is the component of the (\mathcal{G}_t) -intensity that is due to observable covariates and $e^{\eta Y_t}$ is a scaling factor due to the unobservable frailty.

In spirit (see Brémaud (1981), Chapter II, Theorem 14), the econometrician's default intensity for firm i is

$$\bar{\lambda}_{it} = E(\lambda_{it} | \mathcal{F}_t) = e^{\beta \cdot W_{it}} E(e^{\eta Y_t} | \mathcal{F}_t),$$

which is obtained by averaging over the distribution of Y_t given \mathcal{F}_t . We do not rely on this calculation, which need not hold in general settings. Even when this calculation of the econometrician's default intensity is justified, it is *not* generally true⁴ that the conditional probability of survival to a future time T (neglecting the effect of other exits) is given by the “usual formula,”

$$E\left(e^{-\int_t^T \bar{\lambda}_{is} ds} | \mathcal{F}_t\right).$$

Rather, for a firm that has survived to time t , the probability of survival to time T is (again neglecting other exits)

$$E\left(e^{-\int_t^T \lambda_{is} ds} | \mathcal{F}_t\right). \tag{2}$$

Although λ_i is *not* the (\mathcal{F}_t) -intensity of default, (2) is justified by the law of iterated expectations and the fact that, conditional on the complete information filtration, the doubly stochastic property implies that the \mathcal{G}_t -conditional survival probability is

$$E\left(e^{-\int_t^T \lambda_{is} ds} | \mathcal{G}_t\right).$$

Similarly, ignoring other exits, the \mathcal{F}_t -conditional probability of joint survival by two currently alive firms, i and j , until a future time T is

$$E\left(e^{-\int_t^T (\lambda_{is} + \lambda_{js}) ds} | \mathcal{F}_t\right).$$

Before considering the effect of other exits, such as mergers and acquisitions, the maximum likelihood estimators for these \mathcal{F}_t -conditional survival probabilities, and related quantities such as default correlation, are obtained under the usual smoothness conditions by substituting the maximum likelihood estimators for the parameters (γ, θ) into these formulas. An extension that

⁴See Collin-Dufresne, Goldstein, and Huggonier (2004) for another approach to this calculation.

treats other exits is given by Duffie, Saita, and Wang (2006). For example, it is impossible for a firm to default beginning in 2 years if it has already been acquired by another firm within 2 years.

Under the doubly-stochastic assumption, if all covariates determining default intensities are observable, Proposition 2 of Duffie, Saita, and Wang (2006) states that the joint maximum likelihood estimation of the parameter vector γ determining the covariate time-series dynamics and the coefficients θ determining the exit intensities can be decomposed into separate maximum likelihood estimation problems for θ and for γ . This decomposition, however, is not generally feasible for the frailty model.

To further simplify notation, let $W = (W_i : 1 \leq i \leq m)$ denote the vector of observed covariate processes for all companies, and let $D = (D_{is} : 1 \leq i \leq m)$ denote the vector of default indicators of all companies. On the complete-information filtration (\mathcal{G}_t) , the doubly-stochastic property and Proposition 2 of Duffie, Saita, and Wang (2006) states that the likelihood of the data at the parameters (γ, θ) is of the form

$$\begin{aligned} & \mathcal{L}(\gamma, \theta | W, Y, D) \\ &= \mathcal{L}(\gamma | W, Y) \mathcal{L}(\theta | W, Y, D) \\ &= \mathcal{L}(\gamma | W, Y) \prod_{i=1}^m \left(e^{-\sum_{t=t_i}^{T_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{T_i} [D_{it} \lambda_{it} \Delta t + (1 - D_{it})] \right). \end{aligned} \quad (3)$$

We simplify by supposing that the frailty process Y has a fixed probability distribution, and is independent of the observable covariate process W . With respect to the econometrician's filtration (\mathcal{F}_t) , the likelihood is therefore

$$\begin{aligned} & \mathcal{L}(\gamma, \theta | W, D) = \int \mathcal{L}(\gamma, \theta | W, y, D) p_Y(y) dy \\ &= \mathcal{L}(\gamma | W) \int \mathcal{L}(\theta | W, y, D) p_Y(y) dy \\ &= \mathcal{L}(\gamma | W) E_Y \left[\prod_{i=1}^m \left(e^{-\sum_{t=t_i}^{T_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{T_i} [D_{it} \lambda_{it} + (1 - D_{it})] \right) \middle| W, D \right], \end{aligned} \quad (4)$$

where $p_Y(\cdot)$ and E_Y denote the probability density of the path of the unobserved frailty process Y , and expectation with respect to that density,

respectively. (This expression ignores for simplicity the precise intra-month timing of default.)

We will provide the maximum likelihood estimator (MLE) $(\hat{\theta}, \hat{\gamma})$ for (θ, γ) . Extending from Duffie, Saita, and Wang (2006), we can decompose this MLE problem into separate maximum likelihood estimations of γ and θ , by maximization of the first and second factors on the right-hand side of (4), respectively.

Even when considering other exits, $(\hat{\theta}, \hat{\gamma})$ is the full maximum likelihood estimator for (θ, γ) provided other exits are jointly doubly stochastic with defaults on the complete information filtration (\mathcal{G}_t) , as in Duffie, Saita, and Wang (2006). We make this simplifying assumption.

In order to evaluate the expectation in (4) one could simulate sample paths of the frailty process Y . Since our covariate data are monthly observations from 1979 to 2004, evaluating (4) means Monte Carlo integration in a high dimensional space. This is extremely numerically intensive by brute-force Monte Carlo, given the overlying search for parameters. We suppose that Y is a standard Brownian motion and estimate the model parameters using a combination of the EM algorithm and the Gibbs sampler, as described in Section 3 and the appendices.

Brownian motion is a reasonable starting point for this frailty analysis, given its continuity and iid increments properties. Although allowing for jumps and mean reversion could enrich the frailty effects, we have avoided these extensions because of our concern at identifying the additional parameters involved, given the relatively small number of defaults that have occurred. Similarly, we avoid including a drift (time trend) coefficient, allowing the posterior distribution of a zero-drift Brownian motion to inform us of the effect of variation over time of the frailty effect. Without loss of generality, we can fix the volatility parameter of the Brownian motion to be any constant, in our case 1, because scaling the parameter η determining the dependence of the default intensities on Y_t plays precisely the same role in the model as the scaling of the volatility parameter of the Brownian motion Y . The starting value of the Brownian motion is taken to be zero. Although any fixed starting condition for $Y(t)$ can be absorbed into the default intensity intercept coefficient β_1 without loss of generality, we do lose some generality by taking the initial condition for Y to be deterministic. An alternative would be to add one or more additional parameters specifying the initial probability distribution of Y . We found that the posterior of $Y(t)$ tended to be robust to the initial assumed distribution of Y , for t within a

year or two after the initial date in our sample.

2.1 Unobserved Heterogeneity

It may be that a substantial portion of the differences among firms' default risks is due to unobserved heterogeneity. We consider an extension of the model by introducing a firm-specific heterogeneity factor Z_i , so that the complete-information (\mathcal{G}_t) default intensity is assumed to be of the form

$$\lambda_{it} = e^{X_{it}\beta + \gamma Y_t} Z_i = \tilde{\lambda}_{it} e^{\gamma Y_t} Z_i, \quad (5)$$

where Z_1, \dots, Z_m are independently and identically gamma distributed random variables that are jointly independent of the observable covariates W and the common frailty process Y . Again, we assume that defaults and other exits are doubly stochastic on the complete-information filtration (\mathcal{G}_t).

Fixing the mean of the heterogeneity factor Z_i to be 1 without loss of generality, we found that maximum likelihood estimation does not pin down the variance of Z_i to any reasonable precision with the limited set of data available. We anticipate that far larger datasets would be needed, given the already large degree of observable heterogeneity. In the end, we examine the potential role of unobserved heterogeneity for default risk by fixing the standard deviations of Z_i at 0.5.

Letting $Z = (Z_1, \dots, Z_m)$, the complete-information likelihood of the parameters (γ, θ) is

$$\begin{aligned} & \mathcal{L}(\gamma, \theta | W, Y, Z, D) \\ &= \mathcal{L}(\gamma | W) \cdot \mathcal{L}(\theta | W, Y, Z, D) \\ &= \mathcal{L}(\gamma | W) \prod_{i=1}^m \left(e^{-\sum_{t=t_i}^{T_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{T_i} [D_{it} \lambda_{it} + (1 - D_{it})] \right), \end{aligned} \quad (6)$$

where the default intensities λ_{it} are given by (5). Using our independence assumptions, the likelihood of the observed data is therefore

$$\begin{aligned} & \mathcal{L}(\gamma, \theta | W, D) = \mathcal{L}(\gamma | W) \int \int \mathcal{L}(\theta | W, y, z, D) p_Y(y) p_Z(z) dy dz \\ &= \mathcal{L}(\gamma | W) E \left[\prod_{i=1}^m \left(e^{-\sum_{t=t_i}^{T_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{T_i} [D_{it} \lambda_{it} + (1 - D_{it})] \right) \right], \end{aligned} \quad (7)$$

where $p_Y(y)$ and $p_Z(z)$ denote the densities of the frailty path y and the vector z of heterogeneity outcomes, respectively, and where the expectation integrates over the product of the distributions of Y and Z .

3 Parameter Estimation

We now turn to the problem of inference from data. The parameter vector γ determining the time-series model for the observable covariate process W is specified and estimated in Duffie, Saita, and Wang (2006). This model is vector-autoregressive Gaussian, with a number of structural restrictions chosen for parsimony and tractability. We focus here on the estimation of the parameter vector θ of the default intensity model.

3.1 Estimating the Model with Frailty

We use a variant of the expectation maximization (EM) algorithm (Dempster, Laird, and Rubin (1977)), an iterative method for the computation of the maximum likelihood estimator of parameters in models involving missing or incomplete data. See also Cappé, Moulines, and Rydén (2005), who discuss EM in the context of hidden Markov models. In many potential applications, explicitly calculating the conditional expectation required in the E-step of the algorithm may not be possible. Nevertheless, the expectation can be approximated by Monte Carlo integration, which gives rise to the stochastic EM algorithm, as explained for example by Celeux and Diebolt (1986) and Nielsen (2000), or to the Monte-Carlo EM algorithm (Wei and Tanner (1990)).

Maximum likelihood estimation (MLE) of the intensity parameter vector θ involves the following steps:

0. Initialize an estimate of $\theta = (\beta, \eta)$ at $\theta^{(0)} = (\hat{\beta}, 0.05)$, where $\hat{\beta}$ is the maximum likelihood estimator of β in the model without frailty, which can be obtained by maximizing the likelihood function (3) by standard methods such as the Newton-Raphson algorithm.
1. (Monte-Carlo E-step.) Given the current parameter estimate $\theta^{(k)}$ and the observed covariate and default data W and D , respectively, draw sample paths $Y^{(j)} := \{Y_t^{(j)}, 0 \leq t \leq T\}$ for $j = 1, \dots, n$ from the conditional distribution $p(\cdot | W, D, \theta^{(k)})$ of the latent Brownian motion

frailty process Y . We do this with the Gibbs sampler described in Appendix A. We let

$$\begin{aligned} Q(\theta, \theta^{(k)}) &= E_{\theta^{(k)}}(\log \mathcal{L}(\theta | W, Y, D)) \\ &= \int \log \mathcal{L}(\theta | W, y, D) p_Y(y | W, D, \theta^{(k)}) dy, \end{aligned} \quad (8)$$

which is commonly referred to in the EM literature as the “expected complete-data log-likelihood” or “intermediate quantity.” Using the sample paths generated by the Gibbs sampler, (8) can be approximated as

$$\widehat{Q}(\theta, \theta^{(k)}) = \frac{1}{n} \sum_{j=1}^n \log \mathcal{L}(\theta | W, Y^{(j)}, D). \quad (9)$$

2. (M-step.) Maximize $\widehat{Q}(\theta, \theta^{(k)})$ with respect to the parameter vector θ , for example by Newton-Raphson. The maximizing choice of θ is the new parameter estimate $\theta^{(k+1)}$.
3. Replace k with $k + 1$, and return to Step 2, repeating the MC E-step and the M-step until reasonable numerical convergence.

One can show (Demptser, Laird, and Rubin (1977) or Gelman, Carlin, Stern, and Rubin (2004)) that $\mathcal{L}(\gamma, \theta^{(k+1)} | W, D) \geq \mathcal{L}(\gamma, \theta^{(k)} | W, D)$, that is, the observed data likelihood (4) is non-decreasing in each step of the EM algorithm. Under regularity conditions, the parameter sequence $\{\theta^{(k)} : k \geq 0\}$ therefore converges to at least a local maximum. (See Wu (1983) for a precise formulation in terms of stationarity points of the likelihood function.) Nielsen (2000) gives sufficient conditions for global convergence and asymptotic normality of the parameter estimates, which however are usually hard to verify. As with many maximization algorithms, a simple way to mitigate the risk that one misses the global maximum is to start the iterations at many points throughout the parameter space.

One can show under regularity conditions (see for example Proposition 10.1.4. of Cappé, Moulines, and Rydén (2005)) that

$$\nabla_{\theta} \mathcal{L}(\theta' | W, Y, D) = \nabla_{\theta} Q(\theta, \theta') |_{\theta=\theta'},$$

so that in particular

$$\nabla_{\theta} \mathcal{L}(\hat{\theta} | W, Y, D) = \nabla_{\theta} Q(\theta, \hat{\theta}) |_{\theta=\hat{\theta}}.$$

This means that the Hessian matrix of the expected complete-data likelihood can be used to calculate asymptotic standard errors for the MLE parameter estimators.

3.2 Estimation with Unobserved Heterogeneity

An extension of the model that incorporates unobserved heterogeneity can be estimated with the following algorithm:

0. Initialize $Z_i^{(0)} = 1$ for $1 \leq i \leq m$ and initialize $\theta^{(0)} = (\hat{\beta}, 0.05)$, where $\hat{\beta}$ is the maximum likelihood estimator of β , in the model without frailty.
1. (Monte-Carlo E-step.) Given the current parameter estimate $\theta^{(k)}$, draw samples $(Y^{(j)}, Z^{(j)})$ for $j = 1, \dots, n$ from the joint posterior distribution $p_{Y,Z}(\cdot | W, D, \theta^{(k)})$ of the frailty sample path $Y = \{Y_t : 0 \leq t \leq T\}$ and the vector $Z = (Z_i : 1 \leq i \leq m)$ of unobserved heterogeneity variables. This can be done by, for example, using the Gibbs sampler described in Appendix B. Then calculate the expected complete-data log-likelihood

$$\begin{aligned} Q(\theta, \theta^{(k)}) &= E_{\theta^{(k)}}(\log \mathcal{L}(\theta | W, Y, Z, D)) \\ &= \int \log \mathcal{L}(\theta | W, y, z, D) p_{Y,Z}(y, z | W, D, \theta^{(k)}) dy dz. \end{aligned} \quad (10)$$

Using the sample paths generated by the Gibbs sampler, (10) can be approximated by

$$\hat{Q}(\theta, \theta^{(k)}) = \frac{1}{n} \sum_{j=1}^n \log \mathcal{L}(\theta | W, Y^{(j)}, Z^{(j)}, D). \quad (11)$$

2. (M-step.) Maximize $\hat{Q}(\theta, \theta^{(k)})$ with respect to the parameter vector θ , using the Newton-Raphson algorithm. Set the new parameter estimate $\theta^{(k+1)}$ equal to this maximizing value.
3. Replace k with $k + 1$, and return to Step 2, repeating the MC E-step and the M-step until reasonable numerical convergence.

4 Empirical analysis

We fit our models to the data for all matchable U.S. non-financial public firms, as described in Section 1.2. This section presents the basic results.

4.1 The Model with Frailty

Table 2 shows the estimated covariate parameter vector $\hat{\beta}$ and frailty volatility parameter $\hat{\eta}$, with “asymptotic” estimates of standard errors of the coefficients given parenthetically.

Constant	DTD	Trailing Stock Return
−1.625 (0.137)	−1.198 (0.037)	−0.651 (0.075)
3m T-bill	Trailing S&P	Latent Factor Volatility
−0.216 (0.029)	1.677 (0.303)	0.109 (0.015)

Table 2: Maximum likelihood estimates of the intensity parameters in the model with frailty. “DTD” is distance to default. The frailty volatility is the coefficient η of dependence of the default intensity on the standard Brownian frailty process Y . Standard errors, given in parentheses, were computed using the Hessian matrix of the expected complete data log-likelihood at $\theta = \hat{\theta}$.

Our results concerning important roles of both firm-specific and macroeconomic covariates are consistent with prior literature on modeling default intensities. In particular, distance to default, although statistically highly significant, does not on its own determine the default intensity, but does explain a large part of the variation of default risk over time. For example a negative shock of distance to default by one unit increases the default intensity by roughly $e^{1.2} - 1 \approx 230\%$. As in Duffie, Saita, and Wang (2006), the one-year trailing stock return covariate proposed by Shumway (2001) has a highly significant impact on default intensities. Perhaps it is a proxy for firm-specific information that is not captured by distance to default.⁵ The coefficient linking the trailing S&P 500 return to a firm’s default intensity

⁵There is also the potential, with the momentum effects documented by Jegadeesh and Titman (1993) and Jegadeesh and Titman (2001), that trailing return is a forecaster of future distance to default.

is positive at conventional significance levels, and of the unexpected sign by univariate reasoning. Of course, with multiple covariates, the sign need not be evidence that a good year in the stock market is itself bad news for default risk. It could also be the case that, after “boom” years for the stock market, distance-to-default overstates the financial health of a company.

The estimate $\hat{\eta} = 0.109$ of the dependence of the unobservable default intensities on the frailty variable $Y(t)$, corresponds to a monthly volatility of this frailty effect of 10.9%, which translates to an annual volatility of 37.8%. The effect is highly economically and statistically significant.

Constant	DTD	Trailing Stock Return
-2.093 (0.108)	-1.200 (0.035)	-0.681 (0.070)
3m T-bill	Trailing S&P	Latent Factor Volatility
-0.106 (0.020)	1.481 (0.284)	0 (0)

Table 3: Maximum likelihood estimates of the intensity parameters in the model without frailty. Standard errors, given in parenthesis, were computed using the Hessian matrix of the likelihood function at $\theta = \hat{\theta}$.

Table 3 reports the MLE of θ in the model without frailty. We see that the signs, magnitudes and statistical significance of the coefficients on the observable covariates are essentially unaffected. In particular, they almost all lie within one standard error of those of the original estimates for the model with frailty variable.

The Monte Carlo EM algorithm allows us to compute the \mathcal{F}_T -conditional posterior distribution⁶ of the frailty variable Y_t , where T is the final date of our sample. This posterior for the latent Brownian motion frailty Y_t is a by-product of the E-step. Figure 2 visualizes the conditional mean of the latent factor, estimated as the average of 5,000 sample paths for $Y(t)$ drawn from the Gibbs sampler. Also shown are bands around the posterior mean given by standard deviations from the posterior distribution. We see that there are substantial fluctuations in the frailty effect on default risk over time. For example in 2001, average default intensities were estimated to be

⁶With this we mean the conditional distribution of the latent factor given all of the historical default and covariate data through the end of the sample period, and using the estimated parameter vector $\hat{\theta}$.

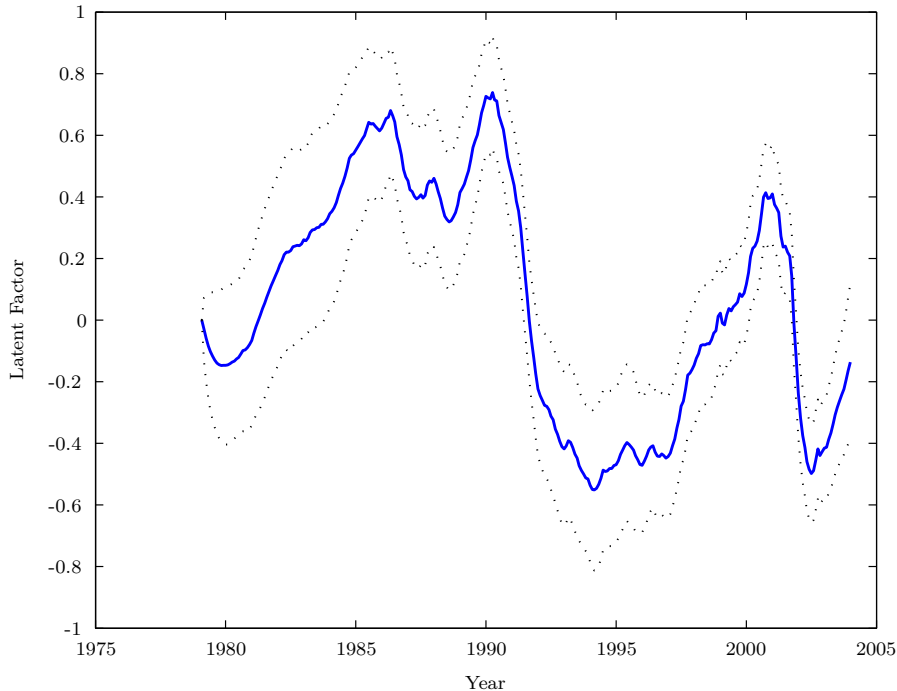


Figure 2: Conditional posterior mean of the scaled latent Brownian motion frailty variable with 1-sigma error bands, that is, $\{E(\eta Y_t | \mathcal{F}_T) : 0 \leq t \leq T\}$

roughly $e^{0.8} \approx 2.2$ times larger than during 1995, ignoring for now the non-linear effects. In order to precisely calculate the increase in the conditional expected default intensity, one compares $E[e^{Y(t)} | \mathcal{F}_T]$ for t in 1995 and in 2001, respectively. The ratio of these two expectations is the ratio of the \mathcal{F}_T -conditional expected default intensities for 1995 and 2001, everything else equal. In our case, the linear approximation works reasonably well because the Jensen effects when calculating the expectations of $e^{Y(t)}$ for the two years are roughly offsetting. The implications of frailty for contemporaneous conditional default probabilities are the subject of Section 6.

Figure 3 shows the conditional density of $Y(t)$ for t at the end of January 1998, conditioning on \mathcal{F}_T (in effect, the entire sample of default times and observable covariates to 2004), and also shows the density of Y_t when conditioning on only \mathcal{F}_t (the data available up to and including January 1998).

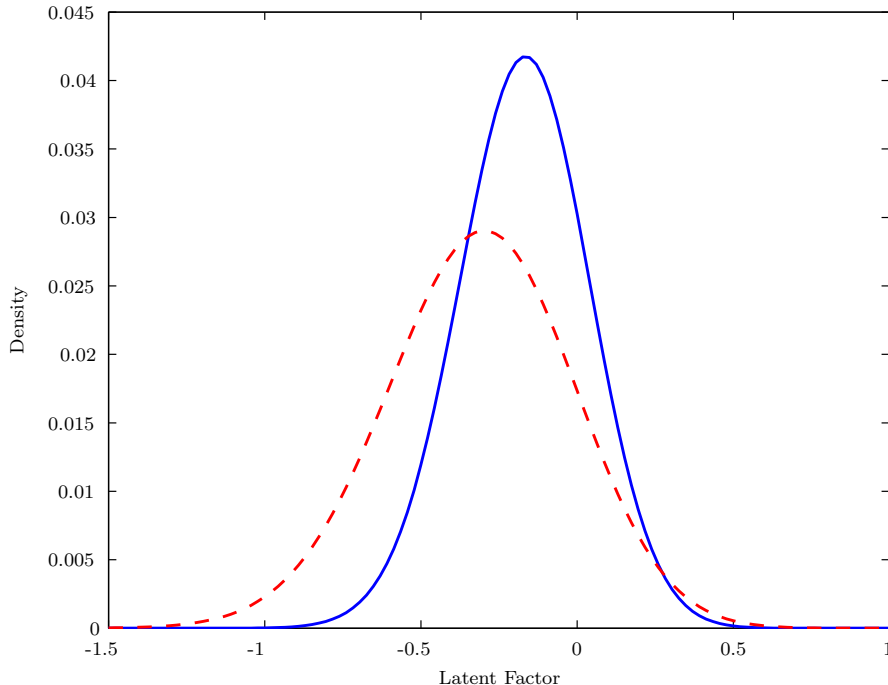


Figure 3: Conditional posterior density of the scaled latent factor in January 1998 using all data, that is, $p(\eta Y_t | \mathcal{F}_T)$ (solid line), and using only past data, that is $p(\eta Y_t | \mathcal{F}_t)$ (dashed line). These densities were calculated using the forward-backward recursions described in Section 5.

One observes that the \mathcal{F}_T -conditional distribution of Y_t is more concentrated than that obtained by conditioning on only the concurrently available information, \mathcal{F}_t . The posterior mean of Y_t given the information available in January 1998 is lower than that given all of the data through 2004, reflecting the sharp rise in corporate defaults in 2001 and 2002 above and beyond that predicted from the observed covariates alone.

Figure 4 shows the cross-sectional average of the observable component $e^{\hat{\beta} \cdot W_{it}}$ of the estimated default intensity. Figure 5 shows the same average covariate-implied default intensity after removing the three most risky firms at each single point in time. The differences between Figure 4 and 5 indicate that the three most risky companies at each point in time explain a large

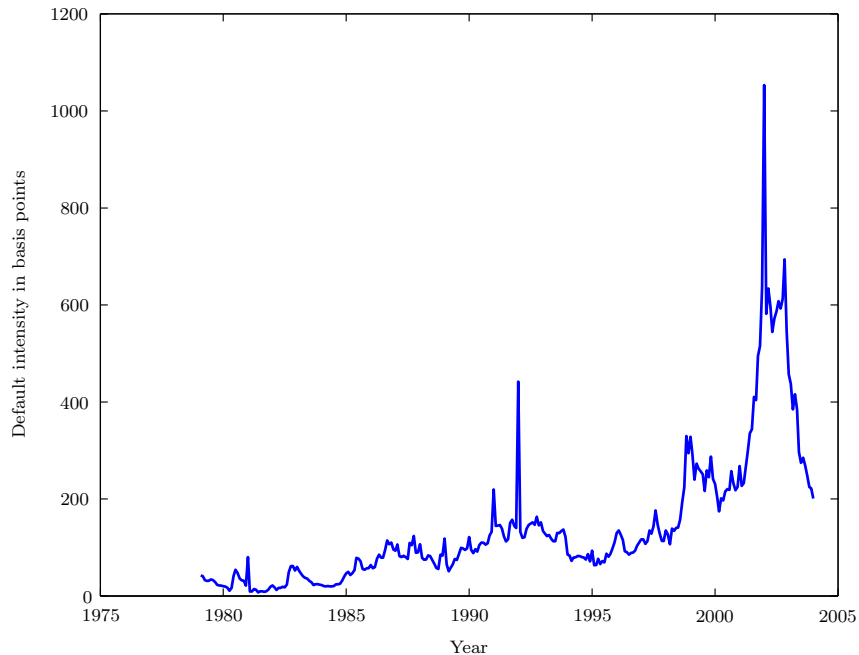


Figure 4: Portfolio average yearly default intensity associated with observable covariates only.

portion of the instantaneous default risk of the whole portfolio. For example, in December 2001 one company alone, Classic Communications, Incorporated, contributed about 300 basis points to the average covariate-implied default intensity of the *whole* portfolio, having had an estimated mean arrival rate of 50 default events per year. Classic Communications defaulted in July 2002.

4.2 With Frailty and Unobserved Heterogeneity

Table 4 shows the MLE of the covariate parameter vector β and the frailty volatility parameter η , with estimated standard errors shown parenthetically. Table 4 and Figure 6 indicate that, while including unobserved heterogeneity decreases the volatility of the latent Brownian motion frailty variable from 10.9% to 9.1% a month, the conclusions from the previous section remain

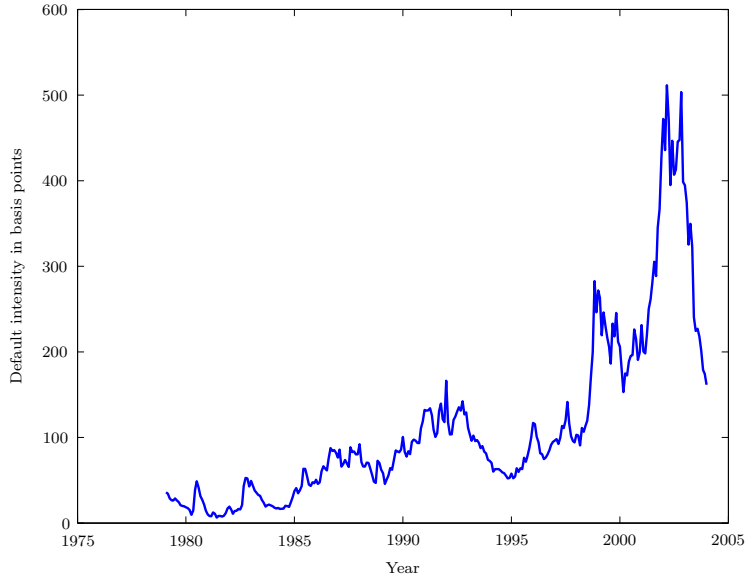


Figure 5: Same as Figure 9, but with three most risky firms removed at each point in time. Note that these three firms will in general vary over time.

the same.

4.3 Model Comparison

Unlike standard tests that evaluate the overall fit of a statistical model (such as the chi-square test), we will compare the marginal likelihoods of the models. This does not rely on large-sample distribution theory and has the intuitive interpretation of attaching weights to the competing models.

We consider a Bayesian approach to comparison of the quality of fit of competing models and assume positive prior weights each of the models “noF” (the model without frailty), “F” (the model with a common frailty variable), “H” (the model with unobserved heterogeneity and no common frailty), and finally “F&H” (the model with a common frailty variable and with unobserved heterogeneity). Consider for example the model with common frailty variable versus the model without. Using the natural informal

Constant	DTD	Trailing Stock Return
-1.551 (0.119)	-1.594 (0.046)	-0.450 (0.074)
3m T-bill	Trailing S&P	Latent Factor Volatility
-0.206 (0.026)	1.655 (0.306)	0.091 (0.013)

Table 4: Maximum likelihood estimates of the intensity parameters in the model with frailty variable and unobserved heterogeneity. Asymptotic standard errors, given in parentheses, were computed using the Hessian matrix of the likelihood function at $\theta = \hat{\theta}$.

notation, the posterior odds ratio is

$$\frac{p(\text{F} | W, D)}{p(\text{noF} | W, D)} = \frac{\mathcal{L}_F(\hat{\theta}_F | W, D)}{\mathcal{L}_{\text{noF}}(\hat{\theta}_{\text{noF}} | W, D)} \frac{p(\text{F})}{p(\text{noF})}, \quad (12)$$

where $\hat{\theta}_M$ and \mathcal{L}_M denote the MLE and the likelihood function for a certain model M , respectively. Plugging (4) into (12) gives

$$\begin{aligned} \frac{p(\text{F} | W, D)}{p(\text{noF} | W, D)} &= \frac{\mathcal{L}(\hat{\gamma}_F | W, Y) \mathcal{L}(\hat{\theta}_F | W, D)}{\mathcal{L}(\hat{\gamma}_{\text{noF}} | W, Y) \mathcal{L}(\hat{\theta}_{\text{noF}} | W, D)} \frac{p(\text{F})}{p(\text{noF})} \\ &= \frac{\mathcal{L}(\hat{\theta}_F | W, D)}{\mathcal{L}(\hat{\theta}_{\text{noF}} | W, D)} \frac{p(\text{F})}{p(\text{noF})}, \end{aligned} \quad (13)$$

using the fact that the time-series model for the covariate process W is the same in both models. The first factor of the right-hand side is sometimes known as the ‘‘Bayes factor.’’

Following Kass and Raftery (1995) and Eraker, Johannes, and Polson (2003), we focus on the size of the statistic Φ given by twice the natural logarithm of the Bayes factor, which is on the same scale as the likelihood ratio test statistic. A value for Φ between 2 and 6 provides positive evidence, a value between 6 and 10 strong evidence, and a value larger than 10 provides very strong evidence for the alternative model. This criterion does not necessarily favor more complex models due to the marginal nature of the likelihood functions in (13). See Smith and Spiegelhalter (1980) for a discussion of the penalizing nature of the Bayes factor, sometimes referred to as the ‘‘fully automatic Occam’s razor.’’ Table 5 shows the Bayes factors

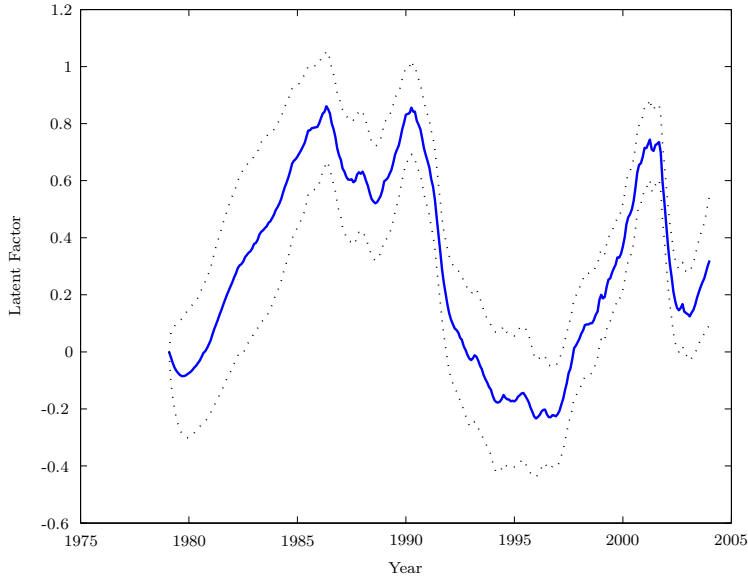


Figure 6: Conditional posterior mean $\{E(\eta Y_t | \mathcal{F}_T) : 0 \leq t \leq T\}$ with 1-sigma error bands for the scaled latent Brownian motion frailty variable in the model that also incorporates unobserved heterogeneity.

for various pairs of models.⁷ In the sense of this approach to model comparison, we see strong evidence in favor of including a time-varying latent frailty variable as well as unobserved heterogeneity.

5 Default Intensity Dynamics

While Figure 2 illustrates the posterior distribution of the frailty Brownian motion Y_t given all information available at the final time T of the sample period, most applications of a default-risk model would call for the posterior distribution of Y_t given the current information \mathcal{F}_t . For example, a bank holding a portfolio of corporate debt could be interested in measuring its current value at risk on this basis.

⁷We currently use the expected complete-data log-likelihood as a crude estimate of the log-likelihood, and will include the precise Bayes factors in the next draft of the paper.

noF vs. F	noF vs. H	F vs. F&H	H vs. F&H
56.2	337.2	337.4	56.4

Table 5: Twice the natural logarithm of the Bayes factor. Here, “noF” is the model without frailty variable, “F” is the model with the common frailty variable, “H” is the model with unobserved heterogeneity, and “F&H” is the model with a common frailty variable and with unobserved heterogeneity.

The standard approach to filtering and smoothing in non-Gaussian state space models is the so-called forward-backward algorithm due to Baum, Petrie, Soules, and Weiss (1970). We let $R(t) = \{i : D_{i,t} = 0, t_i \leq t \leq T_i\}$ denote the set of firms that are alive at time t , and $\Delta R(t) = \{i \in R(t-1) : D_{i,t} = 1, t_i \leq t \leq T_i\}$ be the set of firms that defaulted at time t . A discrete-time approximation of the complete-information likelihood of the observed survivals and defaults at time t is

$$\mathcal{L}_t(\theta | W, Y, D) = \mathcal{L}_t(\theta | W_t, Y_t, D_t) = \prod_{i \in R(t)} e^{-\lambda_{it} \Delta t} \prod_{i \in \Delta R(t)} \lambda_{it} \Delta t.$$

The normalized version of the forward-backward algorithm allows us to calculate the filtered density of the latent Brownian motion frailty variable by the recursion

$$c_t = \int \int p(y_{t-1} | \mathcal{F}_{t-1}) \phi(y_t - y_{t-1}) \mathcal{L}_t(\theta | W_t, y_t, D_t) dy_{t-1} dy_t$$

$$p(y_t | \mathcal{F}_t) = \frac{1}{c_t} \int p(y_{t-1} | \mathcal{F}_{t-1}) \phi(y_t - y_{t-1}) \mathcal{L}_t(\theta | W_t, y_t, D_t) dy_{t-1},$$

where $\phi(\cdot)$ is the Gaussian unconditional density of $Y_t - Y_{t-1}$. For this recursion, we begin with the distribution (Dirac measure) of Y_0 concentrated at 0. Figure 7 shows the path over time of the mean $E(Y_t | \mathcal{F}_t)$ of this posterior density.

Once the filtered density $p(y_t | \mathcal{F}_t)$ is available, the marginal smoothed density $p(y_t | \mathcal{F}_T)$ can be calculated using the normalized backward recursions (Rabiner (1989)). Specifically, for $t = T - 1, T - 2, \dots, 1$, we apply the recursion for the marginal density

$$\bar{\alpha}_{t|T}(y_t) = \frac{1}{c_{t+1}} \int \phi(y_t - y_{t+1}) \mathcal{L}_{t+1}(\theta | W_{t+1}, y_{t+1}, D_{t+1}) \bar{\alpha}_{t+1|T}(y_{t+1}) dy_{t+1}$$

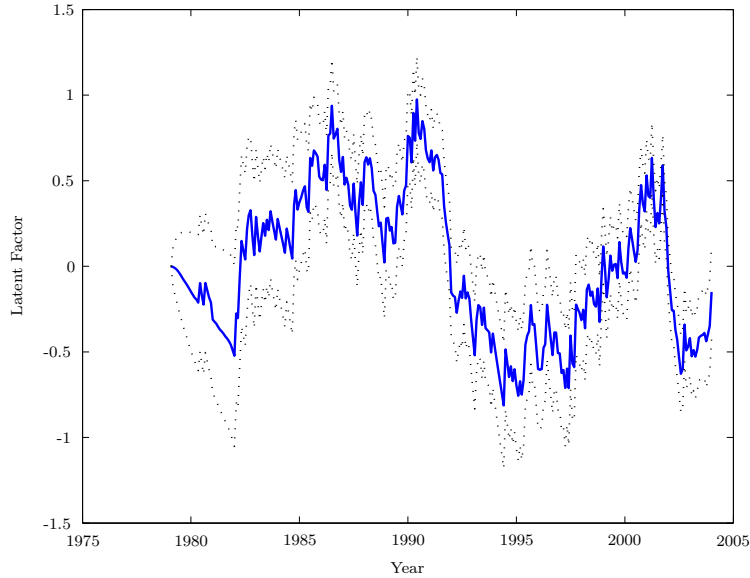


Figure 7: Posterior mean and 1-sigma error bands of the scaled latent Brownian motion frailty variable conditional on only past data, that $\{\eta E(Y_t | \mathcal{F}_t) : 0 \leq t \leq T\}$

$$p(y_t | \mathcal{F}_T) = p(y_t | \mathcal{F}_t) \bar{\alpha}_{t|T}(y_t),$$

beginning with $\bar{\alpha}_{T|T}(y_t) = 1$.

For the joint posterior distribution $p((y_0, y_1, \dots, y_T)' | \mathcal{F}_T)$ of the latent Brownian motion frailty variables, one may employ, for example, the Gibbs sampler described in Appendix A.

6 The Term-Structure of Defaults

We turn to the implications of frailty on the term structure of default risk for a given conditioning date t and currently active issuer i , defined at maturity time u by the hazard rate

$$h_i(t, u) = \frac{1}{p_i(t, u)} \frac{-\partial p_i(t, u)}{\partial u},$$

where (ignoring other exit effects, which are treated in Duffie, Saita, and Wang (2006))

$$p_i(t, u) = E \left(e^{-\int_t^u \lambda_{is} ds} \mid \mathcal{F}_t \right)$$

is the \mathcal{F}_t -conditional probability of survival from t to u . The hazard rate is the time- t conditional mean expected rate of default at time u , given the currently available information \mathcal{F}_t and given as well the event of survival up to time u .

As an illustration, we consider the term structure of default hazard rates of Xerox Corporation for three different models, (i) the basic model in which only observable covariates are considered, (ii) the model with the latent Brownian frailty variable, and (iii) the model with the common Brownian frailty variable as well as unobserved heterogeneity. Figure 8 shows the term structure of default rates for Xerox Corporation in December 2003, given the available information at that time.

7 Default Correlation

As noted before, in the model without frailty, firms' default times are correlated only as implied by the correlation of observable factors determining their default intensities. In this case, model-implied default correlations were found to be significantly lower than the empirically observed ones (DeServigny and Renault (2002), and Das, Duffie, Kapadia, and Saita (2006)). Common dependence on unobservable covariates, as in our model, introduces an additional channel of default correlation.

For a given conditioning date t and maturity date $u > t$, and for two given active firms i and j , the default correlation is the \mathcal{F}_t -conditional correlation between D_{iu} and D_{ju} . Figure 9 shows the effect of the latent Brownian motion frailty variable on the default correlation for two companies in our dataset. We see that the latent factor induces additional correlation and that the effect is increasing as the time horizon increases.

A common frailty also increases the likelihood of a large number of defaults. In order to isolate this effect, we considered a hypothetical portfolio consisting of the 1,813 companies in our data set that were active as of January 1998. We computed the posterior distribution, as of January 1998, of the total number of defaults during the subsequent five years, January

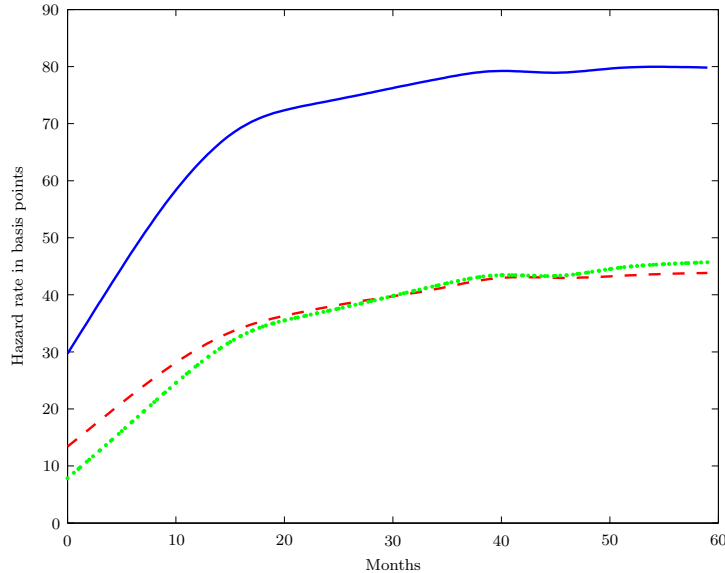


Figure 8: The term structure of hazard rates for Xerox Corporation in December 2003 for a) the model with frailty variable (solid line), b) the model without frailty variable (dashed line) and c) the model with frailty variable and unobserved heterogeneity (dotted line).

1998 to December 2002. Figure 10 shows the density of the total number of defaults in the portfolio for three different models, namely the fitted model with (i) common frailty variable, (ii) a hypothetical model that has the same coefficients but an independently evolving Brownian frailty variable for each company with the same initial value at the beginning time t , January 1998, drawn from the posterior distribution of Y_{t-1} given \mathcal{F}_t and (iii) a hypothetical model, again with the same coefficients, but completely independent frailty variables for each company.

The fatter right tail of the aggregate default distribution for the model with a common frailty variable reflects both the effect of correlation associated with future co-movements of default intensities through their exposure to the common frailty variable, as well as uncertainty regarding the current level of the frailty variable in January 1998.

In particular, we see in Figure 10 that the two hypothetical models that

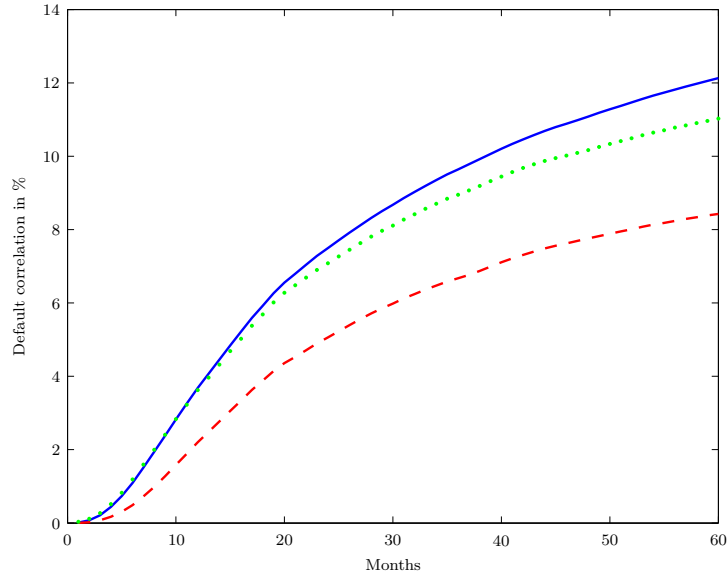


Figure 9: Default correlation of ICO, Incorporated and Xerox Corporation for a) the model with a common frailty variable (solid line), b) the model without a frailty variable (dashed line), and c) the model with frailty variable and unobserved heterogeneity (dotted line)

do not have a common frailty variable assign virtually no probability of more than 175 defaults occurring between 1998/01 and 2002/12. The 95- and 99-percentile for the model with completely independent frailty variable are 117 and 123 defaults, respectively. The model with independently evolving frailty variables with the same initial value in January 1998 has a 95- and 99-percentile of 147 and 167 defaults respectively. The actual number of defaults in our dataset during this time period was 195.

The 95- and 99-percentiles of the model with a common frailty variable are 189 and 245 defaults, respectively. The realized number of defaults during this event horizon, 195, is slightly below the 96-percentile of the distribution implied by the fitted frailty model, and therefore constituting a rather extreme event. On the other hand, taking the hindsight bias into account, in that our analysis was partially motivated by the high number of defaults in the years 2001 and 2002, the occurrence of 195 defaults might not be such

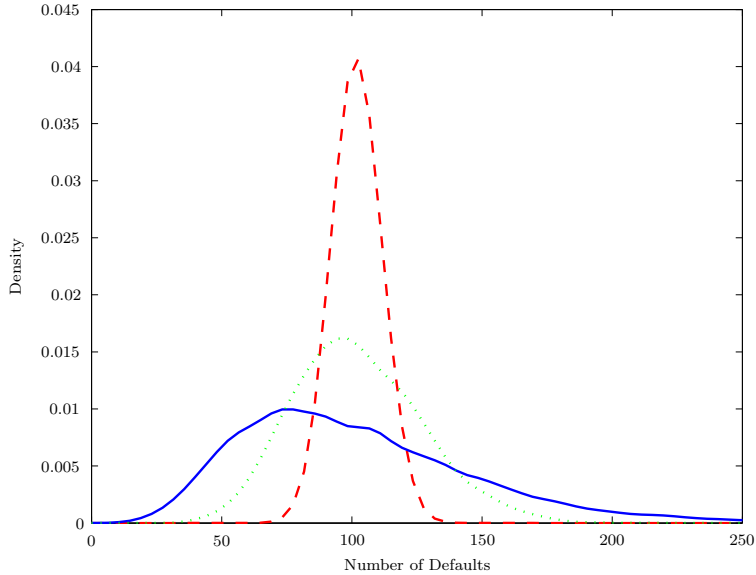


Figure 10: Density of total number of defaults in portfolio in model a) with frailty variable (solid line), b) with independent frailty variable for each company but same initial value drawn from $p(y_{t-1} | \mathcal{F}_t)$, (dashed line), and c) with completely independent frailty variable for each company (dotted line). The density estimates were obtained by applying a Gaussian kernel smoother (bandwidth equal to 5) to the empirical default distributions which in turn were generated by Monte Carlo simulation.

an extreme event.

8 Out-of-Sample Accuracy

Here we examine the out of sample accuracy ratios, computed as explained in Duffie, Saita, and Wang (2006). The overall accuracy is indeed approximately the same as that of Duffie, Saita, and Wang (2006). Accuracy ratios, however, measure only the ability to rank firms by default probability, and do not capture the out-of-sample ability to estimate the magnitudes of the default probabilities, which we will turn to in the next draft.

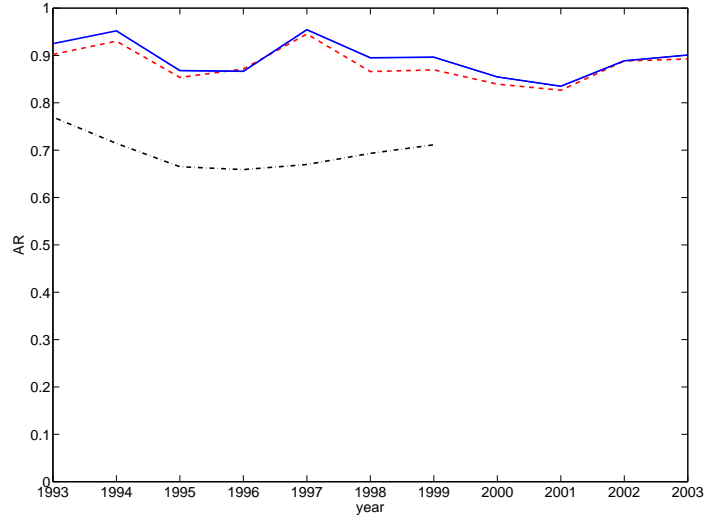


Figure 11: Out-of-sample accuracy ratios (ARs). The model is that estimated with data up to December 1992. The solid line provides one-year-ahead ARs based on the model without frailty. The dashed line shows one-year-ahead ARs for the model with frailty. The dash-dot line shows 5-year-ahead ARs for the model with frailty.

9 Conclusion

This paper finds significant evidence of the presence among U.S. corporates of one or more unobservable common source of default risk, that increases default correlation and extreme portfolio loss risk above and beyond that implied by observable common and correlated macroeconomic and firm-specific sources of default risk. We offer a new model of corporate default intensities in the presence of a time-varying latent “frailty” factor, and with unobserved heterogeneity. We provide a method for fitting the model parameters using a combination of the Monte Carlo EM algorithm and the Gibbs sampler. This also provides the conditional posterior distribution of the Brownian motion frailty variable as a by-product.

Applying this model to data for U.S. firms between January 1979 and March 2004, we find that the level of corporate default rates varies over time well beyond what can be explained by a model that only includes observable

covariates. In particular, the posterior distribution of the frailty variable shows that the rate of corporate defaults was much higher in 1989-1990 and 2001-2002, and much lower during the mid-nineties and in 2003-2004, than those implied by an analogous model without frailty. Moreover, the historically observed number of defaults in our dataset between January 1998 and December 2002 lies above the 99.9-percentile of the aggregate default distribution associated with the model based on observable covariates only, but lies well within the support of the distribution of the total defaults produced by the frailty-based model.

Our methodology could be applied to other situations in which a common unobservable factor is suspected to play an important role in the time-variation of arrivals for certain events, for example mergers and acquisitions, mortgage prepayments and defaults, or leveraged buyouts.

Das, Duffie, Kapadia, and Saita (2006) developed a test that rejected the joint hypothesis of correctly specified default probabilities and the doubly stochastic assumption that defaults are independent conditional on the paths of observable risk factors. We plan to extend that test to this setting, in order to test whether the default clustering in the data can be captured by the frailty effect.

Our results suggests both significant shifts in individual firm default probabilities associated with shifts in the posterior distribution of the frailty factor, as well as significant increases in the likelihood of default clustering.

We estimate that the common frailty variable represents a common unobservable factor in default intensities with an annual volatility of roughly 40%. We are currently investigating the implications of mean reversion of this common factor. In the setting of an Ornstein-Uhlenbeck extension of the frailty model, our preliminary results suggest an estimated annual mean reversion rate of roughly 20%, which means that when defaults cluster in time above and beyond what is suggested by observable default-risk factors, the half life of the impact on the unobservable common default intensity factor is roughly 3 years. Unfortunately, the data do not appear to be sufficiently rich to pin down the mean reversion rate well. While difficult to pin down, mean reversion is nevertheless likely to be an important feature of a model of the type that we suggest, given the unconditionally explosive nature of Brownian motion without mean reversion. While the posterior distribution of the Brownian frailty factor without mean reversion is kept well under control by the mere conditioning effect of the default data, we are finding preliminary evidence in the absence of mean reversion of over-shooting of

portfolio-average default-rate forecasts, out of sample.

Appendices

A Applying the Gibbs Sampler with Frailty

A central quantity of interest for describing and estimating the historical default dynamics is the posterior density $p_Y(\cdot | W, D, \theta)$ of the latent frailty process Y . This is a complicated high-dimensional density. It is prohibitively computationally intensive to directly generate samples from this distribution. Nevertheless, Markov Chain Monte Carlo (MCMC) methods can be used for exploring this posterior distribution by generating a Markov Chain over Y , denoted $\{Y^{(n)}\}_{n \geq 1}^N$, whose equilibrium density is $p_Y(\cdot | W, D, \theta)$. Samples from the joint posterior distribution can then be used for parameter inference and for analyzing the properties of the frailty process Y . For a function $f(\cdot)$ satisfying regularity conditions, the Monte Carlo estimate of

$$E[f(Y) | W, D, \theta] = \int f(y) p_Y(y | W, D, \theta) dy \quad (14)$$

is given by

$$\frac{1}{N} \sum_{n=1}^N f(Y^{(n)}). \quad (15)$$

Under conditions, the ergodic theorem for Markov chains guarantees the convergence of this sum to its expectation as $N \rightarrow \infty$. One such function that will be of interest to us is the identity, $f(y) = y$, so that

$$E[f(Y) | W, D, \theta] = E[Y | W, D, \theta] = \{E(Y_t | \mathcal{F}_T) : 0 \leq t \leq T\},$$

the posterior mean of the latent Brownian motion frailty process.

The linchpin to MCMC is that the joint distribution of the frailty path $Y = \{Y_t : 0 \leq t \leq T\}$ can be broken into a set of conditional distributions. A general version of the Clifford-Hammersley (CH) Theorem (Hammersley and Clifford (1970) and Besag (1974)) provides conditions under which a set of

conditional distributions characterizes a unique joint distribution. For example, in our setting, the CH Theorem indicates that the density $p_Y(\cdot | W, D, \theta)$ is uniquely determined by the following set of conditional distributions,

$$\begin{aligned} & Y_0 | Y_1, Y_2, \dots, Y_T, W, D, \theta \\ & Y_1 | Y_0, Y_2, \dots, Y_T, W, D, \theta \\ & \vdots \\ & Y_T | Y_0, Y_1, \dots, Y_{T-1}, W, D, \theta \end{aligned}$$

using the naturally suggested interpretation of this informal notation. We refer the interested reader to Robert and Casella (2005) for an extensive treatment of Monte Carlo methods, as well as Johannes and Polson (2003) for an overview of MCMC methods applied to problems in financial economics.

In our case, the conditional distribution of Y_t at a single point in time t , given the observable variables (W, D) and given Y_s for all $s \neq t$ in some discrete set $Y^{(-t)}$ of times, is somewhat tractable, as shown below. This allows us to use the Gibbs sampler (Geman and Geman (1984) or Gelman, Carlin, Stern, and Rubin (2004)) to draw whole sample paths from the posterior distribution of $\{Y_t : 0 \leq t \leq T\}$, given the default and covariate data and the parameter vector θ .

First, we derive the conditional distribution of Y_t given (W, D) and given $Y^{(-t)} = \{Y_s : s \neq t\}$, which by the Markov property of Y is the same as the conditional distribution of Y given (W, D) , Y_{t-1} , and Y_{t+1} . Recall that $\mathcal{L}(\theta | W, Y, D)$ denotes the complete-information likelihood of the observed default pattern, where $\theta = (\beta, \eta)$. For $0 < t < T$, Bayes' rule implies that

$$\begin{aligned} p(Y_t | W, D, Y^{(-t)}, \theta) &\propto \mathcal{L}(\theta | W, D, Y^{(-t)}) \cdot p(Y_t | Y^{(-t)}, \theta) \\ &= \mathcal{L}(\theta | W, Y, D) \cdot p(Y_t | Y_{t-1}, \theta) \cdot p(Y_t | Y_{t+1}, \theta). \end{aligned} \tag{16}$$

Combining (4) and (16),

$$\begin{aligned} & \log p(Y_t | W, D, Y^{(-t)}, \theta) \\ &= C_0 + \log \mathcal{L}(\theta | W, Y, D) + \log p(Y_t | Y_{t-1}, \eta) + \log p(Y_t | Y_{t+1}, \eta) \\ &= - \sum_{i=1}^m \lambda_{it} \Delta t + \sum_{i=1}^m \log(\lambda_{it}) D_{it} - \frac{1}{2\eta^2} (Y_t - Y_{t-1})^2 - \frac{1}{2\eta^2} (Y_{t+1} - Y_t)^2 + C_1 \\ &= - \sum_{i=1}^m \tilde{\lambda}_{it} e^{\eta Y_t} \Delta t + \sum_{i=1}^m \log(\tilde{\lambda}_{it}) D_{it} + \eta Y_t \sum_{i=1}^m D_{it} \end{aligned}$$

$$-\frac{1}{2\eta^2} (Y_t - Y_{t-1})^2 - \frac{1}{2\eta^2} (Y_{t+1} - Y_t)^2 + C_2,$$

where C_0 , C_1 , and C_2 are constants. Hence, for $0 < t < T$,

$$\begin{aligned} \log p(Y_t | W, D, Y^{(-t)}, \theta) \\ = c_o + c_1 \cdot ((Y_t - Y_{t-1})^2 + (Y_{t+1} - Y_t)^2) + c_2 \cdot e^{\eta Y_t} + c_3 \cdot Y_t, \end{aligned} \quad (17)$$

where the constants c_0, \dots, c_3 depend on the default and covariate data, but do not depend on the latent factor at any point in time.

Equation (17) determines the conditional density of Y_t given Y_{t-1} and Y_{t+1} in an implicit form. In order to draw numbers from this conditional distribution, we discretize the space of possible outcomes of Y , allowing Y_t to take values in a some finite set $\{y_1, \dots, y_J\}$.⁸ By defining

$$q_j = \exp(c_0 + c_1 \cdot ((y_j - Y_{t-1})^2 + (Y_{t+1} - y_j)^2) + c_2 \cdot e^{\eta y_j} + c_3 \cdot y_j),$$

we adopt an approximation of the posterior distribution of Y_t of the form

$$P(Y_t = y_j | Y_{t-1}, Y_{t+1}, W, D) \simeq \frac{q_j}{q_1 + \dots + q_J}. \quad (18)$$

The Gibbs sampler for drawing paths from the posterior distribution of $\{Y_t : 0 \leq t \leq T\}$ is then given by the algorithm:

0. Initialize $Y_t = 0$ for $t = 0, \dots, T$.
1. For $t \in \{1, \dots, T\}$, draw a new value of Y_t from its conditional distribution⁹ given Y_{t-1} and Y_{t+1} .
2. Store the sample path $\{Y_t : 0 \leq t \leq T\}$ and return to Step 1 until the desired number of paths has been simulated.

⁸As an alternative to discretizing the state space, known as the Griddy Gibbs method (Tanner (1998)), one can use the Metropolis-Hastings algorithm (see Hasting (1970) or Gelman, Carlin, Stern, and Rubin (2004)) to sample from the conditional distribution of Y_t given Y_{t-1} and Y_{t+1} .

⁹The formula (17) applies only for $0 < t < T$. For the two end points, modifications are needed. For $t = T$, it is easy to derive the transition density by using arguments similar to those for the case $0 < t < T$. Finally, we take $Y_0 = 0$ as the starting value of the latent frailty variable.

We usually discard the first several hundred paths as a “burn-in” sample, because initially the Gibbs sampler has not approximately converged¹⁰ to the posterior distribution of $\{Y_t : 0 \leq t \leq T\}$.

In our case, we used $J = 321$ states for the discretized frailty variable Y_t , which should give a reasonably good approximation of a Brownian motion.¹¹ We have also tried using a much higher number of hidden states with little visible improvement. Using a smaller number of states caused some of the graphs of output paper to show artifactual effects of the discretization.

B Gibbs and Unobserved Heterogeneity

The Gibbs sampler for drawing paths from the joint posterior distribution of $\{Y_t : 0 \leq t \leq T\}$ and $\{Z_i : 1 \leq i \leq m\}$ works as follows:

0. Initialize $Y_t = 0$ for $t = 0, \dots, T$. Initialize $Z_i = 1$ for $i = 1, \dots, m$.
1. For $t = 1, \dots, T$ draw a new value of Y_t from its conditional distribution given Y_{t-1}, Y_{t+1} and the current values for Z_i . This can be done using a straightforward modification of the Gibbs sampler described in Appendix A by treating $\log Z_i$ as an additional covariate with corresponding coefficient in (1) equal to 1.
2. For $i = 1, \dots, m$, draw the unobserved heterogeneity variables Z_1, \dots, Z_m from their conditional distributions given the current path of Y . See below.
3. Store the sample path $\{Y_t, 0 \leq t \leq T\}$ and the variables $\{Z_i : 1 \leq i \leq m\}$. Return to Step 1 and repeat until the desired number of scenarios has been drawn, discarding the first several hundred as a burn-in sample.

¹⁰We used various convergence diagnostics, such as trace plots of a given parameter as a function of the number of samples drawn, to assure that the iterations have proceeded long enough for approximate convergence and to assure that our results do not depend markedly on the starting values of the Gibbs sampler. See Gelman, Carlin, Stern, and Rubin (2004), Chapter 11.6, for a discussion of various methods for assessing convergence of MCMC methods.

¹¹A model with discrete state space is an internally consistent model on its own, and need not be viewed as an approximation to the case of Brownian motion. Due to computational limitations, such regime-switching models as in Hamilton (1989) have incorporated only a small number of states, typically two or three, for unobserved variables.

It remains to show how to draw the heterogeneity variables Z_1, \dots, Z_m from their posterior conditional distribution. First, we note that

$$p(Z | W, Y, D, \theta) = \prod_{i=1}^m p(Z_i | W_i, Y, D_i, \theta),$$

by conditional independence of the unobserved heterogeneity variables Z_i . Because we have chosen these to be gamma distributed with mean 1 and standard deviation 0.5, the density parameters a and b are both 4. Applying Bayes' rule,

$$\begin{aligned} p(Z_i | W, Y, D, \theta) &\propto p_{\Gamma}(Z_i; 4, 4) \mathcal{L}(\theta | W_i, Y, Z_i, D_i) \\ &\propto Z_i^3 e^{-4Z_i} e^{-\sum_{t=t_i}^{T_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{T_i} [D_{it} \lambda_{it} \Delta t + (1 - D_{it})], \end{aligned} \quad (19)$$

where $p_{\Gamma}(Z_i; a, b)$ denotes the density function of a Gamma distribution with parameters a and b . Plugging (5) into (19) gives

$$\begin{aligned} p(Z_i | W, Y, D, \theta) &\propto Z_i^3 e^{-4Z_i} \exp\left(-\sum_{t=t_i}^{T_i} \tilde{\lambda}_{it} e^{\gamma Y_t} Z_i\right) \prod_{t=t_i}^{T_i} [D_{it} \lambda_{it} + (1 - D_{it})] \\ &= Z_i^3 e^{-4Z_i} \exp(-A_i Z_i) \cdot \left\{ \begin{array}{ll} B_i Z_i & \text{if company } i \text{ did default} \\ 1 & \text{if company } i \text{ did not default} \end{array} \right\}, \end{aligned} \quad (20)$$

for company specific constants A_i and B_i . The factors in (20) can be combined to give

$$p(Z_i | W_i, Y, D_i, \theta) = \Gamma(Z_i; 3 + D_{i,T_i}, 4 + A_i). \quad (21)$$

This is again a Gamma distribution, but with different parameters, and it is therefore easy to draw samples of Z_i from its conditional distribution.

References

- Altman, E. I. (1968). Financial Ratios, Discriminant Analysis, and the Prediction Of Corporate Bankruptcy. *Journal of Finance* 23, 589–609.
- Baum, L. E., T. P. Petrie, G. Soules, and N. Weiss (1970). A maximization technique occurring In The statistical analysis Of probabilistic functions Of Markov chains. *The Annals of Mathematical Statistics* 41, 164–171.
- Beaver, B. (1968). Market Prices, Financial Ratios, and the Prediction Of Failure. *Journal of Accounting Research Autumn*, 170–192.
- Besag, J. (1974). Spatial Interaction and The Statistical Analysis Of Lattice Systems. *Journal of the Royal Statistical Association: Series B* 36, 192–236.
- Black, F. and M. Scholes (1973). The Pricing Of Options and Corporate Liabilities. *Journal of Political Economy* 81, 637–654.
- Brémaud, P. (1981). *Point Processes and Queues. Martingale Dynamics*. Springer Series in Statistics, New-York.
- Cappé, O., E. Moulines, and T. Rydén (2005). *Inference in Hidden Markov Models*. Springer Series in Statistics.
- Celeux, G. and J. Diebolt (1986). The SEM Algorithm: A Probabilistic Teacher Algorithm Derived From The EM Algorithm For The Mixture Problem. *Computational Statistics Quarterly* 2, 73–82.
- Chava, S. and R. Jarrow (2004). Bankruptcy Prediction with Industry Effects. *Review of Finance* 8, 537–569.
- Collin-Dufresne, P., R. Goldstein, and J. Helwege (2003). Is Credit Event Risk Priced? Modeling Contagion via The Updating Of Beliefs. Working Paper, Haas School, University of California, Berkeley.
- Collin-Dufresne, P., R. Goldstein, and J. Hugonnier (2004). A General Formula For Valuing Defaultable Securities. *Econometrica* 72, 1377–1407.
- Couderc, F. and O. Renault (2004). Times-to-Default: Life Cycle, Global and Industry Cycle Impacts. Working paper, University of Geneva.
- Crosbie, P. J. and J. R. Bohn (2002). Modeling Default Risk. Technical Report, KMV, LLC.

- Das, S. S., D. Duffie, N. Kapadia, and L. Saita (2006). Common Failings: How Corporate Defaults are Correlated. *Journal of Finance*. forthcoming.
- Delloy, M., J.-D. Fermanian, and M. Sbai (2005). Estimation Of a reduced-form credit portfolio model and extensions To dynamic frailties. Preliminary Version, BNP-Paribas.
- Dempster, A., N. Laird, and D. Rubin (1977). Maximum Likelihood Estimation From Incomplete Data via The EM Algorithm (with Discussion). *Journal of the Royal Statistical Society: Series B* 39, 1–38.
- DeServigny, A. and O. Renault (2002). Default Correlation: Empirical Evidence. Working Paper, Standard and Poor's.
- Duffie, D. (2001). *Dynamic Asset Pricing Theory* (Third edition). Princeton, New Jersey: Princeton University Press.
- Duffie, D. and D. Lando (2001). Term Structures Of Credit Spreads with Incomplete Accounting Information. *Econometrica* 69, 633–664.
- Duffie, D., L. Saita, and K. Wang (2006). Multi-Period Corporate Default Prediction with Stochastic Covariates. *Journal of Financial Economics* forthcoming.
- Eraker, B., M. Johannes, and N. Polson (2003). The Impact Of Jumps in Volatility and Returns. *Journal of Finance* 58, 1269–1300.
- Fisher, E., R. Heinkel, and J. Zechner (1989). Dynamic Capital Structure Choice: Theory and Tests. *Journal of Finance* 44, 19–40.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004). *Bayesian Data Analysis* (Second edition). Chapman and Hall.
- Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions and The Bayesian restoration Of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721–741.
- Giesecke, K. (2004). Correlated Default With Incomplete Information. *Journal of Banking and Finance* 28, 1521–1545.
- Gordy, M. (2003). A Risk-Factor Model Foundation For Ratings-Based Capital Rules. *Journal of Financial Intermediation* 12, 199–232.
- Hamilton, J. D. (1989). A New Approach To The Economic Analysis Of Nonstationary Time Series and The Business Cycle. *Econometrica* 57, 357–384.

- Hammersley, J. and P. Clifford (1970). Markov Fields On Finite Graphs and Lattices. Unpublished Manuscript.
- Hasting, W. (1970). Monte-Carlo Sampling Methods using Markov Chains and Their Applications. *Biometrika* 57, 97–109.
- Hillegeist, S. A., E. K. Keating, D. P. Cram, and K. G. Lundstedt (2004). Assessing the Probability Of Bankruptcy. *Review of Accounting Studies* 9, 5–34.
- Jegadeesh, N. and S. Titman (1993). Returns To Buying Winners and Selling Losers: Implications For Stock Market Efficiency. *Journal of Finance* 48, 65–91.
- Jegadeesh, N. and S. Titman (2001). Profitability Of Momentum Strategies: An Evaluation Of Alternative Explanations. *Journal of Finance* 66, 699–720.
- Johannes, M. and N. Polson (2003). MCMC Methods For Continuous-Time Financial Econometrics.
- Kass, R. and A. Raftery (1995). Bayes factors. *Journal of The American Statistical Association* 90, 773–795.
- Kavvathas, D. (2001). Estimating Credit Rating Transition Probabilities for Corporate Bonds. Working paper, University of Chicago.
- Kealhofer, S. (2003). Quantifying Credit Risk I: Default Prediction. *Financial Analysts Journal*, January–February, 30–44.
- Lane, W. R., S. W. Looney, and J. W. Wansley (1986). An Application Of the Cox Proportional Hazards Model to Bank Failure. *Journal of Banking and Finance* 10, 511–531.
- Lee, S. H. and J. L. Urrutia (1996). Analysis and Prediction Of Insolvency in the Property-Liability Insurance Industry: A Comparison Of Logit and Hazard Models. *The Journal of Risk and Insurance* 63, 121–130.
- Leland, H. (1994). Corporate Debt Value, Bond Covenants, and Optimal Capital Structure. *Journal of Finance* 49, 1213–1252.
- McDonald, C. G. and L. M. Van de Gucht (1999). High-Yield Bond Default and Call Risks. *Review of Economics and Statistics* 81, 409–419.
- Merton, R. C. (1974). On the Pricing Of Corporate Debt: The Risk Structure Of Interest Rates. *Journal of Finance* 29, 449–470.

- Nielsen, S. F. (2000). The Stochastic EM algorithm: Estimation and Asymptotic Results. Department of Theoretical Statistics, University of Copenhagen.
- Protter, P. (2004). *Stochastic Integration and Differential Equations, Second Edition*. New York: Springer-Verlag.
- Rabiner, L. R. (1989). A tutorial On hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77, 257–285.
- Robert, C. and G. Casella (2005). *Monte Carlo Statistical Methods* (Second edition). Springer Texts in Statistics.
- Schönbucher, P. (2003). Information Driven Default Contagion. Working Paper, ETH, Zurich.
- Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business* 74, 101–124.
- Smith, A. F. M. and D. J. Spiegelhalter (1980). Bayes Factors and Choice Criteria For Linear Models. *Journal of the Royal Statistical Society: Series B* 42, 213–220.
- Tanner, M. A. (1998). *Tools For Statistical Inference : Methods For The Exploration Of Posterior Distributions and Likelihood Functions* (Third edition). New York: Springer-Verlag.
- Vasicek, O. (1987). Probability Of Loss On Loan Portfolio. Working Paper, KMV Corporation.
- Vassalou, M. and Y. Xing (2004). Default Risk in Equity Returns. *Journal of Finance* 59, 831–868.
- Wei, G. and M. Tanner (1990). A Monte Carlo Implementation Of The EM Algorithm and The Poor Man’s Data Augmentation Algorithm. *Journal of The American Statistical Association* 85, 699–704.
- Wu, C. (1983). On The convergence properties Of The EM algorithm. *Annals of Statistics* 11, 95–103.
- Yu, F. (2005). Accounting Transparency and The Term Structure Of Credit Spreads. *Journal of Financial Economics* 75, 53–84.
- Zhang, G. (2004). Intra-Industry Credit Contagion: Evidence From The Credit Default Swap Market and The Stock Market. Working paper, University of California, Irvine.