Frailty Correlated Default

DARRELL DUFFIE, ANDREAS ECKNER, GUILLAUME HOREL,
and LEANDRO SAITA*

ABSTRACT

The probability of extreme default losses on portfolios of U.S. corporate debt is much greater than would be estimated under the standard assumption that default correlation arises only from exposure to observable risk factors. At the high confidence levels at which bank loan portfolio and collateralized debt obligation (CDO) default losses are typically measured for economic capital and rating purposes, conventionally based loss estimates are downward biased by a full order of magnitude on test portfolios. Our estimates are based on U.S. public nonfinancial firms between 1979 and 2004. We find strong evidence for the presence of common latent factors, even when controlling for observable factors that provide the most accurate available model of firm-by-firm default probabilities.

THIS PAPER PROVIDES a more realistic assessment of the risk of large default losses on portfolios of U.S. corporate debt than had been available with prior methodologies. At the high confidence levels at which portfolio default losses are typically estimated for meeting bank capital requirements and rating collateralized debt obligations (CDOs), our empirical results indicate that conventional estimators are downward biased by a full order of magnitude on typical test portfolios. Our estimates are based on portfolios of U.S. corporate debt existing between 1979 and 2004. For estimating high-quantile portfolio losses, conventional methodologies suffer from their failure to correct for a significant downward omitted variable bias. We find strong evidence that firms are exposed to a common dynamic latent factor driving default, even after controlling for observable factors that on their own provide the most accurate available model of firm-by-firm default probabilities. Both uncertainty about the current level of this variable, as well as joint exposure to future movements of this variable, cause a substantial increase in the conditional probability of large portfolio default losses.

A conventional portfolio loss risk model assumes that borrower-level conditional default probabilities depend on measured firm-specific or marketwide factors. Portfolio loss distributions are typically based on the correlating

*Duffie is at the Graduate School of Business, Stanford University. Eckner and Horel are at Bank of America. Saita is at Barclays Capital. We are grateful for financial support from Moody's Corporation and Morgan Stanley; for data from Moody's and Ed Altman; for research assistance from Sabri Oncu and Vineet Bhagwat; for remarks from Torben Andersen, André Lucas, Richard Cantor, Stav Gaon, Tyler Shumway, Jun S. Liu, Xiao-Li Meng, and especially Michael Johannes; and for suggestions by a referee, an associate editor, and the editor, Campbell Harvey.

influence of such observable factors. For example, rating agencies typically estimate the probability of losses to senior tranches of CDOs, which are intended to occur only when the underlying portfolio losses exceed a high confidence level, by relying on the observable credit ratings of the underlying collateral debt instruments. Modeled co-movement of the ratings of the borrowers represented in the collateral pool is intended to capture default correlation and the tails of the total loss distribution. However, if the underlying borrowers are commonly exposed to important risk factors whose effect is not captured by co-movements of borrower ratings, then the portfolio loss distribution will be poorly estimated. This is not merely an issue of estimation noise; a failure to include risk factors that commonly increase and decrease borrowers' default probabilities will result in a downward biased estimate of tail losses. For instance, in order to receive a triple-A rating, a CDO is typically required to sustain little or no losses at a confidence level such as 99%. Although any model of corporate debt portfolio losses cannot accurately measure such extreme quantiles with the limited available historical data, our model of tail losses avoids a large, downward omitted variable bias, and survives goodness-of-fit tests associated with large portfolio losses.

Whenever it is possible to identify and measure new significant risk factors, they should be included in the model. We do not claim to have identified and included all relevant observable risk factors. Although our observable risk factors include firm-level and macroeconomic variables leading to higher accuracy ratios for out-of-sample default prediction than those offered by any other published model, further research will undoubtedly uncover new significant observable risk factors that should be included. We discuss some proposed inclusions later in this paper.[1] It is inevitable, however, that not all relevant risk factors that are potentially observable by the econometrician will end up being included. There is also a potential for important risk factors that are simply not observable. A downward bias in tail-loss estimates is thus inevitable without some form of bias correction. Our approach is to directly allow for unobserved risk factors whose time-series behavior and whose posterior conditional distribution can both be estimated from the available data by maximum likelihood estimation.

For example, subprime mortgage debt portfolios recently suffered losses in excess of the high confidence levels that were estimated by rating agencies. The losses associated with this debacle that have been reported by financial institutions total approximately $800 billion as of this writing, and are still accumulating. An example of an important factor that was not included in most mortgage portfolio default loss models is the degree to which borrowers and mortgage brokers provided proper documentation of borrowers' credit qualities. With hindsight, more teams responsible for designing, rating, intermediating, and investing in subprime CDOs might have done better by allowing for the possibility that the difference between actual and documented credit qualities would turn out to be much higher than expected, or much lower than expected,

[1] New work by Lando and Nielsen (2008) suggests additional helpful covariates.

in a manner that is correlated across the pool of borrowers. Incorporating this additional source of uncertainty would have resulted in higher prices for CDO "first-loss" equity tranches (a convexity effect). Senior CDOs would have been designed with more conservative over-collateralization, or alternatively would have had lower ratings and lower prices (a concavity effect), on top of any related risk premia effects. Accordingly, more modelers could have improved their models by adding proxies for this moral hazard effect. It seems optimistic to believe that they were prepared to do so, however, for despite the clear incentives, many apparently did not. This suggests that it is not easy, ex ante, to include all important default covariates, and further, that the next event of extreme portfolio loss could be based on a different omitted variable. It therefore seems prudent, going forward, to allow for missing default covariates when estimating tail losses on debt portfolios.

As a motivating instance of missing risk factors in the corporate debt arena on which we focus, the defaults of Enron and WorldCom may have revealed faulty accounting practices that could have been in use at other firms, and thus may have had an impact on the conditional default probabilities of other firms, and therefore on portfolio losses. The basic idea of our methodology is an application of Bayes's Rule to update the posterior distribution of unobserved risk factors whenever defaults arrive with a timing that is more or less clustered than would be expected based on the observable risk factors alone. In the statistics literature on event forecasting, the effect of such an unobserved covariate is called "frailty." In the prior statistics literature, frailty covariates are assumed to be static. It would be unreasonable to assume that latent risk factors influencing corporate default are static over our 25-year data period, so we extend the prior statistical methodology so as to allow a frailty covariate to vary over time according to an autoregressive time-series specification. We use Markov Chain Monte Carlo (MCMC) methods to perform maximum likelihood estimation and to filter for the conditional distribution of the frailty process.

While our empirical results address the arrival of default events, our methodology can be applied in other settings. Recently, for instance, Chernobai, Jorion, and Yu (2008) adopted our methodology to estimate a model of operational risk events. Our model could also be used to treat the implications of missing covariates for mortgage pre-payments, employment events, mergers and acquisitions, and other event-based settings in which there are time-varying latent variables.

The remainder of the paper is organized as follows. Section I gives an overview of our modeling approach and results. Section II places our work in the context of the related literature and clarifies our incremental contribution. Section III specifies the precise probabilistic model for the joint distribution of default times. Section IV describes our data sources, provides the fitted model, and summarizes some of the implications of the fitted model for the distribution of losses on portfolios of U.S. corporate debt. Section V examines the fit of the model and addresses some potential sources of misspecification, providing robustness checks. Section VI concludes. Appendices provide some key technical information, including our estimation methodology, which is based on a combination of the Monte Carlo expectations maximization (EM) algorithm and the Gibbs sampler.

I. Modeling Approach and Results

In order to further motivate our approach and summarize our main empirical results, we briefly outline our specification here, and later provide details. Our objective is to estimate the probability distribution of the number of defaults among m given firms over any prediction horizon. For a given firm i, our model includes a vector U_{it} of observable default prediction covariates that are specific to firm i. These variables include the firm's "distance to default," a widely followed volatility-corrected leverage measure whose construction is reviewed later in this paper, as well as the firm's trailing stock return, an important auxiliary covariate suggested by Shumway (2001). Allowing for unobserved heterogeneity, we include an unobservable firm-specific covariate Z_i. We also include a vector V_t of observable macroeconomic covariates, including interest rates and marketwide stock returns. In robustness checks, we explore alternative and additional choices for observable macro-covariates. Finally, we include an unobservable macroeconomic covariate Y_t whose "frailty" influence on portfolio default losses is our main focus.

If all of these covariates were observable, our model specification would imply that the conditional mean arrival rate of default of firm i at time t is

$$\lambda_{it} = \exp(a + b \cdot V_t + c \cdot U_{it} + Y_t + Z_i),$$

for coefficients a, b, and c to be estimated. If all covariates were observable, this would be a standard proportional hazards specification. The conditional mean arrival rate λ_{it} is also known as a default intensity. For example, a constant annual intensity of 0.01 means Poisson default arrival with an annual probability of default of $1 - e^{-0.01} \simeq 0.01$.

Because Y_t and Z_i are not observable, their posterior probability distributions are estimated from the available information set $\mathcal{F}_t$, which includes the prior history of the observable covariates $\{(U_s, V_s) : s \leq t\}$, where $U_t = (U_{1t}, \ldots, U_{mt})$, and also includes previous observations of the periods of survival and times of defaults of all m firms.

Because public firm defaults are relatively rare, we rely on 25 years of data. We include all 2,793 U.S. public nonfinancial firms for which we were able to obtain matching data from the several data sets on which we rely. Our data, described in Section IV.A, cover over 400,000 firm-months. We specify an autoregressive Gaussian time-series model for (U_t, V_t, Y_t) that will be detailed later. Because Y_t is unobservable, we find that it is relatively difficult to tie down its mean reversion rate with the available data, but the data do indicate that Y has substantial time-series volatility, increasing the proportional volatility of λ_{it} by about 40% above and beyond that induced by time-series variation in U_{it} and V_t.

Our main focus is the conditional probability distribution of portfolio default losses given the information actually available at a given time. For example, consider the portfolio of the 1,813 firms from our data set that were active at the beginning of 1998. For this portfolio, we estimate the probability distribution of the total number of defaulting firms over the subsequent 5 years. This distribution can be calculated from our estimates of the default intensity coefficients α, β, and γ; our estimates of the time-series parameters governing the joint dynamics of (U_t, V_t, Y_t); and the estimated posterior distribution of Y_t and $Z_1, \ldots, Z_m$ given the information $\mathcal{F}_t$ available at the beginning of this 5-year period. The detailed estimation methodology is provided later in the paper. The 95th and 99th percentiles of the estimated distribution are 216 and 265 defaults, respectively. The actual number of defaults during this period turned out to be 195, slightly below the 91% confidence level of the estimated distribution. With hindsight, we know that 2001 to 2002 was a period of particularly severe corporate defaults. In Section IV, we show that a failure to allow for a frailty effect would have resulted in a severe downward bias of the tail quantiles of the portfolio loss distribution, to the point that one would have incorrectly assigned negligible probability to the event that the number of defaults actually realized would have been reached or exceeded.

As a robustness check, we provide a Bayesian analysis of the effect of a joint prior distribution for the mean reversion rate and volatility of Y_t on the posterior distribution of these parameters and on the posterior distribution of portfolio default losses. We find that this parameter uncertainty causes additional "fattening" of the tail of the portfolio loss distribution, notably at extreme quantiles.

More generally, we provide tests of the fit of frailty-based tail quantiles that support our model specification against the alternative of a no-frailty model. We show that there are two important potential channels for the effect of the frailty variable on portfolio loss distributions. First, as with an observable macrovariable, the frailty covariate causes common upward and downward adjustments of firm-level conditional default intensities over time. This causes large portfolio losses to be more likely than would be the case with a model that does not include this additional source of default intensity covariation. Second, because the frailty covariate is not observable, uncertainty about the current level of Y_t at the beginning of the forecast period is an additional source of correlation across firms of the events of future defaults. This second effect on the portfolio loss distribution would be important even if there were certain to be no future changes in this frailty covariate. In an illustrative example, we show that these two channels of influence of the frailty process Y have comparably large impacts on the estimated tail quantiles of the portfolio loss distribution.

After controlling for observable covariates, we find that defaults were persistently higher than expected during lengthy periods of time, for example, 1986 to 1991, and persistently lower in others, for example, during the mid-1990s. From trough to peak, the estimated impact of the frailty covariate Y_t on the average default rate of U.S. corporations during 1980–2004 is roughly a factor of two or more. As a robustness check, and as an example of the impact on the magnitude of the frailty effect of adding an observable factor, we reestimate the model including as an additional observable macro-covariate the trailing average realized rate of default,[2] which could proxy for an important factor that had been omitted from the base-case model. We show that this trailing default rate covariate is statistically significant, but that there remains an important role for frailty in capturing the tails of portfolio loss distributions.

II. 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. An estimate of this default covariate, using market equity data and accounting data for liabilities, has been adopted in industry practice by Moody's Analytics, a leading provider of estimates of default probabilities for essentially all publicly traded firms.[3] Based on this theoretical foundation, we include distance to default as a covariate.

In the context of a standard structural default model of this type, 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 on other covariates, both firm-specific and macroeconomic, that may reveal additional information about the firm's condition. If, across firms, there is correlation in the observation noises of the various firms' distances to default, then there is frailty. For reasons of tractability, we have chosen a reduced-form specification of frailty.

Altman (1968) and Beaver (1968) are among the first to estimate reduced-form statistical models of the likelihood of default of a firm within one accounting period, using accounting data.[4] Although the voluminous subsequent empirical literature addressing the statistical modeling of default probabilities has typically not allowed for unobserved covariates affecting default probabilities, the topic of hidden sources of default correlation has recently received some attention. Collin-Dufresne, Goldstein, and Helwege (2003) and Zhang and Jorion

[2] We are grateful to a referee for suggesting this.

[3] See Crosbie and Bohn (2002) and Kealhofer (2003).

[4] 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 et al. (2004) predict bankruptcy. Shumway (2001) uses a discrete duration model with time-dependent covariates. Duffie, Saita, and Wang (2007) 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.

(2007) find that a major credit event at one firm is 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 Huggonier (2004), Giesecke (2004), and Schönbucher (2003) explore learning from default interpretations, based on the expected effect of unobservable covariates. 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. Das et al. (2007), using roughly the same data studied here, provide evidence that defaults are significantly more correlated than would be suggested by the assumption that default risk is captured by the observable covariates. They do not, however, estimate a model with unobserved covariates.

Here, we depart from traditional duration-based specifications of default prediction, such as those of Couderc and Renault (2004), Shumway (2001), and Duffie, Saita, and Wang (2007), by allowing for dynamic unobserved covariates. Independent of our work, and with a similar thrust, Delloy, Fermanian, and Sbai (2005) and Koopman, Lucas, and Monteiro (2008) estimate dynamic frailty models of rating transitions. They suppose that the only observable firm-specific default covariate is an agency credit rating, and that all intensities of downgrades from one rating to the next depend on a common unobservable factor. Because credit ratings are incomplete and lagging indicators of credit quality, as shown, for example, by Lando and Skødeberg (2002), one would expect to find substantial frailty in ratings-based models such as these. As shown by Duffie, Saita, and Wang (2007), who estimate a model without frailty, the observable covariates that we propose offer substantially better out-of-sample default prediction than does prediction based on credit ratings. Even with the benefit of these observable covariates, however, in this paper we explicitly incorporate the effect of additional unincluded sources of default correlation and show that they have statistically and economically significant implications for the tails of portfolio default loss distributions.

III. A Dynamic Frailty Model

The introduction has given a basic outline of our model. This section provides a precise specification of the joint probability distribution of covariates and default times. We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an information filtration $\{\mathcal{G}_t : t \geq 0\}$. For a given borrower whose default time is τ, we say that a nonnegative progressively measurable process λ is the default intensity of the borrower if, as of time t, the borrower has not yet defaulted; λ_t is the conditional mean arrival rate of default, measured in events per unit of time.[5]

We suppose that all firms' default intensities at time t depend on a Markov state vector X_t of firm-specific and macroeconomic covariates. We suppose, however, that X_t is only partially observable to the econometrician. With complete observation of X_t, the default intensity of firm i at time t would be of the form

[5] Precisely, a martingale is defined by $1_{\tau \leq t} - \int_0^t \lambda_s 1_{\tau > s}\, ds$.

$\lambda_{it} = \Lambda(S_i(X_t), \theta)$, where θ is a parameter vector to be estimated and $S_i(X_t)$ is the component of the state vector that is relevant to the default intensity of firm i.

We assume that, conditional on the path of the underlying state process X determining default and other exit intensities, the exit times of firms are the first event times of independent Poisson processes with time-varying intensities determined by the path of X. This "doubly stochastic" assumption means that, given the path of the state vector process X, the merger and failure times of different firms are conditionally independent. While this conditional-independence assumption is traditional for duration models, we depart in an important way from the traditional setting by assuming that X is not fully observable to the econometrician. Thus, from the viewpoint of the econometrician's information, defaults are not doubly stochastic, and we cannot use standard estimation methods.

One may entertain various alternative approaches. For example, there is the possibility of "contagion," by which the default of one firm could have a direct influence on the revenues (or expenses or capital-raising opportunities) of another firm. In this paper, we examine instead the implications of "frailty," 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.

We let U_{it} be a firm-specific vector of covariates that are observable for firm i from when it first appears in the data at some time t_i until its exit time T_i. We let V_t denote a vector of macroeconomic variables that are observable at all times, and let Y_t be a vector of unobservable frailty variables. The complete state vector is then $X_t = (U_{1t}, \ldots, U_{mt}, V_t, Y_t)$, where m is the total number of firms in the data set. A time-series model of X, to be described, is determined by a vector γ of parameters to be estimated.

We let $W_{it} = (1, U_{it}, V_t)$ be the vector of observed covariates for company i (including a constant).[6] The last observation time T_i of company i could be the time of a default or another form of exit, such as a merger or acquisition. While we take the first appearance time t_i to be deterministic, our results are not affected by allowing t_i to be a stopping time under additional technical conditions.

The econometrician's information filtration $(\mathcal{F}_t)_{0\leq t\leq T}$ is that generated by the observed variables

$$\{V_s : 0 \leq s \leq t\} \cup \{(D_{is}, U_{is}) : 1 \leq i \leq m, t_i \leq s \leq \min(t, T_i)\},$$

where D_i is the default indicator process of company i (which is zero before default, one afterwards). The complete-information filtration $(\mathcal{G}_t)_{0\leq t\leq T}$ is generated by the variables in $\mathcal{F}_t$ as well as the frailty process $\{Y_s : 0 \leq s \leq t\}$.

[6] Because 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 assume that $\lambda_{it} = \Lambda(S_i(X_t), \theta)$, where $S_i(X_t) = (W_{it}, Y_t)$. We take the proportional hazards form

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

for a parameter vector $\theta = (\beta, \eta, \kappa)$ common to all firms, where κ is a parameter whose role will be defined later.[7]

Before considering the effect of other exits such as mergers and acquisitions, the maximum likelihood estimators (MLE) of $\mathcal{F}_t$-conditional survival probabilities, portfolio loss distributions, and related quantities are obtained under the usual smoothness conditions by treating the MLE of the parameters as though they are the true parameters (γ, θ).[8] We will also examine the implications of Bayesian uncertainty regarding certain key parameters.

To further simplify notation, let $W = (W_1, \ldots, W_m)$ denote the vector of observed covariate processes for all companies, and let $D = (D_1, \ldots, D_m)$ denote the vector of default indicators of all companies. If the econometrician were to be given complete information, Proposition 2 of Duffie, Saita, and Wang (2007) would imply a likelihood of the data at the parameters (γ, θ) of the form

$$\begin{aligned} &\mathcal{L}(\gamma, \theta \mid W, Y, D) \\ &\quad = \mathcal{L}(\gamma \mid W)\mathcal{L}(\theta \mid W, Y, D) \\ &\quad = \mathcal{L}(\gamma \mid W) \prod_{i=1}^{m} \left(e^{-\sum_{t=t_i}^{T_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{T_i} [D_{it}\lambda_{it}\Delta t + (1 - D_{it})] \right). \end{aligned} \tag{2}$$

We simplify by supposing that the frailty process Y is independent of the observable covariate process W. With respect to the econometrician's limited

[7] In the sense of Proposition 4.8.4 of Jacobsen (2006), the econometrician's default intensity for firm i is

$$\bar{\lambda}_{it} = E(\lambda_{it} \mid \mathcal{F}_t) = e^{\beta \cdot W_{it}} E\left(e^{\eta Y_t} \mid \mathcal{F}_t\right).$$

It is *not* generally true that the conditional probability of survival to a future time T (neglecting the effect of mergers and other exits) is given by the "usual formula" $E(e^{-\int_t^T \bar{\lambda}_{is}\, ds} \mid \mathcal{F}_t)$. Rather, for a firm that has survived to time t, the probability of survival to time T (again neglecting other exits) is $E(e^{-\int_t^T \lambda_{is}\, ds} \mid \mathcal{F}_t)$. This is justified by the law of iterated expectations and the doubly stochastic property on the complete-information filtration $(\mathcal{G}_t)$, which implies that the $\mathcal{G}_t$-conditional survival probability is $E(e^{-\int_t^T \lambda_{is}\, ds} \mid \mathcal{G}_t)$. See Collin-Dufresne, Goldstein, and Huggonier (2004) for another approach to this calculation.

[8] If other exits, for example, due to mergers and acquisitions, are jointly doubly stochastic with default exits, and other exits have the intensity process μ_i, then the conditional probability at time t that firm i will not exit before time $T > t$ is $E(e^{-\int_t^T (\mu_{is} + \lambda_{is})\, ds} \mid \mathcal{F}_t)$. 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.

filtration $(\mathcal{F}_t)$, the likelihood is then

$$\begin{aligned}\mathcal{L}(\gamma,\theta \mid W,D) &= \int \mathcal{L}(\gamma,\theta \mid W,y,D)p_Y(y)\,dy \\ &= \mathcal{L}(\gamma \mid W)\int \mathcal{L}(\theta \mid W,y,D)p_Y(y)\,dy \\ &= \mathcal{L}(\gamma \mid W)\,E\left[\prod_{i=1}^{m}\left(e^{-\sum_{t=t_i}^{T_i}\lambda_{it}\Delta t}\prod_{t=t_i}^{T_i}[D_{it}\lambda_{it}\Delta t+(1-D_{it})]\right)\,\Bigg|\, W,D\right],\end{aligned} \tag{3}$$

where $p_Y(\cdot)$ is the unconditional probability density of the path of the unobserved frailty process Y. The final expectation of (3) is with respect to that density.[9]

Most of our empirical results are properties of the MLE $(\hat{\gamma},\hat{\theta})$ for (γ,θ). Even when considering other exits such as those due to acquisitions, $(\hat{\gamma},\hat{\theta})$ is the full MLE for (γ,θ) because we have assumed that all forms of exit are jointly doubly stochastic on the artificially enlarged information filtration $(\mathcal{G}_t)$.

In order to evaluate the expectation in (3), one could simulate sample paths of the frailty process Y. Since our covariate data are monthly observations from 1979 to 2004, evaluating (3) by direct simulation would then mean 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 now turn to a special case of the model that can be feasibly estimated.

We suppose that Y is an Ornstein–Uhlenbeck (OU) process, in that

$$dY_t = -\kappa Y_t\,dt + dB_t, \qquad Y_0 = 0, \tag{4}$$

where B is a standard Brownian motion with respect to $(\Omega,\mathcal{F},\mathbb{P},(\mathcal{G}_t))$, and where κ is a nonnegative constant, the mean-reversion rate of Y. Without loss of generality, we have fixed the volatility parameter of the Brownian motion to be unity because scaling the parameter η, which determines in (1) the dependence of the default intensities on Y_t, plays precisely the same role in the model as scaling the frailty process Y.

The OU model for the frailty variable Y_t could capture the accumulative effect over time of various different unobserved fundamental common shocks to default intensities. For example, as suggested in the introduction, a borrower's measured credit qualities could be subject to a common source of reporting noise. While such an accounting failure could be mitigated over time with improved corporate governance and accounting standards, some new form of common unobserved shift in default intensities could arise, such as the incentive effects of a change in bankruptcy law that the econometrician failed to consider,

[9] For notational simplicity, expression (3) ignores the precise intramonth timing of default, although it was accounted for in the parameter estimation by replacing Δt with $\tau_i - t_{i-1}$ in the case that company i defaults in the time interval $(t_{t-1}, t_i]$.

or a correlated shift in the liquidity of balance sheets that went unobserved, and so on. The mean-reversion parameter κ is intended to capture the expected rate of decay of the cumulative effect of past unobserved shocks to default intensities.

Although an OU process is a reasonable starting model for the frailty process, one could allow much richer frailty models. From the Bayesian analysis reported in Section IV, however, we have found that even our relatively large data set is too limited to identify much of the time-series properties of frailty. This is not so surprising, given that the sample paths of the frailty process are not observed, and given the relatively sparse default data. For the same reason, we have not attempted to identify sector-specific frailty effects.

The starting value and long-run mean of the OU process Y are taken to be zero, since any change (of the same magnitude) of these two parameters can be absorbed into the default intensity intercept coefficient β_1. However, we do lose some generality by taking the initial condition for Y to be deterministic and to be equal to the long-run mean. An alternative would be to add one or more additional parameters specifying the initial probability distribution of Y. We have found that the posterior of Y_t tends to be robust to the assumed initial distribution of Y, for points in time t that are a year or two after the initial date of our sample.

We estimate the model parameters using a combination of the EM algorithm and the Gibbs sampler that is described in Appendix A.

IV. Major Empirical Results

This section describes our data, presents the estimated model, and provides its implications for the distribution of portfolio default losses relative to a model without frailty.

A. Data

Our data set, drawing elements from Bloomberg, Compustat, CRSP, and Moody's, is almost the same as that used to estimate the no-frailty models of Duffie, Saita, and Wang (2007) and Das et al. (2007). We have slightly improved the data by using *The Directory of Obsolete Securities* and the SDC database to identify additional mergers, defaults, and failures. We have checked that the few additional defaults and mergers identified through these sources do not change significantly the results of Duffie, Saita, and Wang (2007). Our data set 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 data set, Table I 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. We refer the interested reader to Section 3.1 of Duffie, Saita, and Wang (2007) for an in-depth description of the construction of the data set and an exact definition of these event types.

Table I
Number of Firm Exits of Each Type between 1979 and 2004

Exit Type	Number
Bankruptcy	176
Other default	320
Merger-acquisition	1,047
Other exits	671

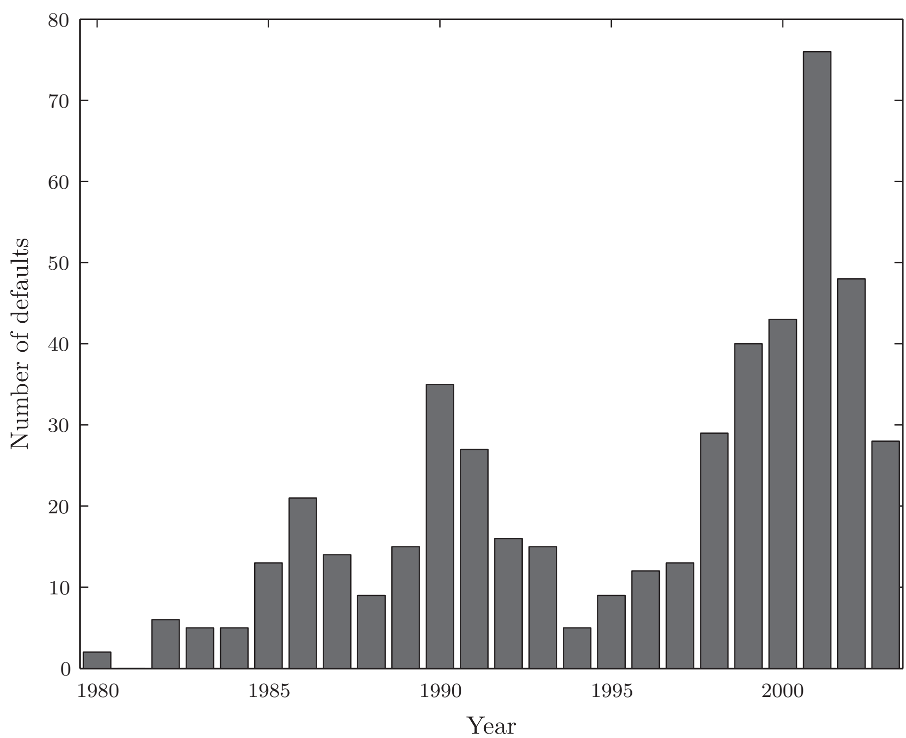

Figure 1. Yearly number of defaults. The number of defaults in our data set for each year between 1980 and 2003.

Figure 1 shows the total number of defaults (bankruptcies and other defaults) in each year. Moody's 13th annual corporate bond default study[10] 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 set of observable firm-specific and macroeconomic covariates:

1. Distance to default, a volatility-adjusted measure of leverage. Our method of construction, based on market equity data and Compustat book liability

[10] Moody's Investor Service, "Historical Default Rates of Corporate Bond Issuers, 1920–1999."

data, is that used by Vassalou and Xing (2004), Crosbie and Bohn (2002), and Hillegeist et al. (2004). Although the conventional approach to measuring distance to default involves some rough approximations, Bharath and Shumway (2008) provide evidence that default prediction is relatively robust to varying the proposed measure with some relatively simple alternatives.

2. The firm's trailing 1-year stock return, a covariate suggested by Shumway (2001). Although we do not have in mind a particular structural interpretation for this covariate, like Shumway, we find that it offers significant incremental explanatory power, perhaps as a proxy for some unobserved factor that has an influence on default risk beyond that of the firm's measured distance of default.
3. The 3-month Treasury bill rate, which plays a role in the estimated model consistent with the effect of a monetary policy that lowers short-term interest rates when the economy is likely to be performing poorly.
4. The trailing 1-year return on the S&P 500 index. The influence of this covariate, which is statistically significant but, in the presence of distance to default, of only moderate economic importance, will be discussed later.

Duffie, Saita, and Wang (2007) give a detailed description of these covariates and discuss their relative importance in modeling corporate default intensities. As robustness checks, we examine the influence of GDP growth rates, industrial production growth rates, average BBB–AAA corporate bond yield spreads, industry average distance to default, and firm size, measured as the logarithm of the model-implied assets.[11] Each of these is found to be at best marginally significant after controlling for our basic covariates, distance to default, trailing returns of the firm and the S&P 500, and the 3-month Treasury bill rate. Later in this paper, we also consider the implications of augmenting our list of macro-covariates with the trailing average default rate, which could proxy for important missing common covariates. This variable might also capture a direct source of default contagion, in that when a given firm defaults, other firms that had depended on it as a source of sales or inputs may also be harmed. This was the case, for example, in the events surrounding the collapse of Penn Central in 1970 to 1971. Another example of such a contagion effect is the influence of the bankruptcy of auto parts manufacturer Delphi in November 2005 on the survival prospects of General Motors. We do not explore the role of this form of contagion, which cannot be treated within our modeling framework.

B. The Fitted Model

Table II shows the estimated covariate parameter vector $\hat{\beta}$ and frailty parameters $\hat{\eta}$ and $\hat{\kappa}$, together with estimates of asymptotic standard errors.

[11] 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 renegotiate 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, however, this covariate does not play a statistically significant role.

Table II

Maximum Likelihood Estimates of Intensity Model Parameters

The frailty volatility is the coefficient η of dependence of the default intensity on the OU frailty process Y. Estimated asymptotic standard errors are computed using the Hessian matrix of the expected complete data log likelihood at $\theta = \hat{\theta}$. The mean reversion and volatility parameters are based on monthly time intervals.

	Coefficient	Std. Error	t-Statistic
Constant	−1.029	0.201	−5.1
Distance to default	−1.201	0.037	−32.4
Trailing stock return	−0.646	0.076	−8.6
3-month T-bill rate	−0.255	0.033	−7.8
Trailing S&P 500 return	1.556	0.300	5.2
Latent-factor volatility η	0.125	0.017	7.4
Latent-factor mean reversion κ	0.018	0.004	4.8

Our results show important roles for both firm-specific and macroeconomic covariates. Distance to default, although a highly significant covariate, does not on its own determine the default intensity, but does explain a large part of the variation of default risk across companies and over time. For example, a negative shock to distance to default by one standard deviation increases the default intensity by roughly $e^{1.2} - 1 \approx 230\%$. The 1-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.[12] The coefficient linking the trailing S&P 500 return to a firm's default intensity 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 in the stock market, a firm's distance to default overstates its financial health.

The estimate $\hat{\eta} = 0.125$ of the dependence of the unobservable default intensities on the frailty variable Y_t corresponds to a monthly volatility of this frailty effect of 12.5%, which translates to an annual volatility of 43.3%, which is highly economically and statistically significant.

Table III reports the intensity parameters of the same model after removing the role of frailty. The signs, magnitudes, and statistical significance of the coefficients of the observable covariates are similar to those with frailty, with the exception of the coefficient on the 3-month Treasury bill rate, which is smaller without frailty but remains statistically significant.

[12] There 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.

Table III

Maximum Likelihood Estimates of the Intensity Parameters in the Model without Frailty

Estimated asymptotic standard errors were computed using the Hessian matrix of the likelihood function at $\theta = \hat{\theta}$.

	Coefficient	Std. Error	t-Statistic
Constant	−2.093	0.121	−17.4
Distance to default	−1.200	0.039	−30.8
Trailing stock return	−0.681	0.082	−8.3
3-month T-bill rate	−0.106	0.034	−3.1
Trailing S&P 500 return	1.481	0.997	1.5

C. The Posterior of the Frailty Path

In order to interpret the model and apply it to the computation of portfolio loss distributions, we calculate the posterior distribution of the frailty process Y given the econometrician's information.

First, we compute the $\mathcal{F}_T$-conditional posterior distribution of the frailty process Y, where T is the final date of our sample. This is the conditional distribution of the latent factor given all of the historical default and covariate data through the end of the sample period. For this computation, we use the Gibbs sampler described in Appendix B. Figure 2 shows the conditional mean of the scaled latent factor, ηY_t, estimated as the average of 5,000 samples of Y_t drawn from the Gibbs sampler. One-standard deviation bands are shown around the posterior mean. We see substantial fluctuations in the frailty effect over time. For example, the multiplicative effect of the frailty factor on default intensities in 2001 is roughly $e^{1.1}$, or approximately three times larger than during 1995.[13]

While Figure 2 illustrates the posterior distribution of the frailty effect ηY_t given all information $\mathcal{F}_T$ available at the final time T of the sample period, most applications of a default risk model would call for the posterior distribution of ηY_t given the current information $\mathcal{F}_t$. For example, this is the relevant information for measurement by a bank of the risk of a portfolio of corporate debt. Although the covariate process is Gaussian, we also observe survivals and defaults, so we are in a setting of filtering in non-Gaussian state space models, to which we apply the "forward–backward algorithm" of Baum et al. (1970), as explained in Appendix C.

Figure 3 compares the conditional density of the frailty effect ηY_t for t at the end of January 2000, conditioning on $\mathcal{F}_T$ (in effect, the entire sample of default times and observable covariates up to 2004), with the density of ηY_t when conditioning on only $\mathcal{F}_t$ (the data available up to and including January 2000). Given

[13] A comparison that is based on replacing $Y(t)$ in $E[e^{\eta Y(t)} \mid \mathcal{F}_t]$ with the posterior mean of $Y(t)$ works reasonably well because the Jensen effects associated with the expectations of $e^{\eta Y(t)}$ for times in 1995 and 2001 are roughly comparable.

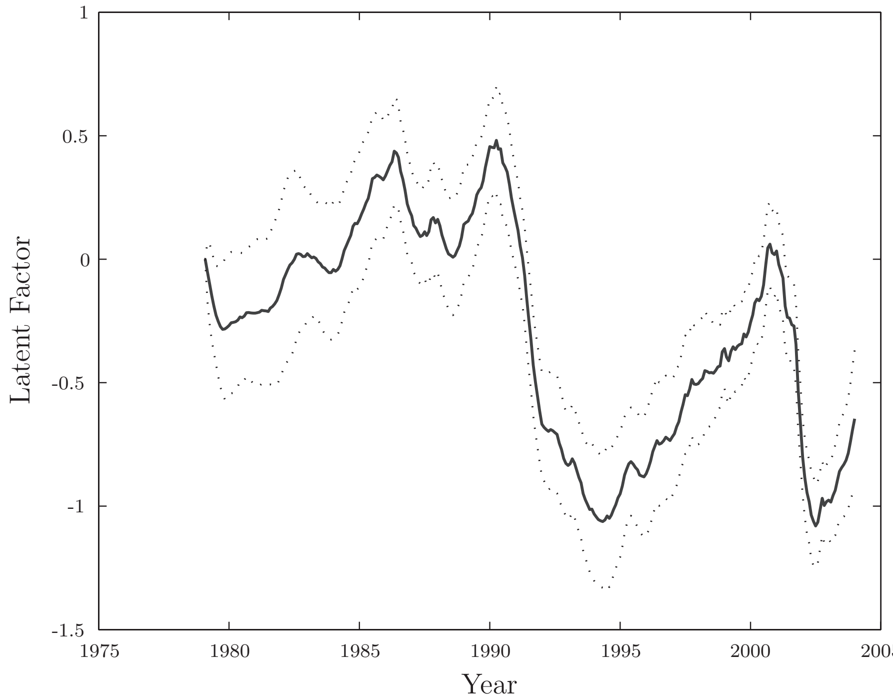

Figure 2. Frailty posterior distribution. Conditional posterior mean $E(\eta Y_t \mid \mathcal{F}_T)$ of the scaled latent Ornstein-Uhlenbeck frailty variable, with one standard deviation bands based on the $\mathcal{F}_T$-conditional variance of Y_t.

the additional information available at the end of 2004, the $\mathcal{F}_T$-conditional distribution of ηY_t is more concentrated than that obtained by conditioning on only the concurrently available information $\mathcal{F}_t$. The posterior mean of ηY_t given the information available in January 2000 is lower than that given all of the data through 2004, reflecting the sharp rise in corporate defaults in 2001 above and beyond that predicted from the observed covariates alone.

Figure 4 shows the path over time of the mean $E(\eta Y_t \mid \mathcal{F}_t)$ of this posterior density.

D. Portfolio Loss Risk

In order to illustrate the role of the common frailty effect on the tail risk of portfolio losses, we consider the distribution of the total number of defaults from a hypothetical portfolio consisting of all 1,813 companies in our data set that were active as of January 1998. We computed the posterior distribution, conditional on the information $\mathcal{F}_t$ available for t in January 1998, of the total number of defaults during the subsequent 5 years, January 1998 to December 2002. Figure 5 shows the probability density of the total number of defaults in this portfolio for three different models. All three models have the same

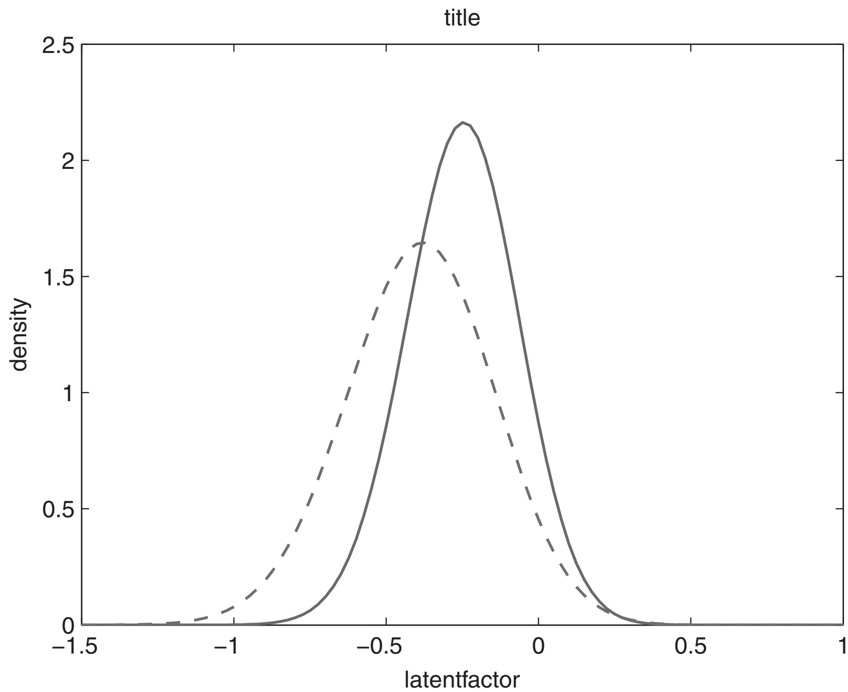

Figure 3. Conditional frailty posterior, January 2000. Conditional posterior density of the scaled frailty factor, ηY_t, for t in January 2000, given $\mathcal{F}_T$, that is, given all data (solid line), and given only contemporaneously available data in $\mathcal{F}_t$ (dashed line). These densities are calculated using the forward–backward recursions described in Appendix C.

posterior marginal distribution for each firm's default intensity process and default time, but the joint distribution of default times varies across the three models. Model (a) is the actual fitted model with a common frailty variable. For Models (b) and (c), however, we examine the hypothetical effects of reducing the effect of frailty. For both Models (b) and (c), the default intensity λ_{it} is changed by replacing the dependence of λ_{it} on the actual frailty process Y with dependence on a firm-specific process Y_i that has the same $\mathcal{F}_t$-conditional distribution as Y. For model (b), the initial condition Y_{it} of Y_i is common to all firms, but the future evolution of Y_i is determined not by the common OU process Y, but rather by an OU process Y_i that is independent across firms. Thus, Model (b) captures the common source of uncertainty associated with the current posterior distribution of Y_t, but has no common future frailty shocks. For Model (c), the hypothetical frailty processes of the firms, $Y_1, \ldots, Y_m$, are independent. That is, the initial condition Y_{it} is drawn independently across firms from the posterior distribution of Y_t, and the future shocks to Y_i are those of an OU process Y_i that is independent across firms.

One can see that the impact of the frailty effect on the portfolio loss distribution is substantially affected both by uncertainty regarding the current level Y_t

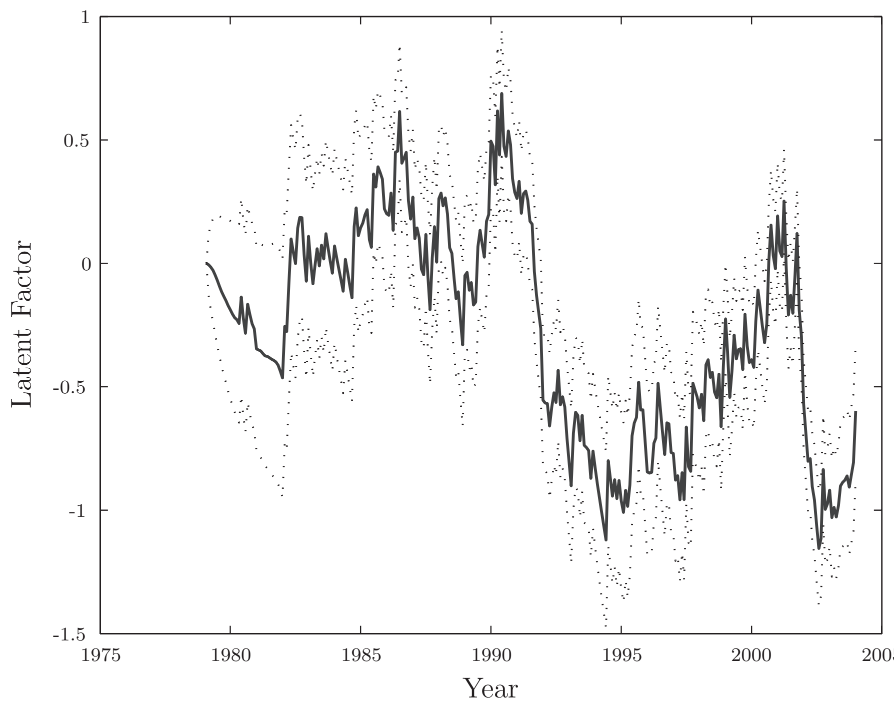

Figure 4. Filtered frailty. Conditional mean $E(\eta Y_t \mid \mathcal{F}_t)$ and conditional one-standard deviation bands of the scaled frailty variable, given only contemporaneously available data ($\mathcal{F}_t$).

of common frailty in January 1998, and also by common future frailty shocks to different firms. Both of these sources of default correlation are above and beyond those associated with exposure of firms to observable macroeconomic shocks, and exposure of firms to correlated observable firm-specific shocks (especially correlated changes in leverage).

In particular, we see in Figure 5 that the two hypothetical models that do not have a common frailty variable assign virtually no probability to the event of more than 200 defaults between January 1998 and December 2002. The 95th and 99th percentile losses of Model (c) with completely independent frailty variables are 144 and 150 defaults, respectively. Model (b), with independently evolving frailty variables with the same initial value in January 1998, has a 95th and 99th percentile of 180 and 204 defaults, respectively. The actual number of defaults in our data set during this time period was 195.

The 95th and 99th percentile of the loss distribution of the actual estimated model (a), with a common frailty variable, are 216 and 265 defaults, respectively. The realized number of defaults during this event horizon, 195, is slightly below the 91st percentile of the distribution implied by the fitted frailty model, therefore constituting a relatively severe event.

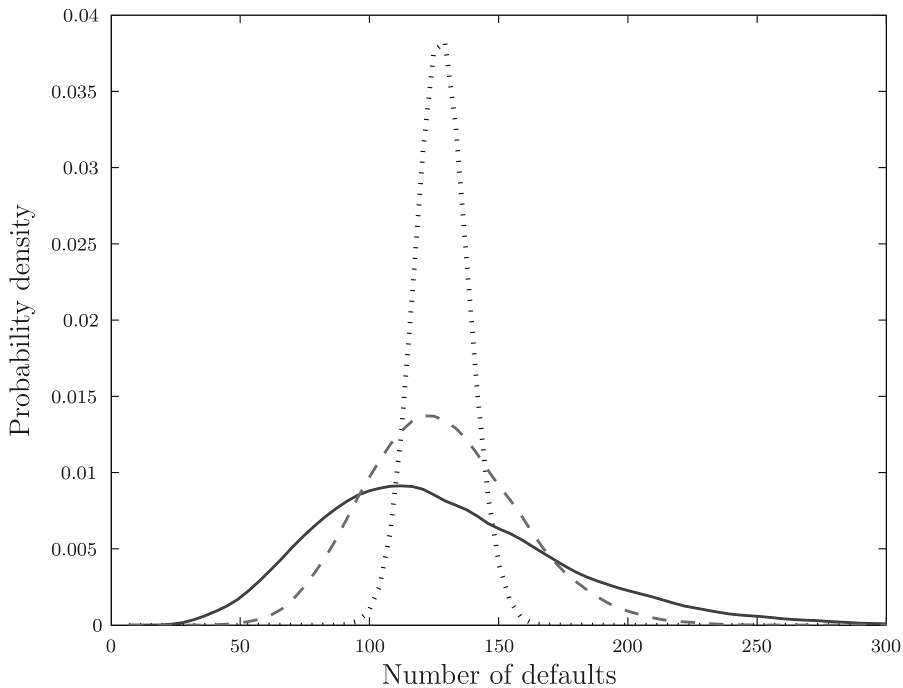

Figure 5. Conditional 5-year portfolio loss distribution in 1998. The conditional probability density, given $\mathcal{F}_t$ for t in January 1998, of the total number of defaults within 5 years from the portfolio of all active firms at January 1998, in (a) the fitted model with frailty (solid line), (b) a hypothetical model in which the common frailty process Y is replaced with firm-by-firm frailty processes with initial condition at time t equal to that of Y_t, but with common Brownian motion driving frailty for all firms replaced with firm-by-firm independent Brownian motions (dashed line), and (c) a hypothetical model in which the common frailty process Y is replaced with firm-by-firm independent frailty processes having the same posterior probability distribution as Y (dotted line). The density estimates are obtained with a Gaussian kernel smoother (bandwidth equal to five) applied to a Monte Carlo-generated empirical distribution.

V. Analysis of Model Fit and Specification

This section examines the ability of our model to survive tests of its fit. We also examine its out-of-sample accuracy, and its robustness to some alternative specifications.

A. Frailty versus No Frailty

In order to judge the relative fit of the models with and without frailty, we do not use standard tests, such as the chi-square test. Instead, we compare the marginal likelihoods of the models. This approach does not rely on large-sample distribution theory and has the intuitive interpretation of attaching prior probabilities to the competing models.

Specifically, we consider a Bayesian approach to comparing the quality of fit of competing models and assume positive prior probabilities for the two models "noF" (the model without frailty) and "F" (the model with a common frailty variable). The posterior odds ratio is

$$\frac{\mathbb{P}(\mathrm{F} \mid W, D)}{\mathbb{P}(\mathrm{noF} \mid W, D)} = \frac{\mathcal{L}_F(\hat{\gamma}_F, \hat{\theta}_F \mid W, D)}{\mathcal{L}_{noF}(\hat{\gamma}_{noF}, \hat{\theta}_{noF} \mid W, D)} \, \frac{\mathbb{P}(\mathrm{F})}{\mathbb{P}(\mathrm{noF})}, \tag{5}$$

where $\hat{\theta}_M$ and $\mathcal{L}_M$ denote the MLE and the likelihood function for a certain model M, respectively. Plugging (3) into (5) gives

$$\begin{aligned} \frac{\mathbb{P}(\mathrm{F} \mid W, D)}{\mathbb{P}(\mathrm{noF} \mid W, D)} &= \frac{\mathcal{L}(\hat{\gamma}_F \mid W)\mathcal{L}_F(\hat{\theta}_F \mid W, D)}{\mathcal{L}(\hat{\gamma}_{noF} \mid W)\mathcal{L}_{noF}(\hat{\theta}_{noF} \mid W, D)} \, \frac{\mathbb{P}(\mathrm{F})}{\mathbb{P}(\mathrm{noF})} \\ &= \frac{\mathcal{L}_F(\hat{\theta}_F \mid W, D)}{\mathcal{L}_{noF}(\hat{\theta}_{noF} \mid W, D)} \, \frac{\mathbb{P}(\mathrm{F})}{\mathbb{P}(\mathrm{noF})}, \end{aligned} \tag{6}$$

using the fact that the time-series model for the covariate process W is the same in both models. The first factor on the right-hand side of (6) 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 (6). Smith and Spiegelhalter (1980) discuss the penalizing nature of the Bayes factor, sometimes referred to as the "fully automatic Occam's razor." In our case, the outcome of the test statistic is 22.6. In the sense of this approach to model comparison, we see strong evidence in favor of including a frailty variable.[14]

B. Misspecification of Proportional Hazards

A comparison of Figures 1 and 2 shows that the frailty effect is generally higher when defaults are more prevalent. In light of this, one might suspect misspecification of the proportional hazards intensity model (1), which would automatically induce a measured frailty effect if the true intensity model has a higher-than-proportional dependence on distance to default, which is by far the most economically and statistically significant covariate. If the response of the true log intensity to variation in distance to default is faster than linear, then the estimated latent variable in our current formulation would be higher when distances to default are well below normal, as in 1991 and 2003. In an

[14] Unfortunately, the Bayes factor cannot be used for comparing the model with frailty to the model with frailty and unobserved heterogeneity, because for the latter model evaluating the likelihood function is computationally prohibitively expensive.

Internet Appendix,[15] we provide an extension of the model that incorporates non-parametric dependence of default intensities on distance to default. The results indicate that the proportional hazards specification is unlikely to be a significant source of misspecification in this regard. The response of the estimated log intensities is roughly linear in distance to default, and the estimated posterior of the frailty path has roughly the appearance shown in Figure 2.

C. Unobserved Heterogeneity

It may be that a substantial portion of the differences across firms' default risks is due to heterogeneity in the degree to which different firms are sensitive to the covariates, perhaps through additional firm-specific omitted variables. Failure to allow for this could result in biased and inefficient estimation. We consider an extension of the model by introducing a firm-specific heterogeneity factor Z_i for firm i, so that the complete-information ($\mathcal{G}_t$) default intensity of firm i is of the form

$$\lambda_{it} = e^{\beta \cdot W_{it} + \gamma Y_t} Z_i \tag{7}$$

where $Z_1, \ldots, Z_m$ are independently and identically gamma-distributed[16] random variables that are jointly independent of the observable covariates W and the common frailty process Y.

Fixing the mean of the heterogeneity factor Z_i to be one without loss of generality, we find that maximum likelihood estimation does not pin down the variance of Z_i to any reasonable precision with our limited set of data. We anticipate that far larger data sets would be needed, given the already large degree of observable heterogeneity and the fact that default is, on average, relatively unlikely. In the end, we examine the potential role of unobserved heterogeneity for default risk by fixing the standard deviation of Z_i at 0.5. It is easy to check that the likelihood function is again given by (3), where in this case the final expectation is with respect to the product of the distributions of Y and $Z_1, \ldots, Z_n$.

In an Internet Appendix, we show that our general conclusions regarding the economic significance of the covariates and the importance of including a time-varying frailty variable remain in the presence of unobserved heterogeneity. Moreover, the posterior mean path of the time-varying latent factor is essentially unchanged.

[15] This Internet Appendix is available online in the "Supplements and Data Sets" section at http://www.afajof.org/supplements.asp.

[16] Pickles and Crouchery (1995) show in simulation studies that it is relatively safe to make concrete parametric assumptions about the distribution of static frailty variables. Inference is expected to be similar whether the frailty distribution is modeled as gamma, log normal, or some other parametric family, but for analytical tractability we choose the gamma distribution.

D. Parameter Uncertainty

Until this point, our analysis is based on maximum likelihood estimation of the frailty mean reversion and volatility parameters, κ and σ. Uncertainty regarding these parameters, in a Bayesian sense, could lead to an increase in the tail risk of portfolio losses, which we investigate next. We are also interested in examining our ability to learn these parameters, in a Bayesian sense. Among other implications of our Bayesian analysis, we will see that the mean reversion parameter κ is particularly hard to tie down.

The stationary variance of the frailty variable Y_t is

$$\sigma_\infty^2 \equiv \lim_{s\to\infty} \mathrm{var}(Y_s \mid \mathcal{G}_t) = \lim_{s\to\infty} \mathrm{var}(Y_s \mid Y_t) = \frac{\sigma^2}{2\kappa}.$$

Motivated by the historical behavior of the posterior mean of the frailty, we take the prior density of the stationary standard deviation, σ_∞, to be Gamma distributed with a mean of 0.5 and a standard deviation of 0.25. The prior distribution for the mean reversion rate κ is also assumed to be Gamma, with a mean of $\log 2/36$ (which corresponds to a half-life of 3 years for shocks to the frailty variable) and a standard deviation of $\log 2/72$. The joint prior density of σ and κ is therefore of the form

$$p\,(\sigma, \kappa) \propto \left(\frac{\sigma}{\sqrt{2\kappa}}\right)^3 \exp\left(-\frac{8\sigma}{\sqrt{2\kappa}}\right) \kappa^3 \exp\left(-\kappa\frac{144}{\log 2}\right).$$

Figure 6 shows the marginal posterior densities of the volatility and mean reversion parameters of the frailty variable. Figure 7 shows their joint posterior density. These figures indicate considerable posterior uncertainty regarding these parameters. From the viewpoint of subjective probability, estimates of the tail risk of the portfolio loss distribution that are obtained by fixing these common frailty parameters at their maximum likelihood estimates might significantly underestimate the probability of certain extreme events.

Although parameter uncertainty has a minor influence on the portfolio loss distribution at intermediate quantiles, Figure 8 reveals a moderate impact of parameter uncertainty on the extreme tails of the distribution. For example, when fixing the frailty parameters η and κ at their maximum likelihood estimates, the 99th percentile of the portfolio default distribution is 265 defaults. Taking posterior parameter uncertainty into account, this quantile rises to 275 defaults.

E. Do Trailing Defaults Proxy for Unobserved Covariates?

Table IV reports the fitted model coefficients for a model without frailty, but with trailing 1-year average yearly default rate as a covariates. We emphasize that this model violates the assumptions that justify our likelihood function, for the obvious reason that defaults cannot be independent across different firms conditional on the path of the covariate process if we include average realized

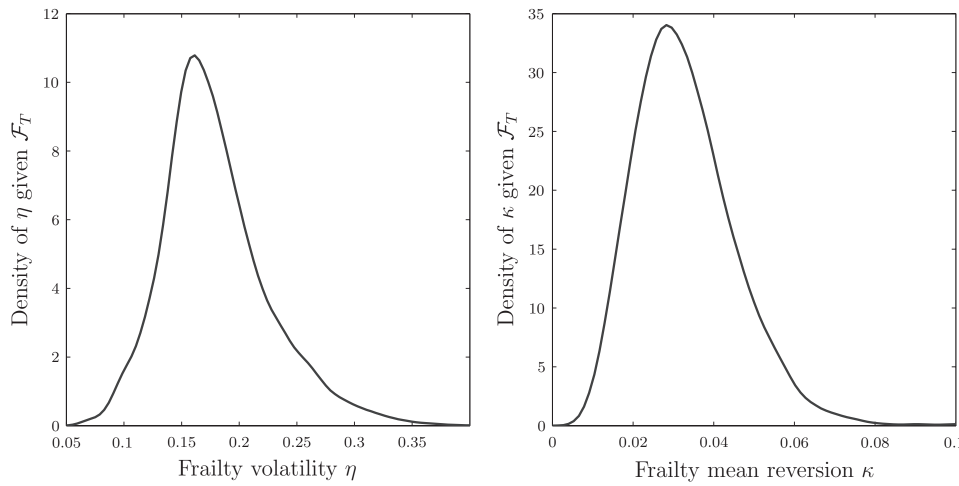

Figure 6. Marginal frailty parameter posterior distribution. Marginal posterior densities, given $\mathcal{F}_T$, of the frailty volatility parameter η and the frailty mean reversion rate κ in the Bayesian approach of Section V.D.

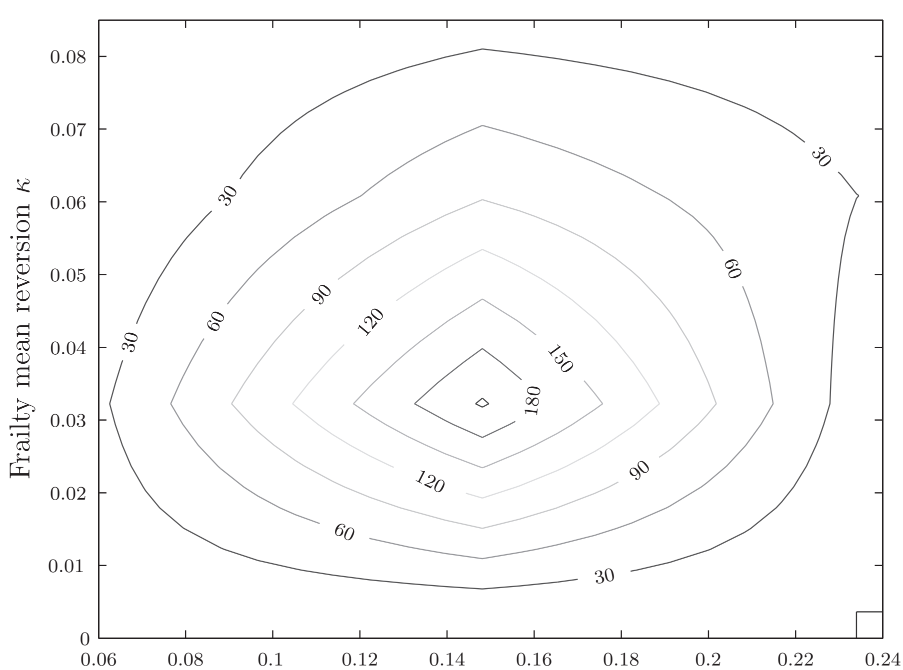

Figure 7. Joint frailty parameter posterior distribution. Isocurves of the joint posterior density, given $\mathcal{F}_T$, of the frailty volatility parameter η and mean reversion rate κ.

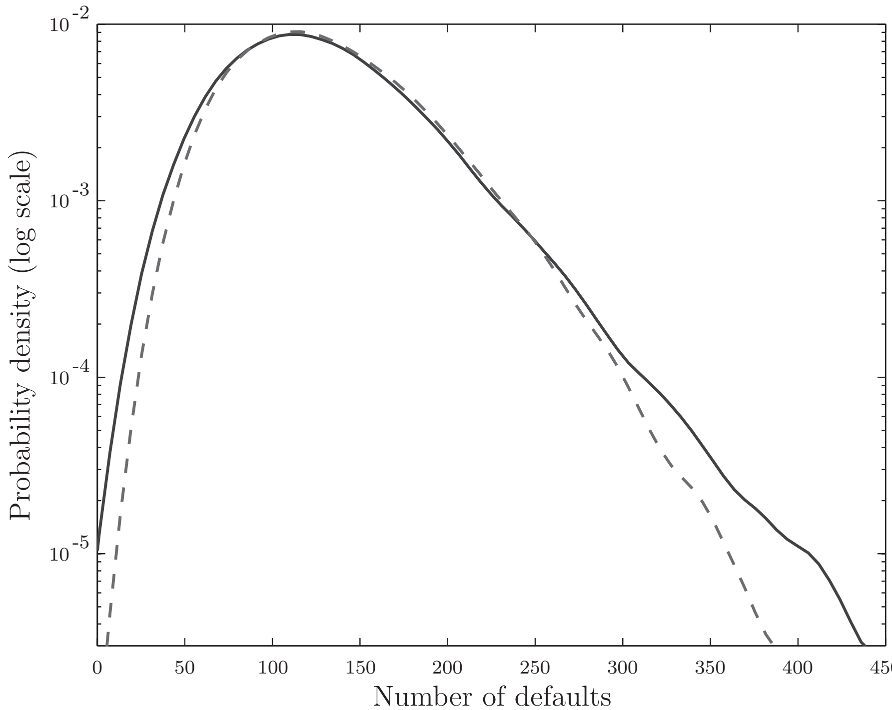

Figure 8. Portfolio loss density comparison. Density, on a logarithmic scale, of the number of defaults in the portfolio when fixing the volatility and mean reversion parameter at their MLE estimates (dashed line), and in the Bayesian estimation framework (solid line). The density estimates were obtained by applying a Gaussian kernel smoother (with a bandwidth of 10) to the Monte Carlo-generated empirical distribution.

Table IV

Maximum Likelihood Estimates of the Intensity Parameters in the Model without Frailty but with Trailing 1-Year Average Yearly Default Rate as a Covariate

Estimated asymptotic standard errors were computed using the Hessian matrix of the likelihood function at $\theta = \hat{\theta}$.

	Coefficient	Std. Error	t-Statistic
Constant	−2.364	0.955	−2.5
Distance to default	−1.189	0.052	−23.1
Trailing stock return	−0.678	0.301	−2.3
3-month T-bill rate	−0.086	0.135	−0.6
Trailing S&P 500 return	1.766	1.001	1.8
Trailing 1-year default rate	7.154	1.000	7.2

default rates as a covariate. It may be, however, that trailing default rates will proxy for an important source of default risk covariation that is otherwise unobserved, and reduce the relative importance of frailty.

The signs, magnitudes, and statistical significance of the coefficients on the observable covariates are similar to those of the model that does not include the trailing default rate as a covariate. The trailing default rate plays a moderately important auxiliary role. For example, fixing other covariates, if the trailing average default rate were to increase by 1% per year, a large but plausible shift given our data set, the model estimates imply a proportional increase in the conditional mean arrival rates of all firms of about 7.1%. This would cause a shift in the default intensity of a particular firm from, say, 2% to about 2.14%.

For the reason described above (the distribution of trailing default is an endogenous property of the default intensity model), we cannot examine the influence of trailing default on the posterior of the frailty process. We are able, though, to see whether including trailing default rates is an effective alternative to frailty in capturing the distribution of portfolio tail losses. In the sense of the tests described in Section V.F, it is not.

F. Portfolio Default Quantile Tests

We turn to the realism with which the frailty-based model estimates the quantiles of portfolio defaults. We will focus on the quantiles of the conditional distributions of the total number of defaults over 1-year horizons, from the portfolios of all active firms at the beginning of the respective years.

In terms of firm-by-firm default prediction, Duffie, Saita, and Wang (2007) show that the observable covariates of our basic model already provide the highest out-of-sample accuracy ratios documented in the default prediction literature. Allowing for frailty does not add significantly to firm-by-firm default prediction. In an Internet Appendix, we show that accuracy ratios with frailty are essentially the same as those without. Likewise, accuracy ratios are roughly unaffected by adding the trailing average default rate as a covariate. At the level of individual firms, most of our ability to sort firms according to default probability is coming from the firm-level covariates, particularly distance to default. The coefficients on these variables are relatively insensitive to the alternative specifications that we have examined.

Our main focus is the distribution of portfolio losses. In order to gauge the ability of our model to capture this distribution, we proceed as follows. At the beginning of each year between 1980 and 2003, we calculate for the companies in our data set the model-implied distribution of the number of defaults during the subsequent 12 months. We then determine the quantile of the realized number of defaults with respect to this distribution.

Figure 9 shows these quantiles for (i) our benchmark model with frailty, (ii) our benchmark model adjusted by removing frailty, and (iii) the model without frailty but including the trailing 1-year average default rate as an additional covariate. The quantiles of the two models without frailty seem to cluster around zero and one, which suggests that these models underestimate the probabilities

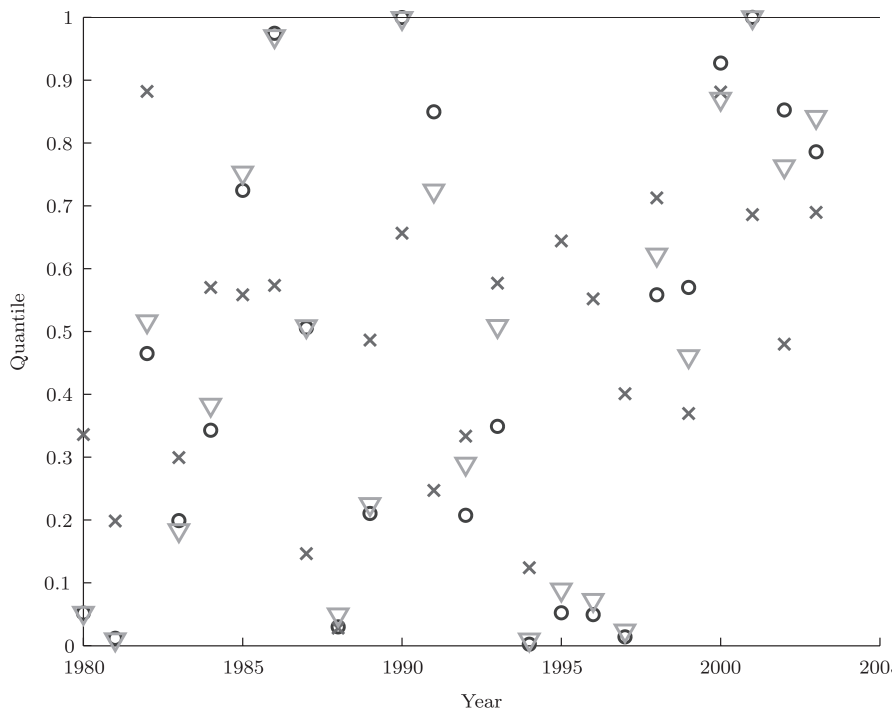

Figure 9. Realized portfolio loss quantiles. Quantile of the realized number of defaults with respect to the predicted 1-year portfolio loss distribution as implied by the model with the frailty variable (crosses), without the frailty variable (circles), and without the frailty variable but with the trailing 1-year average default rate as the covariate (triangles).

of unusually low portfolio losses and of unusually high portfolio losses. For example, in 1994 the realized number of defaults lies below the estimated 1st percentile of the portfolio default distribution for the model without frailty, while in 1990 and 2001 the realized number of defaults lies above the 99.9th percentile of the estimated distribution. For the model that also includes the trailing 1-year average default rate as a covariate, these quantiles are only slightly less extreme. On the other hand, the quantiles for the model with frailty are distributed relatively evenly in the unit interval, indicating a more accurate assessment of credit risk on the portfolio level.

Moreover, the forecasting errors for the two models without frailty tend to be serially correlated over time, which is most evident for the periods 1994 to 1997 as well as 2000 to 2003. The null hypothesis of no serial correlation in the quantiles is indeed rejected at the 1% significance level for the model without frailty (p-value of 0.004). For the model without the frailty variable but with the trailing 1-year average default rate as a covariate, the null hypothesis of no serial correlation in the quantiles can still be rejected at the 5% significance level (p-value of 0.019). On the other hand, with a p-value of 0.62, the null hypothesis of no serial correlation in the quantiles cannot be rejected for the model with frailty.

VI. Concluding Remarks

We find substantial evidence among U.S. public corporates of a common unobserved source of default risk, relative to the information provided by a powerful set of observable factors for predicting individual firm defaults. According to our estimates, failure to allow for unobserved factors in this setting leads to dramatic downward biases in value-at-risk estimates for large corporate debt portfolios.

Our results have important implications 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 Vasicek (1987) and Gordy (2003). 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 defaults, such as CDOs and options on portfolios of default swaps. While we do not address the pricing of credit risk in this paper, frailty could play a useful role in the market valuation of relatively senior tranches of CDOs, which suffer a loss of principal only when the total default losses of the underlying portfolio of bonds is extreme.

We estimate our model on data for U.S. firms between January 1979 and March 2004. We find that realized corporate default rates vary over time well beyond levels that can be explained by a model that includes only our observable covariates. In goodness-of-fit and quantile tests, the models without frailty that we examine are rejected and significantly underestimate the probability of extreme positive as well as negative events in portfolios of corporate credits.

For our data and model, we estimate that unobserved frailty has an impact on default intensities that adds a proportional annual volatility of roughly 40%. The estimated rate of mean reversion of the frailty factor is approximately 1.8% per month, although this mean reversion rate is difficult to pin down with the available data.

Our methodology can 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 a given class of events, for example, operational risk events, mergers and acquisitions, or mortgage prepayments and defaults.

Appendix A: Parameter Estimation

This appendix provides our estimation methodology. The parameter vector γ determining the time-series model for the observable covariate process W is specified and estimated in Duffie, Saita, and Wang (2007). This model, summarized in an Internet Appendix, 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.

We use a variant of the EM algorithm (see Demptser, Laird, and Rubin (1977)), an iterative method for the computation of the MLE of parameters of 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 of the intensity parameter vector θ involves the following steps:

1. Initialize an estimate of $\theta = (\beta, \eta, \kappa)$ at $\theta^{(0)} = (\hat{\beta}, 0.05, 0)$, where $\hat{\beta}$ is the maximum likelihood estimator of β in the model without frailty, which can be obtained by maximizing the likelihood function (2) by standard methods such as the Newton–Raphson algorithm.
2. (E-step) Given the current parameter estimate $\theta^{(k)}$ and the observed covariate and default data W and D, respectively, draw n independent sample paths $Y^{(1)}, \ldots, Y^{(n)}$ from the conditional density $p_Y(\cdot \mid W, D, \theta^{(k)})$ of the latent OU frailty process Y. We do this with the Gibbs sampler described in Appendix B. We let

$$Q(\theta, \theta^{(k)}) = E_{\theta^{(k)}}(\log \mathcal{L}(\theta \mid W, Y, D)) \tag{A1}$$

$$= \int \log \mathcal{L}(\theta \mid W, y, D) p_Y(y \mid W, D, \theta^{(k)})\, dy, \tag{A2}$$

where E_θ denotes expectation with respect to the probability measure associated with a particular parameter vector θ. This "expected complete-data log likelihood" or "intermediate quantity," as it is commonly called in the EM literature, can be approximated with the sample paths generated by the Gibbs sampler as

$$\hat{Q}(\theta, \theta^{(k)}) = \frac{1}{n} \sum_{j=1}^{n} \log \mathcal{L}(\theta \mid W, Y^{(j)}, D). \tag{A3}$$

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

One can show (Demptser, Laird, and Rubin (1977) or Gelman et al. (2004)) that $\mathcal{L}(\gamma, \theta^{(k+1)} \mid W, D) \geq \mathcal{L}(\gamma, \theta^{(k)} \mid W, D)$. That is, the observed data likelihood

(3) 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, although these conditions 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.

Under regularity conditions, the Fisher and Louis identities[17] imply that

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

and

$$\nabla^2_\theta \mathcal{L}(\hat{\theta} \mid W, Y, D) = \nabla^2_\theta Q(\theta, \hat{\theta})|_{\theta=\hat{\theta}}.$$

The Hessian matrix of the expected complete-data likelihood (A2) can therefore be used to estimate asymptotic standard errors for the MLE parameter estimators.

We also estimate a generalization of the model that incorporates unobserved heterogeneity, using an extension of this algorithm that is provided in the Internet Appendix.

Appendix B: Applying the Gibbs Sampler with Frailty

A central quantity of interest for describing and estimating the historical default dynamics is the posterior density $p_Y(\cdot \mid W, D, \theta)$ 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, 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 $p_Y(\cdot \mid W, D, \theta)$. 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

$$E[f(Y) \mid W, D, \theta] = \int f(y) p_Y(y \mid W, D, \theta)\, dy \tag{B1}$$

is given by

$$\frac{1}{N} \sum_{n=1}^{N} f(Y^{(n)}). \tag{B2}$$

Under conditions, the ergodic theorem for Markov chains guarantees the convergence of this average to its expectation as $N \to \infty$. One such function of

[17] See, for example, Proposition 10.1.6 of Cappé, Moulines, and Rydén (2005).

interest is the identity $f(y) = y$, so that

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

the posterior mean of the latent OU frailty process.

The linchpin to MCMC is that the joint distribution of the frailty path $Y = \{Y_t : 0 \le t \le T\}$ can be broken down 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 $p_Y(\cdot \mid W, D, \theta)$ is uniquely determined by the conditional distributions:

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

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 $Y_{(-t)} = \{Y_s : s \ne t\}$, is somewhat tractable, as shown further. This allows us to use the Gibbs sampler (Geman and Geman (1984) or Gelman et al. (2004)) to draw whole sample paths from the posterior distribution of $\{Y_t : 0 \le t \le T\}$ by the following algorithm:

1. Initialize $Y_t = 0$ for $t = 0, \ldots, T$.
2. For $t = 1, 2, \ldots, T$, draw a new value of Y_t from its conditional distribution given $Y_{(-t)}$. For a method, see later.
3. Store the sample path $\{Y_t : 0 \le t \le T\}$ and return to Step 2 until the desired number of paths has been simulated.

We usually discard the first several hundred paths as a "burn-in" sample because initially the Gibbs sampler has not approximately converged[18] to the posterior distribution of $\{Y_t : 0 \le t \le T\}$.

It remains to show how to sample Y_t from its conditional distribution given $Y_{(-t)}$. Recall that $\mathcal{L}(\theta \mid W, Y, D)$ denotes the complete-information likelihood

[18] We use 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 et al. (2004), Chapter 11.6, for a discussion of various methods for assessing convergence of MCMC methods.

of the observed covariates and defaults, where $\theta = (\beta, \eta, \kappa)$. For $0 < t < T$, we have

$$\begin{aligned}
p(Y_t \mid W, D, Y_{(-t)}, \theta) &= \frac{p(W, D, Y, \theta)}{p(W, D, Y_{(-t)}, \theta)} \\
&\propto p(W, D, Y, \theta) \\
&= p(W, D \mid Y, \theta) p(Y, \theta) \\
&\propto \mathcal{L}(\theta \mid W, Y, D) p(Y, \theta) \\
&= \mathcal{L}(\theta \mid W, Y, D) p(Y_t \mid Y_{(-t)}, \theta) p(Y_{(-t)}, \theta) \\
&\propto \mathcal{L}(\theta \mid W, Y, D) p(Y_t \mid Y_{(-t)}, \theta),
\end{aligned}$$

where we repeatedly make use of the fact that terms not involving Y_t are constant.

From the Markov property it follows that the conditional distribution of Y_t given $Y_{(-t)}$ and θ is the same as the conditional distribution of Y_t given Y_{t-1}, Y_{t+1}, and θ. Therefore,

$$\begin{aligned}
p(Y_t \mid Y_{(-t)}, \theta) &= p(Y_t \mid Y_{t-1}, Y_{t+1}, \theta) \\
&= \frac{p(Y_{t-1}, Y_t, Y_{t+1} \mid \theta)}{p(Y_{t-1}, Y_{t+1} \mid \theta)} \\
&\propto p(Y_{t-1}, Y_t, Y_{t+1} \mid \theta) \\
&= p(Y_{t-1}, Y_t \mid \theta) p(Y_{t+1} \mid Y_{t-1}, Y_t, \theta) \\
&\propto \frac{p(Y_{t-1}, Y_t \mid \theta)}{p(Y_{t-1} \mid \theta)} p(Y_{t+1} \mid Y_t, \theta) \\
&= p(Y_t \mid Y_{t-1}, \theta) p(Y_{t+1} \mid Y_t, \theta),
\end{aligned}$$

where $p(Y_t \mid Y_{t-1}, \theta)$ is the one-step transition density of the OU process (4). Hence,

$$p(Y_t \mid W, D, Y_{(-t)}, \theta) \propto \mathcal{L}(\theta \mid W, Y, D) \cdot p(Y_t \mid Y_{t-1}, \theta) \cdot p(Y_{t+1} \mid Y_t, \theta). \quad \text{(B3)}$$

Equation (B3) determines the desired conditional density of Y_t given Y_{t-1} and Y_{t+1} in an implicit form. Although it is not possible to directly draw samples from this distribution, we can employ the Random Walk Metropolis–Hastings algorithm (Metropolis and Ulam (1949), and Hastings (1970)).[19] We use the proposal density $q(Y_t^{(n)} \mid W, D, Y^{(n-1)}, \theta) = N(Y_t^{(n-1)}, 4)$, that is, we take the mean to be the value of Y_t from the previous iteration of the Gibbs sampler, and the variance to be twice the variance of the standard Brownian motion

[19] Alternatively, we could discretize the sample space and approximate the conditional distribution by a discrete distribution, an approach commonly referred to as the Griddy Gibbs method (Tanner (1998)). However, the Metropolis–Hastings algorithm is usually a couple of times faster in cases in which the conditional density is not known explicitly.

increments[20]. The Metropolis–Hastings step to sample Y_t in the nth iteration of the Gibbs sampler therefore works as follows:

1. Draw a candidate $y \sim N(Y_t^{(n-1)}, 4)$.
2. Compute

$$\alpha\big(y, Y_t^{(n)}\big) = \min\left(\frac{\mathcal{L}\big(\theta \mid W, Y_{(-t)}^{(n-1)}, Y_t = y, D\big)}{\mathcal{L}\big(\theta \mid W, Y^{(n-1)}, D\big)}, 1\right). \tag{B4}$$

3. Draw U with the uniform distribution on (0, 1), and let

$$Y_t^{(n)} = \left\{\begin{array}{ll} y & \text{if } U < \alpha\big(y, Y_t^{(n)}\big) \\ Y_t^{(n-1)} & \text{otherwise.} \end{array}\right\}.$$

The choice of the acceptance probability (B4) ensures that the Markov chain $\{Y_t^{(n)} : n \geq 1\}$ satisfies the detailed balance equation

$$\begin{aligned} & p(y_1 \mid W, D, Y_{(-t)}, \theta)\phi_{y_1,4}(y_2)\alpha(y_1, y_2) \\ & = p(y_2 \mid W, D, Y_{(-t)}, \theta)\phi_{y_2,4}(y_1)\alpha(y_2, y_1), \end{aligned}$$

where ϕ_{μ,σ^2} denotes the density of a normal distribution with mean μ and variance σ^2. Moreover, $\{Y_t^{(n)} : n \geq 1\}$ has $p(Y_t \mid W, D, Y_{(-t)}, \theta)$ as its stationary distribution (see, for example, Theorem 7.2 in Robert and Casella (2005)).

Appendix C: Forward–Backward Filtering for Frailty

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 $\Delta R(t) = \{i \in R(t-1) : D_{it} = 1, t_i \leq t \leq T_i\}$ 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

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

The normalized version of the forward–backward algorithm allows us to calculate the filtered density of the latent Ornstein–Uhlenbeck frailty variable by the recursion

$$\begin{aligned} c_t &= \int\int p(y_{t-1} \mid \mathcal{F}_{t-1})\phi(y_t - y_{t-1})\mathcal{L}_t(\theta \mid W_t, y_t, D_t)\, dy_{t-1}\, dy_t \\ p(y_t \mid \mathcal{F}_t) &= \frac{1}{c_t}\int p(y_{t-1} \mid \mathcal{F}_{t-1})p(y_t \mid y_{t-1}, \theta)\mathcal{L}_t(\theta \mid W_t, y_t, D_t)\, dy_{t-1}, \end{aligned}$$

[20] We calculate the conditional density for various points in time numerically to assure that it does not have any fat tails. This was indeed the case so that using a normal proposal density does not jeopardize the convergence of the Metropolis–Hastings algorithm. See Mengersen and Tweedie (1996) for technical conditions.

where $p(Y_t \mid Y_{t-1}, \theta)$ is the one-step transition density of the OU process (4). For this recursion, we begin with the distribution (Dirac measure) of Y_0 concentrated at zero.

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

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

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

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

In order to explore the joint posterior distribution $p((y_0, y_1, \ldots, y_T)' \mid \mathcal{F}_T)$ of the latent frailty variable, one may employ, for example, the Gibbs sampler described in Appendix B.

REFERENCES

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

Baum, Leonard E., Ted P. Petrie, George Soules, and Norman Weiss, 1970, A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *Annals of Mathematical Statistics* 41, 164–171.

Beaver, William, 1968, Market prices, financial ratios, and the prediction of failure, *Journal of Accounting Research* Autumn, 170–192.

Besag, Julian, 1974, Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Association: Series B* 36, 192–236.

Bharath, Sreedhar, and Tyler Shumway, 2008, Forecasting default with the Merton distance to default model, *Review of Financial Studies* 21, 1339–1369.

Black, Fischer, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–654.

Cappé, Olivier, Eric Moulines, and Tobias Rydén, 2005, *Inference in Hidden Markov Models* (Springer-Verlag, New York).

Celeux, Gilles, and Jean Diebolt, 1986, The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorith for the mixture problem, *Computational Statistics Quaterly* 2, 73–82.

Chava, Sudheer, and Robert Jarrow, 2004, Bankruptcy prediction with industry effects, *Review of Finance* 8, 537–569.

Chernobai, Anna, Philippe Jorion, and Fan Yu, 2008, The determinants of operational losses, Working paper. Syracuse University, UC-Irvine, and Michigan State University.

Collin-Dufresne, Pierre, Robert Goldstein, and Jean Helwege, 2003, Is credit event risk priced? Modeling contagion via the updating of beliefs, Working paper, Haas School, University of California, Berkeley.

Collin-Dufresne, Pierre, Robert Goldstein, and Julien Huggonier, 2004, A general formula for valuing defaultable securities, *Econometrica* 72, 1377–1407.

Couderc, Fabien, and Olivier Renault, 2004, Times-to-default: Life cycle, global and industry cycle impacts, Working paper, University of Geneva.

Crosbie, Peter J., and Jeffrey R. Bohn, 2002, Modeling default risk, Technical report, KMV, LLC.

Das, Sanjiv, Darrell Duffie, Nikunj Kapadia, and Leandro Saita, 2007, Common failings: How corporate defaults are correlated, *Journal of Finance* 62, 93–117.

Delloy, Martin, Jean-David Fermanian, and Mohammed Sbai, 2005, Estimation of a reduced-form credit portfolio model and extensions to dynamic frailties, Working paper, BNP-Paribas.

Demptser, Arthur P., Nan M. Laird, and Donald B. 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.

Duffie, Darrell, and David Lando, 2001, Term structures of credit spreads with incomplete accounting information, *Econometrica* 69, 633–664.

Duffie, Darrell, Leandro Saita, and Ke Wang, 2007, Multi-period corporate default prediction with stochastic covariates, *Journal of Financial Economics* 83, 635–665.

Eraker, Bjørn, Michael Johannes, and Nicholas Polson, 2003, The impact of jumps in volatility and returns, *Journal of Finance* 58, 1269–1300.

Fisher, Edwin, Robert Heinkel, and Josef Zechner, 1989, Dynamic capital structure choice: Theory and tests, *Journal of Finance* 44, 19–40.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin, 2004, *Bayesian Data Analysis*, 2nd Edition (Chapman and Hall, New York).

Geman, Stuart, and Donald Geman, 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721–741.

Giesecke, Kay, 2004, Correlated default with incomplete information, *Journal of Banking and Finance* 28, 1521–1545.

Gordy, Michael, 2003, A risk-factor model foundation for ratings-based capital rules, *Journal of Financial Intermediation* 12, 199–232.

Hammersley, John, and Peter Clifford, 1970, Markov fields on finite graphs and lattices, Working paper, Oxford University.

Hastings, W. Keith, 1970, Monte-Carlo sampling methods using Markov chains and their applications, *Biometrika* 57, 97–109.

Hillegeist, Stephen A., Elizabeth K. Keating, Donald P. Cram, and Kyle G. Lundstedt, 2004, Assessing the probability of bankruptcy, *Review of Accounting Studies* 9, 5–34.

Jacobsen, Martin, 2006, *Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes* (Birkhäuser, Boston).

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jegadeesh, Narasimhan, and Sheridan Titman, 2001, Profitability of momentum strategies: An evaluation of alternative explanations, *Journal of Finance* 66, 699–720.

Johannes, Michael, and Nicholas Polson, 2003, MCMC methods for continuous-time financial econometrics, in Yacine Aït-Sahalia and Lars Hansen, eds.: *Handbook of Financial Econometrics*, Working paper, Columbia University, forthcoming.

Kass, Robert, and Adrian Raftery, 1995, Bayes factors, *Journal of The American Statistical Association* 90, 773–795.

Kavvathas, Dimitrios, 2001, Estimating credit rating transition probabilities for corporate bonds, Working paper, University of Chicago.

Kealhofer, Stephen, 2003, Quantifying credit risk I: Default prediction, *Financial Analysts Journal*, January–February, 30–44.

Koopman, Siem, André Lucas, and André Monteiro, 2008, The multi-state latent factor intensity model for credit rating transitions, *Journal of Econometrics* 142, 399–424.

Lando, David, and Torben Skødeberg, 2002, Analyzing rating transitions and rating drift with continuous observations, *Journal of Banking and Finance* 26, 423–444.

Lando, David, and Mads Stenbo Nielsen, 2008, Correlation in corporate defaults: Contagion or conditional independence?, Working paper, Copenhagen Business School.

Lane, William R., Stephen W. Looney, and James W. Wansley, 1986, An application of the Cox proportional hazards model to bank failure, *Journal of Banking and Finance* 10, 511–531.

Lee, Suk Hun, and Jorge L. Urrutia, 1996, Analysis and prediction of insolvency in the property-liability insurance industry: A comparison of logit and hazard models, *Journal of Risk and Insurance* 63, 121–130.

Leland, Hayne, 1994, Corporate debt value, bond covenants, and optimal capital structure, *Journal of Finance* 49, 1213–1252.

McDonald, Cynthia G., and Linda M. Van de Gucht, 1999, High-yield bond default and call risks, *Review of Economics and Statistics* 81, 409–419.

Mengersen, Kerrie, and Richard L. Tweedie, 1996, Rates of convergence of the Hastings and Metropolis algorithms, *Annals of Statistics* 24, 101–121.

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

Metropolis, Nicholas, and Stanislaw Ulam, 1949, The Monte Carlo method, *Journal of The American Statistical Association* 44, 335–341.

Nielsen, Søren Feodor, 2000, The stochastic EM algorithm: Estimation and asymptotic results, Working paper, University of Copenhagen.

Pickles, Andrew, and Robert Crouchery, 1995, A comparison of frailty models for multivariate survival data, *Statistics in Medicine* 14, 1447–1461.

Rabiner, Lawrence R., 1989, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE* 77, 257–285.

Robert, Christian, and George Casella, 2005, *Monte Carlo Statistical Methods*, 2nd edition (Springer-Verlag, New York).

Schönbucher, Philipp, 2003, Information driven default contagion, Working Paper, ETH, Zurich.

Shumway, Tyler, 2001, Forecasting bankruptcy more accurately: A simple hazard model, *Journal of Business* 74, 101–124.

Smith, Adrian F. M., and David J. Spiegelhalter, 1980, Bayes factors and choice criteria for linear models, *Journal of the Royal Statistical Society: Series B* 42, 213–220.

Tanner, Martin A., 1998, *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions*, 3rd edition (Springer-Verlag, New York).

Vasicek, Oldrich, 1987, Probability of loss on loan portfolio, Working paper, KMV Corporation.

Vassalou, Maria, and Yuhang Xing, 2004, Default risk in equity returns, *Journal of Finance* 59, 831–868.

Wei, Greg C. G., and Martin A. 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, Chien Fu Jeff, 1983, On the convergence properties of the EM algorithm, *Annals of Statistics* 11, 95–103.

Yu, Fan, 2005, Accounting transparency and the term structure of credit spreads, *Journal of Financial Economics* 75, 53–84.

Zhang, Gaiyan, and Philippe Jorion, 2007, Good and bad credit contagion: Evidence from credit default swaps, *Journal of Financial Economics* 84, 860–883.