MANAGEMENT SCIENCE
Vol. 57, No. 7, July 2011, pp. 1267–1287
ISSN 0025-1909 | EISSN 1526-5501 | 11 | 5707 | 1267

informs®

DOI 10.1287/mnsc.1110.1345

Modeling the Loss Distribution

Sudheer Chava
College of Management, Georgia Institute of Technology, Atlanta, Georgia 30308, sudheer.chava@mgt.gatech.edu

Catalina Stefanescu
ESMT European School of Management and Technology, 10178 Berlin, Germany,
catalina.stefanescu-cuntze@esmt.org

Stuart Turnbull
Bauer College of Business, University of Houston, Houston, Texas 77204, sturnbull@uh.edu

In this paper, we focus on modeling and predicting the loss distribution for credit risky assets such as bonds and loans. We model the probability of default and the recovery rate given default based on shared covariates. We develop a new class of default models that explicitly accounts for sector specific and regime dependent unobservable heterogeneity in firm characteristics. Based on the analysis of a large default and recovery data set over the horizon 1980–2008, we document that the specification of the default model has a major impact on the predicted loss distribution, whereas the specification of the recovery model is less important. In particular, we find evidence that industry factors and regime dynamics affect the performance of default models, implying that the appropriate choice of default models for loss prediction will depend on the credit cycle and on portfolio characteristics. Finally, we show that default probabilities and recovery rates predicted *out of sample* are negatively correlated and that the magnitude of the correlation varies with seniority class, industry, and credit cycle.

Key words: loss distribution; default prediction; recovery rates; Basel II
History: Received May 11, 2009; accepted February 10, 2011, by Wei Xiong, finance. Published online in *Articles in Advance* May 17, 2011.

1. Introduction

The predicted loss distribution is a basic input for calculating the loan loss reserves and the economic capital and for computing portfolio risk metrics such as value-at-risk and expected shortfall. For an individual asset such as a bond or loan, the loss distribution depends on the probability of default and on the recovery rate given default. The negative correlation between these two factors has been documented empirically by academics and recognized by regulators in Basel II. The 2007–2009 credit crisis, when the default frequency increased and the recovery rates decreased simultaneously, provided a painful reminder to investors that both dimensions are very important.

In this paper, we generate the loss distribution by modeling the default probabilities and the recovery rates with shared covariates, and we analyze their interdependence. Although there is a large literature on predicting defaults and a separate emergent literature on modeling recovery rates, ours appears to be the first study that explicitly models the loss distribution by modeling defaults and recoveries and that investigates the impact of the choice of default and recovery models on the loss distribution. This paper makes three contributions.

The paper's main contribution is to elucidate the impact that the choice of default and recovery models has on the predicted loss distribution. Using an extensive default and recovery data set, we investigate four default models and three recovery models inspired by extant finance literature and use them to predict the out-of-sample loss distributions in different portfolios of bonds. We choose a one-year horizon, as suggested by regulatory requirements. We first find that based on the standard performance metrics for default and recovery prediction used in the literature, there is virtually no performance difference either between the four default models or between the three recovery models. This suggests that the choice of any particular combination of default and recovery models should have little impact on the predicted loss distribution.

We show, however, that the specification of the default model has a crucial impact on the predicted loss distribution and that the relative predictive performance of default models depends on the credit cycle. For example, in years with high default frequency, the default model from Duffie et al. (2007) predicts loss distributions where the expected loss is closest to the actual realized loss, whereas in years with low default frequency, other default models lead to expected loss estimates that are closer to the realized loss. We further show that controlling for industry-level heterogeneity has a significant positive impact on forecasting performance and that the choice of default model for loss prediction should depend on the industry characteristics of the firms in the portfolio. Although the Duffie et al. (2007) model has good prediction performance for telecommunication firms over the out-of-sample horizon 1996–2008, for the sample of manufacturing firms, the best model is a hybrid obtained by adding two macroeconomic covariates to Shumway's (2001) specification. Additionally, we find evidence that the choice of recovery model has a smaller impact on the predicted loss distribution than does the choice of default model.

Our second contribution is to propose a new class of default models that explicitly accounts for the dynamic effects of unobservable measurement errors, incomplete information, and industry-level heterogeneity on default probabilities. In practice investors usually have only incomplete information about the true state of a firm. There are differences between firms that affect their default probabilities but are not directly observable, such as variations in managerial styles, in the skill sets of workers, and in firm culture. Even differences in such areas as production skills, resource usage, cost control, and risk management are only partially revealed in accounting statements. This unobservable firm heterogeneity can be modeled using shared frailties models, where the hazard rate is multiplied by a latent random variable common to all firms in a given industry group (Gagliardini and Gourieroux 2003, Duffie et al. 2007) and which can be assumed to follow a stochastic diffusion process (Yashin and Manton 1997, Duffie et al. 2009). Motivated by Dai et al. (2007) and Li et al. (2010) who demonstrate the importance of using regimes to model changes in the economy, we extend the shared frailty approach by developing regime frailty models at industry level. In these models, a multiplicative factor magnifies the impact of group-specific frailties during periods of economic distress, and potential distress is assessed annually at industry level, thus allowing a dynamic change in the industry default risk. In a range of portfolios over different industries and different years during the out-of-sample horizon, we find that accounting for regime changes in the unobserved heterogeneity significantly improves the performance of default models for predicting the loss distribution and the total number of defaults in a portfolio.

Our regime based frailty approach complements the latent frailty approach in Duffie et al. (2009). As Li et al. (2010) point out in the context of dynamic term structure models, it is useful to relate the latent state variables to fundamental macroeconomic variables. In the context of frailty, we take a step in that direction by constructing a frailty factor at industry level with two regimes based on whether or not the industry is distressed. Our model with a different frailty for each industry complements the model in Duffie et al. (2009) with a single economy-wide frailty and is computationally easier to implement. The frailty dynamics in our model, although limited on some dimensions as compared to Duffie et al. (2009), vary over industry economic cycles and in addition convey useful industry-specific information about defaults, which is not possible from a model with an economy-wide frailty. Indeed, through extensive empirical analysis we demonstrate that controlling for industry-level heterogeneity significantly improves out-of-sample forecasting performance.

Our third contribution is to analyze the dependence of predicted default probabilities and predicted recovery rates in our joint modeling framework based on shared covariates and to generate new insights related to this dependence. Basel II recognizes that changes in the probability of default and in the loss given default are generally related for most assets, and it requires financial institutions using the advanced internal ratings based approach to recognize this dependence (Basel Committee on Banking Supervision 2005a). We show that the default probability and recovery rate predicted *out of sample* are negatively correlated, consistent with regulatory requirements. The magnitude of the correlation varies with industry, seniority, and macroeconomic conditions.

The testing of default and recovery rate prediction models is particularly relevant in light of Basel II, which allows banks to develop their own estimates of default probabilities and of recovery rates so that these reflect the nature of their portfolios. Banks now have an incentive to use their own estimates to model the loss distributions; however, they face challenges raised by the plethora of existing default and recovery models. Basel II stipulates that banks' estimates of default probabilities and recovery rates are subject to supervisory review, but it doesn't explicitly indicate how financial institutions can show that their quantitative models are reasonable in order to gain regulatory approval for the resulting estimates. In fact, regulators themselves are also unsure how to assess whether the models that an institution uses are reasonable (Basel Committee on Banking Supervision 2005b). This issue has assumed greater importance since the 2007–2009 credit crisis, which sent a forceful reminder to regulators about the importance of modeling and validating the loss distribution for credit risky assets. The crisis has also prompted a new round of regulations designed to reduce systemic risk, Basel III, which increased the minimum capital ratios to hold against future losses. With its methodological focus and analysis insights, this paper contributes to the Basel II and Basel III debate by addressing the challenging issues related to predicting loss distributions.

This paper is related to several different strands of previous research. There is a large and growing literature devoted to the modeling of the probability of default—see Shumway (2001), Chava and Jarrow (2004), Campbell et al. (2008), Duffie et al. (2007), Bharath and Shumway (2008), and Duffie et al. (2009). An extensive survey of methodologies is given in Altman and Hotchkiss (2005). There is also an emerging literature addressing the modeling of the determinants of the recovery rate given default. A survey of empirical evidence regarding the properties of recovery rates is given in Schuermann (2004)—see also Acharya et al. (2003). Several studies model the dependence between the probability of default and the recovery rate given default by assuming there is a common latent factor affecting both (Frye 2000, Dullmann and Trapp 2004). However, to the best of our knowledge, there are no empirical studies that attempt to explicitly model the covariates affecting the probability of default, the recovery rate given default, their dependence, and the impact on the loss distribution.

This paper is structured as follows. In §2 we develop our modeling methodology, and in §3 we describe the data set used in this study. The empirical results for the estimation of the probability of default and of the recovery rate are given in §4. In §5 we investigate the modeling of the loss distribution, and §6 concludes the paper with a summary.

2. The Default and Recovery Models

In this section, we first describe the specification and estimation of default models with unobservable heterogeneity, and then we discuss several specifications of recovery rate models.

2.1. The Default Models

The sample data contains firms in G groups or industries. Let n_i be the number of firms in the ith group over the entire observation period $[0, T]$, and $n = \sum_{i=1}^{G} n_i$ be the total number of firms in the sample. During $[0, T]$ a firm may experience a default, may leave the sample before time T for reasons other than default (for example a merger, an acquisition, or a liquidation), or may survive in the sample until time T. Some firms may enter the sample during the observation period. A firm's lifetime is censored if either default does not occur by the end of the observation period or if the firm leaves the sample because of a nondefault event. Let T_{ij} denote the observed and possibly censored lifetime (duration) of the jth firm in the ith group, and let $\perp_{ij}$ be the censoring indicator, where $\perp_{ij} = 1$ if T_{ij} is a default time and $\perp_{ij} = 0$ if T_{ij} is a censoring time. The total number of failures in group i is $\perp_{i.} = \sum_{j=1}^{n_i} \perp_{ij}$. Let $X_{ij}(t)$ be a $1 \times K$ vector of covariates at time t. The vector includes an intercept, as well as macroeconomic and firm-specific variables observed at discrete time intervals.

Let $\lambda_{ij}(t)$ be the default intensity (the hazard function) for the jth firm in the ith group. We assume that the unobservable heterogeneity at time t can be represented by a latent nonnegative random variable $\tilde{Y}_{ti}$ common to all firms in the same industry, which we shall refer to as frailty and which represents the effects of the unobservable measurement errors and missing variables (Hougaard 2000). The frailty $\tilde{Y}_{ti}$ acts multiplicatively on the intensity functions $\lambda_{ij}(t)$. We model $\tilde{Y}_{ti}$ as a combined effect of an industry-specific frailty factor Y_i and a time-varying indicator of industry distress. Specifically, let $\tilde{Y}_{ti} = Y_i \Delta^{Z_i(t)}$, where $Z_i(t)$ is an industry-specific distress indicator, defined as $Z_i(t) = 1$ if industry i is in distress at time t and 0 otherwise, and the parameter $\Delta > 0$ increases the impact of frailty in periods of economic distress. The hazard rates are modeled as

$$\begin{aligned}\lambda_{ij}(t) &= \tilde{Y}_{ti} \exp(X_{ij}(t)\beta) \\ &= \begin{cases} Y_i \Delta \exp(X_{ij}(t)\beta) & \text{if industry } i \text{ is distressed,} \\ Y_i \exp(X_{ij}(t)\beta) & \text{otherwise,} \end{cases}\end{aligned} \tag{1}$$

where β denotes the $K \times 1$ vector of regression parameters. In the remainder of this section, we condition on all the observed data—the lifetimes $\{T_{ij}\}$, censoring indicators $\{\perp_{ij}\}$, covariates $\{X_{ij}(t)\}$, and distress indicators $\{Z_i(t)\}$—but for the sake of simplicity we omit the conditioning from the formulae.

The lifetimes of firms in the ith group are independent conditional on the unobserved Y_i. When the unknown Y_i is integrated out, the lifetimes become dependent due to the common value of Y_i. Let us denote

$$H(T_{ij}) = \int_0^{T_{ij}} \Delta^{Z_i(t)} \exp(X_{ij}(t)\beta)\, dt. \tag{2}$$

From (1) and (2) it follows that

$$\int_0^{T_{ij}} \lambda_{ij}(t)\, dt = \int_0^{T_{ij}} Y_i \Delta^{Z_i(t)} \exp(X_{ij}(t)\beta)\, dt = Y_i H(T_{ij}). \tag{3}$$

The frailty model specified by (1) is a natural approach for modeling dependence and taking into account unobservable heterogeneity. In this paper, we denote the specification given by (1) as the *multiplicative frailty* model, because the Δ factor acts multiplicatively on the hazard rate. A special case of the multiplicative frailty model is obtained when $\Delta = 1$. This corresponds to a *shared frailty* model, which does not account for regime differences in the impact of frailties. In our subsequent analysis in §§4 and 5, we show that the multiplicative frailty has significantly better prediction performance than the shared frailty, implying that regime-specific differences are important for default and loss modeling.

The multiplicative frailty model can be easily extended to the case where frailties are multivariate rather than univariate or obligor specific rather than shared by all obligors in the same sector. Such extensions allow modeling of more flexible patterns of default dependence and different levels of contagion. For example, the multiplicative frailty model implies positive correlation of defaults within an industry, whereas some degree of negative correlation can be conceivable in practice due to competition. The multivariate lognormal frailty model (Stefanescu and Turnbull 2006) can accommodate negative default dependence as well and could thus be a reasonable alternative. In practice, however, the feasibility of different model extensions will be substantially constrained by the availability of data for estimation. For example, Duffie et al. (2009) account for time variation in a single common frailty factor by specifying a stochastic process for the frailty; their model cannot capture sector-specific contagion, but it does have a flexible time dynamic for the frailty. However, they note that "…we have found that even our relatively large data set is too limited to identify much of the time-series properties of frailty. This is not so surprising, given that the sample paths of the frailty process are not observed, and given the relatively sparse default data. For the same reason, we have not attempted to identify sector-specific frailty effects" (p. 2099).

The multiplicative frailty model (1) is an approach to modeling default contagion that is structurally tractable in order to make estimation feasible, and yet it has the complexity required to account for both sector-specific effects and time dynamics. Indeed, the frailties $\tilde{Y}_{ti}$ act at sector level rather than economy level and thus allow different levels of default contagion in different industries. This is an important feature because it may not be reasonable to assume that the contagion in the financial industry, for example, is similar to the contagion in manufacturing. Moreover, time dynamics are captured out of sample through updating the prior distribution of the frailties as the default information set evolves over time.

We assume that the sector frailties Y_i are independent and identically distributed with a gamma distribution $G(1/\theta, 1/\theta)$, with $\theta > 0$. This is a popular choice due to mathematical convenience. The gamma density function of Y is given by

$$f(y) = \frac{y^{1/\theta-1}}{\theta^{1/\theta}\Gamma(1/\theta)} \exp(-y/\theta), \tag{4}$$

where $\Gamma(\cdot)$ is the gamma function. The expected value is $E[Y] = 1$ and the variance is $\mathrm{Var}(Y) = \theta$.

Over time as more default information becomes available, we update dynamically the parameter estimates and the frailty distribution for each sector. For each year t, let $E[\tilde{Y}_{ti} \mid \mathcal{F}_t] = E[Y_i \Delta^{Z_i(t)} \mid \mathcal{F}_t] = \Delta^{Z_i(t)} E[Y_i \mid \mathcal{F}_t]$ be the expectation of the frailty term $\tilde{Y}_{ti}$ conditional on the information set $\mathcal{F}_t$ up to time t. Let θ_t, β_t, and Δ_t be the values of the parameters estimated on data up to time t. Conditional on $\mathcal{F}_t$, the frailty Y_i has a gamma distribution $G(A_{ti}, C_{ti})$ with scale parameter $C_{ti} = 1/\theta_t + \sum_{j=1,\, T_{ij} \le t}^{n_{ij}} H(T_{ij})$ and shape parameter $A_{ti} = 1/\theta_t + \sum_{j=1,\, T_{ij} \le t}^{n_{ij}} \perp_{ij}$ and with conditional mean given by

$$E[Y_i \mid \mathcal{F}_t] = A_{ti}/C_{ti}. \tag{5}$$

As t varies, the dynamic Bayesian updating of the frailty distributions for each sector is shown in Figure 1, which plots the annual expected frailties aggregated across all industry sectors. If no firms within a particular sector default, this might increase confidence in the credit worthiness of the firms in this sector and decrease the frailty. Conversely, if there is a failure in a particular sector or the aggregate number of defaults in the economy increases, this might adversely affect the assessment of credit worthiness and increase the frailty.

The parameters of the multiplicative frailty model (the regression coefficients β, the frailty variance θ, and the multiplicative factor Δ) can be estimated through maximum likelihood. Because the lifetimes in each group are independent conditional on the group frailty, we obtain that the conditional likelihood for group i is

$$\begin{aligned} &L_i(\Delta, \beta \mid Y_i = y_i) \\ &\quad = \prod_{j=1}^{n_i} [y_i \Delta^{Z_i(T_{ij})} \exp(X_{ij}(T_{ij})\beta)]^{\perp_{ij}} \exp(-y_i H(T_{ij})). \end{aligned} \tag{6}$$

The marginal log-likelihood for group i is derived by integrating out the frailties:

$$L_i(\theta, \Delta, \beta) = \int_0^\infty L_i(\Delta, \beta \mid y_i) f(y_i)\, dy_i, \tag{7}$$

where $f(y_i)$ is the gamma density of Y_i. Replacing (4) and (6) in (7), we develop the integral

$$\begin{aligned} &L_i(\theta, \Delta, \beta) \\ &\quad = \int_0^\infty L_i(\Delta, \beta \mid y_i) f(y_i)\, dy_i \\ &\quad = \int_0^\infty \left[\prod_{j=1}^{n_i} [y_i \Delta^{Z_i(T_{ij})} \exp(X_{ij}(T_{ij})\beta)]^{\perp_{ij}} \exp(-y_i H(T_{ij})) \right] \\ &\qquad \cdot \frac{y_i^{1/\theta-1}}{\theta^{1/\theta}\Gamma(1/\theta)} \exp(-y_i/\theta)\, dy_i \end{aligned}$$

$$= \frac{\prod_{j=1}^{n_i} \Delta^{\perp_{ij} Z_i(T_{ij})} \exp(\perp_{ij} X_{ij}(T_{ij})\beta)}{\theta^{1/\theta}\Gamma(1/\theta)} \cdot \int_0^\infty y_i^{1/\theta - 1 + \sum_{j=1}^{n_i} \perp_{ij}} \exp\left(-y_i\left(\frac{1}{\theta} + \sum_{j=1}^{n_i} H(T_{ij})\right)\right) dy_i. \quad (8)$$

Let us denote $w = 1/\theta + \sum_{j=1}^{n_i} H(T_{ij})$. The integral in (8) becomes

$$\int_0^\infty y_i^{1/\theta - 1 + \sum_{j=1}^{n_i} \perp_{ij}} \exp\left(-y_i\left(\frac{1}{\theta} + \sum_{j=1}^{n_i} H(T_{ij})\right)\right) dy_i$$
$$= \int_0^\infty y_i^{1/\theta - 1 + \perp_{i.}} \exp(-y_i w)\, dy_i$$
$$= \frac{1}{w^{1/\theta + \perp_{i.}}} \int_0^\infty (y_i w)^{1/\theta - 1 + \perp_{i.}} \exp(-y_i w)\, d(y_i w)$$
$$= \frac{\Gamma(1/\theta + \perp_{i.})}{(1/\theta + \sum_{j=1}^{n_i} H(T_{ij}))^{1/\theta + \perp_{i.}}}. \quad (9)$$

Replacing (9) in (8) and taking logarithms, we obtain

$$\log L_i(\theta, \Delta, \beta) = \log\Gamma(\perp_{i.} + 1/\theta) - \log\Gamma(1/\theta) - (1/\theta)\log(\theta) + \sum_{j=1}^{n_i} \perp_{ij} [X_{ij}(T_{ij})\beta + Z_i(T_{ij})\log(\Delta)] - (\perp_{i.} + 1/\theta)\log\left(1/\theta + \sum_{j=1}^{n_i} H(T_{ij})\right). \quad (10)$$

Because the groups are independent, the marginal sample likelihood as a function of the parameters is given by the sum of the log-likelihoods for all groups,

$$\log L(\theta, \Delta, \beta) = \sum_{i=1}^{G} \log L_i(\theta, \Delta, \beta), \quad (11)$$

and we used the *fmincon* optimization routine in MATLAB to maximize the likelihood numerically. After convergence, we computed standard errors of the estimates based on the information matrix.

The model can be extended to assume that the covariates $\{X_{ij}(t)\}$ follow a stochastic process with parameter vector γ, for example an autoregressive time series process. The sample likelihood would then also include the likelihood function $L_X(\gamma)$ of the covariates; the maximization program separates, implying that γ is estimated separately from β and θ. In general, the estimation of γ is the standard numerical procedure of fitting a multivariate time series process to the covariate vectors $\{X(t)\}$. This methodology can also be extended to the case of competing risks (Crowder 2001, Lawless 2003, Duffie et al. 2007). Firms may exit the sample for reasons other than default, such as a merger or an acquisition, and these nondefault events are all competing risks that can cause censoring of a firm's lifetime. With multiple causes for exit, we may consider a multivariate frailty model with one frailty component for each cause of exit. The likelihood function is separable under the assumption that the frailty components are independent with gamma distributions. Maximum likelihood estimates of the frailty variances and of the covariate parameters can be computed using the Expectation-Maximization (EM) algorithm (Dempster et al. 1977).

2.2. The Recovery Rate Models

In this subsection, we briefly describe several specifications of recovery rate models. Let $R_i(t)$ be the recovery rate of bond i at time t. We assume that $R_i(t)$ depends on a set of covariates $X_i(t)$ through a function of the linear form $X_i(t)\beta_r$, where β_r is a vector of regression coefficients. Note that the covariate vector $X_i(t)$ can include macroeconomic, industry-, firm-, and bond-specific variables, and thus the recovery rates will vary with bond, firm, and industry characteristics.

Many extant studies assume that recovery rates depend linearly on the available covariates (Acharya et al. 2003, Varma and Cantor 2005), so that $R_i(t) = X_i(t)\beta_r$. Note, however, that in practice recovery rates are always nonnegative and usually less than one. Because the linear specification implies that the recovery rates are unconstrained, it can lead to predicted recovery rates that are negative or greater than one and it is thus not appropriate for modeling recoveries. We investigate instead two other specifications.

The probit transformation gives $R_i(t) = \Phi(X_i(t)\beta_r)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and it implies that $X_i(t) = \Phi^{-1}(R_i(t))$ (Andersen and Sidenius 2005). The logit transformation gives $R_i(t) = 1/(1 + \exp(X_i(t)\beta_r))$, implying that $X_i(t) = \log(L_i(t)/R_i(t))$, where $L_i(t) = 1 - R_i(t)$ is the loss given default (Schönbucher 2003). In practice, the logit and probit models lead to very similar insights.

3. Data Description

In this section, we first describe the data sources and then discuss the covariates used at different stages of the analysis. Summary statistics for all covariates are available in Table 1.

3.1. Data Sources

Our primary data source is Moody's Ultimate Recovery Database (Moody's URD) that contains information on all bonds rated by Moody's during our sample period 1980–2008. In particular, Moody's URD has information on default history of the bonds and recovery rates in the event of default (Covitz and Han 2004, Varma and Cantor 2005, Duffie

Table 1 Data Descriptive Statistics

	Mean	25th percentile	50th percentile	75th percentile	Std. dev.
Macroeconomic variables					
Term spread	1.106	0.280	1.030	2.025	1.217
Credit spread	1.106	0.745	1.030	1.345	0.440
T-bill three-month yield	5.780	3.400	5.150	7.700	3.307
S&P 500 return	0.109	0.017	0.124	0.260	0.147
Logarithm(total defaulted debt)	2.022	0.843	1.962	3.374	1.945
Firm-specific variables					
Excess return	0.026	−0.231	−0.015	0.211	0.454
Relative size	−8.832	−9.983	−8.781	−7.552	1.785
Volatility	0.105	0.063	0.090	0.130	0.060
Net income to total assets	0.024	0.007	0.033	0.060	0.080
Total liabilities to total assets	0.665	0.539	0.648	0.789	0.199
Distance-to-default	6.808	3.542	6.294	9.477	4.998
Stock return	0.160	−0.113	0.114	0.359	0.471
Logarithm(total assets)	7.594	6.360	7.572	8.757	1.811
Tangible assets to total assets ratio	0.358	0.125	0.313	0.579	0.268
Market-to-book ratio	1.520	1.052	1.239	1.635	0.990
Bond-specific variables					
Recovery rate	0.352	0.145	0.280	0.548	0.254
Coupon rate	9.58	8.00	10.00	12.00	3.09

et al. 2007). We restrict our attention to only those firms that are in the intersection of Moody's URD, CRSP, and COMPUSTAT databases during 1980–2008. The COMPUSTAT (active and research) files for this period provide the firm level balance sheet data, and CRSP provides the market data. The sample includes 3,555 firms, 46,605 firm-years, and 477 defaults.

We use Moody's definition of default in our analysis. Moody's defines default as the event in which one or more of the following happen: (a) there is a missed or delayed disbursement of interest and/or principal, including delayed payments made within a grace period; (b) the company files for bankruptcy, administration, legal receivership, or other legal blocks to the timely payment of interest or principal; (c) a distressed exchange takes place. This happens either when the exchange has the apparent purpose of helping the borrower avoid default or when the issuer offers bondholders a new security or a package of securities that represents a diminished financial obligation (such as preferred or common stock, or debt with a lower coupon or par amount, lower seniority, or longer maturity).

The default data contain one record for each year of each firm's existence, from the year of listing to the year when the firm has left the sample through default, merger, or other event. We group firms into industry groups based on four digit DNUM codes. Our data contain 385 groups ranging in size from 1 to 2,582 firms, with a mean size of 107 firms and a median size of 54 firms. The number of defaults in each group ranges from 0 to 25.

The recovery data contain one record for each defaulted bond, where recovery rate on a bond is measured as the bond price within a month after default as given by Moody's URD.

3.2. Covariates

3.2.1. Macroeconomic Variables. In this study, we report the results of our investigation on the effects of five macroeconomic variables. These include the *term spread*, computed as the difference between the 10-year Treasury yield and the 1-year Treasury yield: the *credit spread*, computed as the difference between AAA and BAA yields; and the *three-month Treasury yield*, all taken from Federal Reserve's H.15 statistical release. We also use the *S&P 500 index trailing one year return* computed from CRSP, and the *logarithm of the trailing amount of total defaulted debt (in billions of USD)* taken from Moody's (2009) default study.

3.2.2. Industry-Level Variables. For the frailty default models we use industry groups based on four digit DNUM industry codes from COMPUSTAT. For the recovery rate models we identify four broad industry classes: utilities, transportation, industrials, and financials. We take the utilities class as a baseline and construct dummy variables for the other three industry classes.

For the default analysis, we also use an industry-specific variable that shows whether the industry is in distress in a particular time period. We follow Gilson et al. (1990), Opler and Titman (1994), and Acharya et al. (2007) in defining an industry as distressed. We construct an indicator variable that takes the value of one if the median stock return in that industry during the year is less than −20% and the value of 0 otherwise. The number of distressed industries by year varies between 0 in 1980 and 227 in 2008, with a mean of 32 and a median of 12. The number of years during 1980–2008 in which any particular industry has been distressed varies between zero and nine, with a mean of 2.5 and a standard deviation of 1.8.

3.2.3. Firm Level Variables. We follow Shumway (2001) in constructing the following firm level variables: the *relative size*, defined as the logarithm of each firm's equity value divided by the total NYSE/AMEX/NASDAQ market capitalization; the *excess return*, defined as the return on the firm minus the value-weighted CRSP NYSE/AMEX/NASDAQ index return; the *ratio of net income to total assets*, extracted from COMPUSTAT; the *ratio of total liabilities to total assets*, also extracted from COMPUSTAT; and the *volatility*, defined as the idiosyncratic standard deviation of the firm's monthly stock returns computed from CRSP.

Additionally, we also use the firm's *trailing one year stock return*, computed by cumulating the firm's monthly return from CRSP, and the *distance-to-default*, essentially a volatility corrected measure of leverage based on Merton (1974) and constructed as in Bharath and Shumway (2008).

We also construct the following firm level variables from COMPUSTAT for use in the recovery models: the *logarithm of the total assets*, the *market-to-book ratio* (a proxy for the firm's growth prospects), and the *ratio of property plant and equipment to total assets* (a measure of the firm's tangible assets). To avoid any outlier effects, all variables are winsorized at the 1% and 99% of the cross-sectional distributions.

To investigate the extent of collinearity within the spectrum of firm-specific information relevant for default, we perform an exploratory factor analysis on all seven covariates. We find that there are three significant factors corresponding to the eigenvalues greater than one of the correlation matrix, and these factors together explain 78% of the total variance. Table 2 gives the factor loadings for the seven variables; the first factor loads highly on the excess return and on the stock return; the second factor is strongly correlated with volatility, relative size, and distance-to-default; and the third factor loads highly on the accounting variables net income to total assets and total liabilities to total assets. These results suggest that two dimensions are probably not sufficient for capturing the range of firm-specific information relevant for default and that at least three dimensions of firm-specific covariates can be necessary in a default model.

Table 2 Factor Loadings for Firm-Specific Variables

	Factor 1	Factor 2	Factor 3
Excess return	0.981	0.025	−0.005
Volatility	0.143	−0.841	−0.029
Relative size	0.120	0.784	0.078
Net income to total assets	−0.010	0.350	0.696
Total liabilities to total assets	0.014	0.054	−0.909
Distance-to-default	0.286	0.584	0.440
Stock return	0.980	0.062	0.009

3.2.4. Bond Level Variables. Our recovery rate models include the *coupon rate*, the *logarithm of the time to maturity*, and the *seniority* as bond level variables. We identify four classes of seniority in ascending order of claim priority: subordinate, senior subordinate, senior unsecured, and senior secured. We take subordinate bonds as baseline and construct dummy variables for the other three seniority classes.

4. Default and Recovery—Empirical Results

In this section, we discuss the estimation results for the default and recovery models described in §2 and investigate their predictive performance.

4.1. Default Models: In-Sample Estimates

We report here the results for four default models, a subset of the many models that we investigated. In particular, the first model MD1 includes the same covariates as in Shumway (2001), whereas the fourth model MD4 is the specification from Duffie et al. (2007).

Table 3 reports the estimation results. For comparison, we fitted the four default models first with a shared frailty and then with a multiplicative frailty in the hazard rate.[1] In the multiplicative frailty models the frailty variance θ is statistically significant, providing evidence of default clustering. The frailty multiplicative factor Δ is also highly statistically significant, and its magnitude indicates that default contagion is much stronger in bad than in good economic periods. Almost all the covariate effects parameters are statistically significant and of the expected sign. Exceptions are the effects of credit spread in model MD3 and the coefficient for net income to total assets which is not significant in models MD1 and MD3, consistent with insights from Shumway (2001). Note also that contrary to expectation but similar with results from Duffie et al. (2007), the coefficient for the S&P 500 return is positive in model MD4. The sign and order of magnitude of all estimated coefficients are consistent with other published default studies—for example, Campbell et al. (2008), Duffie et al. (2007), Shumway (2001), and Zmijewski (1984).

[1] We also fitted a shared frailty model with industry distress as a covariate—this model had lower log-likelihoods than those of the multiplicative frailty model in Table 3 for all four covariate specifications. For parsimony, the details are not shown here but are available from the authors.

Table 3 Frailty Default Models: Estimation Results

	Shared frailty				Multiplicative frailty			
	MD1	MD2	MD3	MD4	MD1	MD2	MD3	MD4
Frailty variance θ	0.088	0.117	0.093	0.142	0.022	0.073	0.029	0.111
	(0.050)	(0.038)	(0.051)	(0.035)	(0.011)	(0.032)	(0.012)	(0.046)
Multiplicative factor Δ					2.568	2.456	2.577	2.265
					(0.281)	(0.284)	(0.282)	(0.266)
Intercept	−9.903	−6.457	−10.082	−3.268	−10.181	−6.728	−10.376	−3.469
	(0.507)	(0.702)	(0.530)	(0.126)	(0.376)	(0.421)	(0.416)	(0.159)
Excess return	−2.201	−1.845	−2.280		−2.059	−1.719	−2.138	
	(0.179)	(0.304)	(0.183)		(0.169)	(0.179)	(0.175)	
Relative size	−0.258	−0.247	−0.263		−0.272	−0.260	−0.278	
	(0.054)	(0.041)	(0.054)		(0.037)	(0.036)	(0.037)	
Volatility	2.061		1.969		2.0327		1.933	
	(0.331)		(0.321)		(0.208)		(0.213)	
Net income to total assets	−0.426		−0.366		−0.393		−0.347	
	(0.747)		(0.731)		(0.516)		(0.519)	
Total liabilities to total assets	1.709		1.738		1.728		1.759	
	(0.463)		(0.448)		(0.231)		(0.231)	
Distance-to-default		−0.308		−0.372		−0.305		−0.371
		(0.061)		(0.050)		(0.022)		(0.020)
Stock return				−2.017				−1.935
				(0.363)				(0.196)
Term spread			0.099				0.100	
			(0.050)				(0.046)	
Credit spread			0.022				0.035	
			(0.149)				(0.135)	
T-bill three-month yield				−0.078				−0.062
				(0.014)				(0.021)
S&P 500 return				1.365				1.240
				(0.342)				(0.318)
Log-likelihood	−1,735	−1,579	−1,732	−1,595	−1,703	−1,553	−1,701	−1,573
N	3,308	3,140	3,308	3,140	3,308	3,140	3,308	3,140

For the default models with shared frailty, the signs of all estimated coefficients remain unchanged, with only small changes in the magnitudes of the coefficients. For each specification there is a deterioration in the log-likelihood function compared to that of the corresponding default model with multiplicative frailty.

Figure 1 illustrates the dynamic updating of the frailty distributions for each sector. The figure plots the annual variation of the average frailties estimated from default models MD1 and MD4. The frailty values plotted here are computed as annual averages of the estimated frailties for each industry group, weighted by the number of firms in the group:

$$\bar{Y}_t = \frac{\sum_{i=1}^{G} n_{ti} E[\tilde{Y}_{ti} \mid \mathcal{F}_t]}{\sum_{i=1}^{G} n_{ti}},$$

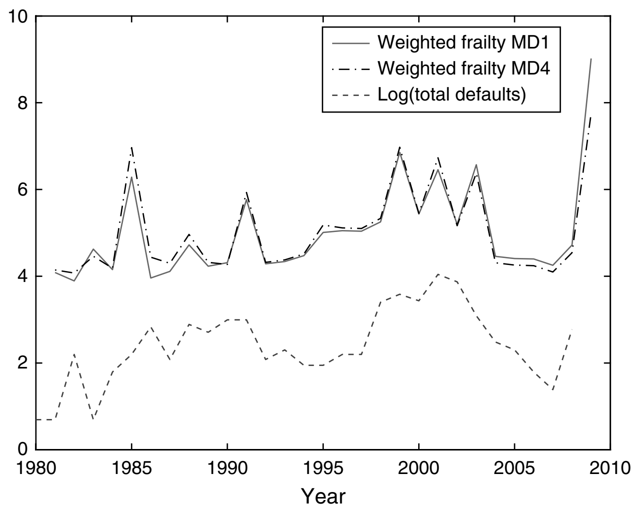
where n_{ti} is the number of firms in industry group i at time t and $E[\tilde{Y}_{ti} \mid \mathcal{F}_t]$ is the expected frailty value conditional on the information set $\mathcal{F}_t$ up to time t given by Equation (5). The $\bar{Y}_t$ values can thus be interpreted as annual aggregate measures of default risk across all firms. For comparison, the plot also presents the logarithm of the annual number of defaults. The annual variation of estimated frailties mirrors closely the actual realized default frequency. The patterns of variation are broadly similar on the comparable horizon to those of Figure 2 in Duffie et al. (2009), which gives the conditional posterior means of their single common frailty.

Figure 2 plots the annual frailty values estimated from models MD1 and MD4 for specific industry groups in three different sectors: telecommunications (television broadcasting stations), manufacturing (railroad equipment), and retail (shoe stores). The plots show that the patterns of annual variation are very different across industries, emphasizing the importance of accounting for industry-specific effects of contagion and unobservable heterogeneity.

4.2. Default Models: Out-of-Sample Performance

In this subsection, we investigate the out-of-sample forecasting performance of the four default models, using a one-year horizon as suggested by regulatory

Figure 1 Annual Frailty Values



requirements. We take several approaches to assessing forecasting performance and study their relative ability to differentiate between models.

4.2.1. Individual Firm Defaults. We first focus on predicting individual firm defaults, an approach extensively employed in extant literature. We define the out-of-sample horizon to be the period 1996–2008. For each year t during 1996–2008, we compute parameter estimates from data between 1980 and $t-1$; then we forecast default probabilities during year t for each firm that is alive at the beginning of that year. The firms are then ranked into deciles in descending order of their forecasted default probabilities, and for each year we compute the cumulative percentages and counts of defaults in the top two deciles since 1996.

Table 4 summarizes the cumulative percentages and counts in the top two deciles. For each model and each year, the table presents the percentages and counts of defaults classified in the top two deciles since 1996. For example, between 1996 and 1998 model MD1 with shared frailty correctly ranked 46 out of the 48 defaulted firms (or 95.83%) in the top two deciles. Overall, the predictive performance of all models is very similar, correctly identifying around 92–93% of the defaulting firms in the first two deciles over the horizon 1996–2008. This is consistent with the results for model MD4 reported in Duffie et al. (2007).

The results in Table 4 show that models with multiplicative and shared frailty are virtually identical in their prediction performance for individual firm defaults, both at annual level and at aggregate level over the entire horizon. In fact, the performance of shared frailty models as measured by this standard metric is in all years slightly superior to that of multiplicative frailty models, although the differences are negligible. Therefore, modeling regime changes in frailty effects through a multiplicative factor does not bring obvious improvements when individual firm default is the focus of the analysis. As we show in the next section, however, the benefits of using multiplicative frailties are substantial when the focus is on modeling portfolio defaults. The multiplicative frailty outperforms the shared frailty in this case, and this

Figure 2 Annual Frailty Values, Specific Industries

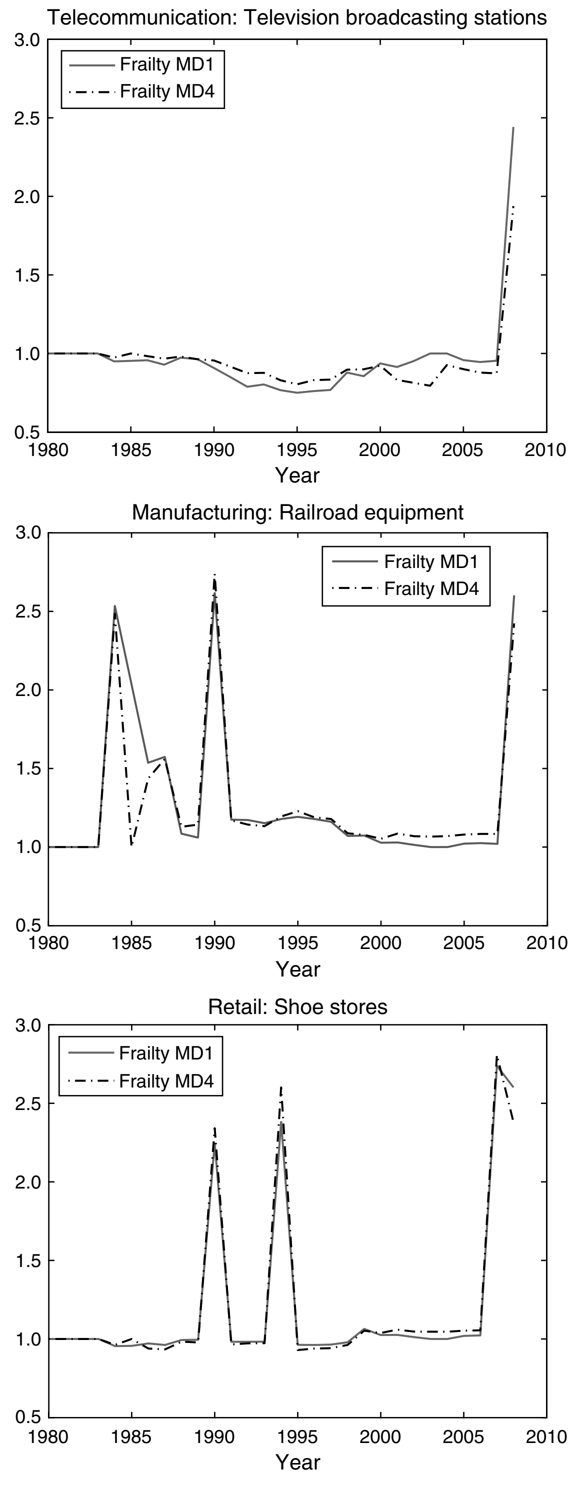


Table 4 Default Forecasting: Cumulatively Correctly Classified Defaults

	Shared frailty				Multiplicative frailty			
	MD1	MD2	MD3	MD4	MD1	MD2	MD3	MD4
1996	88.89% 8/9	100% 8/8	88.89% 8/9	100% 8/8	88.89% 8/9	100% 8/8	88.89% 8/9	100% 8/8
1997	94.44% 17/18	100% 16/16	88.89% 16/18	100% 16/16	94.44% 17/18	100% 16/16	88.89% 16/18	100% 16/16
1998	95.83% 46/48	95.35% 41/43	93.75% 45/48	95.35% 41/43	93.75% 45/48	95.35% 41/43	91.67% 44/48	95.35% 41/43
1999	94.05% 79/84	93.24% 69/74	92.86% 78/84	91.89% 68/74	92.86% 78/84	93.24% 69/74	91.67% 77/84	91.89% 68/74
2000	93.91% 108/115	92.31% 96/104	93.04% 107/115	92.31% 96/104	90.43% 104/115	91.35% 95/104	89.57% 103/115	90.38% 94/104
2001	94.19% 162/172	91.77% 145/158	93.60% 161/172	91.77% 145/158	91.28% 157/172	91.14% 144/158	90.12% 155/172	90.51% 143/158
2002	93.64% 206/220	92.04% 185/201	93.18% 205/220	92.54% 186/201	91.82% 202/220	92.04% 185/201	90.91% 200/220	90.55% 182/201
2003	94.21% 228/242	92.83% 207/223	93.80% 227/242	92.83% 207/223	92.56% 224/242	92.83% 207/223	91.74% 222/242	91.03% 203/223
2004	93.70% 238/254	92.34% 217/235	93.31% 237/254	92.34% 217/235	92.52% 235/254	92.77% 218/235	91.73% 233/254	90.64% 213/235
2005	93.94% 248/264	92.28% 227/246	93.56% 247/264	92.68% 228/246	92.80% 245/264	92.68% 228/246	92.05% 243/264	91.06% 224/246
2006	93.70% 253/270	92.46% 233/252	93.33% 252/270	92.86% 234/252	92.59% 250/270	92.86% 234/252	91.85% 248/270	91.27% 230/252
2007	93.80% 257/274	92.55% 236/255	93.43% 256/274	92.94% 237/255	92.70% 254/274	92.94% 237/255	91.97% 252/274	91.37% 233/255
2008	92.76% 269/290	91.85% 248/270	92.41% 268/290	92.96% 251/270	91.38% 265/290	92.22% 249/270	90.34% 262/290	91.48% 247/270

further translates into better performance for predicting loss distributions in §5.2.

4.2.2. Portfolio Defaults. The loss distribution for a portfolio of bonds or loans depends on the distribution of defaults within the portfolio and on the losses associated with each default. The ability to predict the number of defaults in a portfolio is thus critical for generating the loss distribution. Consequently, we next focus on forecasting the total number of defaults in a portfolio, an approach of major importance for risk and portfolio managers. Predicting the total number of defaults involves the actual magnitudes of default probabilities rather than just the ordinal risk ranking of firms as in §4.2.1.

We consider the entire portfolio of all firms, and we predict out of sample the total number of defaults in this portfolio for each year during 1996–2008. Table 5 summarizes the expectations of the annual default distributions predicted by the four default models with either a shared frailty or a multiplicative frailty. The table also gives the root mean squared errors of prediction (RMSE) over the horizon 1996–2008, based on the realized default counts relative to the expected defaults for each year. The RMSE is computed as $\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n}(pred_i - actual_i)^2/n}$, where $n = 13$ is the number of years in the out-of-sample horizon and $pred_i$ and $actual_i$ are the predicted and realized aggregate default numbers in year i. For all four models, the prediction errors are substantially smaller when the hazard rate includes a multiplicative frailty rather than a shared frailty factor. In §5.2 we shall see that this better default prediction performance of multiplicative frailty models translates into better loss prediction performance as well.

Table 5 reports performance measures aggregated over the entire horizon and all industries. We next pursue a disaggregate analysis by year; for conciseness, we henceforth focus on default models with multiplicative frailties. Figure 3 gives the predicted distributions of the annual number of defaults during 2001–2008 in the entire portfolio of all firms. In most years, the four models predict quite different distributions; also, both the expectation and the variability of the predicted distributions can change substantially between years.

We conclude the disaggregate analysis by year with an assessment of the time dynamics of default rate predictions. We consider all models with multiplicative frailty and focus first on the entire portfolio of all firms and second on separate portfolios of firms from the telecommunications, manufacturing, and retail sectors. The four plots in Figure 4 show the

Table 5 Portfolio Predicted Default

		Expected number of defaults							
		Shared frailty				Multiplicative frailty			
Year	Defaults realized	MD1	MD2	MD3	MD4	MD1	MD2	MD3	MD4
1996	9	15.45	12.99	16.19	15.62	13.89	11.45	13.11	11.84
1997	9	11.10	10.07	10.66	9.80	10.26	9.13	9.40	8.70
1998	30	13.84	12.24	12.15	12.91	16.63	14.72	13.73	14.31
1999	36	28.87	34.00	28.96	43.28	26.73	31.64	25.66	38.93
2000	31	27.99	26.96	28.90	23.99	32.92	38.36	29.13	29.17
2001	57	24.36	27.88	24.55	22.28	24.29	25.82	29.88	25.39
2002	48	14.55	17.31	12.24	21.40	19.54	23.50	17.35	31.58
2003	22	17.88	19.78	19.93	35.94	14.53	16.17	16.04	27.67
2004	12	6.78	5.38	7.74	4.60	6.33	4.78	6.95	3.75
2005	10	5.62	5.10	5.65	5.51	5.06	4.25	6.07	4.56
2006	6	6.49	6.31	5.88	6.01	5.41	5.14	4.88	4.94
2007	4	5.03	4.27	4.42	3.97	5.16	4.65	4.61	4.51
2008	16	8.23	11.20	8.30	15.17	16.04	20.39	16.38	25.61
Out-of-sample RMSE		3.79	3.62	3.85	3.77	3.64	3.53	3.58	3.41

annual realized default rates in each of these portfolios as well as the out-of-sample predictions from all models. The patterns of annual variation of default rates in the four portfolios are very different. In most years the predictions from all four models are fairly close to the realized default rates in all portfolios. The few exceptions are the large increases in realized default rates around the crisis years 2001–2002.

As discussed in the previous subsection, all four default models have almost identical performance according to standard metrics. They predict, however, very different distributions for the total portfolio defaults. When interest centers exclusively on predicting individual firm default, the traditional metrics can be an adequate measure of model performance. When interest, however, lies in predicting quantities that involve the actual magnitudes of default probabilities rather than just their ranking (for example, predicting the number of defaults in a portfolio), the ordinal ranking is no longer an adequate measure of model performance. This issue will be particularly relevant for loss distributions in §5.

4.3. Recovery Rate Models

In this subsection, we discuss three recovery rate models that we later use for modeling loss. Our goal is not to derive new predictive models for recovery rates or to investigate exhaustively the determinants of recoveries because extant literature has already focused on this topic (Acharya et al. 2003, 2007). The modeling of recovery rates in this paper is only a preliminary stage toward the ultimate objective of assessing the impact of default and recovery models on predicted loss.

Extant literature showed that contract characteristics, firm-specific variables, and macroeconomic variables are important factors that affect recovery rates. We experimented with many model specifications using the variables described in §3. The models that we retained during the model selection process have little or no redundant information in the form of covariates that are not statistically significant. We next briefly discuss the general insights from our analysis of many different covariate specifications; then we describe the three models that we retain for our subsequent analysis of predicted loss.

Among the contract characteristics, no variable was consistently statistically significant. The coupon rate has a positive and marginally significant coefficient in some models, which is to be expected given the results in Acharya et al. (2003). The impact of the logarithm of maturity outstanding is negative and also marginally significant in some models, whereas the logarithm of the issue size is not significant, similar to results from Acharya et al. (2003). To model the seniority class of a bond, we use three seniority dummies and take the subordinate class as baseline. All three dummies (senior subordinate, senior unsecured, and senior secured) were generally statistically significant. Although the differences between their estimated coefficients are relatively small, the ordering of the coefficients is as expected, implying that senior secured bonds earn on average higher recoveries than do senior unsecured bonds, which in turn have larger recoveries than senior subordinated bonds.

Among the firm characteristics, the relative size, the logarithm of total assets, the market-to-book ratio, the volatility, and the ratio of total liabilities to total assets were statistically significant in most models. We use four industry dummies following Moody's classification: financials, industrials, transportation, and utilities. We take utilities as baseline and find that the coefficients of financials, industrials, and

Figure 3 **Distribution of Annual Number of Defaults for Multiplicative Frailty Default Models**

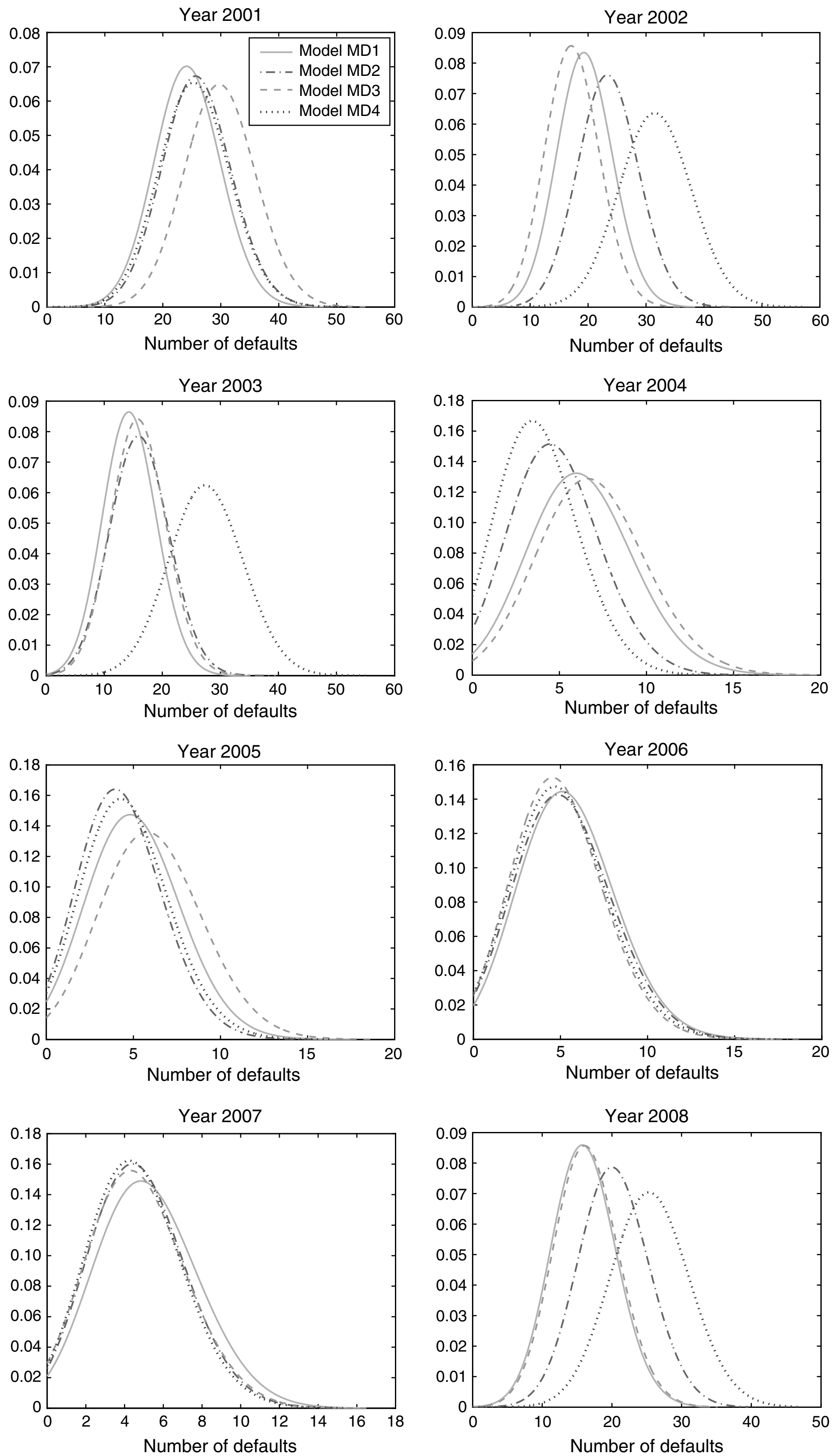
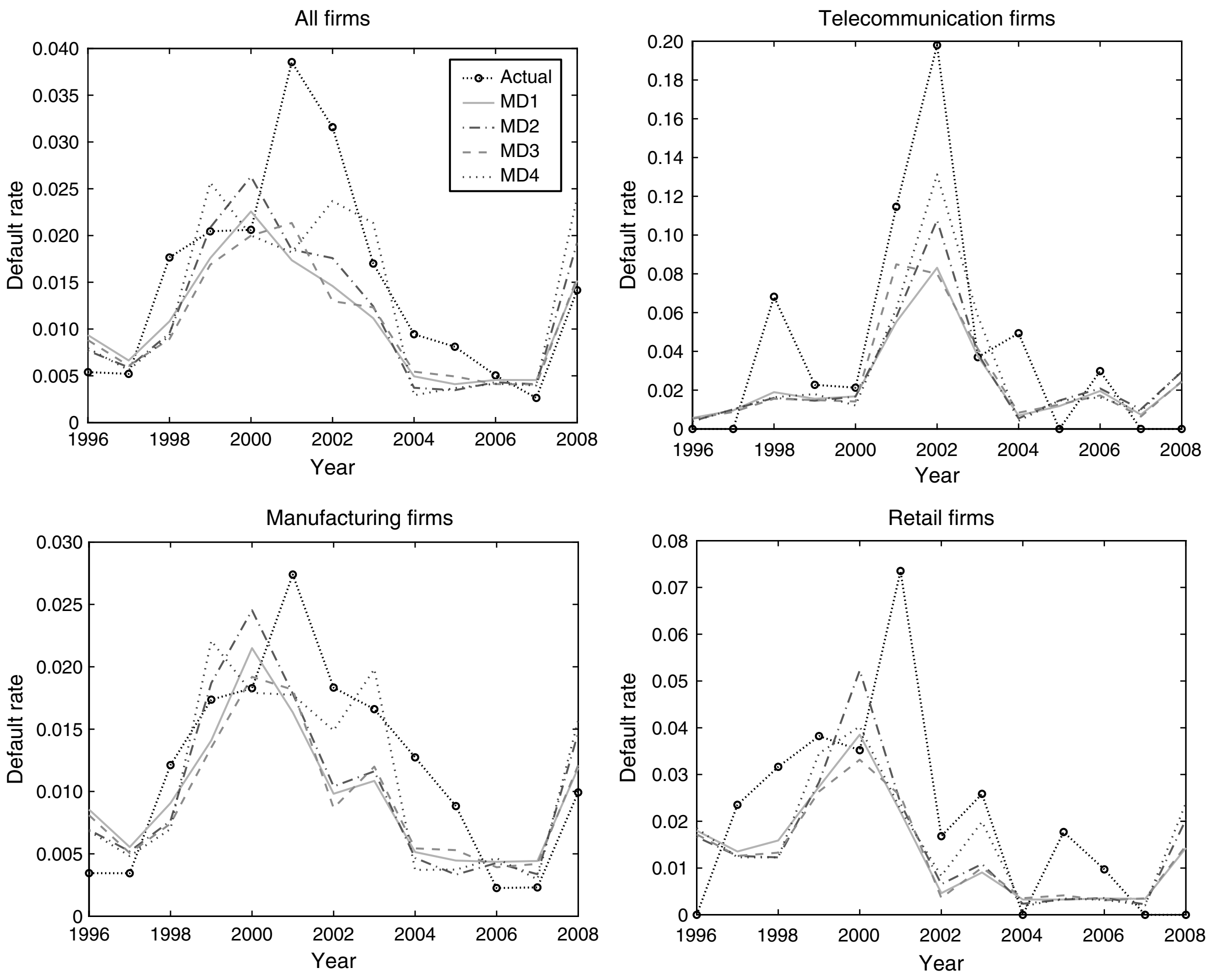


Figure 4 Annual Realized and Predicted Default Rates



transportation dummies are generally significant and negative in most models.

Among the macroeconomic variables, the Treasury three-month yield and the logarithm of the amount of all defaulted debt are statistically significant in most models. The credit spread and term spread are not significant in any model that we investigated.

To account for the potential effect of unobserved common factors on recoveries, all models include year dummies. We have also investigated whether the estimated annual frailties from the default models fitted in §4.1 have explanatory power for recoveries by including them as additional covariates. Their effect, however, was consistently not significant in all recovery models.

These insights are summarized in Table 6, which reports the estimates from fitting three selected recovery models among the multitude of models that we investigated. For each model we estimate both the logit and probit specifications described in §2.2. The results were similar, so we only report in Table 6 the estimates for the logit models.

We are not aware of any studies that examine the out-of-sample prediction performance of recovery rate models. We assess the performance of models MR1, MR2, and MR3 by using a rolling horizon calibration method. Similar to the methodology for default prediction, we define the out-of-sample horizon to be 1996–2008. For each year t, we estimate parameters for all recovery models from data between 1980 and $t-1$; then we forecast recovery rates during year t for all outstanding bonds of each firm that is alive at the beginning of that year. These forecasts are then compared with the actual realized recovery rates for bonds defaulted during year t. The average RMSEs are reported in the last line of Table 6. The out-of-sample RMSE values are virtually identical across all three recovery rate models.

5. Loss Distributions

In this section, we first investigate the correlation between default probabilities and recovery rates predicted out of sample. Next, we describe the methodology for predicting loss and we investigate

Table 6 Recovery Rate Models: Estimation Results and Out-of-Sample Prediction

	MR1	MR2	MR3
Intercept	4.8935	1.1574	5.4528
	(1.6127)	(0.5115)	(1.0877)
Coupon rate		0.0162	0.0290
		(0.0167)	(0.0169)
Logarithm(*maturity outstanding*)	−0.1869		
	(0.0955)		
Senior subordinate	0.6174	0.3316	0.2948
	(0.2130)	(0.2220)	(0.2088)
Senior unsecured	0.9734	0.7132	0.7667
	(0.1536)	(0.1394)	(0.1524)
Senior secured	1.1463	0.7593	0.8048
	(0.2615)	(0.3044)	(0.2822)
Financials	−0.9994	−1.2209	−0.4829
	(0.4840)	(0.5700)	(0.5842)
Industrials	−0.8836	−0.9939	−0.9383
	(0.3613)	(0.4340)	(0.3981)
Transportation	−1.2385	−1.3978	−1.2403
	(0.3709)	(0.4727)	(0.4278)
Excess return	0.1175		
	(0.1819)		
Volatility	−1.6446		
	(0.8215)		
Net income to total assets	0.3005		
	(0.6826)		
Total liabilities to total assets	0.6398		
	(0.2694)		
Relative size	0.1391		0.2522
	(0.0704)		(0.0584)
Logarithm(*total assets*)	−0.2033		−0.3348
	(0.0808)		(0.0628)
Market-to-book ratio	−0.2502		−0.4416
	(0.1334)		(0.1550)
Tangible assets to total assets ratio			0.3450
			(0.3086)
Distance-to-default		0.0479	
		(0.0303)	
T-bill three-month yield		−0.0871	−0.0688
		(0.0345)	(0.0356)
S&P 500 return		0.0861	0.5058
		(0.4113)	(0.4167)
Logarithm(*total defaulted debt*)	−0.1620	−0.3033	
	(0.0627)	(0.0558)	
N	849	794	830
Out-of-sample RMSE	0.2487	0.2584	0.2559

the out-of-sample loss distributions obtained under different default and recovery models.

5.1. Default and Recovery Correlation

Empirical evidence shows that ex post the frequency of default and the recovery rate given default are negatively correlated (Altman et al. 2005). We investigate whether this empirical relation between actual *realized* default frequencies and average recoveries is also apparent in the *predicted* default probabilities and recovery rates. Note that the empirical relation is based on aggregate data because it involves frequencies of multiple firm defaults. Our study, however, investigates the correlation at individual firm level because our methodology allows the computation of individual firm default probabilities.

Table 7 summarizes the aggregate correlations between predicted out-of-sample default probabilities and recovery rates for each pair of frailty default and recovery models. For firms with multiple bonds of the same seniority, we take the average of the predicted recovery rates within the same seniority class. We stratify the firms by industry and the bonds by seniority, and within each industry and seniority class we compute the bivariate Pearson correlation coefficient between the predicted default probability and the recovery rate across all firms and all years. The correlations are not computed for utilities senior subordinate bonds and financials senior secured bonds where the sample sizes are fewer than 10.

Almost all the aggregate correlations are negative and highly statistically significant. They vary with industry group and seniority level and also with the choice of default and recovery models. For any given choice of recovery model, the correlations obtained with default models MD1 and MD3 are similar, usually smaller in absolute value than are those obtained with default model MD2, which in turn are generally smaller than are those obtained with default model MD4.

Are the patterns detected in aggregate correlations over the entire horizon 1996–2008 preserved for disaggregate annual correlations? Figure 5 gives the annual correlations between out-of-sample predicted default probabilities and recovery rates for each pair of frailty default models MD1 and MD4 and recovery models MR1 and MR2. The firms are all industrials, and the recovery rates are all for senior unsecured bonds (the class with the largest number of recoveries). It is apparent from Figure 5 that the annual variation of correlations is related to the credit cycle. The correlations increase in absolute value as the state of the economy worsens in 2001, then decrease as the economy improves during 2003–2004. This is consistent with insights from Das and Hanouna (2009), who also find that the correlations of default probabilities and recovery rates become increasingly negative with increasing default risk. The patterns of annual variation differ considerably between the correlations computed with model MR1 and those computed with model MR2. For any specific default model, the correlations implied by MR1 are generally smaller in absolute value than are those from MR2. The main difference between models MR1 and MR2 is in the way in which firm-specific information is taken into account because the distance-to-default is the only

Table 7 Aggregate Correlations of Default Probabilities and Recovery Rates

		Recovery models								
		Utilities			Industrials			Financials		
Seniority	Default models	MR1	MR2	MR3	MR1	MR2	MR3	MR1	MR2	MR3
Subordinate	MD1	−0.17	−0.30	−0.20	−0.27	−0.33	−0.29	−0.22	−0.20	−0.07
	MD2	−0.17	−0.34	−0.17	−0.35	−0.32	−0.29	−0.32	−0.29	−0.13
	MD3	−0.22	−0.32	−0.23	−0.29	−0.32	−0.29	−0.23	−0.20	−0.07
	MD4	−0.24	−0.39	−0.23	−0.44	−0.27	−0.30	−0.33	−0.28	−0.10
Sample size			139			3,010			830	
Senior subordinate	MD1				−0.83	−0.55	−0.51	−0.36	−0.11	0.08
	MD2				−0.82	−0.54	−0.49	−0.49	−0.13	0.04
	MD3				−0.82	−0.55	−0.50	−0.41	−0.12	0.07
	MD4				−0.82	−0.53	−0.48	−0.55	−0.23	−0.12
Sample size						2,105			29	
Senior unsecured	MD1	−0.16	−0.24	−0.12	−0.34	−0.27	−0.31	−0.33	−0.29	−0.13
	MD2	−0.20	−0.28	−0.16	−0.47	−0.35	−0.49	−0.36	−0.32	−0.15
	MD3	−0.19	−0.24	−0.13	−0.34	−0.24	−0.29	−0.35	−0.30	−0.13
	MD4	−0.25	−0.34	−0.19	−0.50	−0.33	−0.45	−0.40	−0.34	−0.16
Sample size			669			6,795			1,221	
Senior secured	MD1	−0.18	−0.13	−0.11	−0.16	−0.22	−0.30			
	MD2	−0.21	−0.20	−0.11	−0.24	−0.34	−0.35			
	MD3	−0.22	−0.16	−0.13	−0.18	−0.24	−0.30			
	MD4	−0.35	−0.30	−0.22	−0.27	−0.33	−0.35			
Sample size			506			589				

firm-level covariate included in model MR2. Figure 5 thus shows that the modeling of firm-specific information in the recovery models has a significant impact on both the level and the annual variation of the correlation values. Figure 5 also shows that for any specific recovery model the correlations implied by default model MD4 are generally larger in absolute value than are those implied by MD1, although the difference is marginal. These results are consistent with the patterns in aggregate correlations from Table 7.

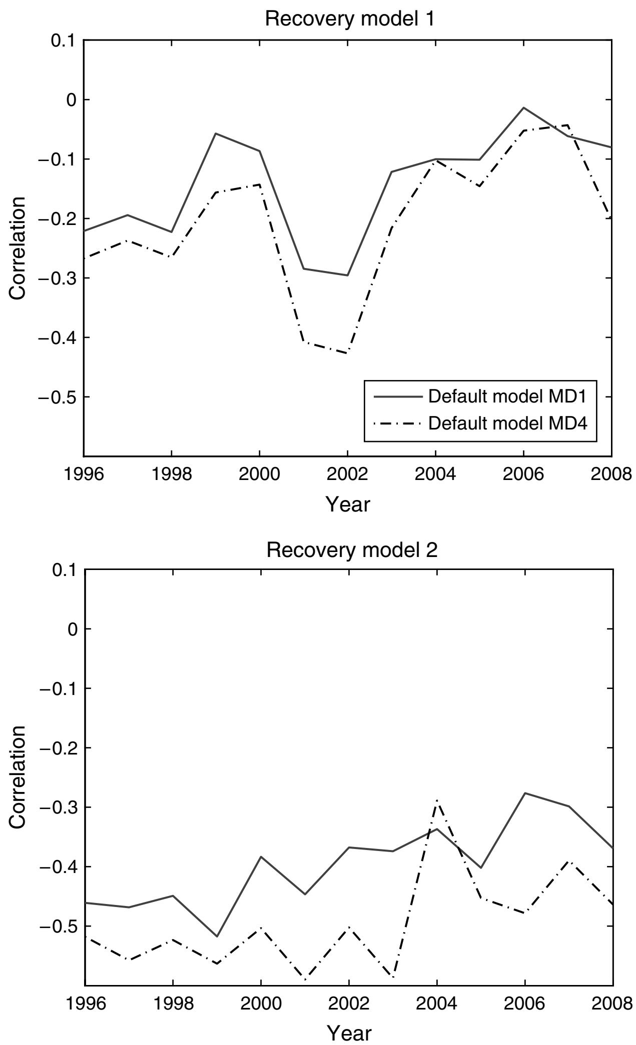
5.2. Predicted Loss Distributions

We next investigate the impact that the choice of default and recovery models has on predicted loss distributions. We assume that the face value of each bond is one. With the notation from previous sections, let us denote by $\perp_i(t)$ an indicator function that equals 1 if firm i defaults in period t conditional on survival up to period t, and 0 otherwise. Let us also denote by $R_i(t)$ the recovery rate for bonds of firm i at time t. The loss $L_i(t)$ from obligor i in period t is then $L_i(t) = 1 - R_i(t)$ if $\perp_i(t) = 1$, and $L_i(t) = 0$ if $\perp_i(t) = 0$. The expected loss over the next period computed at t is $E_t[L_i(t+1)] = E_t[\perp_i(t+1) \cdot (1 - R_i(t+1))]$, where both $\perp_i(t+1)$ and $R_i(t+1)$ depend on a set of covariates $X_i(t+1)$. It is also possible to compute the expected loss over multiple-period horizons. This entails modeling the evolution of the covariates, for example by means of a stochastic process such as an autoregressive time series.

We exemplify the methodology by predicting loss distributions in different portfolios. We first focus on the entire portfolio of all firms. Using the predicted default probabilities and recovery rates, for every year during 1996–2008 we generated the out-of-sample loss distributions from each default and recovery model. The loss distributions are based on 10,000 simulated loss scenarios. When a firm defaults, we assume that all its outstanding bonds default and we compute the loss using the predicted recovery rate for each bond. The overall loss in a simulation scenario is the loss from all outstanding bonds of all firms defaulting in the scenario.

We first investigate the aggregate performance in loss prediction over the entire out-of-sample horizon 1996–2008. We consider default models with multiplicative and with shared frailty in order to assess whether the multiplicative frailty factor improves prediction performance. Because the number of bonds outstanding changes every year, we consider two different loss measures: the total aggregate loss in the portfolio (the sum of losses on all bonds) and the loss per bond (equal to the aggregate loss divided by the number of bonds each year). Table 8 reports the RMSE over annual loss predictions for each combination of default and recovery models. The table shows that for all recovery models, all default models, and prediction of both aggregate loss and loss per bond, the default model with multiplicative frailty has a significantly smaller RMSE than does the default model

Figure 5 **Annual Correlations of Default Probabilities and Recovery Rates**



with shared frailty. It is also apparent that for any recovery model, default model MD4 has consistently the lowest RMSE, followed by models MD2 and MD3 with similar performance. This is true both for prediction of the aggregate loss in the entire portfolio and for prediction of loss per bond. Recovery model MR1 has in most cases the smallest RMSE, although the error differences across the three recovery models are small. This implies that the choice of recovery model has a smaller impact on predicted loss performance than does the choice of default model.

We next investigate the predicted loss distributions for bonds stratified by industry. We focus separately on the telecommunication, manufacturing, and retail sectors and consider all firms in these sectors with debt outstanding. Using a similar analysis as for the entire portfolio for all firms, we report in Table 9 the mean squared errors of out-of-sample predictions for all combinations of recovery models and default models with multiplicative and shared frailty. The numbers from Table 9 support at industry level the insights from the aggregate analysis in Table 8. Default models with multiplicative frailty perform better than do default models with shared frailty, particularly for telecommunication firms. Table 9 also shows that no single default model is uniformly best for all industries; indeed, MD4 is generally best for telecommunication firms, whereas model MD3 is best for retail and manufacturing firms, both for predicting aggregate loss and loss per bond. In terms of recovery specifications, model MR1 again has lower prediction errors than do MR2 and MR3 in almost all cases, although the differences among the three recovery models in all three industry groups are relatively small.

The aggregate analysis over the entire horizon 1996–2008 has so far shown that the default models with multiplicative frailties perform better for loss prediction than do those with shared frailties and that recovery model MR1 is marginally better than are models MR2 and MR3. We next pursue a disaggregate analysis by year; for conciseness, we henceforth focus on default models with multiplicative frailties and on recovery model MR1. For all the subsequent analyses, the results from recovery models MR2 and MR3 are similar to the ones from recovery model MR1, emphasizing that the specific choice of recovery model has much less of an impact on predicted loss distributions than does the choice of default model.

Table 8 **Loss Prediction Errors, All Firms**

	Multiplicative frailty				Shared frailty			
Model	MD1	MD2	MD3	MD4	MD1	MD2	MD3	MD4
Portfolio aggregate loss error								
MR1	25.4382	23.4363	22.2222	19.6478	29.4810	27.7474	29.8769	27.7540
MR2	26.1152	24.0247	23.2348	19.3375	30.0079	28.2652	30.3478	27.0297
MR3	26.6799	24.7662	23.7334	20.8743	30.3155	28.6837	30.6795	28.4023
Loss per bond error								
MR1	0.0226	0.0205	0.0204	0.0163	0.0263	0.0247	0.0266	0.0236
MR2	0.0229	0.0208	0.0209	0.0159	0.0266	0.0250	0.0268	0.0230
MR3	0.0236	0.0217	0.0216	0.0175	0.0269	0.0254	0.0272	0.0243

Table 9 Loss Prediction Errors, Various Industries

Model	Multiplicative frailty				Shared frailty			
	MD1	MD2	MD3	MD4	MD1	MD2	MD3	MD4
				Telecommunication firms				
Portfolio aggregate loss error								
MR1	6.1183	4.7418	5.2950	3.3174	8.1779	7.2245	8.7093	6.9793
MR2	6.9506	5.8211	5.9144	4.3877	8.7159	7.9429	9.1789	7.6117
MR3	6.6384	5.2960	5.6994	3.9531	8.5233	7.5969	9.0103	7.3661
Loss per bond error								
MR1	0.0674	0.0525	0.0586	0.0389	0.0893	0.0790	0.0951	0.0781
MR2	0.0761	0.0637	0.0653	0.0491	0.0949	0.0865	0.0999	0.0839
MR3	0.0729	0.0583	0.0630	0.0449	0.0930	0.0830	0.0983	0.0819
				Manufacturing firms				
Portfolio aggregate loss error								
MR1	3.0267	2.9238	2.1984	3.7115	3.1412	3.1529	3.1117	4.5154
MR2	3.1748	3.0993	2.5553	3.6936	3.2572	3.2792	3.2255	4.3725
MR3	3.2197	3.1011	2.4564	3.6895	3.3304	3.3317	3.3000	4.4695
Loss per bond error								
MR1	0.0063	0.0061	0.0046	0.0077	0.0066	0.0065	0.0065	0.0092
MR2	0.0066	0.0065	0.0053	0.0077	0.0068	0.0068	0.0067	0.0090
MR3	0.0067	0.0065	0.0051	0.0076	0.0070	0.0069	0.0069	0.0092
				Retail firms				
Portfolio aggregate loss error								
MR1	7.6429	7.5489	7.2019	7.2202	7.5667	7.3603	7.5661	7.1488
MR2	7.6924	7.6009	7.3576	7.3162	7.6276	7.4355	7.6334	7.2493
MR3	7.6895	7.6120	7.2762	7.3043	7.6271	7.4517	7.6283	7.2421
Loss per bond error								
MR1	0.0486	0.0480	0.0458	0.0460	0.0480	0.0467	0.0480	0.0452
MR2	0.0489	0.0484	0.0468	0.0467	0.0484	0.0472	0.0484	0.0459
MR3	0.0488	0.0484	0.0463	0.0465	0.0484	0.0473	0.0484	0.0458

Figure 6 gives the probability density functions of the annual loss distributions during 2001–2008, predicted by all four multiplicative frailty default models and recovery model MR1. In most years the loss distributions are relatively similar. The main exceptions are distributions generated with MD3 in 2001 and with MD4 in 2001–2003 and 2008, which have both significantly higher expectations and variances. These are also the years with the highest default frequency and realized aggregate loss; this suggests that to a larger extent than the other default models, model MD4 captures more of the uncertainty and high volatility related to the negative state of the economy in those years and allows for a larger upward adjustment of the predicted expected loss in periods of economic distress. The comparison with the corresponding annual distributions of total default numbers between 2001 and 2008 from Figure 3 shows that the annual patterns of default predictions are largely conserved in loss prediction. In particular, this holds for the differences in default and loss distributions predicted by MD4 in 2002–2003 and 2008. This is consistent with the earlier insight that the choice of default model is the main factor for differences in loss predictions.

Table 10 summarizes the annual aggregate loss distributions for the entire portfolio of all firms from Figure 6. The table reports the expectation and standard deviation of the loss distributions, the total number of bonds outstanding, and the actual realized loss every year. Consistent with the insights from Figure 6, the loss distributions from default model MD4 have often larger variability than do the other loss distributions, particularly in those years with high loss and default frequency, 2001–2002 and 2008.

To conclude the disaggregate analysis, we next assess the time dynamics of the loss per bond predictions from all default models and recovery model MR1, first in the entire portfolio of all firms and then separately in portfolios of firms from the telecommunications, manufacturing, and retail sectors. The plots in Figure 7 show the annual realized loss per bond in each of these portfolios as well as the loss per bond predictions over the horizon 1996–2008. The patterns of annual variation of the loss per bond in the four portfolios are very different. In most years the predictions from all models are fairly close to the realized loss per bond in all portfolios. The few exceptions are the large increases in realized loss around

Figure 6 Loss Distributions for Multiplicative Frailty Default Models

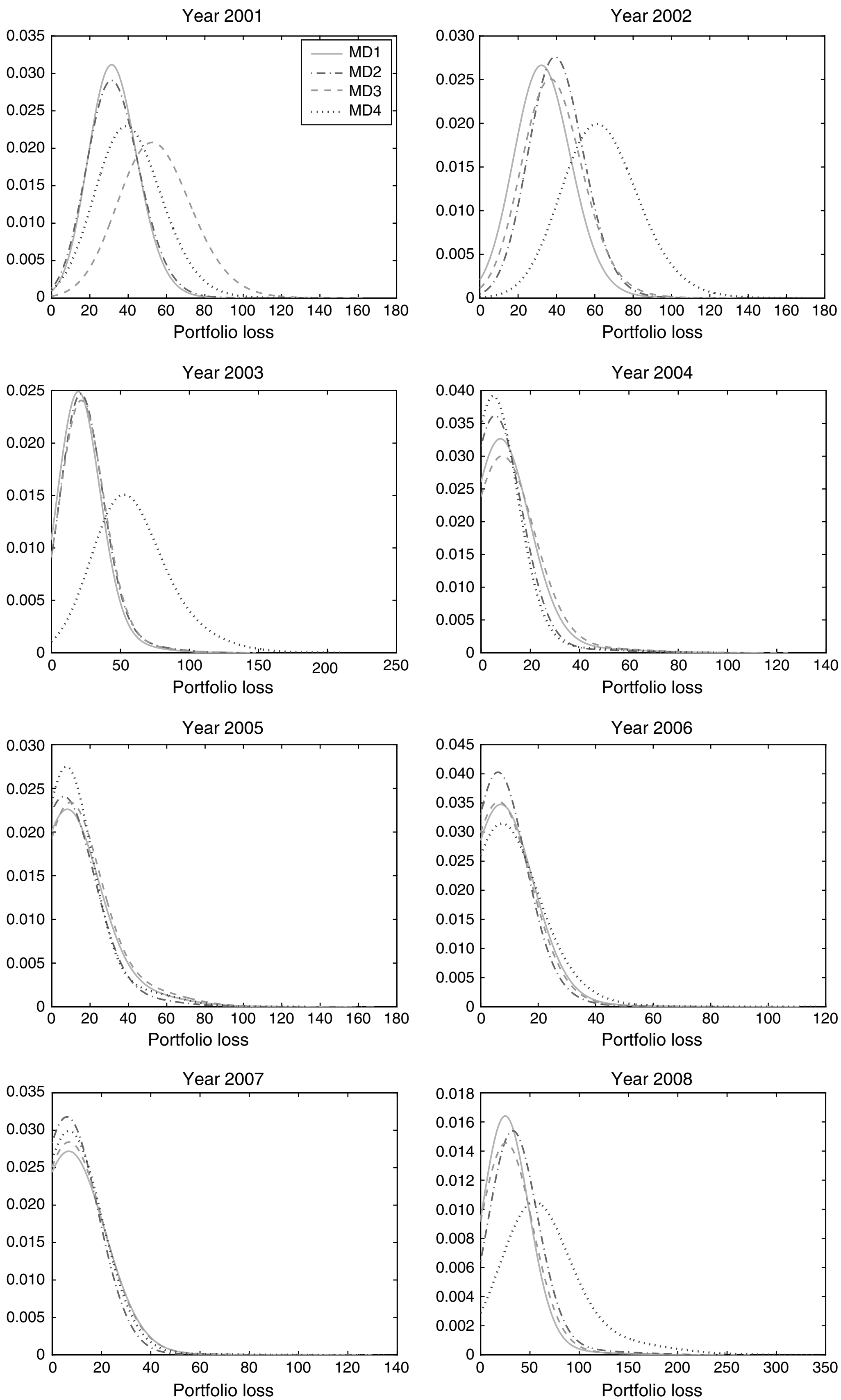


Table 10 Annual Predicted Loss, Multiplicative Frailty

			Expected value				Standard deviation			
Year	Loss	*N*	MD1	MD2	MD3	MD4	MD1	MD2	MD3	MD4
1996	0.65	1,205	9.04	6.28	8.62	7.82	6.18	5.21	5.88	6.75
1997	2.81	1,305	7.49	6.09	6.85	6.45	6.75	5.56	6.33	6.24
1998	18.13	1,367	13.11	11.15	10.58	12.45	5.92	5.08	5.42	5.65
1999	21.54	1,375	24.18	29.47	23.07	42.24	9.45	9.79	9.42	12.93
2000	30.15	1,309	33.69	40.42	29.54	40.82	11.45	12.50	10.68	14.56
2001	80.18	1,299	31.97	32.10	53.35	39.86	9.84	10.38	14.49	13.26
2002	59.89	1,242	32.32	39.37	37.43	61.96	10.07	10.76	11.92	15.75
2003	23.20	1,272	19.47	21.49	21.79	54.52	11.56	12.79	12.23	22.54
2004	8.93	1,286	7.27	5.17	8.21	4.57	8.49	7.06	8.79	8.19
2005	21.84	1,270	7.43	5.59	9.55	7.52	12.24	9.51	13.86	12.59
2006	3.43	1,227	6.60	5.72	5.82	6.97	5.89	5.63	5.66	7.91
2007	1.50	1,152	6.15	5.56	6.17	6.39	5.63	4.73	7.13	6.13
2008	94.68	1,047	24.43	32.67	24.86	55.85	15.95	19.39	15.69	33.21

the crisis years 2001–2002 and 2008; although the predictions succeed in capturing the jump in realized loss for telecommunication firms in 2002, they are less successful at capturing the large loss for retail firms in 2001 or for the portfolio of all firms in 2008. Again, the annual patterns in loss per bond from Figure 7 mirror closely the corresponding annual patterns in default rate predictions from Figure 4.

We further test the different loss distributions adapting a methodology from Egorov et al. (2006)

Figure 7 Annual Realized and Predicted Loss per Bond

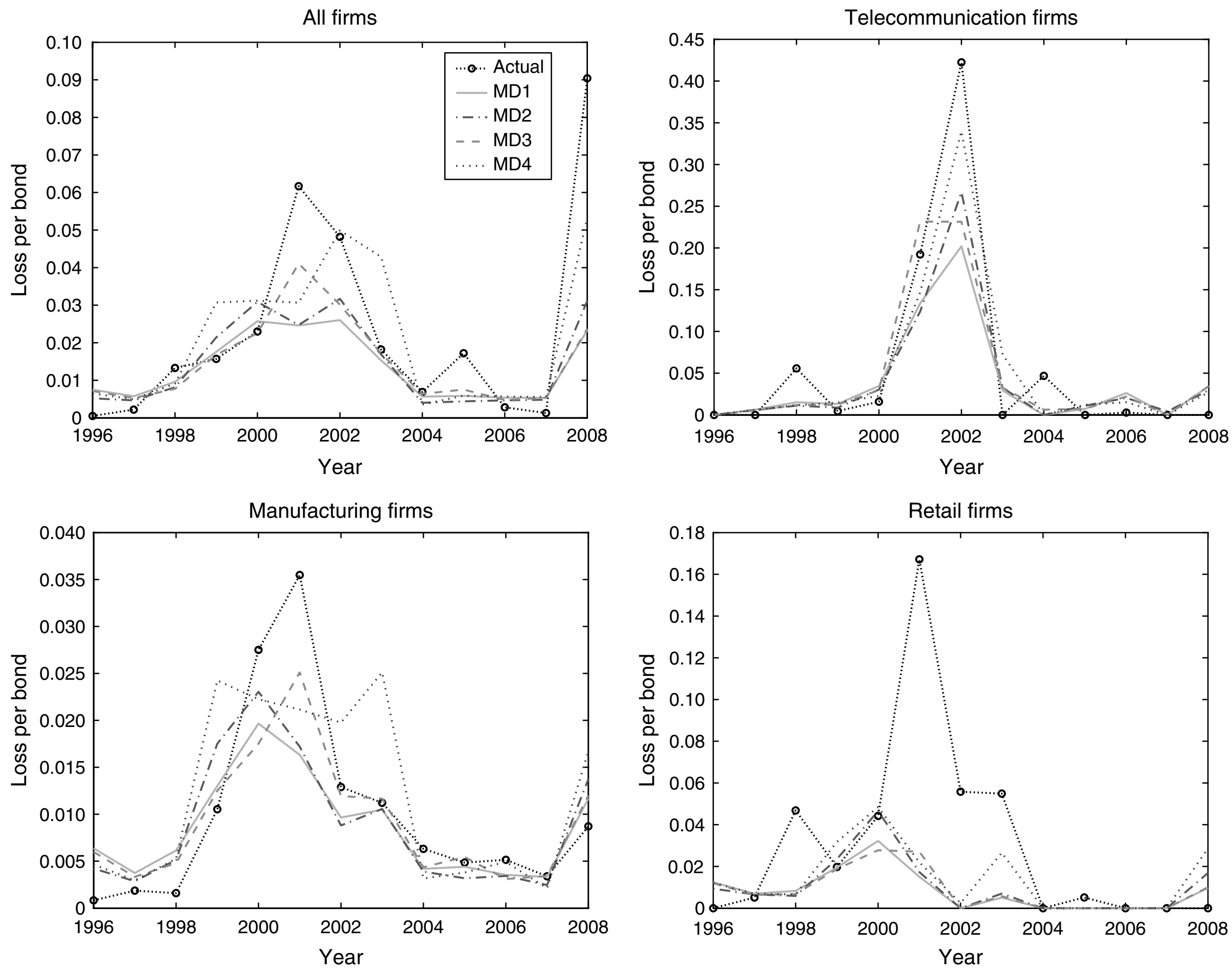
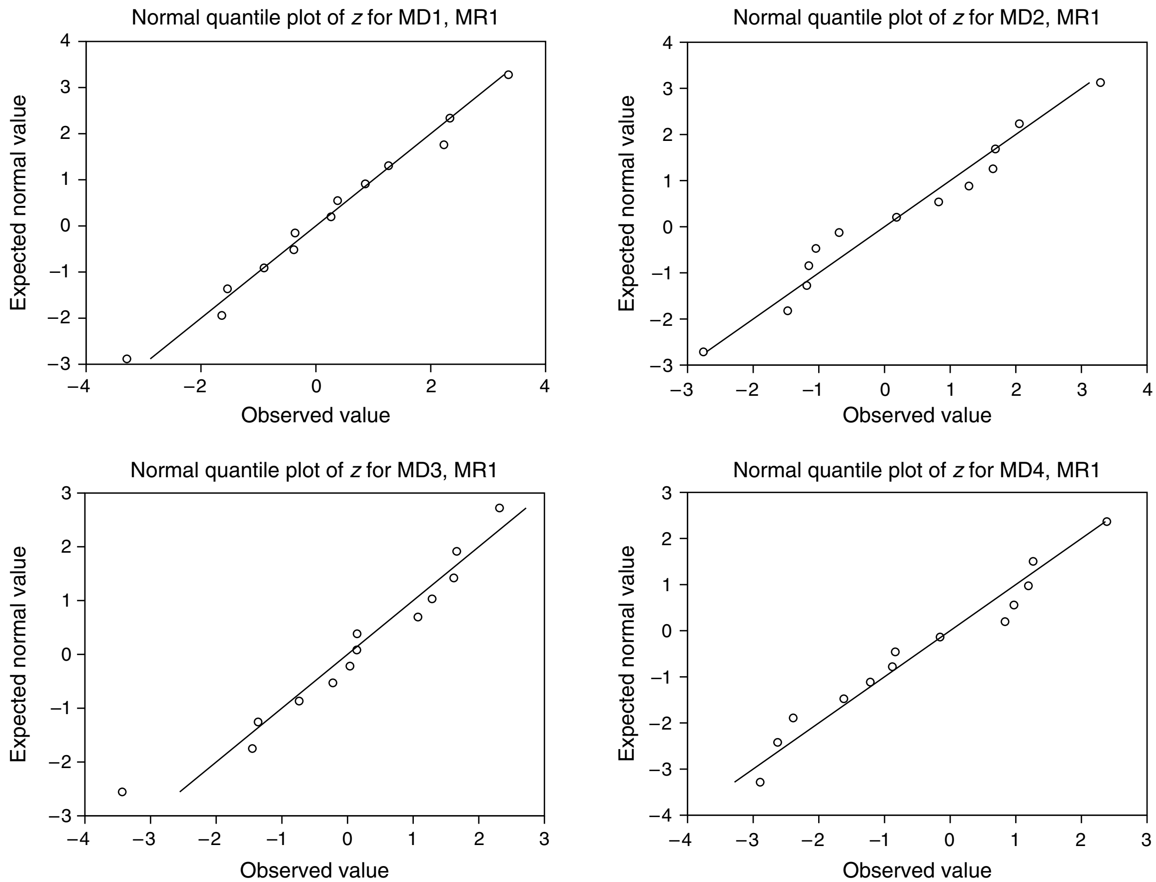


Figure 8 Normal Quantile Plots for Transformed Predicted Loss



and Hong et al. (2007). Let L_t be the loss on a specific portfolio at time t, and let l_t be the realized loss. Given information at time t, the density function of the loss for the next period $t+1$ is denoted by $f_t(u)$. Let $X_t = P[L_{t+1} \leq l_{t+1}] = \int_0^{l_{t+1}} f_t(u)\,du$ be the cumulative probability function for the loss at time $t+1$. The random variables X_t are independent and uniformly distributed on $[0, 1]$. Following Berkowitz (2001), let $Z_t = \Phi^{-1}(X_t)$, where $\Phi(\cdot)$ is the standard normal cdf. If the densities $f_t(u)$ fit the actual realized loss, then Z_t are independent and normally distributed.

We consider the portfolio of all bonds and the horizon 1996–2008, and we assess normality of Z_t using quantile plots. Figure 8 gives the quantile plots for annual loss predicted out of sample by all four default models with multiplicative frailty and recovery model MR1. The plots suggest that the transformed predicted loss values from model MD1 are closer to normality than are those from the other three default models. However, the Jarque-Bera and Lilliefors tests fail to reject the normality assumption for Z_t computed with all default models.

6. Conclusions

This paper focuses on modeling and validating the loss distribution with sector specific and regime dependent unobservable heterogeneity in firm characteristics. To generate the loss distribution, it is necessary to model the probability of default and the recovery rate given default. In this paper, we address two main issues—the modeling of sector specific and regime dependent unobservable heterogeneity and the appropriate choice of default and recovery models for loss prediction.

We first focus on default prediction and account for unobservable heterogeneity using a multiplicative frailty model with industry specific and regime dependent factors. Based on the analysis of a large default data set over 1980–2008, we show that four default models inspired by extant literature have very similar performance according to standard metrics that use relative ordinal rankings of default probabilities. We present a different approach that compares the actual realized number of defaults in a given portfolio with the total number of defaults predicted out of sample and find that under this metric the differences between the four default models' predictions can be substantial. We show that controlling for industry-level heterogeneity significantly improves out-of-sample forecasting performance and that the appropriate choice of default model subsequently has a crucial impact on loss prediction.

We next address the modeling of recovery rates and find that three different specifications for recovery in the event of default have similar out of sample performance. We show that the default probabilities and recovery rates predicted out of sample are negatively correlated and that the magnitude of the correlation is related to the credit cycle and varies with industry and seniority.

Finally, we investigate the impact that the choice of default and recovery models has on the predicted loss distribution. We show that the default model specification significantly affects the predicted loss distribution, whereas the choice of recovery model has only a marginal impact. We prove that accounting for regime dynamics in the default models is crucial for loss predictions because regime-dependent multiplicative frailty models perform significantly better than do shared frailty models. We also find strong industry effects that must guide the choice of an appropriate default model for loss prediction; for example, the Duffie et al. (2007) model performs best for telecommunication firms and in years with high default frequency. In contrast, the Shumway (2001) model enhanced with two macroeconomic covariates performs best for manufacturing firms and in years with low default frequency. These regime and industry effects imply that for any practical application, the appropriate choice of default models for loss prediction will depend on the credit cycle and on portfolio characteristics.

Acknowledgments

The authors are grateful to the associate editor and the two referees for constructive feedback on the paper.

References

Acharya, V. V., S. T. Bharath, A. Srinivasan. 2003. Understanding the recovery rates on defaulted securities. Working paper, London Business School, London.

Acharya, V. V., S. T. Bharath, A. Srinivasan. 2007. Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. *J. Financial Econom.* **85**(3) 787–821.

Altman, E. I., E. Hotchkiss. 2005. *Corporate Financial Distress and Bankruptcy*, 3rd ed. Third Edition, Wiley, New York.

Altman, E. I., B. Brady, A. Resti, A. Sironi. 2005. The link between default and recovery rates: Theory, empirical evidence and implications. *J. Bus.* **78**(6) 2203–2227.

Andersen, L., J. Sidenius. 2005. Extensions to the Gaussian copula: Random recovery and random factor loadings. *J. Credit Risk* **1**(1) 29–70.

Basel Committee on Banking Supervision. 2005a. Guidance on paragraph 468 of the framework document. (July). Bank for International Settlements, Basel, Switzerland.

Basel Committee on Banking Supervision. 2005b. Studies on the validation of the internal rating systems. BCBS Working Paper 14, Bank for International Settlements, Basel, Switzerland.

Berkowitz, J. 2001. Testing density forecasts, with applications to risk management. *J. Bus. Econom. Statist.* **19**(4) 465–474.

Bharath, S. T., T. Shumway. 2008. Forecasting default with the Merton distance-to-default model. *Rev. Financial Stud.* **21**(3) 1339–1369.

Campbell, J. Y., J. Hilscher, J. Szilagyi. 2008. In search of distress risk. *J. Finance* **63**(6) 2899–2939.

Chava, S., R. A. Jarrow. 2004. Bankruptcy prediction with industry effects. *Rev. Finance* **8**(4) 537–569.

Covitz, D., S. Han. 2004. An empirical analysis of bond recovery rates: Exploring a structural view of default. Working paper, The Federal Reserve Board of Washington, Washington, DC.

Crowder, M. 2001. *Classical Competing Risks*. Chapman & Hall, New York.

Dai, Q., K. J. Singleton, W. Yang. 2007. Regime shifts in a dynamic term structure model of U.S. treasury bond yields. *Rev. Financial Stud.* **20**(5) 1669–1706.

Das, S. P. Hanouna. 2009. Implied recovery. *J. Econom. Dynam. Control* **33**(11) 1837–1857.

Dempster, A. P., N. M. Laird, D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. B* **39**(1) 1–38.

Duffie, D., L. Saita, K. Wang. 2007. Multi-period corporate failure prediction with stochastic covariates. *J. Financial Econom.* **83**(3) 635–665.

Duffie, D., A. Eckner, G. Horel, L. Saita. 2009. Frailty correlated default. *J. Finance* **64**(5) 2089–2123.

Dullmann, K., M. Trapp. 2004. Systematic risk in recovery rates—An empirical analysis of U.S. corporate credit exposure. Working paper, Deutsche Bundesbank, Frankfurt, Germany.

Egorov, A. V., Y. Hong, H. Li. 2006. Validating forecasts of the joint probability density of bond yields: Can affine models beat random walk? *J. Econometrics* **135**(1–2) 255–284.

Frye, J. 2000. Depressing recoveries. *Risk Magazine* **13**(11) 108–111.

Gagliardini, P., C. Gourieroux. 2003. Spread term structure and default correlation. Working paper, University of Toronto, Toronto.

Gilson, S., K. John, L. Lang. 1990. Troubled debt restructurings: An empirical study of private reorganization of firms in default. *J. Financial Econom.* **27**(2) 315–355.

Hong, Y., H. Li, F. Zhao. 2007. Can the random walk model be beaten in out-of-sample density forecasts? Evidence from intraday foreign exchange rates. *J. Econometrics* **141**(2) 736–776.

Hougaard, P. 2000. *Analysis of Multivariate Survival Data*. Springer, New York.

Lawless, J. L. 2003. *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, Hoboken, NJ.

Li, H., T. Li, C. Yu. 2010. No-arbitrage Taylor rules with switching regimes. Working paper, Ross School of Business, University of Michigan, Ann Arbor.

Merton, R. C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *J. Finance* **29**(2) 449–470.

Moody's. 2009. Moody's corporate default and recovery rates, 1920–2008. Report, Moody's Corporation, New York.

Opler, T. C., S. Titman. 1994. Financial distress and corporate performance. *J. Finance* **49**(3) 1015–1040.

Schönbucher, P. J. 2003. *Credit Derivatives Pricing Models*. John Wiley & Sons, Hoboken, NJ.

Schuermann, T. 2004. What do we know about loss given default? D. Shimko, ed. *Credit Risk: Models and Management*, 2nd ed. Risk Books, London, 249–274.

Shumway, T. 2001. Forecasting bankruptcy more accurately: A simple hazard model. *J. Bus.* **74**(1) 101–124.

Stefanescu, C., B. W. Turnbull. 2006. Multivariate frailty models for exchangeable survival data with covariates. *Technometrics* **48**(3) 411–418.

Varma, P., R. Cantor. 2005. Determinants of recovery rates on defaulted bonds and loans for North American corporate issuers: 1983–2003. *J. Fixed Income* **14**(4) 29–44.

Yashin, A. I., K. G. Manton. 1997. Effects of unobserved and partial observed covariate processes on system failures: A review of models and estimation strategies. *Statist. Sci.* **12**(1) 20–34.

Zmijewski, M. E. 1984. Methodological issues related to the estimation of financial distress prediction models. *J. Accounting Res.* **22**(Supplement) 58–82.