Frailty Correlated Default*

Darrell Duffie, Andreas Eckner,
Guillaume Horel, and Leandro Saita

Current version: January 27, 2008

Abstract

This paper shows that 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 CDO default losses are typically measured for economic-capital and rating purposes, our empirical results indicate that conventionally based estimates are downward biased by a full order of magnitude on test portfolios. Our estimates are based on U.S. public non-financial firms existing 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.

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

*We are grateful for financial support from Moody's Corporation and Morgan Stanley, and for research assistance from Sabri Oncu and Vineet Bhagwat. We are also grateful for remarks from Torben Andersen, André Lucas, Richard Cantor, Stav Gaon, Tyler Shumway, and especially Michael Johannes. This revision is much improved because of suggestions by a referee, an associate editor, and Campbell Harvey. We are thankful to Moodys and to Ed Altman for generous assistance with data. Duffie is at The Graduate School of Business, Stanford University. Eckner and Horel are at Merrill Lynch. Saita is at Lehman Brothers.

1 Introduction

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 bank capital requirements and for 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. Uncertainty about the current level of this variable, as well as exposure to future movements of this variable, both 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 market-wide factors. Portfolio loss distributions are typically based on the correlating influence of such observable factors. For example, rating agencies typically estimate the probability of losses to senior collateralized debt obligations (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. If the underlying borrowers are commonly exposed to important risk factors whose effect is not captured by co-movements of borrower ratings, however, 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 default losses at a confidence level such as 99.9%. 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. 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. It is inevitable, however, that not all relevant risk factors that are potentially observable by the econometrician will end up being included in the model. 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, sub-prime 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 a mere handful of major commercial banks total in excess of $80 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 sub-prime 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, 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 have had lower ratings and lower prices (a concavity effect), on top of any related effects of risk premia. Perhaps more modelers should have thought to look for, might have found, and might have included in their models proxies for this moral-hazard effect. It seems optimistic to believe that they would have done so, for despite the clear incentives, many apparently did not. Presumably it is not easy, ex ante, to include all important default covariates. The next event of extreme portfolio loss could be based on a different omitted variable. It 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 Rule to update the posterior distribution of unobserved risk factors whenever defaults arrive with a timing that is more clustered or less clustered than would be expected based on the observable risk factors alone. In the statistics literature treating 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 have extended the prior statistical methodology so as to allow a frailty covariate to vary over time according to an autoregressive time-series specification, and using 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 (2007) have 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. The rest of this introductory section gives an overview of our modeling approach and results, a summary of the related literature, and a description of our dataset. Section 2 specifies the precise probabilistic model for the joint distribution of default times. Section 3 summarizes some of the properties of the fitted model and of the posterior distribution of the frailty variable, given the entire sample. Section 4 examines the fit of the model and addresses some potential sources of misspecification, providing robustness checks. Section 5 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.

1.1 Summary of Model 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 volatility-corrected leverage measure whose construction is described 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 macro-economic covariates, including interest rates and market-wide 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

$$\lambda_{it} = \exp\left(\alpha + \beta \cdot W_{it} + \gamma \cdot U_t + Y_t + Z_i\right),$$

for coefficients α, β, and γ 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 non-financial firms for which we were able to obtain matching data from the several data sets on which we rely. Our data 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 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 1813 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 from 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 95-percentile and 99-percentile 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 2001-2002 was a period of particularly severe corporate defaults. In Section 3, 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 macro-variable, 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 1986-1991, and persistently lower in others, for example during the mid-nineties. 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 re-estimate the model including as an additional observable macro-covariate the trailing average realized rate of default,[1] 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.

1.2 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

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

by Moody's KMV, a leading provider of estimates of default probabilities for essentially all publicly traded firms (see Crosbie and Bohn (2002) and Kealhofer (2003)). Based on this theoretical foundation, we include distance to default as a covariate into our model for default risk.

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 one has the effect of frailty. For reasons of tractability, we have chosen a reduced-form specification of frailty.

Altman (1968) and Beaver (1968) were among the first to estimate reduced-form statistical models of the likelihood of default of a firm within one accounting period, using accounting data.[2] 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 (2004) 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 statistical modeling of frailty, under which default intensities include the expected effect of unobservable covariates. Yu (2005) finds empirical evidence that, other things equal, a reduction in the measured precision of

[2]Early in the empirical literature on default time distributions is the work of Lane, Looney, and Wansley (1986) on bank default prediction, using time-independent covariates. Lee and Urrutia (1996) used a duration model based on a Weibull distribution of default times. Duration models based on time-varying covariates include those of McDonald and Van de Gucht (1999), in a model of the timing of high-yield bond defaults and call exercises. Related duration analysis by Shumway (2001), Kavvathas (2001), Chava and Jarrow (2004), and Hillegeist, Keating, Cram, and Lundstedt (2004) predict bankruptcy. Shumway (2001) uses a discrete duration model with time-dependent covariates. Duffie, Saita, and Wang (2006) provide maximum likelihood estimates of term structures of default probabilities by using a joint model for default intensities and the dynamics of the underlying time-varying covariates.

accounting variables is associated with a widening of credit spreads. Das, Duffie, Kapadia, and Saita (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-model specifications of default prediction, such as those of Couderc and Renault (2004), Shumway (2001), and Duffie, Saita, and Wang (2006), by allowing for dynamic unobserved covariates. Independently of our work, and with a similar thrust, Delloy, Fermanian, and Sbai (2005) and Koopman, Lucas, and Monteiro (2005) estimate dynamic frailty models of rating transitions. They suppose that the only observable firm-specific default covariate is an agency credit rating, and assume 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 (2006), 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.

1.3 Data

Our dataset, 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 (2006) and Das, Duffie, Kapadia, and Saita (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 (2006). Our dataset contains 402,434 firm-months of data between January 1979 and March 2004. Because of the manner in which we define defaults, it is appropriate to use data only up to December 2003.

For the total of 2,793 companies in this improved dataset, Table 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 (2006) for an in-depth description of the construction of the dataset and an exact definition of these event types.

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

Table I: Number of firm exits of each type between 1979 and 2004.

Figure 1 shows the total number of defaults (bankruptcies and other defaults) in each year. Moody's 13th annual corporate bond default study[3] 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:

- Distance to default, a volatility-adjusted measure of leverage. Our method of construction, based on market equity data and Compustat book liability data, is that used by Vassalou and Xing (2004), Crosbie and Bohn (2002), and Hillegeist, Keating, Cram, and Lundstedt (2004). Although the conventional approach to measuring distance to default involves some rough approximations, Bharath and Shumway (2004) provide evidence that default prediction is relatively robust to varying the proposed measure with some relatively simple alternatives.

- 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

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

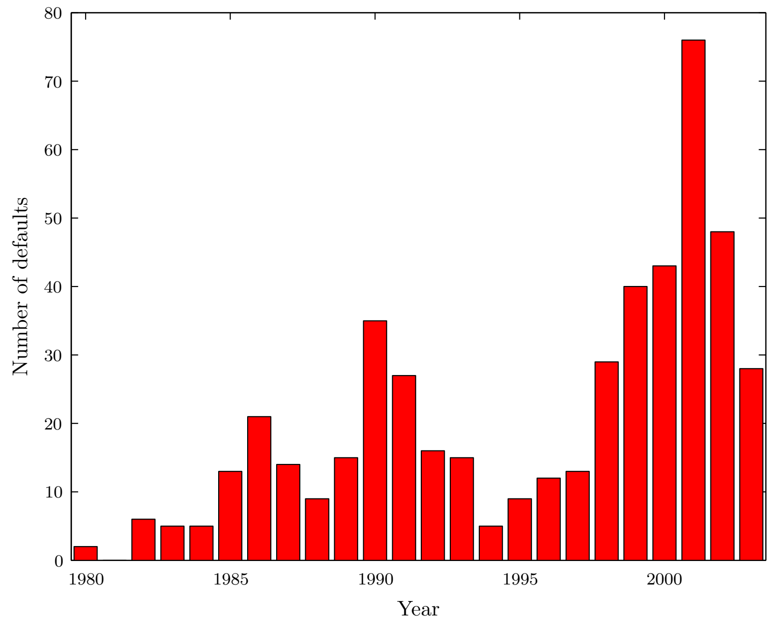

Figure 1: The number of defaults in our dataset for each year between 1980 and 2003.

unobserved factor that has an influence on default risk beyond that of the firm's measured distance of default.

- 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 performing poorly (and defaults are high).

- 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 (2006) give a detailed description of these covariates and discuss their relative importance in modeling corporate default intensities. As robustness checks, we have examined 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.[4] Each of these was 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 1970-71. 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.

2 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 non-negative progressively-measurable process λ is the default intensity of the borrower if a martingale is defined by $1_{\tau \leq t} - \int_0^t \lambda_s 1_{\tau > s}\, ds$. This means that, as of time t, if the borrower has not yet defaulted, λ_t is the conditional mean arrival rate of default, measured in events per unit of time.

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

[4] Size may be associated with market power, management strategies, or borrowing ability, all of which may affect the risk of failure. For example, it might be easier for a big firm to re-negotiate with its creditors to postpone the payment of debt, or to raise new funds to pay old debt. In a "too-big-to-fail" sense, firm size may also negatively influence failure intensity. The statistical significance of size as a determinant of failure risk has been documented by Shumway (2001). For our data and our measure of firm size, however, this covariate did not play a statistically significant role.

of the form $\lambda_{it} = \Lambda\left(S_i(X_t), \theta\right)$, 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, we cannot use standard estimation methods.

We depart from the traditional complete-information doubly-stochastic assumption because it has been shown by Das, Duffie, Kapadia, and Saita (2007) to understate default correlation for our dataset. One may entertain various alternatives. For example, we have mentioned the possibility of "contagion," by which the default by 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 macro-economic 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 dataset.

We let $W_{it} = (1, U_{it}, V_t)$ be the vector of observed covariates for company i (including a constant).[5] We let T_i be the last observation time of company i, which could be the time of a default or another form of exit. While we take the first appearance time t_i to be deterministic, our results are not affected

[5] 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.

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

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

where D_i is the default indicator process of company i (which is 0 before default, 1 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 $\{Y_s : 0 \leq s \leq t\}$.

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 below.[6]

Before considering the effect of other exits such as mergers and acquisitions, the maximum likelihood estimators of $\mathcal{F}_t$-conditional survival probabilities, portfolio-loss distributions, and related quantities such as default correlations, are obtained under the usual smoothness conditions by treating the maximum likelihood estimators of the parameters as though they are the true parameters (γ, θ).[7] We will also examine the implications of Bayesian uncertainty regarding certain key parameters.

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

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

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" $E\left(e^{-\int_t^T \overline{\lambda}_{is}\, ds} \,|\, \mathcal{F}_t\right)$. Rather, for a firm that has survived to time t, the probability of survival to time T (again neglecting other exits) is $E\left(e^{-\int_t^T \lambda_{is}\, ds} \,|\, \mathcal{F}_t\right)$. 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 $E\left(e^{-\int_t^T \lambda_{is}\, ds} \,|\, \mathcal{G}_t\right)$. See Collin-Dufresne, Goldstein, and Huggonier (2004) for another approach to this calculation.

[7]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 $E\left(e^{-\int_t^T (\mu_{is} + \lambda_{is})\, ds} \,|\, \mathcal{F}_t\right)$. 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.

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 observation, Proposition 2 of Duffie, Saita, and Wang (2006) would imply a likelihood of the data at the parameters (γ, θ) of the form

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

We simplify by supposing that the frailty process Y is independent of the observable covariate process W. With respect to the econometrician's limited filtration $(\mathcal{F}_t)$, the likelihood is then

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

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.[8]

Most of our empirical results are properties of the maximum likelihood estimator (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 maximum likelihood estimator for (γ, θ) because we have assumed that all forms of exit are jointly doubly-stochastic on the artificially enlarged information filtration $(\mathcal{G}_t)$.

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

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 non-negative 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 types of unobserved fundamental common shocks to default intensities, each of which has an impact that decays over time. For example, as suggested in the introduction, 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, 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 impact of such successive 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 4, 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 their distribution can be inferred only from relatively sparse default time 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 the appendix.

3 Major Empirical Results

This section shows the estimated model and its implications for the distribution of portfolio default losses relative to a model without frailty.

3.1 The Fitted Model

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

	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

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 $\theta = \widehat{\theta}$. The mean reversion and volatility parameters are based on monthly time intervals.

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 one-year trailing stock return covariate proposed by Shumway (2001) has a highly significant impact on default intensities. Perhaps it is a proxy for firm-specific information that is not captured by distance to default.[9] 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 for the 3-month Treasury bill rate, which is smaller without frailty, but remains statistically significant.

	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

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 $\theta = \widehat{\theta}$.

[9] 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.

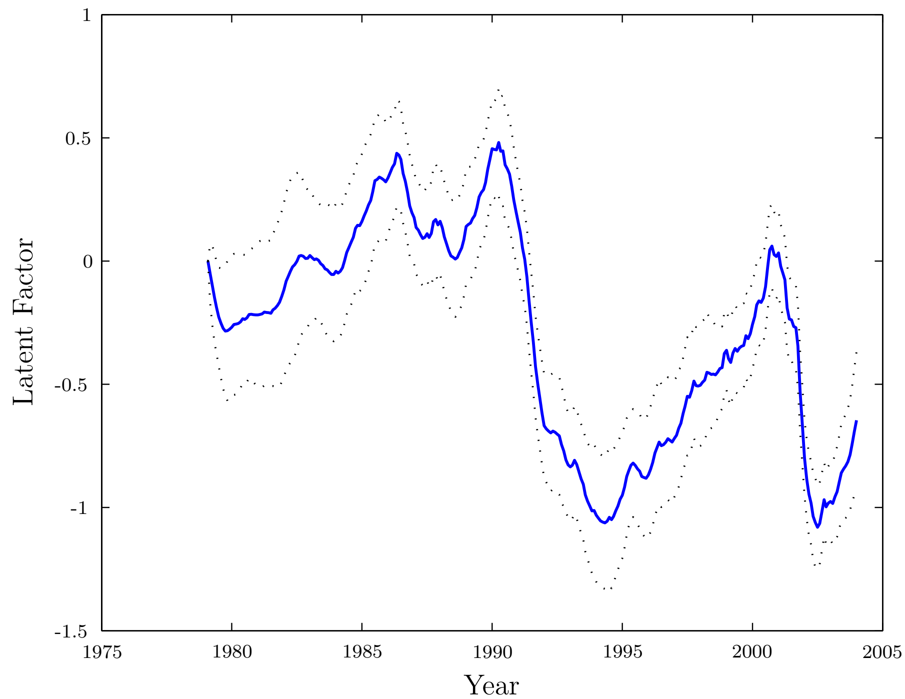

Figure 2: Conditional posterior mean $E(\eta Y_t \,|\, \mathcal{F}_T)$ of the scaled latent Ornstein-Uhlenbeck frailty variable, with one-standard-deviation bands based on the $\mathcal{F}_T$-conditional variance of Y_t.

3.2 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 the appendix. Figure 2 shows the conditional mean of the latent factor, 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.[10]

While Figure 2 illustrates the posterior distribution of the frailty variable Y_t given all information available $\mathcal{F}_T$ at the final time T of the sample period, most applications of a default-risk model would call for the posterior distribution of Y_t given the current information $\mathcal{F}_t$. For example, 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 can apply the "forward-backward algorithm" due to Baum, Petrie, Soules, and Weiss (1970). Appendix D explains how we apply this algorithm in our setting.

Figure 3 compares the conditional density of Y_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 Y_t when conditioning on only $\mathcal{F}_t$ (the data available up to and including January 2000). Given the additional information available at the end of 2004, the $\mathcal{F}_T$-conditional distribution of Y_t is more concentrated than that obtained by conditioning on only the concurrently available information $\mathcal{F}_t$. The posterior mean of Y_t given the information available in January 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.

3.3 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

[10] 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.

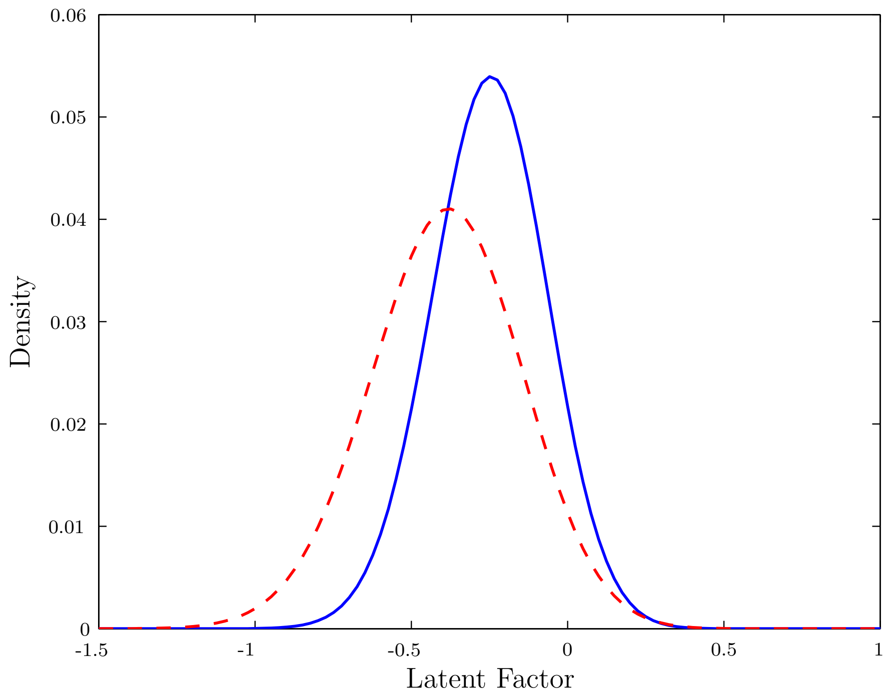

Figure 3: Conditional posterior density of the scaled frailty factor, ηY_t, for t in January 2000, given $\mathcal{F}_T$, that is, given all data, (solid line), and given only contemporaneously available data in $\mathcal{F}_t$ (dashed line). These densities are calculated using the forward-backward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

number of defaults during the subsequent five years, January 1998 through 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 posterior marginal distribution for each firm's default time, but the joint distribution of default times varies among 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 that has the same $\mathcal{F}_t$-conditional distribution as Y. For model (b), the initial condition Y_{it} of Y_i is

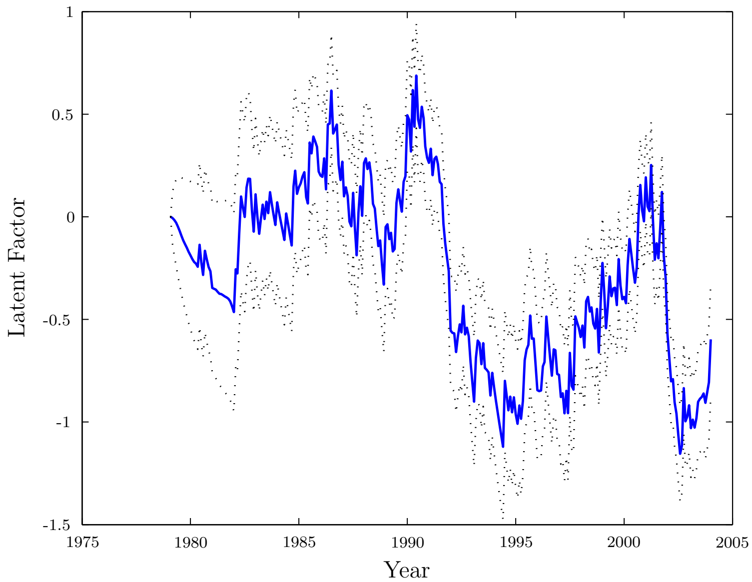

Figure 4: Conditional mean $E(\eta Y_t \mid \mathcal{F}_t)$ and conditional one-standard-deviation bands of the scaled frailty variable, given only contemporaneously available data ($\mathcal{F}_t$).

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 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).

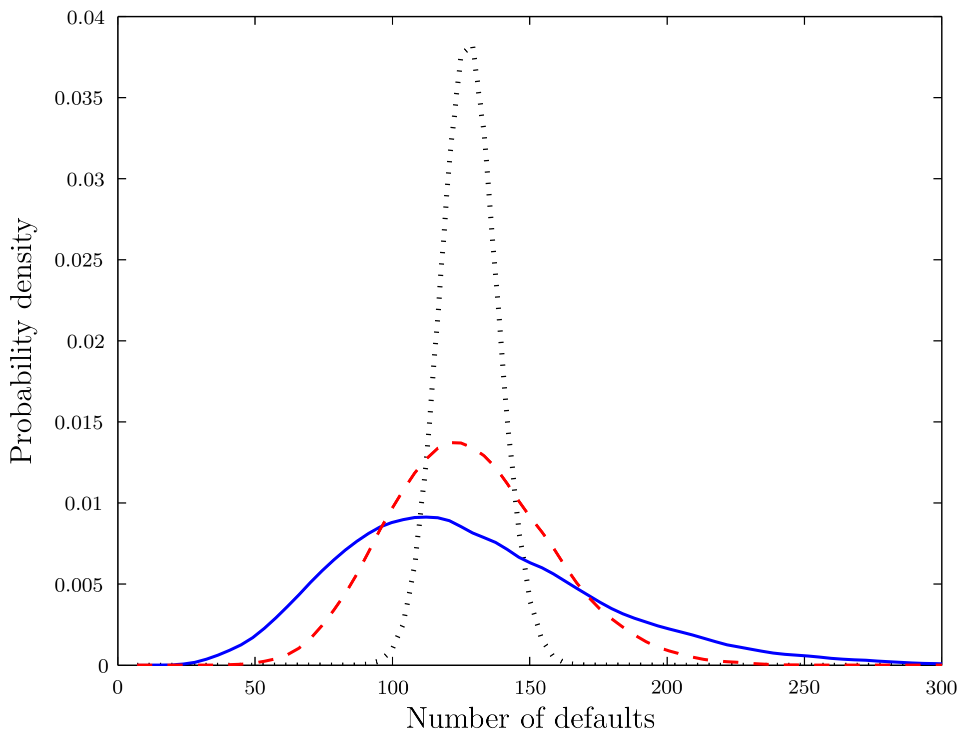

Figure 5: The conditional probability density, given $\mathcal{F}_t$ for t in January 1998, of the total number of defaults within five years from the portfolio of all active firms at January 1998, in (a) the fitted model with frailty (solid line), (b) a hypothetical model in which the common frailty 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 5) applied to a Monte-Carlo generated empirical distribution.

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 95-percentile and 99-percentile losses of the 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 95-percentile and 99-percentile of 180 and 204 defaults, respectively. The actual number of defaults in our dataset during this time period was 195.

The 95-percentile and 99-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 91-percentile of the distribution implied by the fitted frailty model, therefore constituting a quite extreme event. On the other hand, taking the hindsight bias into account, in that our analysis was partially motivated by the high number of defaults in the years 2001 and 2002, the occurrence of 195 defaults might be viewed as an only moderately extreme event for the frailty model.

4 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.

4.1 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

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

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

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

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). See Smith and Spiegelhalter (1980) for a discussion of the penalizing nature of the Bayes factor, sometimes referred to as the "fully automatic Occam's razor." 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.[11]

4.2 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

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

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. Appendix E provides 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.

4.3 Unobserved Heterogeneity

It may be that a substantial portion of the differences among 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

$$\lambda_{it} = e^{X_{it}\beta + \gamma Y_t} Z_i = \widetilde{\lambda}_{it} e^{\gamma Y_t} Z_i, \tag{7}$$

where $Z_1, \ldots, Z_m$ are independently and identically gamma-distributed[12] 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 1 without loss of generality, we found that maximum likelihood estimation does not pin down the variance of Z_i to any reasonable precision with our limited set of data. We anticipate that far larger datasets 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

[12]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 chose the gamma distribution.

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 $Z_1, \ldots, Z_n$.

Appendix C shows 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.

4.4 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 next investigate. We are also interested in examining our ability to learn these parameters, in a Bayesian sense. 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

$$\sigma_\infty^2 \equiv \lim_{s \to \infty} \operatorname{var}\left(Y_s \,|\, \mathcal{G}_t\right) = \lim_{s \to \infty} \operatorname{var}\left(Y_s \,|\, Y_t\right) = \frac{\sigma^2}{2\kappa}.$$

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 three 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

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

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

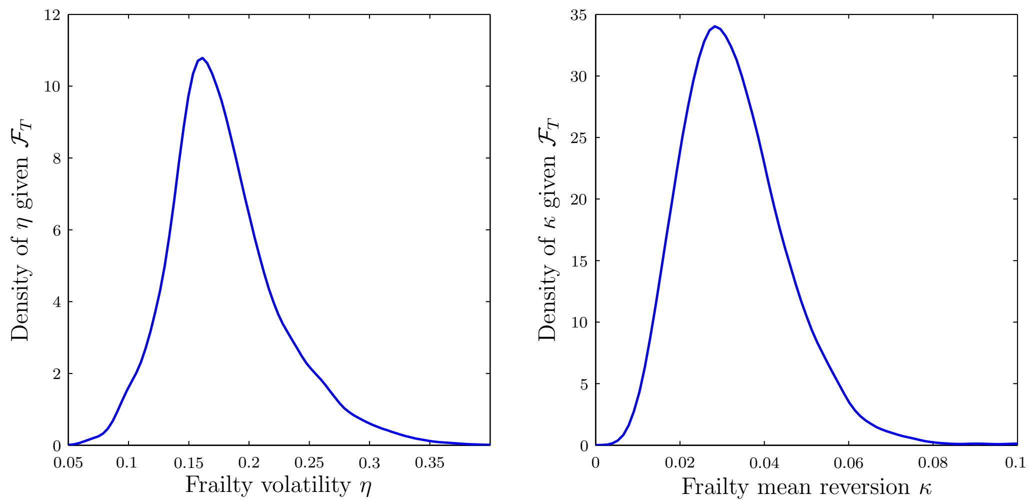

Figure 6: Marginal posterior densities, given $\mathcal{F}_T$, of the frailty volatility parameter η and the frailty mean reversion rate κ in the Bayesian approach of Section 4.4.

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 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 99-percentile of the portfolio default distribution is 265 defaults. Taking posterior parameter uncertainty into account, this quantile rises to 275 defaults.

4.5 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

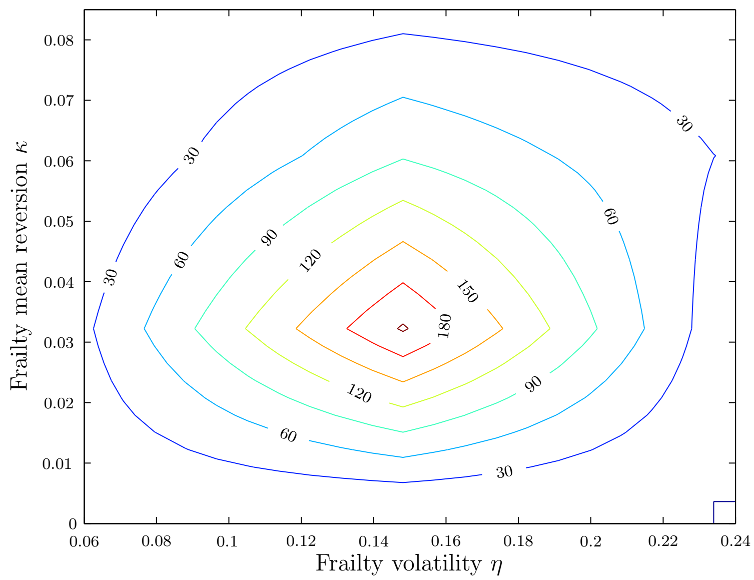

Figure 7: Isocurves of the joint posterior density, given $\mathcal{F}_T$, of the frailty volatility parameter η and mean reversion rate κ.

firms conditional on the path of the covariate process if we include average realized 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 significant auxiliary role. For example, when trailing average default rates 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

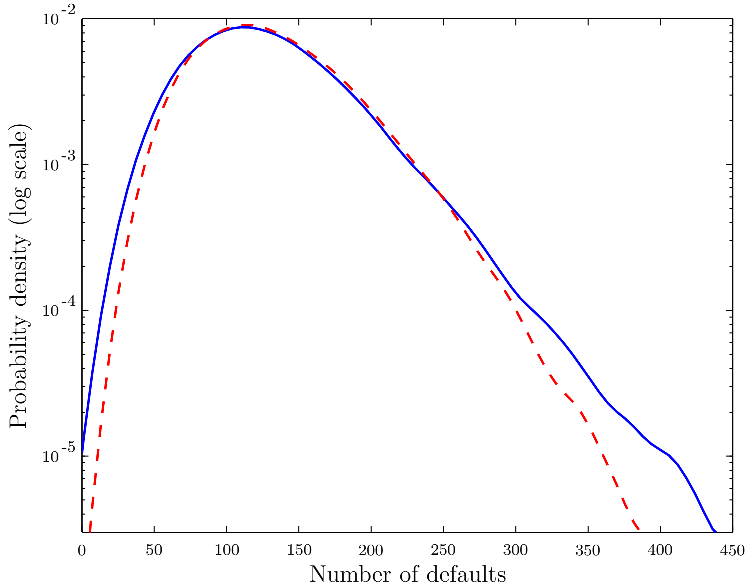

Figure 8: Density, on a logarithmic scale, of the number of defaults in the portfolio when fixing the volatility and mean reversion parameter at their MLE estimates (dashed line), and in the Bayesian estimation framework (solid line). The density estimates were obtained by applying a Gaussian kernel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

alternative to frailty in capturing the distribution of portfolio tail losses. In the sense of the tests described in the next sub-section, it is not.

4.6 Model Accuracy

We turn to the ability of our model to capture portfolio default risk.

In terms of firm-by-firm default prediction, Duffie, Saita, and Wang (2006) showed 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. Results reported in Appendix F show that accuracy ratios with frailty are essentially the same as those without. Likewise, accuracy ratios are roughly unaffected by adding trailing average default rate as a

	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

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 $\theta = \widehat{\theta}$.

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 dataset the model-implied distribution of the number of defaults during the subsequent twelve months, and 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 variable but including trailing one-year average default rate as an additional covariate. The quantiles of the two models without frailty variable seem to cluster around 0 and 1, which suggests that these models underestimate the probabilities of unusually low portfolio losses and of unusually high portfolio losses. For example, in 1994 the realized number of defaults lies below the estimated 1-percentile of the portfolio default distribution for the model without frailty, while in 1990 and 2001 the realized number of defaults lies above the 99.9-percentile of estimated distribution. For the model that in addition includes the trailing one-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.

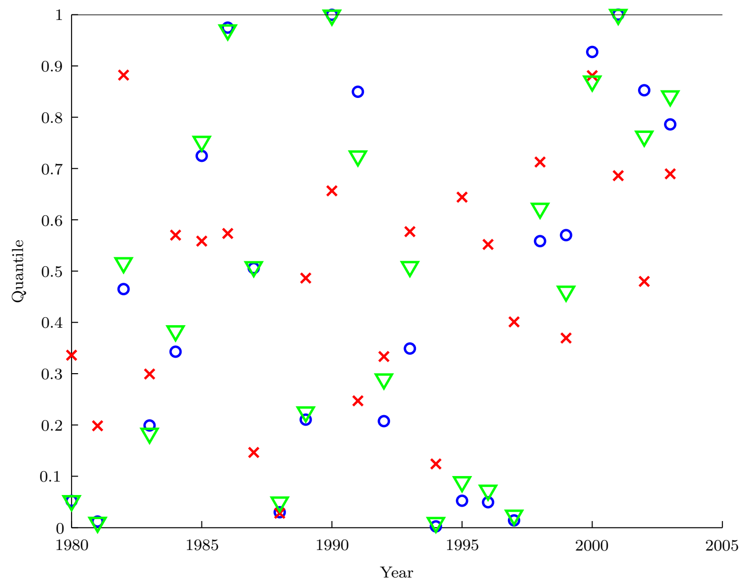

Figure 9: Quantile of the realized number of defaults with respect to the predicted one-year portfolio loss distribution as implied by the model with frailty variable (crosses), without frailty variable (circles), and without frailty variable but with trailing one-year average default rate as covariate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

Moreover, the forecasting errors for the two models without frailty tend to be serially correlated over time, which is most evident for the periods 1994-1997 as well as 2000-2003. 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 frailty variable but with trailing one-year average default rate as 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.

5 Concluding Remarks

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 collateralized debt obligations (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.

This paper finds significant evidence among U.S. corporates of a common unobserved source of default risk that increases default correlation and extreme portfolio loss risk above and beyond that implied by observable common and correlated macroeconomic and firm-specific sources of default risk. We offer a new model of corporate default intensities in the presence of a time-varying latent frailty factor, and with unobserved heterogeneity.

Applying this model to data for U.S. firms between January 1979 and March 2004, we find that corporate default rates vary over time well beyond levels that can be explained by a model that includes only observable covariates. In particular, the posterior distribution of the frailty variable shows that the expected rate of corporate defaults was much higher in 1989-1990 and 2001-2002, and much lower during the mid-nineties and in 2003-2004, than those implied by an analogous model without frailty. An out-of-sample test for data between 1980 and 2003 indicates that a model without frailty significantly underestimates the probability of extreme positive as well as negative events in portfolios of corporate credits, while a model with frailty gives a more accurate assessment of credit risk on the portfolio level.

We estimate that the frailty variable represents a common unobservable factor in default intensities with an annual volatility of over 40%. The estimated rate of mean reversion of the frailty factor, 1.8% per month, implies that when defaults cluster in time to a degree that is above and beyond that suggested by observable default-risk factors, the half life of the impact of this unobservable factor is roughly 3 years. We show that the mean-reversion rate is difficult to pin down with the available data. Without mean reversion, however, the variance of the frailty effect would explode over time.

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

Appendices

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 (2006). This model, summarized in Appendix G, 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 expectation-maximization (EM) algorithm (see Demptser, Laird, and Rubin (1977)), an iterative method for the computation of the maximum likelihood estimator 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 (MLE) of the intensity parameter vector θ involves the following steps:

0. Initialize an estimate of $\theta = (\beta, \eta, \kappa)$ at $\theta^{(0)} = (\widehat{\beta}, 0.05, 0)$, where $\widehat{\beta}$ 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.

1. (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 Ornstein-Uhlenbeck frailty process Y. We do this with the Gibbs sampler described in Appendix B. We let

$$Q\left(\theta, \theta^{(k)}\right) = E_{\theta^{(k)}}\left(\log \mathcal{L}\left(\theta \mid W, Y, D\right)\right) \tag{8}$$

$$= \int \log \mathcal{L}\left(\theta \mid W, y, D\right) p_Y\left(y \mid W, D, \theta^{(k)}\right)\, dy, \tag{9}$$

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

$$\widehat{Q}\left(\theta, \theta^{(k)}\right) = \frac{1}{n} \sum_{j=1}^{n} \log \mathcal{L}\left(\theta \,|\, W, Y^{(j)}, D\right). \tag{10}$$

2. (M-step) Maximize $\widehat{Q}(\theta, \theta^{(k)})$ with respect to the parameter vector θ, for example by Newton-Raphson. The maximizing choice of θ is the new parameter estimate $\theta^{(k+1)}$.

3. Replace k with $k+1$, and return to Step 1, repeating the E-step and the M-step until reasonable numerical convergence is achieved.

One can show (Demptser, Laird, and Rubin (1977) or Gelman, Carlin, Stern, and Rubin (2004)) that $\mathcal{L}(\gamma, \theta^{(k+1)} \,|\, W, D) \geq \mathcal{L}(\gamma, \theta^{(k)} \,|\, W, D)$. That is, the observed data likelihood (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 (see for example Proposition 10.1.6 of Cappé, Moulines, and Rydén (2005)) imply that

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

and

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

The Hessian matrix of the expected complete-data likelihood (9) can therefore be used to estimate asymptotic standard errors for the MLE parameter estimators.

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

B Applying the Gibbs Sampler with Frailty

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

$$E\left[f(Y)\,|\,W,D,\theta\right] = \int f(y)\,p_Y(y\,|\,W,D,\theta)\,dy \tag{11}$$

is given by

$$\frac{1}{N}\sum_{n=1}^{N} f\left(Y^{(n)}\right). \tag{12}$$

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

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

the posterior mean of the latent Ornstein-Uhlenbeck frailty process.

The linchpin to MCMC is that the joint distribution of the frailty path $Y = \{Y_t : 0 \leq t \leq T\}$ can be broken 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 following set of 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 $Y_{(-t)} = \{Y_s : s \neq t\}$, is somewhat tractable, as shown below. This allows us to use the Gibbs sampler (Geman and Geman (1984) or Gelman, Carlin, Stern, and Rubin (2004)) to draw whole sample paths from the posterior distribution of $\{Y_t : 0 \leq t \leq T\}$ by the algorithm:

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

We usually discard the first several hundred paths as a "burn-in" sample, because initially the Gibbs sampler has not approximately converged[13] to the posterior distribution of $\{Y_t : 0 \leq t \leq T\}$.

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

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

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

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

where we repeatedly made 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

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

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

$$p\left(Y_t \,|\, W, D, Y_{(-t)}, \theta\right) \propto \mathcal{L}\left(\theta \,|\, W, Y, D\right) \cdot p\left(Y_t \,|\, Y_{t-1}, \theta\right) \cdot p\left(Y_{t+1} \,|\, Y_t, \theta\right) \quad (13)$$

Equation (13) 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)).[14]

[14] 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 where the conditional density is not known explicitly.

We use the proposal density $q(Y_t^{(n)} \,|\, W, D, Y^{(n-1)}, \theta) = N(Y_t^{(n-1)}, 4)$, 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 increments[15]. The Metropolis-Hastings step to sample Y_t in the $n-th$ iteration of the Gibbs sampler therefore works as follows:

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

2. Compute

$$\alpha\left(y, Y_t^{(n)}\right) = \min\left(\frac{\mathcal{L}\left(\theta \,|\, W, Y_{(-t)}^{(n-1)}, Y_t = y, D\right)}{\mathcal{L}\left(\theta \,|\, W, Y^{(n-1)}, D\right)}, 1\right). \tag{14}$$

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

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

The choice of the acceptance probability (14) ensures that the Markov chain $\{Y_t^{(n)} : n \geq 1\}$ satisfies the detailed balance equation

$$p\left(y_1 | W, D, Y_{(-t)}, \theta\right) \phi_{y_1,4}\left(y_2\right) \alpha\left(y_1, y_2\right) = p\left(y_2 | W, D, Y_{(-t)}, \theta\right) \phi_{y_2,4}\left(y_1\right) \alpha\left(y_2, y_1\right),$$

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 $p\left(Y_t | W, D, Y_{(-t)}, \theta\right)$ as its stationary distribution (see for example Theorem 7.2 in Robert and Casella (2005)).

C With Unobserved Heterogeneity

The Monte Carlo EM algorithm described in Appendix A and the Gibbs sampler described in Appendix B are extended to treat unobserved heterogeneity as follows.

The extension of the Monte Carlo EM algorithm is:

[15] We calculated 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.

0. Initialize $Z_i^{(0)} = 1$ for $1 \le i \le m$ and initialize $\theta^{(0)} = (\widehat{\beta}, 0.05, 0)$, where $\widehat{\beta}$ is the maximum likelihood estimator of β in the model without frailty.

1. (Monte-Carlo E-step.) Given the current parameter estimate $\theta^{(k)}$, draw samples $\left(Y^{(j)}, Z^{(j)}\right)$ for $j = 1, \ldots, n$ from the joint posterior distribution $p_{Y,Z}(\,\cdot \mid W, D, \theta^{(k)})$ of the frailty sample path $Y = \{Y_t : 0 \le t \le T\}$ and the vector $Z = (Z_i : 1 \le i \le m)$ of unobserved heterogeneity variables. This can be done, for example, by using the Gibbs sampler described below. The expected complete-data log-likelihood is now given by

$$\begin{aligned} Q\left(\theta, \theta^{(k)}\right) &= E_{\theta^{(k)}}\left(\log \mathcal{L}\left(\theta \mid W, Y, Z, D\right)\right) \\ &= \int \log \mathcal{L}\left(\theta \mid W, y, z, D\right) p_{Y,Z}\left(y, z \mid W, D, \theta^{(k)}\right) \, dy \, dz. \end{aligned} \tag{15}$$

   Using the sample paths generated by the Gibbs sampler, (15) can be approximated by

$$\widehat{Q}\left(\theta, \theta^{(k)}\right) = \frac{1}{n} \sum_{j=1}^{n} \log \mathcal{L}\left(\theta \mid W, Y^{(j)}, Z^{(j)}, D\right). \tag{16}$$

2. (M-step.) Maximize $\widehat{Q}(\theta, \theta^{(k)})$ with respect to the parameter vector θ, using the Newton-Raphson algorithm. Set the new parameter estimate $\theta^{(k+1)}$ equal to this maximizing value.

3. Replace k with $k+1$, and return to Step 2, repeating the MC E-step and the M-step until reasonable numerical convergence.

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

0. Initialize $Y_t = 0$ for $t = 0, \ldots, T$. Initialize $Z_i = 1$ for $i = 1, \ldots, m$.

1. For $t = 1, \ldots, T$ draw a new value of Y_t from its conditional distribution given Y_{t-1}, Y_{t+1} and the current values for Z_i. This can be done using a straightforward modification of the Metropolis-Hastings algorithm described in Appendix B by treating $\log Z_i$ as an additional covariate with corresponding coefficient in (1) equal to 1.

2. For $i = 1, \ldots, m$, draw the unobserved heterogeneity variables $Z_1, \ldots, Z_m$ from their conditional distributions given the current path of Y. See below.

3. Store the sample path $\{Y_t, 0 \le t \le T\}$ and the variables $\{Z_i : 1 \le i \le m\}$. Return to Step 1 and repeat until the desired number of scenarios has been drawn, discarding the first several hundred as a burn-in sample.

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

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

by conditional independence of the unobserved heterogeneity variables Z_i. In order to draw Z from its conditional distribution, it therefore suffices to show how to draw the Z_i's from their conditional distributions. Recall that we have chosen the heterogeneity variables Z_i to be gamma distributed with mean 1 and standard deviation 0.5. A short calculation shows that in this case the density parameters a and b are both 4. Applying Bayes' rule,

$$\begin{aligned} p\left(Z_i \mid W, Y, D, \theta\right) &\propto p_\Gamma\left(Z_i; 4, 4\right) \mathcal{L}\left(\theta \mid W_i, Y, Z_i, D_i\right) \\ &\propto Z_i^3 e^{-4Z_i} e^{-\sum_{t=t_i}^{T_i} \lambda_{it} \Delta t} \prod_{t=t_i}^{T_i} \left[D_{it} \lambda_{it} \Delta t + (1 - D_{it})\right], \end{aligned} \tag{17}$$

where $p_\Gamma\left(\,\cdot\,; a, b\right)$ is the density function of a Gamma distribution with parameters a and b. Plugging (7) into (17) gives

$$p\left(Z_i \mid W, Y, D, \theta\right) \propto Z_i^3 e^{-4Z_i} \exp\left(-\sum_{t=t_i}^{T_i} \tilde{\lambda}_{it} e^{\gamma Y_t} Z_i\right) \prod_{t=t_i}^{T_i} \left[D_{it} \lambda_{it} \Delta_t + (1 - D_{it})\right]$$

$$= Z_i^3 e^{-4Z_i} \exp\left(-A_i Z_i\right) \cdot \left\{ \begin{array}{cl} B_i Z_i & \text{if company } i \text{ did default} \\ 1 & \text{if company } i \text{ did not default} \end{array} \right\}, \tag{18}$$

for company specific constants A_i and B_i. The factors in (18) can be combined to give

$$p\left(Z_i \mid W_i, Y, D_i, \theta\right) = p_\Gamma\left(Z_i; 4 + D_{i,T_i}, 4 + A_i\right). \tag{19}$$

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

Table V shows the MLE of the covariate parameter vector β and the frailty parameters η and κ together with estimated standard errors. We see that, while including unobserved heterogeneity decreases the coefficient η of dependence (sometimes called volatility) of the default intensity on the OU frailty process Y from 0.125 to 0.112, our general conclusions regarding the economic significance of the covariates and the importance of including a time-varying frailty variable remain. Moreover, Figure 10 shows that the posterior distribution of the frailty qualitatively remains essentially the same.

	Coefficient	Std. error	t-statistic
constant	−0.895	0.134	−6.7
distance to default	−1.662	0.047	−35.0
trailing stock return	−0.427	0.074	−5.8
3-month T-bill rate	−0.241	0.027	−9.0
trailing S&P 500 return	1.507	0.309	4.9
latent factor volatility	0.112	0.022	5.0
latent factor mean reversion	0.061	0.017	3.5

Table V: Maximum likelihood estimates of the intensity parameters in the model with frailty and unobserved heterogeneity. Asymptotic standard errors are computed using the Hessian matrix of the likelihood function at $\theta = \widehat{\theta}$.

D Forward-Backward Filtering for Frailty

For this, we let $R(t) = \{i : D_{i,t} = 0,\, t_i \leq t \leq T_i\}$ denote the set of firms that are alive at time t, and $\Delta R(t) = \{i \in R(t-1) : D_{it} = 1,\, t_i \leq t \leq T_i\}$ 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

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

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

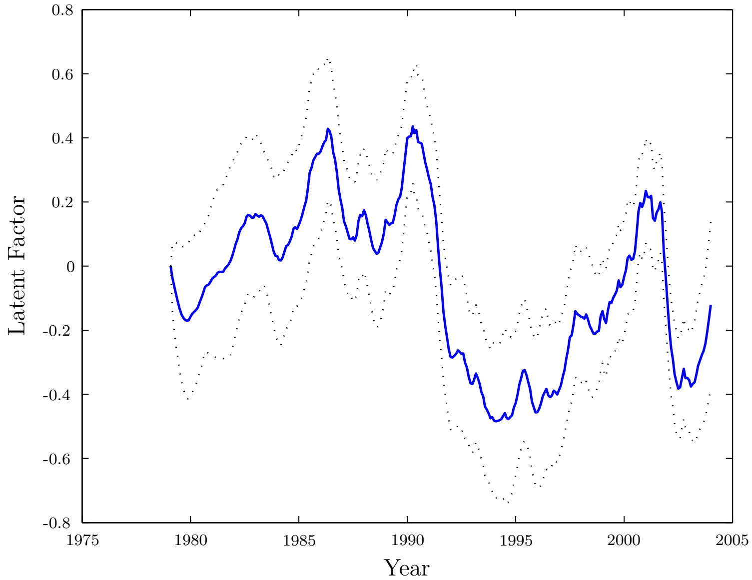

Figure 10: Conditional posterior mean $\{E(\eta Y_t \,|\, \mathcal{F}_T) : 0 \le t \le T\}$ with one-standard-deviation bands, for the scaled Ornstein-Uhlenbeck frailty variable ηY_t in the model that also incorporates unobserved heterogeneity.

by the recursion

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

where $p(Y_t \,|\, Y_{t-1}, \theta)$ 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 0.

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

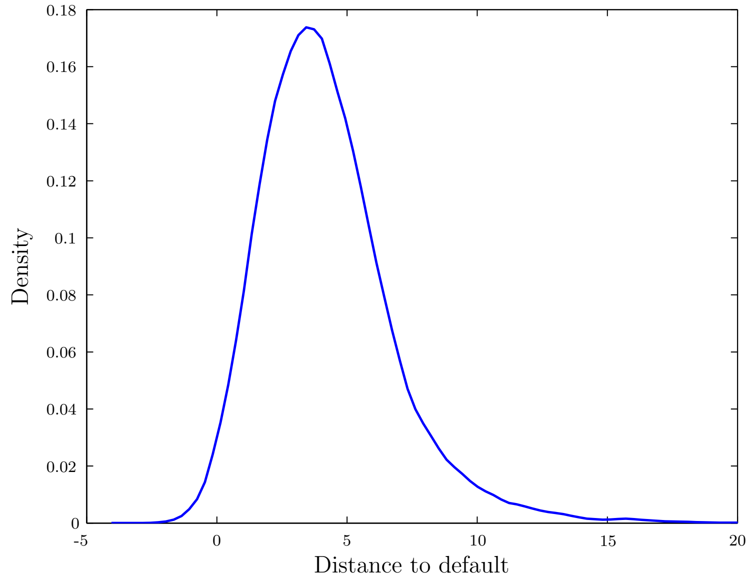

Figure 11: Population density estimate of distance to default for 402,434 firm-months between January 1979 and March 2004. The estimate was obtained by applying a Gaussian kernel smoother (bandwidth equal to 0.2) to the empirical distribution.

recursion for the marginal density

$$\begin{aligned}\overline{\alpha}_{t|T}\left(y_t\right) &= \frac{1}{c_{t+1}}\int p\left(y_t\,|\,y_{t-1},\theta\right)\mathcal{L}_{t+1}\left(\theta\,|\,W_{t+1},y_{t+1},D_{t+1}\right)\overline{\alpha}_{t+1|T}\left(y_{t+1}\right)dy_{t+1}\\ p\left(y_t\,|\,\mathcal{F}_T\right) &= p\left(y_t\,|\,\mathcal{F}_t\right)\overline{\alpha}_{t|T}\left(y_t\right),\end{aligned}$$

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

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

E Non-Linearity Check

So far, see (1), we have assumed a linear dependence of the log-intensity on the covariates. This assumption might be overly restrictive, especially in the

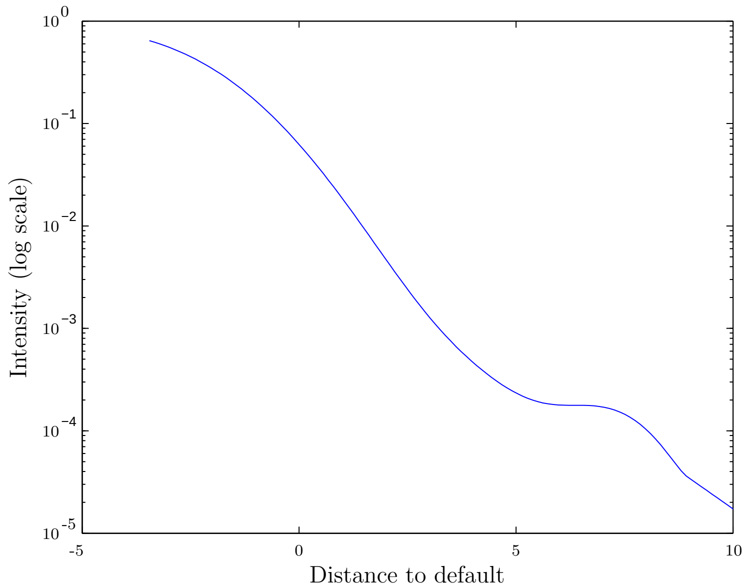

Figure 12: Non-parametric estimate of the dependence of annual default frequency on the current level of distance to default (DTD). For values of distance to default less than 9, a Gaussian kernel smoother with bandwidth of 1 was used to obtain the intensity estimate. For DTD larger than 9, a log-linear relationship was assumed.

case of the distance to default (DTD), which explains most of the variation of default intensities across companies and across time. It is indeed possible that, if the response of the true log-intensity to DTD is faster than linear, then the latent variable in our current formulation would be higher when DTDs go well below normal and vice versa.

To check the robustness of our findings with respect to the linearity assumptions, we therefore re-estimate the model using a non-parametric model for the contribution of distance to default, replacing $DTD(t)$ with $-\log U(t)$ in (1), where $U(t) = f(DTD(t))$, and $f(x)$ is the non-parametric kernel-smoothed fit of 1-year frequency of default in our sample at distance to default of x. Figure 11 shows the historical occurrence of different levels of distance-to-default for 402,434 firm-months, while Figure 12 shows the estimated relationship between the current level of DTD and the annualized default intensity. For values of $DTD \leq 9$, a Gaussian kernel smoother with bandwidth equal to one was used to obtain the intensity estimate, whereas due to lack of data the tail of the distribution was approximated by a log-linear relationship, smoothly extending the graph in Figure 11.

Using this extension, we re-estimate the model parameters as before. Table VI shows the estimated covariate parameter vector $\widehat{\beta}$ and frailty parameters $\widehat{\eta}$ and $\widehat{\kappa}$ together with "asymptotic" estimates of standard errors.

	Coefficient	Std. Error	t-statistic
Constant	2.279	0.194	11.8
$-\log(f(DTD))$	-1.198	0.042	-28.6
Trailing stock return	-0.618	0.075	-8.3
3-month T-bill rate	-0.238	0.030	-8.1
Trailing S&P 500 return	1.577	0.312	5.1
Latent factor volatility	0.128	0.020	6.3
Latent factor mean reversion	0.043	0.009	4.8

Table VI: Maximum likelihood estimates of the intensity parameters θ in the model with frailty, replacing distance to default with $-\log(f(DTD))$, where DTD is distance to default and $f(\,\cdot\,)$ is the non-parametric kernel estimated mapping from DTD to annual default frequency, illustrated in Figure 12. The frailty volatility is the coefficient η of dependence of the default intensity on the standard Ornstein-Uhlenbeck frailty process Y. Estimated asymptotic standard errors were computed using the Hessian matrix of the expected complete data log-likelihood at $\theta = \widehat{\theta}$.

Comparing Tables II and VI, we see that none of the coefficients linking a firm's covariates to its default intensity has changed noteworthily. In particular, the coefficient now linking the default intensity and $-\log U(t)$ is virtually the same as the coefficient for DTD in the original model. Note however that the intercept has changed from -1.20 to 2.28. This difference is due to the fact that $-\log U(t) \approx DTD - 3.5$. Indeed, for the intercept at $DTD = 0$ in Figure 12 we have $10^{-1.5} \approx 0.032 \approx \exp(-1.20 - 2.28)$. In addition, the posterior path of the latent Ornstein-Uhlenbeck frailty variable looks as before (not shown here). In view of these findings we decided the keep the model with a log-linear relationship between a firm's DTD and its default intensity.

F Out-of-Sample Accuracy Ratios

This appendix provides out-of-sample accuracy ratios for our model and some variants.

Given a future time horizon and a particular default prediction model, the "power curve" for out-of-sample default prediction is the function f that maps any x in $[0,1]$ to the fraction $f(x)$ of firms that default before the time horizon that were initially ranked by the model in the lowest fraction x of the population. For example, for the model without frailty, on average over 1993 to 2004, the highest quintile of firms ranked by estimated default probability at the beginning of a year accounted for 92% of firms defaulting within one year. Power curves for the model without frailty are provided in Duffie, Saita, and Wang (2006).

The "accuracy ratio" of a model with power curve f is defined as

$$2\int_0^1 (f(x) - r(x))\, dx,$$

where $x \mapsto r(x) = x$, the identity, is the expected power curve of a completely uninformative model, one that sorts firms randomly. So, a random-sort model has an expected accuracy ratio of 0. A "crystal ball" perfect-sort model has an accuracy ratio of 1 minus the total ex-post default rate. The accuracy ratio is a benchmark for comparing the default prediction accuracy of different models.

Duffie, Saita, and Wang (2006), who do not allow for frailty, already find accuracy ratios are an improvement on those of any other model in the available literature. A comparison of the accuracy ratios found in Duffie, Saita, and Wang (2006) with those for the frailty model shown in Figure 13 shows that accuracy ratios are essentially unaffected by allowing for frailty. This may be due to the fact that, because of the dominant role of the distance-to-default covariate, firms generally tend to be ranked roughly in order of their distances to default, which of course do not depend on the intensity model. Accuracy ratios, however, measure ordinal (ranking) quality, and do not fully capture the out-of-sample ability of a model to estimate the magnitudes of default probabilities. Our results, not reported here, suggest that the frailty model that we have proposed does not improve the out-of-sample accuracy of the magnitudes of firm-level estimates of default probabilities, over the model without frailty.

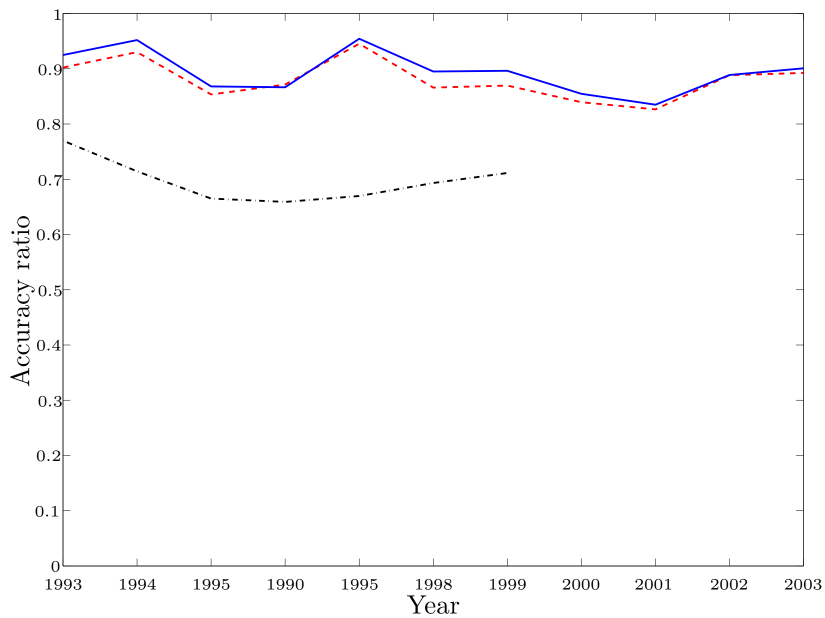

Figure 13: Out-of-sample accuracy ratios (ARs), based on models estimated with data up to December 1992. The solid line provides one-year-ahead ARs based on the model without frailty. The dashed line shows one-year-ahead ARs for the model with frailty. The dash-dot line shows 5-year-ahead ARs for the model with frailty.

Figure 14 shows accuracy ratios for the variant of our model that replaces the unobserved frailty variable Y with the one-year trailing average default rate. The accuracy ratios are comparable to those of the model with frailty.

G Summary of Covariate Time-Series Model

We summarize here the particular parameterization of the time-series model for the covariates that we adopt from Duffie, Saita, and Wang (2006). Because of the high-dimensional state-vector, which includes the macroeconomic covariates as well as the distance to default and size of each of almost 3000 firms, we have opted for a Gaussian first-order vector auto-regressive time series model, with the following simple structure.

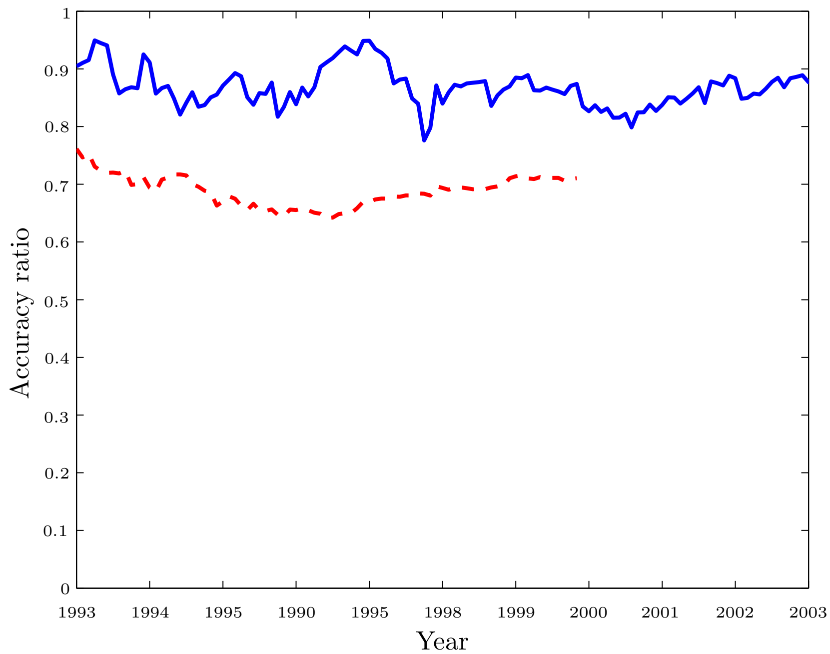

Figure 14: Out-of-sample one-year (solid line) and five-year (dashed line) accuracy ratios (ARs), based on the DSW model enhanced with the trailing one-year default rate as an additional covariate.

The 3-month and 10-year Treasury rates, r_{1t} and r_{2t}, respectively, are modeled by taking $r_t = (r_{1t}, r_{2t})'$ to satisfy

$$r_{t+1} = r_t + k_r(\theta_r - r_t) + C_r \epsilon_{t+1} \, ,$$

where $\epsilon_1, \epsilon_2, \ldots$ are independent standard-normal vectors, C_r is a 2×2 lower-triangular matrix, and the time step is one month. Provided C_r is of full rank, This is a simple arbitrage-free two-factor affine term-structure model. Maximum-likelihood parameter estimates and standard errors are reported in Duffie, Saita, and Wang (2006).

For the distance to default D_{it} and log-assets V_{it} of firm i, and the trailing one-year S&P500 return, S_t, we assume that

$$\begin{bmatrix} D_{i,t+1} \\ V_{i,t+1} \end{bmatrix} = \begin{bmatrix} D_{it} \\ V_{it} \end{bmatrix} + \begin{bmatrix} k_D & 0 \\ 0 & k_V \end{bmatrix} \left(\begin{bmatrix} \theta_{iD} \\ \theta_{iV} \end{bmatrix} - \begin{bmatrix} D_{it} \\ V_{it} \end{bmatrix} \right) + \\ + \begin{bmatrix} b \cdot (\theta_r - r_t) \\ 0 \end{bmatrix} + \begin{bmatrix} \sigma_D & 0 \\ 0 & \sigma_V \end{bmatrix} \eta_{i,t+1} \, , \tag{20}$$

$$S_{t+1} = S_t + k_S(\theta_S - S_t) + \xi_{t+1}, \tag{21}$$

where

$$\begin{aligned} \eta_{it} &= Az_{it} + Bw_t \, , \\ \xi_t &= \alpha_S u_t + \gamma_S w_t, \end{aligned} \tag{22}$$

for $\{z_{1t}, z_{2t}, \ldots, z_{nt}, w_t : t \geq 1\}$ that are *iid* 2-dimensional standard-normal, all independent of $\{u_1, u_2, \ldots\}$, which are independent standard normals. The 2×2 matrices A and B have $A_{12} = B_{12} = 0$, and are normalized so that the diagonal elements of $AA' + BB'$ are 1. For estimation, some such standardization is necessary because the joint distribution of η_{it} (over all i) is determined by the 6 (non-unit) entries in $AA' + BB'$ and BB'. Our standardization makes A and B equal to the Cholesky decompositions of AA' and BB', respectively. For simplicity, although this is unrealistic, we assume that ϵ is independent of (η, ξ). The maximum-likelihood parameter estimates, with standard errors, are provided in Duffie, Saita, and Wang (2006), and are relatively unsurprising.

References

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis, and the Prediction Of Corporate Bankruptcy. *Journal of Finance 23*, 589–609.

Baum, L. E., T. P. Petrie, G. Soules, and N. Weiss (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics 41*, 164–171.

Beaver, B. (1968). Market Prices, Financial Ratios, and the Prediction Of Failure. *Journal of Accounting Research Autumn*, 170–192.

Besag, J. (1974). Spatial Interaction and The Statistical Analysis Of Lattice Systems. *Journal of the Royal Statistical Association: Series B 36*, 192–236.

Bharath, S. and T. Shumway (2004). Forecasting Default with the KMV-Merton Model. Working Paper, University of Michigan.

Black, F. and M. Scholes (1973). The Pricing Of Options and Corporate Liabilities. *Journal of Political Economy 81*, 637–654.

Cappé, O., E. Moulines, and T. Rydén (2005). *Inference in Hidden Markov Models*. Springer Series in Statistics.

Celeux, G. and J. Diebolt (1986). The SEM Algorithm: A Probabilistic Teacher Algorithm Derived From The EM Algorith For The Mixture Problem. *Computational Statistics Quaterly 2*, 73–82.

Chava, S. and R. Jarrow (2004). Bankruptcy Prediction with Industry Effects. *Review of Finance 8*, 537–569.

Chernobai, A., P. Jorion, and F. Yu (2007). The Determinants of Operational Losses. Unpublished Manuscript, University of Michigan.

Collin-Dufresne, P., R. Goldstein, and J. Helwege (2003). Is Credit Event Risk Priced? Modeling Contagion via The Updating Of Beliefs. Working Paper, Haas School, University of California, Berkeley.

Collin-Dufresne, P., R. Goldstein, and J. Huggonier (2004). A General Formula For Valuing Defaultable Securities. *Econometrica 72*, 1377–1407.

Couderc, F. and O. Renault (2004). Times-to-Default: Life Cycle, Global and Industry Cycle Impacts. Working paper, University of Geneva.

Crosbie, P. J. and J. R. Bohn (2002). Modeling Default Risk. Technical Report, KMV, LLC.

Das, S., D. Duffie, N. Kapadia, and L. Saita (2007). Common Failings: How Corporate Defaults are Correlated. *Journal of Finance 64*. forthcoming.

Delloy, M., J.-D. Fermanian, and M. Sbai (2005). Estimation of a Reduced-Form Credit Portfolio Model and Extensions to Dynamic Frailties. Preliminary Version, BNP-Paribas.

Demptser, A., N. Laird, and D. Rubin (1977). Maximum Likelihood Estimation From Incomplete Data via The EM Algorithm (with Discussion). *Journal of the Royal Statistical Society: Series B 39*, 1–38.

Duffie, D. and D. Lando (2001). Term Structures Of Credit Spreads with Incomplete Accounting Information. *Econometrica 69*, 633–664.

Duffie, D., L. Saita, and K. Wang (2006). Multi-Period Corporate Default Prediction with Stochastic Covariates. *Journal of Financial Economics forthcoming.*

Eraker, B., M. Johannes, and N. Polson (2003). The Impact Of Jumps in Volatility and Returns. *Journal of Finance 58*, 1269–1300.

Fisher, E., R. Heinkel, and J. Zechner (1989). Dynamic Capital Structure Choice: Theory and Tests. *Journal of Finance 44*, 19–40.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004). *Bayesian Data Analysis* (Second edition). Chapman and Hall.

Geman, S. and D. Geman (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 6*, 721–741.

Giesecke, K. (2004). Correlated Default With Incomplete Information. *Journal of Banking and Finance 28*, 1521–1545.

Gordy, M. (2003). A Risk-Factor Model Foundation For Ratings-Based Capital Rules. *Journal of Financial Intermediation 12*, 199–232.

Hammersley, J. and P. Clifford (1970). Markov Fields On Finite Graphs and Lattices. Unpublished Manuscript.

Hastings, W. (1970). Monte-Carlo Sampling Methods using Markov Chains and Their Applications. *Biometrika 57*, 97–109.

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

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

Jegadeesh, N. and S. Titman (1993). Returns To Buying Winners and Selling Losers: Implications For Stock Market Efficiency. *Journal of Finance 48*, 65–91.

Jegadeesh, N. and S. Titman (2001). Profitability Of Momentum Strategies: An Evaluation Of Alternative Explanations. *Journal of Finance 66*, 699–720.

Johannes, M. and N. Polson (2003). MCMC Methods For Continuous-Time Financial Econometrics. Unpublished Manuscript, Columbia University.

Kass, R. and A. Raftery (1995). Bayes factors. *Journal of The American Statistical Association 90*, 773–795.

Kavvathas, D. (2001). Estimating Credit Rating Transition Probabilities for Corporate Bonds. Working paper, University of Chicago.

Kealhofer, S. (2003). Quantifying Credit Risk I: Default Prediction. *Financial Analysts Journal*, January–February, 30–44.

Koopman, S., A. Lucas, and A. Monteiro (2005). The Multi-State Latent Factor Intensity Model for Credit Rating Transitions. Working Paper, Tinbergen Institute.

Lando, D. and T. Skødeberg (2002). Analyzing Rating Transitions and Rating Drift with Continuous Observations. *Journal of Banking and Finance 26*, 423–444.

Lane, W. R., S. W. Looney, and J. W. Wansley (1986). An Application Of the Cox Proportional Hazards Model to Bank Failure. *Journal of Banking and Finance 10*, 511–531.

Lee, S. H. and J. L. Urrutia (1996). Analysis and Prediction Of Insolvency in the Property-Liability Insurance Industry: A Comparison Of Logit and Hazard Models. *The Journal of Risk and Insurance 63*, 121–130.

Leland, H. (1994). Corporate Debt Value, Bond Covenants, and Optimal Capital Structure. *Journal of Finance 49*, 1213–1252.

McDonald, C. G. and L. M. Van de Gucht (1999). High-Yield Bond Default and Call Risks. *Review of Economics and Statistics 81*, 409–419.

Mengersen, K. and R. L. Tweedie (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Annals of Statistics 24*, 101–121.

Merton, R. C. (1974). On the Pricing Of Corporate Debt: The Risk Structure Of Interest Rates. *Journal of Finance 29*, 449–470.

Metropolis, N. and S. Ulam (1949). The Monte Carlo method. *Journal of The American Statistical Association 44*, 335–341.

Nielsen, S. F. (2000). The Stochastic EM algorithm: Estimation and Asymptotic Results. Department of Theoretical Statistics, University of Copenhagen.

Pickles, A. and R. Crouchery (1995). A Comparison of Frailty Models for Multivariate Survival Data. *Statistics in Medicine 14*, 1447–1461.

Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE 77*, 257–285.

Robert, C. and G. Casella (2005). *Monte Carlo Statistical Methods* (Second edition). Springer Texts in Statistics.

Schönbucher, P. (2003). Information Driven Default Contagion. Working Paper, ETH, Zurich.

Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business 74*, 101–124.

Smith, A. F. M. and D. J. Spiegelhalter (1980). Bayes Factors and Choice Criteria For Linear Models. *Journal of the Royal Statistical Society: Series B 42*, 213–220.

Tanner, M. A. (1998). *Tools For Statistical Inference : Methods For The Exploration Of Posterior Distributions and Likelihood Functions* (Third edition). New York: Springer-Verlag.

Vasicek, O. (1987). Probability Of Loss On Loan Portfolio. Working Paper, KMV Corporation.

Vassalou, M. and Y. Xing (2004). Default Risk in Equity Returns. *Journal of Finance 59*, 831–868.

Wei, G. and M. Tanner (1990). A Monte Carlo Implementation Of The EM Algorithm and The Poor Man's Data Augmentation Algorithm. *Journal of The American Statistical Association 85*, 699–704.

Wu, C. (1983). On the Convergence Properties of the EM Algorithm. *Annals of Statistics 11*, 95–103.

Yu, F. (2005). Accounting Transparency and The Term Structure Of Credit Spreads. *Journal of Financial Economics 75*, 53–84.

Zhang, G. (2004). Intra-Industry Credit Contagion: Evidence From The Credit Default Swap Market and The Stock Market. Working paper, University of California, Irvine.