Predicting Defaults with Regime Switching Intensity: Model and Empirical Evidence

Hui-Ching Chuang* Chung-Ming Kuan
Department of Finance Department of Finance
National Taiwan University National Taiwan University

January 16, 2011

Abstract

In this paper, we propose a default intensity model with regime switching, in which the intensity is affected by both observed risk factors and a latent state variable characterizing different intensity regimes. In particular, both the level of default intensity and risk exposures to observable covariates are regime dependent. We provide an estimation algorithm when the state variable is Markovian and illustrate the proposed model using the default data of US-listed companies during 1990-2009. Our test indicates that the regime switching effect in the intensity function is significant. Regime-switching intensity model outperforms classical intensity model in both in-sample and out-of-sample default predictions.

Keywords: credit risk modeling, Markov switching, frailty correlated default, bankruptcy prediction, ROC analysis, distance to default.

JEL Classification: C35, C41, G32, G33

*Correspondence: Hui-Ching Chuang, Department of Finance, National Taiwan University, No.1, Sec. 4, Roosevelt Rd., Taipei 106, Taiwan. Tel: +886 23366-1072; E-mail: hchuang@ntu.edu.tw. We thank Jin-Chuan Duan, Bruce Grundy, Baeho Kim, Tao Wang and participants at 2010 Asia FMA D.S.C., Financial Engineering and Risk Management International Symposium 2010, and the Fourth Annual Risk Management Conference for helpful discussions and comments. H.-C. Chuang acknowledges research supports from the National Science Council of Taiwan (NSC 97-2917-I-002-138) and the hospitality of professor Jin-Chuan Duan and Risk Management Institute, National University of Singapore during her visit.
1 Introduction

It has been well documented that default clustering depends on firm specific characteristics and macro fundamentals, e.g., Das et al. (2006), Pesaran et al. (2006) and Koopman et al. (2009). A leading model that characterizes time-varying default rates is the doubly-stochastic intensity model proposed by Duffie, Saita, and Wang (2007). They assume that, conditioning on the path of observable risk factors (e.g., profitability measure, leverage proxy, and common macroeconomics variables) that determine default intensities, the default times of different firms are the first jump times of independent Poisson processes. As such, these default times are correlated, and hence default clusters, due to the exposure to risk factors.

The empirical data show, however, that the actual default rates usually exceed the intensities implied by the doubly-stochastic model, e.g., Das et al. (2007) and Lando and Nielsen (2010). A possible explanation is that researchers fail to include all relevant, observable factors in their model specification for default intensity, as shown in Lando and Nielsen (2010). Moreover, not all risk factors are readily observed. For example, the fraudulent accounting practices of Enron and WorldCom are unobservable, so that their true default risk may be underestimated; see Giesecke (2004) and Jorion and Zhang (2007) for more discussions. To account for the latent factor, Duffie et al. (2009) introduce a frailty variable (latent stochastic process) in the level of default intensity function and find strong evidence for the presence of a common latent variable in the data of U.S. nonfinancial firms. Koopman, Lucas, and Schwaab (2010) further show that macro factors, common frailty and industry-specific frailty are three main driven forces of default events. Both studies indicate that including frailty in an intensity model results in more probability mass at the right tail of portfolio loss distribution (cluster of defaults).

In this paper, we propose a default prediction model with regime-switching effects in a doubly-stochastic intensity model, such that the default intensity is affected by both observable risk factors and unobservable regime indicators. In particular, the level of the intensity function and the risk exposures to observable risk factors in the intensity function are specified as state dependent. The regime-dependent level of intensity may be understood as a common/industry frailty, whereas the regime-dependent risk exposures (factor loadings) to observable factors can characterize the linkages between financial intermediaries and firms. For instance, banks tend to take a loose lending policy in a tranquil period and a tight policy during an economy-wide distress. Thus, default risk increases due to the increase of firms’ risk exposure to observable risk factors.

In the empirical study, the proposed model, RS intensity model hereafter, is applied to 10,950 US-listed nonfinancial firms during the period of 1990 to 2009. To ensure relevant, observable risk factors are included in the intensity function, we first adopt Least Absolute
Shrinkage and Selection Operator (LASSO) method of [Tibshirani (1997)](#) to select variables from more than forty firm specific and macro factors; see also [Friedman, Hastie, and Tibshirani (2010)](#). We then estimate the RS intensity model with those selected factors. It is found that the intercept parameter and the risk exposure parameter to firm distance to default (DtD) are significantly regime dependent. Moreover, incorporating regime switching parameters in a well-specified intensity model can better capture the fat right tail effect in real data. The in-sample and out-of-sample precision of default prediction are also found to be superior to existing doubly-stochastic intensity models.

This paper is organized as follows. We review related literature and existing intensity-based models in Section 2. We discuss the proposed RS intensity model and its estimation procedure in Section 3. An empirical application of the proposed model to US-listed companies during 1990-2009 is presented in Section 4. Section 5 concludes this paper.

2 Related Literature

Intensity-based models have gained more and more popularity in recent studies of bankruptcy/default prediction. Given default time of the \(i\)-th firm \(\tau_i\) and all available information \(\mathcal{F}_t\) at time \(t\), the default intensity \(\lambda_{i,t}\) is the instantaneous default probability per unit of time, condition on this firm has not yet defaulted at time \(t\), i.e.,

\[
\lambda_{i,t} = \lim_{\Delta t \to 0} \frac{P(t < \tau_i \leq t + \Delta t | \tau_i > t, \mathcal{F}_t)}{\Delta t} = \Lambda(W_{i,t}; \mu),
\]

where \(\Lambda(\cdot; \cdot)\) is a non-negative real-valued function, and \(W_{i,t}\) is a vector of risk factors/covariates determining firm \(i\)'s default intensity with unknown parameters \(\mu\). In other words, the probability of default within a small time interval \(\Delta t\) after \(t\) is approximately \(\lambda_{i,t}\Delta t\). In this setting, modeling the default probability of \(i\)-th firm reduces to specify the intensity function.

The theory of doubly-stochastic intensity is mostly stated in continuous time, however, the estimators are mainly based on the discrete-time hazard regression with small time period. [Shumway (2001)](#) uses hazard (intensity) function as:

\[
\Lambda(W_{i,t}; \mu) = \frac{1}{1 + e^{\mu W_{i,t}}},
\]

where \(W_{i,t}\) are fimspecific variables. [Shumway (2001)](#) shows that hazard model is more appropriate than single-period logistic regression model for forecasting bankruptcy. He also finds that hazard model with market size, past stock returns and return variability and two accounting ratios have more prediction power than [Altman (1968)](#) which includes less fimspecific, market information.
Duffie, Saita, and Wang (2007) formalize the doubly-stochastic intensity model assumptions and its ML estimation approach. They set up the intensity function as:

$$\Lambda(W_{i,t}; \mu) = \exp \left( \sum_{p=0}^{4} \mu_p W_{i,p,t} \right)$$

where $W_{i,t}$ include constant, one-year trailing stock return, distance to default, S&P500 index return and three month Treasury bill rate. Duffie, Saita, and Wang (2007) incorporate important macroeconomics variables and firm-specific market information in their intensity setting. Their results show that predictive performance of this model improves over Bharath and Shumway (2008), Beaver, McNichols, and Rhie (2005) based on accounting ratios.

Recent literature indicates that condition on observable firm-specific and macro factors can not fully characterize the degree of default clustering. Das et al. (2007) utilize the test of time-transform of the counting process to examine different specifications of intensity function. They reject the joint assumption of well-specified intensity function and doubly stochastic intensity of Duffie, Saita, and Wang (2007). From this finding, Lando and Nielsen (2010) consider 3 macro economic variables: S&P 500 return, change in U.S. industrial production and spread between the 10-year and 1-year treasury rate and additional firm-specific variables, quick ratio, short-term debt to total debt and book asset value (log). They cannot reject the same hypothesis of Das et al. (2007). This result suggests that relevant risk factors may be omitted in the intensity function setting.

Frailty/latent variables have been introduced to intensity models to overcome the missing variable problem. Duffie et al. (2009) specify a latent variable in the level of intensity function, i.e.,

$$\Lambda(W_{i,t}, y_t; \mu) = \exp \left( \sum_{p=0}^{k} \mu_p W_{i,p,t} + \gamma y_t \right)$$

where $y_t$ is an unobservable variable following Ornstein-Uhlenbeck (OU) process,

$$dy_t = -\kappa y_t dt + dB_t, \quad y_0 = 0.$$ 

$B$ is a standard Brownian motion. $\gamma$ and $\kappa$ are parameters. $y_t$ can capture those variation in default rates beyond observable factors. Duffie et al. (2009) demonstrate that frailty factor has statistically and economically significant implications for the tails of portfolio default loss distributions.

Some recent reduced-form credit risk models also consider frailty as an important factor in default prediction. McNeil and Wendin (2007) consider the unobserved random effects in Bayesian generalized linear mixed model. Koopman, Lucas, and Schwaab (2010) construct a
dynamic factor model to handle the non-linear, non-gaussian filtering problem when introducing the frailties in default process. They show that the common factors to all firms (macro and frailty) account for approximately 75% of the default clustering. The extreme tail clustering in defaults cannot be captured using observable variables alone. The frailty variable can not be discarded.

3 Methodology

In this section, we will briefly discuss the notation and assumptions of doubly-stochastic intensity model in section 3.1. Section 3.2 introduces regime-switching intensity model and its estimation approach. LASSO methodology to choose covariates used in RS intensity model will be described in section 3.3.

3.1 Doubly-stochastic intensity model

Following the intensity model literature, denote covariates as $W_{i,t}$, and default indicator as $D_{i,t}$, for $1 \leq i \leq N$ and $1 \leq t \leq T$. In classical doubly-stochastic model, $W_{i,t}$ are all observable variables. We will generalize this notation later to include latent regime indicator as one element of $W_{i,t}$. Default indicator $D_{i,t}$ equals to zero before default, and equals to one afterwards. To further simplify notations, let $W_t := (W'_{1,t}, \ldots, W'_{N,t})'$ denote the vector of observed covariate processes for all companies at time $t$, and let $D_t := (D_{1,t}, \ldots, D_{N,t})'$ denote the vector of default indicators. It is noted when $W_{i,t}$ is a firm-specific variable, its observable period is from time $t_{0,i}$ to $T_i$, the first and last observation time of company $i$. $T_i$ is the minimum of default time or other exit time of firm $i$.

Let $A_t := \{i | D_{i,t} = 0, t_{0,i} \leq t \leq T_i \}$ denote the set of firms alive at time $t$ and $\triangle A_t := \{i \in A_{t-1} | D_{i,t} = 1, t_{0,i} \leq t \leq T_i \}$ be the set of firms defaulted at time $t$. A discrete-time approximation of the classical doubly-intensity log likelihood of the observed survivals and defaults is

$$L(\mu | F_T) = \sum_{t=1}^{T} \log \left[ l_t(\mu | W_t, D_t) \right],$$

where

$$l_t(\mu | W_t, D_t) = \prod_{i \in A_t} e^{-\Lambda(W_{i,t}; \mu) \triangle t} \prod_{i \in \triangle A_t} \left( 1 - e^{-\Lambda(W_{i,t}; \mu) \triangle t} \right).$$

That is, for firm $i$, the likelihood of defaulting at time $t$ conditioned on no default before is $(1 - e^{-\Lambda(X_{i,t}; \mu) \triangle t})$; otherwise, $e^{-\Lambda(X_{i,t}; \mu) \triangle t}$ is the probability of survive at time $t$. We denote $\hat{\mu}$ as the MLEs of classical doubly-intensity model and consider the smallest time interval $\triangle t$
3.2 Regime-switching intensity model

Depart from doubly-stochastic intensity model, assume the state variable $W_{i,t}$ can be partitioned into observable variable $X_{i,t}$ and unobservable regime indicator $s_t$. $W_{i,t} = (X_{i,t}, s_t)$. Observable $X_{i,t}$ consists of firm-specific and common covariates. We do not restrict $X_{i,t}$ to be directly observed but they can be constructed with confidence through statistical methods such as diffusion indexing or other dimension reduction techniques applied to large panel of cross-sectional or time series data.

Regime indicator, $s_t$, follows $N$ states ($N$ regimes) first order Markov chain with transition probability $p_{ij}$, for $1 \leq i, j \leq N$. $p_{ij} := P(s_t = j | s_{t-1} = i, s_{t-2} = k, ... ) = P(s_t = j | s_{t-1} = i)$. For simplicity, we consider that $X_{i,t}$ and $s_t$ are mutually independent processes and $s_t$ is a one dimensional discrete Markov chain with time-invariant transition probabilities. Multiple regime indicators and time-varying transition probability can be incorporated under our framework; see Kim and Nelson (1999) for further discussions.

Conditional on regime $j$, the intensity function takes the proportional hazard form:

$$
\Lambda(X_{i,t}, s_t = j; \mu_j) = \exp \left( \mu_{0,j} + \mu_{1,j} X_{i,1,t} + \mu_{2,j} X_{i,2,t} + \cdots + \mu_{k,j} X_{i,k,t} \right),
$$

(5)

where $X_{i,t}$ is a $k \times 1$ vector observable covariates for $i$-th firm and $\mu_j$ is a $k \times 1$ vector of parameters specific to regime $j$. In our most general specification of RS model, both the level parameter of intensity function ($\mu_{0,j}$) and all the risk exposure/factor loading parameters ($\mu_{j}$) are regime dependent.

A specific model is with 4 observable covariates. The intensity function is

$$
\Lambda(X_{i,t}, s_t = j; \mu_j) = \exp \left( \mu_{0,j} + \mu_{1,j} X_{i,1,t} + \mu_{2,j} X_{i,2,t} + \mu_{3,j} X_{3,t} + \mu_{4,j} X_{4,t} \right),
$$

where $X_{i,1,t}$ and $X_{i,2,t}$ are firm-specific variables such as DtD measure and stock return; $X_{3,t}$ and $X_{4,t}$ are macroeconomic variables, such as S&P 500 index return and three month Treasury bill rate. The intercept, one factor loading to firm-specific risk factor are state dependent. Compare to equation [1], this model is a generalization of Duffie, Saita, and Wang (2007) model with regime-switching effect in both intercept and factor loading of intensity function.

Another specific case is regime-switching in intercept of intensity function.

$$
\Lambda(X_{i,t}, s_t = j; \mu_j) = \exp \left( \sum_{p=1}^{k} \mu_{p,j} X_{i,p,t} + \mu_{0,j} \right).
$$

5
If the true parameters $\mu_{0,1} \geq \mu_{0,j}$, $\forall j$, we have regime 1 as the highest intensity level among all regimes. If the number of regimes is large, this model can be compared to common/industry-specific frailty model as Duffie et al. (2009) in equation [2].

Due to the unobservable regime indicator $s_t$, the ML estimation becomes a filtering problem. Unlike classical doubly-stochastic problem [1] the log-likelihood of regime-switching intensity model is

$$L(\theta|\mathcal{F}_T) = \sum_{t=1}^{T} \log \left[ 1'(\xi_{t|t-1} \odot \eta_t) \right],$$

where $\theta$ consists of all regime-specific parameters and transition probabilities of regime indicator. $1$ is $(N \times 1)$ vector of ones and $\odot$ is element-by-element multiplication. $\xi_{t|t-1}$ is a $N \times 1$ vector of predicting probability of regime indicator with elements $P(s_t = j|\mathcal{F}_{t-1})$. $\eta_t$ is a $N \times 1$ vector of regime-specific default likelihood. We initial our estimation procedure with unconditional probabilities of regime indicator. We list detail estimation algorithm in Appendix A.

Transition probabilities are useful quantities when inferring empirical results of regime-switching intensity model. Higher $p_{ij}$ estimate means higher probability of the regime $j$ followed by regime $i$. The expected duration of regime $j$ can be estimated via $1/(1 - p_{jj})$. Predicting probability, $\xi_{t|t-1}$, uses the information at time $t - 1$ to predict the probabilities of regimes in next period. Smoothed probability, $\xi_{t|T}$, infers the regimes using all available information till time $T$. In our empirical study of US default data later, we will use predicting probabilities in out-of sample prediction ability evaluation; and use smoothed probability to obtain the in-sample performance. Predicting probabilities is the by-product of MLE algorithms, and smoothed probabilities will be estimated using Kim (1994)’s algorithm; see Appendix A.

### 3.3 Covariates selection via LASSO approach

LASSO is a shrinkage and variable selection method. It minimizes the log likelihood subject to the sum of the absolute values of parameters being constrained by a constant. LASSO concept has been widely applied to linear regression, logistic regression, multinomial and Cox regression models; see Tibshirani (1997), Friedman, Hastie, and Tibshirani (2010).

LASSO estimates solve the problem

$$\max_{\mu} L(\mu|\mathcal{F}_T) = \sum_{t=1}^{T} \log \left[ l_t(\mu|\tilde{W}_t, D_t) \right]$$

subject to $\sum_{p=1}^{k} |\mu_p| \leq s$ (6)
where $\tilde{W}_t$ is the standardized value of $W_{t,p}$. We denote LASSO estimates as $\tilde{\mu}(s)$ to show that it is varying with respect to user-specified constant $s$.

The shrinkage level is controlled by a pre-specified constant constant $s > 0$. Suppose $\mu$ is the maximizer of equation 4 with standardized covariates, and $s_0 = \sum_{p=1}^k |\hat{\mu}_p|$. Then if $s \geq s_0$, the solution to equation 6 are the solutions of usual doubly-stochastic intensity model. If $s = s_0/2$, then the original intensity coefficients are shrunk by about 50% on average. By specify sufficiently small $s$, some of the coefficient estimates will be exactly zero, thus the LASSO does a kind of continuous subset selection.

The usual practice of using LASSO to choose covariates is, first, computing the whole path of estimates as $s$ is varied. Then, choosing the suitable cutoff value $s$ by means of generalized cross-validation (GCV) statistics or Akaike-style criterion. The final covariates chosen in intensity function are those variables with nonzero coefficient estimates under the suitable cutoff value. We employ the GCV style statistics to determine $s$ as suggested by Tibshirani [1997] in Cox regression content. Define the number of nonzero coefficient as $p(s)$, and GCV is

$$GCV(s) = \frac{-L(\tilde{\mu}(s))}{T^2 [1 - p(s)/T]^2}. \tag{7}$$

Intuitively, GCV criterion is to balance model accuracy and model parsimony. GCV style statistics penalizes the log likelihood with large $p(s)$ to achieve parsimony in model selection.

### 4 Empirical analysis

This section is organized as follows: In section 4.1 we illustrate our data set and covariates chosen by LASSO approach. Next, rigorous examination of the existence of regime switching effect in intensity model will be presented in section 4.2. Finally, we compare the default prediction accuracy among various classical and RS intensity models in section 4.3 and 4.4.

#### 4.1 Data and covariates selection

Our default data are collected from CRSP, COMPUSTAT and Bloomberg. We include US-listed companies between 1990-2009. Financial firms (SIC 6001-6999) and utility firms (SIC 4900-4999) are excluded. Our data has 10,950 firms, and among them, 1,319 went default. The total firm-month observations is 947,132 months. We consider a firm as defaulted if any of

---

1. Given covariate $p$, $\tilde{W}_{i,p,t} = (W_{i,p,t} - \bar{W}_p)/\hat{\sigma}_{W_p}$, where $\bar{W}_p = \sum_i \sum_t W_{i,p,t} / (N(T_i - t_0^i))$ and $\hat{\sigma}_{W_p}^2 = \sum_i \sum_t (W_{i,p,t} - \bar{W}_p)^2 / (N(T_i - t_0^i))$.

2. Part of this research was done during H.-C. Chuang’s visit to Risk Management Institute, National University of Singapore. We greatly thank professor Jin-Chuan Duan and Tao Wang’s help on providing the Bloomberg default indicators and DtD measure for this study. Any responsibility for errors in fact or judgments is ours.
three criteria is satisfied: de-listed firm with CRSP code 574, firm with COMPUSTAT delist reason 02 or firm marked as Default Corp Action in Bloomberg CACS function.

Table 1 reports the annual number of default firms and year-end number of active firms (survive or newly existence) in our sample period. Figure 1 plots monthly number of default firms, VIX index and S&P500 index return series. It is evident that defaults cluster during the financial crisis 1997-1998 and internet bubble 2001; however, during 1992-1996 and 2006-2007, the default numbers are relatively low. The default number during 2008-2009 is not as high as expected may due to the highly government intervention during this period. Those figures suggest the mean arrival rates of default events of US-listed companies are highly time-varying.

We construct 13 firm-specific variables and 29 macro factors when using LASSO method to choose covariates. Table 5 reports the variables we used in LASSO estimation. Firm-specific variables are constructed form COMPUSTAT quarterly files, and CRSP monthly files. They includes Dtd measure, net income to total asset (NITA), one year trailing stock return (RET), total liability to total asset (TLTA), return volatility measure (STD) and others. All firm-specific variables are winsorized using a 5/95 percentile interval to prevent outliers as Shumway (2001) and Duffie, Saita, and Wang (2007) noted. We also follow Campbell, Hilscher, and Szigeti (2008)’s method to adjust book equity values to prevent mis-measurement in the values in total asset.

Macro factors are collected from website of Federal Reserve Bank of Saint Louis. They include short-term, and long-term rate, term spread, credit spread, VIX index, S&P 500 index return and other macro fundamental factors. We lag some macroeconomic variables and all accounting variables for three months when estimation LASSO coefficients in order to mimic the real default prediction scenario. Market price data such as stock return and interest rates are assumed to be observable in real time.

Figure 2 shows the estimated coefficients from the LASSO fit as the function of the constraint parameter s. We use US default data during 1990 to 2008 in LASSO estimation. It indicates that Dtd has the most dominant effect in intensity function no matter what s is. TLTA, NITA, STD and VIX index also has nonzero coefficients among most of the shrinkage levels. It is noted that most firm-specific variables have larger influences to default intensity than macro variables. In small s, most of coefficients of macro factors are zeros except VIX, which reflect its character as fear gauge in stock market.

GCV plot of our LASSO analysis is shown in Figure 3. It shows the typical V shape in most model selection criterion. The small constrain s will render higher GCV statistics due to the small log likelihood; higher constrain s will include too many explanatory variables in model and increases p(s), the number of effective coefficients. The minimum GCV statistic occurs at s = 5. This is our LASSO shrinkage level, which is about 35.71% proportion to sum
of absolute coefficients of model without LASSO constrain. Therefore, the variables we choose as covariates in the intensity function are DtD, TLTA, NITA, STD and VIX index. We denote this model as $M_L$ model thereafter.

4.2 Is regime-switching effect statistically significant?

We employ Hansen (1992, 1996) supreme likelihood ratio test to validate the existence of regime-switching effect in the level and in the factor loadings of $M_L$ model. The null hypothesis is no regime-switching effect model is correct model specification. For example, if we want to test the existence of regime-dependent coefficients in the factor loading of DtD, the alternative specifications of intensity function are

$$H_A: \Lambda(X_{i,t}, s_t = j, \mu_j) = e^{(\mu_0+\mu_{1,j}DtD_{i,t}+\mu_2NITA_{i,t}+\mu_3TLTA_{i,t}+\mu_4STD_{i,t}+\mu_5VIX_{i,t})},$$

for $j = 1, 2$. We denote this alternative model as $H_A: RS_{DtD}$ model. Under the null hypothesis of no RS effect in loading coefficient, i.e., $\mu_{1,1} = \mu_{1,2}$, the log likelihood value can not be affected by the transition probability $p_{11}$ and $p_{22}$. This is noted as nuisance parameter problem, and the standard hypothesis testing procedures such as Wald test or Likelihood Ratio (LR) test can not be used in this type of hypothesis testing problem.

Hansen (1992, 1996) proposes a conservative testing procedure to deal with this issue. Likelihood function is viewed as an empirical process of the unknown parameters. He derives a bound for the asymptotic distribution of a standardized LR statistics, and uses simulation to construct the distribution (p-values) of this empirical process. In this test, we need to decide an autocorrelation lag which accounts for the autocorrelation structure in likelihood difference of null and alternative models. It is very computing intensive procedure and requires optimizing each RS models for predetermined grids of each combination of nuisance parameters and constrained coefficients. Appendix B provides detail procedures to implement this test in the framework of intensity model.

Our null hypothesis is that no regime-switching effect LASSO model ($H_0: M_L$ model) is preferred to regime-switching model. Due to the computing limitation, we construct 6 different alternatives RS models with one-parameter switching only i.e., $H_A: RS_I$, or $H_A: RS_{DtD}$, or $H_A: RS_{VIX}$, or $H_A: RS_{NITA}$, or $H_A: RS_{TLTA}$ or $H_A: RS_{STD}$, where $RS_I$ stands for regime-dependent intercept model. For each model, we specify about 1,200 grid points and use 1,000 bootstrapped samples to construct the empirical distribution of each test statistics.

Table 2 reports p-values in different autocorrelation lag. For each column in table 2, the null hypothesis is $M_L$ model is more suitable than $RS_x$ model (regime-dependent coefficient in $x$ variable). For example, we want to compare $M_L$ model to $RS_{DtD}$ model. The test statistics is
6.634 and p-value is 0.000 no matter in which level of autocorrelation lag parameter. Standard LR test statistics is 273.461 which is much greater than usual asymptotic chi-square rejection level. (e.g. $\chi^2$ is 11.34 with probability 99% in 3 degree of freedom). This suggests that regime-switching effect in loadings of DtD is very significant.

Test results in table 2 favor the models with regime-switching effect in intercept, DtD and VIX loading coefficients under the p-values less than 5%. However, it shows weak evidence of RS effect in TLTA factor loading and no RS effect in NITA and STD variables. Significance of the RS effect in intercept provides another evidence to the existence of frailty factor in Duffie et al. (2009) and Koopman et al. (2009). We also show that factor loading to observable risk factors also needs to considered in empirical studies of US-listed nonfinancial default data during 1990 to 2008.

4.3 Time series model fit and predicted default frequency

We construct a series of regime-switching models for our empirical study on US-listed companies, 1990-2008. Specifically, the default intensity function of firm $i$ at regime $j$ using the risk factors chosen by LASSO method is of the form:

$$\Lambda_{i,t}(s_t = j) = \exp \left( \mu_{0,j} + \mu_{1,j}DtD_{i,t} + \mu_{2,j}NITA_{i,t} + \mu_{3,j}TLTA_{i,t} + \mu_{4,j}STD_{i,t} + \mu_{5,j}VIX_{t} \right).$$

For simplicity, we use the notation RS$_{ijk}$ to indicate the model with regime-switching effect in all $i$-th, $j$-th and $k$-th parameters in intensity function above. For instance, RS$_{015}$ indicates the model with RS effect in intercept, risk exposure of DtD, and VIX index. In order to comparable to Duffie et al. (2009) frailty model, the model we estimated needs including RS effect in intercept. In other words, we estimate RS$_{0m}$, RS$_{0mn}$, RS$_{0mnl}$, RS$_{0mnlk}$ and RS$_{012345}$, for $1 \leq m,n,l,k \leq 5$, but we do not consider model without subscript zero, e.g. RS$_{234}$ model.

We use AIC and BIC as criterion to choose among RS intensity models. Result indicates that RS$_{015}$ is the best model specification among all RS intensity models estimated, however, the coefficient of $\mu_{5,1}$ is highly insignificant (p-value is 22.50%). For model parsimony, we use RS$_{01}$ model as our empirical model in evaluating in-sample and out-of sample performance of default prediction. We denote this model as RS$_{I,DtD}$ model to signify that regime-dependent coefficients in both intercept and loading of DtD measure.

Table 3 reports maximum likelihood estimates (MLEs) for various alternative specifications. Standard error estimates, shown in parentheses, are obtained from outer product of gradient. Log likelihood value and corresponding model selection criteria, AIC and BIC of each model are also reported in Table 3. In the first columns, we follow Duffie, Saita, and Wang (2007) and estimate a model (model $M_D$) with four standard variables: DtD, firm one-year trailing
return (RET), three month treasury rate (G3M), and S&P 500 index return. The second column is the MLEs of LASSO model, denoted as $M_L$ model. $M_L$ model has five risk factors, DtD, NITA, TLTA, STD and VIX. Comparing $M_L$ to $M_D$ model, we find that LASSO methods improves the log likelihood from -7827.874 to -7313.331. This demonstrates that importance of well-specified risk factors in doubly-stochastic intensity model.

We report MLEs of $RS_{I,DtD}$ model in the last column of Table 3. MLEs of regime-switching intercept ($RS_I$), and regime-switching DtD ($RS_{DtD}$) models are also listed in column three and four for comparison purpose. There are several interesting results shown in this table. First, it is evident that all RS intensity models outperform the classical intensity model in terms of both AIC and BIC. $RS_{I,DtD}$ model has more than 150 log likelihood value than $M_L$ model. Second, NITA and TLTA have large impact on the default intensity. and the coefficients are -8.081 and 3.002 respectively in $RS_{I,DtD}$ model. NITA and TLTA almost maintain the same levels no matter in which model specifications. This result show that standard accounting variables such as net income and total liability explains large part of default rate variation.

Third, $RS_{I,DtD}$ model outperforms $RS_I$ model implies that not only the level of intensity function is regime-dependent, but also the the factor loading to DtD measure. Regime-dependent factor loading to DtD measure is highly significant. In $RS_{I,DtD}$ model, for high-default rate regime (regime 1), coefficients of DtD is -0.625, however, in low-default rate regime (regime 2), the coefficient drops to -3.901. The expected durations of high- and low-intensity regimes are 3 and 2.8 months respectively.

In the rest of this section, we compare the time series of monthly default number to the model-implied aggregate intensity. Moreover, we also examine the cross-sectional sample fitness of RS intensity model. Figure 4 plots monthly defaults number during January 1990 through December 2008 (bars), and the aggregate (across firms) estimated default intensities of $M_L$ model (dotted black line) and regime-specific aggregate intensities of $RS_{I,DtD}$ model (red lines). At month $t$, we define $j$-th regime-specific aggregate (across firms) intensity as $\Lambda_j^t$, 

$$\Lambda_j^t := \sum_{i=1}^{N_t} [\Lambda(X_{i,t}, s_t = j; \hat{\mu}_j) \Delta t].$$

---

3 We also use EM and Gibb’s sampler proposed in [Duffie et al.] (2009) to estimate frailty model. Following the notations in equations 3 and 4, the parameter estimates for constant, DtD, NITA, TLTA, STD and VIX are $\hat{\mu} = (-4.829, -0.959, -7.777, 3.254, 0.641, -0.009)'$, with corresponding standard error as $(0.161, 0.0361, 0.364, 0.133, 0.078, 0.004)'$. The parameters associate with OU process is $(\hat{\gamma}, \hat{\kappa}) = (0.259, 0.118)'$, with standard errors $(0.019, 0.072)'$. It is noted that $\hat{\gamma}$ is strongly significant and $\hat{\kappa}$ is also significant at 95% confidence level. The log likelihood of frailty model is −7214.616. In this exercise, we find strong evidence of the existence of frailty process in US default data even with different covariates to [Duffie et al.] (2009). The parameter estimates of observable factors are close to estimates of $RS_I$ model in table 3. Log likelihood of frailty model is slightly higher than $RS_I$ model and lower than $RS_{I,DtD}$ model. This result is consistent to our argument that the risk exposure to observable factors needs to be considered in default prediction modelling.
The dashed red line represents the high-default-rate regime aggregate intensity, and the solid line is the low-default-rate regime one. Gray zone means that smoothed probabilities of high-default-rate regime are greater than 0.5. In most of gray zones, it can be seen that high-default-rate regime aggregate intensity is more suitable than that of $M_L$ model to describe the actual month default rate. In other aspect, the non-gray time zone, the low-default-rate regime aggregate intensity is better fit for default numbers than $M_L$ model. Figure 4 demonstrates that $RS_{I,D,t,D}$ model capture the time variation of mean default rates better than $M_L$ model.

In additional to time series aspect, we also employ the predict default frequency plot to examine the distributional properties of model fitness. Conditional on intensity regime $j$ at time $t$, probability of zero and one defaulters out of $N_t$ risky firms over the period $[t, t + \Delta t]$ can be expressed as

$$P_j^t(0) := P\left(\sum_{i=1}^{N_t} D_{i,t} = 0 | s_t = j\right) = \frac{N_t}{\prod_{i=1}^{N_t} e^{-\Lambda(X_{i,t}, s_t=j; \hat{\mu}_j)\Delta t}};$$

$$P_j^t(1) := P\left(\sum_{i=1}^{N_t} D_{i,t} = 1 | s_t = j\right) = \frac{N_t}{\prod_{i=1}^{N_t} \left(1 - e^{-\Lambda(X_{i,t}, s_t=j; \hat{\mu}_j)\Delta t}\right) \prod_{l=1, l\neq i} e^{-\Lambda(X_{l,t}, s_t=j; \hat{\mu}_j)\Delta t}}.$$

If the default number is more than one company, the analytical expression is complex. Therefore, we use the convolution algorithm proposed by Duan (2010) to create the predicted default frequency of $RS_{I,D,t,D}$ model at each month and calculate the average (cross time) monthly regime-specific default frequency plot.

Figure 5 presents the $RS_{I,D,t,D}$ model-implied average (across time) default frequency of each regime as well as monthly default frequency of US-listed companies. We also draw the $M_L$ model implied average default frequency plot to comparison purpose. Bars in this figure are monthly default frequencies. For example, 3.51% month in the sample period (8 out of 228 months) has zero default event within a month; 15.35% (35 out of 228) month has 4 defaulters within a month. $RS_{I,D,t,D}$ model seizes different characteristics of the frequency plot with different regime-specific distributions. High-default rate regime default frequency captures the distribution of highly clustered default events such as more than 10 defaults per month and low-default rate regime capture the left skewness of actual monthly default frequency plot.

4.4 Default prediction performance

We compare the in-sample and out-of-sample default prediction performance of regime-switching models to classical doubly-stochastic model. Receiver operating characteristic (ROC) curve and area under ROC (AUC) measures are our main tool to compare one-step-ahead prediction ability of different models. Brier (1950)'s score (square prediction error) and absolute
prediction errors are also reported in out-of-sample evaluation.

ROC diagram is a summary for evaluating the trade off between false positive rate and true positive rate. Each default prediction model can be viewed as a binary classifier which assign a default probability risky firms. Given a predicted PD as a threshold value, a confusion matrix (or contingency table) is defined as:

<table>
<thead>
<tr>
<th>Prediction</th>
<th>Actual Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Default</td>
</tr>
<tr>
<td></td>
<td>True Positive (TP)</td>
</tr>
<tr>
<td></td>
<td>True Negative (TN)</td>
</tr>
<tr>
<td>Total</td>
<td>D</td>
</tr>
</tbody>
</table>

where D and S (\(\hat{D}\) and \(\hat{S}\)) are actual default number and survive number (predicted default number and predicted survive number) in the evaluation period. True positive rate (TPR) is TP/D and false positive rate (FPR) is FP/S. Through varying the threshold PD value, we have a series of TPR and FPR values and ROC curve summaries those information. It is noted that the best prediction model yields a upper left corner of the ROC diagram, which represents no false positive rate and all are true positive cases. A completely random flipping coin design would give a 45\(^\circ\) line to show its no-discrimination nature. Therefore, the area under ROC curve (AUC) is a measure for comparing different models. Friedman, Tibshirani, and Hastie (2009) provides detailed introduction about ROC analysis.

We plot the annual aggregate AUC curves for \(M_D\), \(M_L\), \(RS_I\) and \(RS_{I,D|D}\) model in figure 6. In every month, we estimate each model using all data available up to time \(t\) and record each firms estimated default probability. The default probability for regime-switching models is the weighting-average by smoothed probability at each month. In the end of year, we aggregate past 12 months default probabilities and actual default events to calculate annual aggregate AUC. We do not calculate monthly AUCs because the default number of each month is extremely low. Figure 6 shows that the \(M_D\) model has the lowest AUCs and \(M_L\) and \(RS_I\) model have similar AUCs. This finding suggest that carefully choose covariates into default intensity function can improve the in-sample default prediction precision.

The \(RS_{I,D|D}\) outperforms all other models in every year AUCs through 1991 to 2008. For instance, in 1997, AUCs are 84.76\%, 87.81\%, and 87.87\% for \(M_D\), \(M_L\) and \(RS_I\) respectively. AUC of the \(RS_{I,D|D}\) model in 1997 is as high as 89.67\%. This result suggest regime-switching factor loading specification contributes to default prediction precision beyond common frailty-type model and well-specified doubly-stochastic intensity model.

In addition to in-sample prediction performance, we also use 2009 default data to evaluate the out-of-sample prediction ability of RS models. We estimated each RS models using the
information at the beginning of each month, then record their prediction values month by month. Aggregating 12 month prediction PD and corresponding month-end actual default indicators, we plot the ROC curve and calculate Brier score (BS) and absolute error (AE) as:

\[ BS = \sum_{i=1}^{N^*} (D_i - \hat{PD}_i)^2; \quad AE = \sum_{i=1}^{N^*} |D_i - \hat{PD}_i|. \]

\( N^* \) denotes aggregate number of survival firms at end of each 12 months, and \( D_i \) is the default indicator equals to one if firm defaulted within the month. \( \hat{PD}_i \) is the begin-of-month model implied default probability of firm \( i \).

The one-month ahead prediction ROC curve during January 2009 to December 2009 is shown in Figure 7 and corresponding performance measures are listed in Table 4. We plot the ROC curve for \( M_D \) and \( RS_{I,DtD} \) for comparison purposes. ROC curve of \( RS_{I,DtD} \) is more toward upper-left corner than ROC curve of \( M_D \) model which shows in average, the prediction sensitivity (the proportion of actual positives which are correctly identified) of \( RS_{I,DtD} \) model is higher than that of \( M_D \) model. Table 4 shows the AUC increases form 93.142% to 94.832%. Brier score (absolute prediction error) are also decreases form 0.415% (0.103%) to 0.339% (0.101%) when we using \( RS_{I,DtD} \) model as default predictor instead of \( M_D \) model.

5 Conclusion

In this paper, we propose regime-switching intensity model to characterize the time-varying nature of coefficients in default intensity model. This paper makes several contributions to the literature on financial distress. First of all, to avoid model mis-specification in the classical intensity model framework, we use LASSO method to do variable selection among more than forty firmspecific and marco risk factors. Those chosen variables have better in-sample and out-of-sample prediction power than standard risk factors considered in Duffie, Saita, and Wang [2007].

Second, we adopt statistical test to validate the existence of regime-switching effect in the double-stochastic intensity model framework. Hansen [1992, 1996] test demonstrates that the intercept of intensity function as well as the factor loading of DtD measure have significant regime-dependent effect in the US-listed default data during 1990 to 2008. This result aligns to Duffie et al. [2009] and Koopman, Lucas, and Schwaab [2010] who claim the existence of frailty factors in reduced form default prediction models. In addition, it points out that factor loadings are also state-dependent. Therefore, it is not sufficient to only consider the common/industry frailty process in intensity modelling. Time-varying risk exposure/factor loadings to observable risk factors should also be taken into consideration when constructing default prediction models.

Third, we develop an estimation algorithm when the state variable is Markovian and illustrate the proposed regime-switching intensity model using the default data of US-listed

Appendix

A. Maximum likelihood estimation

The log-likelihood function of default process needs to be expressed as the cross sectional product of firm’s default likelihoods first and then aggregate the conditional default likelihood over time in contrast to the original setting of Duffie, Saita, and Wang (2007). In this representation, we can utilize Hamilton (1994) algorithm to estimate the parameters of the unobservable regime indicator.

Assume regime indicator, \( s_t \), follows \( N \) states (\( N \) regimes) first order Markov chain with transition probability \( p_{ij} \), for \( 1 \leq i, j \leq N \). The transition matrix is

\[
\Pi = \begin{pmatrix}
p_{11} & p_{21} & \cdots & p_{N1} \\
p_{12} & p_{22} & \cdots & p_{N2} \\
\vdots & \vdots & \ddots & \vdots \\
p_{1N} & p_{2N} & \cdots & p_{NN}
\end{pmatrix},
\]

where \( p_{ij} := P(s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots) = P(s_t = j | s_{t-1} = i) \). For simplicity, we consider that \( X_{i,t} \) and \( s_t \) are mutually independent processes and \( s_t \) is a one dimensional discrete Markov chain with time-invariant transition probabilities.

The log-likelihood of default process is

\[
\mathcal{L}(\theta | D, X_T) = \sum_{t=1}^{T} \log \left[ 1' (\xi_{t|t-1} \odot \eta_t) \right], \tag{8}
\]

where \( \theta \) consists of all regime-specific parameters and transition probabilities of regime indicator, i.e., \( \theta = (\mu'_j, p_{i,j}, 1 \leq i, j \leq N) \). \( 1 \) is \( (N \times 1) \) vector of ones and \( \odot \) is element-by-element multiplication. \( \xi_{t|t-1} \) is a \( N \times 1 \) vector of predicting probability of \( s_t \) with element \( P(s_t = j | F_{t-1}) \). Conditional on each regime, each element of \( \eta_t \) is default likelihood of risky firms (surviving, defaulted and newly existent firms) at time \( t \), i.e.,

\[
\eta_t = \begin{bmatrix}
l_t(\mu_1 | X_t, D_t, s_t = 1) \\
l_t(\mu_2 | X_t, D_t, s_t = 2) \\
\vdots \\
l_t(\mu_N | X_t, D_t, s_t = N)
\end{bmatrix},
\]
where

\[ l_t(\mu_j | X_t, D_t, s_t = j) = \prod_{i \in A_t} \frac{e^{-\Lambda(X_{i,t}, s_t = j; \mu_j) \Delta t}}{1 - e^{-\Lambda(X_{i,t}, s_t = j; \mu_j) \Delta t}}. \]

for \( j = 1, \ldots, N \). That is, conditional on the state as regime \( j \), the likelihood of the firm defaulting or surviving at time \( t \) is the same as classical doubly-stochastic intensity model [4].

To obtain the log-likelihood function \( \mathcal{L} \), we use Hamilton (1994) filter to infer the unobservable regime indicator. Given the starting value \( \xi_{1|0} \) as unconditional probability of Markov chain, \( \pi \), we iterate the following two equations for \( t = 1, \ldots, T \) to obtain log-likelihood function \( \mathcal{L} \):

\[ \xi_{t|t} = \frac{\xi_{t|t-1} \odot \eta_t}{1'(\xi_{t|t-1} \odot \eta_t)}; \quad (9) \]
\[ \xi_{t+1|t} = \Pi \xi_{t|t}. \quad (10) \]

\( \xi_{t|t} \) is known as filtering probability of regime indicator with element \( P(s_t = j|\mathcal{F}_t) \). The unconditional probability of regime is defined as \( \pi := (A'A)^{-1}A'e_{N+1} \), where \( A = [I_N - \Pi; 1'] \). \( I_N \) is identity matrix and \( e_{N+1} \) denotes the last column of identity matrix \( I_{N+1} \).

In order to obtain more precise inference about unobservable regime indicator, we adopt Kim (1994) algorithm to generate the smoothed probability of \( s_t \) after obtaining MLEs. Smoothed probability of regime \( j \) is the probability of being in this regime at time \( t \) conditional on all information available at time \( T \). Denote \( \xi_{t|T} \) as \( N \times 1 \) smoothed probabilities with element \( P(s_t = j|\mathcal{F}_T) \). \( \xi_{t|T} \) can be evaluated via:

\[ \xi_{t|T} = \xi_{t|t} \odot \left\{ \Pi' \left[ \xi_{t+1|T}(\div)\xi_{t+1|t} \right] \right\}, \]

where \( (\div) \) denotes element-by-element division. This algorithm starts with filtering probability \( \xi_{T|T} \) and predicting probability \( \xi_{T|T-1} \) and backward iterates for \( t = T - 1, T - 2, \ldots, 1 \).

B. Supremum likelihood ratio test

We briefly discussed the how to implement the test proposed by Hansen (1992, 1996) in the framework of intensity model. Define the null hypothesis as classical doubly-stochastic model. The alternative hypothesis is regime-switching in one parameters with 2 regimes. We can re-parameterize default intensity of firm \( i \) as

\[ \Lambda(X_{i,t}, s_t = j; \mu_j) = \exp \left( \sum_{p=0}^{k-1} \mu_p X_{i,p,t} + [\mu_{k,1} + \mu_{k,2}(s_t - 1)]X_{i,k,t} \right), \quad j = 1, 2. \]
In our case, $X_{i,k,t}$ can be constant, DtD, VIX index, or other covariates chosen by LASSO in section ???. Partition the parameter vector as $\theta = (\mu_{k,2}, p', \mu_p')$, where $p = (p_{11}, p_{22})'$. $\mu_p = (\mu_0, \cdots, \mu_{k-1}; \mu_{k,1})'$. The null and alternative hypothesis is therefore,

$$H_0 : \mu_{k,2} = 0, \quad H_A : \mu_{k,2} \neq 0.$$  

Note that $p$ is not identified under $H_0$. That is the log likelihood $L(\mu_p, 0, p)$ does not depend on $p$ under the null hypothesis. Given $\gamma = (\mu_{k,2}, p')'$, the concentrated MLEs of $\mu_p$ and concentrated likelihood function are

$$\hat{\mu}_p(\gamma) = \arg \max L(\mu_p, \gamma); \quad \hat{L}(\gamma) = L(\gamma, \hat{\mu}_p(\gamma)).$$

Define the likelihood ratio process and its corresponding sample variance are

$$\hat{L}R(\gamma) = \hat{L}(\gamma) - \hat{L}(0, p); \quad \hat{V}(\gamma, \hat{\mu}_p(\gamma)) = \sum_{t=1}^T q_t(\gamma, \hat{\mu}_p(\gamma))^2.$$  

where the form of $q_t$ is the difference of unlikelihood under the null and alternative. Precisely, suppose the log-likelihood of regime-switching intensity model can be rewritten in the form $L(\gamma, \mu_p) = \sum_{t=1}^T L_t(\gamma, \mu_p)$. The definition of $q_t$ is

$$q_t(\gamma, \hat{\mu}_p(\gamma)) = L_t(\gamma, \hat{\mu}_p(\gamma)) - L_t(0, p, \hat{\mu}_p(0, p)) - \frac{1}{T} \hat{L}R(\gamma).$$

The standardized supremum LR statistics is

$$\hat{L}R^* = \sup_{\gamma \in E} \frac{\hat{L}R(\gamma)}{\hat{V}(\gamma, \hat{\mu}_p(\gamma))^{1/2}}.$$  

where $E : R \times I \times I$ is parameter space of $\gamma$, $I$ is $[0,1]$ interval. Hansen (1992) proves the following result

$$P(\hat{L}R^* \geq x) \leq P(\sup_{\gamma \in E} \hat{Q}(\gamma) \geq x) \rightarrow P(\sup Q^* \geq x),$$  

where $\sup_{\gamma \in E} \hat{Q}(\gamma)$ is the large sample counterpart of $\hat{L}R^*$. Hansen (1992) provides a bound for the standarized standardized supremum LR statistics in terms of the distribution of the random variables $\sup Q^*$.

The distribution of the random variable $\sup Q^*$ is non-standard and can not be tabulated. Hansen (1992) suggest use simulation method to generate this distributions. It is noted that
$Q^*$ is a Gaussian process with mean zero and the covariance function $K(\gamma_1, \gamma_2)$. The exact form of $K(\gamma_1, \gamma_2)$ and its estimator, $\hat{K}(\gamma_1, \gamma_2)$, can be found in Hansen (1996). As a Gaussian process is completely determined by its covariance function, the supremum of each generated process can approximate the distribution of sup $Q^*$. Critical values and p-values of supremum LR test statistic can be calculated from simulated distribution. Hansen (1996) suggests that generating a sample of i.i.d. standard normal random variables $\{\varepsilon_t\}_{t=1}^{t+M}$ and computing

$$\frac{\sum_{k=0}^{M} \sum_{t=1}^{T} q_t(\gamma, \hat{\mu}_p(\gamma))\varepsilon_{t+k}}{\sqrt{1 + \hat{M} V(\gamma, \hat{\mu}_p(\gamma))^{1/2}}}.$$  \hspace{1cm} (11)

$M$ is the autocorrelation parameter to account for the serial correlation of $q_t(\gamma)$. It can be shown that equation (11) is a zero mean Gaussian process with covariance function $\hat{K}(\gamma_1, \gamma_2)$ which is an asymptotic approximate to $K(\gamma_1, \gamma_2)$. As the theory does not give any specific guidance for choose of $M$, we follow Hansen (1996)'s suggestion, calculate the tests for $M = 0, \cdots, 5$.

Above simulation methods need to determine the grid of $\gamma$. For each value of $\gamma$, the constrained likelihood needs to be optimization. Take $RS_{DLD}$ model for example, we set up the 20 grid points of $\mu_{k,2} : -1.900, -1.874, -1.848, \cdots, -1.4$. Eight grid points for each $p_{11}$ and $p_{22}$. $p_{11} : 0.3, 0.38, 0.46, \cdots, 0.9$. $p_{22} : 0.5, 0.555, 0.61, \cdots, 0.9$. For each combination, we estimate constrained optimization of $RS_{DLD}$ model and compute $q_t$. There are 1,280 optimizations in total. This is a huge computational burden, even we only have three parameters in $\gamma$. Therefore, we do not consider more than one regime-switching parameter case in our supermen likelihood ratio test study.
References


Figure 1: Time series of monthly default numbers (bars), and VIX index (line) and S&P500 index return multiplied by 100 (dashed line).

Table 1: Annual default number during 1990 to 2009

<table>
<thead>
<tr>
<th>year</th>
<th>default firm no.</th>
<th>active firm</th>
<th>default to active firm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>31</td>
<td>3,695</td>
<td>0.008 %</td>
</tr>
<tr>
<td>1991</td>
<td>47</td>
<td>3,665</td>
<td>0.012 %</td>
</tr>
<tr>
<td>1992</td>
<td>43</td>
<td>3,767</td>
<td>0.011 %</td>
</tr>
<tr>
<td>1993</td>
<td>27</td>
<td>4,045</td>
<td>0.006 %</td>
</tr>
<tr>
<td>1994</td>
<td>30</td>
<td>4,394</td>
<td>0.006 %</td>
</tr>
<tr>
<td>1995</td>
<td>42</td>
<td>4,554</td>
<td>0.009 %</td>
</tr>
<tr>
<td>1996</td>
<td>37</td>
<td>4,771</td>
<td>0.007 %</td>
</tr>
<tr>
<td>1997</td>
<td>77</td>
<td>5,088</td>
<td>0.015 %</td>
</tr>
<tr>
<td>1998</td>
<td>116</td>
<td>4,959</td>
<td>0.023 %</td>
</tr>
<tr>
<td>1999</td>
<td>117</td>
<td>4,574</td>
<td>0.025 %</td>
</tr>
<tr>
<td>2000</td>
<td>129</td>
<td>4,459</td>
<td>0.028 %</td>
</tr>
<tr>
<td>2001</td>
<td>185</td>
<td>4,260</td>
<td>0.043 %</td>
</tr>
<tr>
<td>2002</td>
<td>118</td>
<td>3,934</td>
<td>0.029 %</td>
</tr>
<tr>
<td>2003</td>
<td>89</td>
<td>3,661</td>
<td>0.024 %</td>
</tr>
<tr>
<td>2004</td>
<td>42</td>
<td>3,538</td>
<td>0.011 %</td>
</tr>
<tr>
<td>2005</td>
<td>43</td>
<td>3,423</td>
<td>0.012 %</td>
</tr>
<tr>
<td>2006</td>
<td>32</td>
<td>3,339</td>
<td>0.009 %</td>
</tr>
<tr>
<td>2007</td>
<td>25</td>
<td>3,217</td>
<td>0.007 %</td>
</tr>
<tr>
<td>2008</td>
<td>48</td>
<td>3,160</td>
<td>0.015 %</td>
</tr>
<tr>
<td>2009</td>
<td>41</td>
<td>2,533</td>
<td>0.016 %</td>
</tr>
</tbody>
</table>
Table 2: p-values of supremum LR test

<table>
<thead>
<tr>
<th>Lag</th>
<th>$RS_I$</th>
<th>$RS_{PLD}$</th>
<th>$RS_{VIX}$</th>
<th>$RS_{NITA}$</th>
<th>$RS_{TLTA}$</th>
<th>$RS_{STD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.323</td>
<td>0.061</td>
<td>0.133</td>
</tr>
<tr>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.333</td>
<td>0.072</td>
<td>0.113</td>
</tr>
<tr>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.299</td>
<td>0.067</td>
<td>0.096</td>
</tr>
<tr>
<td>3</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.271</td>
<td>0.086</td>
<td>0.108</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.276</td>
<td>0.099</td>
<td>0.151</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.262</td>
<td>0.094</td>
<td>0.134</td>
</tr>
<tr>
<td>S-LR</td>
<td>6.909</td>
<td>6.634</td>
<td>6.024</td>
<td>1.578</td>
<td>2.337</td>
<td>2.335</td>
</tr>
<tr>
<td>LR</td>
<td>160.197</td>
<td>273.461</td>
<td>152.403</td>
<td>63.721</td>
<td>144.265</td>
<td>79.176</td>
</tr>
</tbody>
</table>

1. For each column, the null hypothesis is $M_L$ model (LASSO model without RS effect) is more suitable than $RS_x$ model (regime-dependent coefficient in x variable. $I$ stands for intercept).

2. Lag is the autocorrelation lag in Hansen (1992, 1996) test. S-LR is the supremum LR ratio test statistics and LR is the standard likelihood ratio statistics.

Figure 2: Parameter estimates of intensity model as a function of LASSO constraint parameter. The vertical dotted line is the cutoff value determined by generalized cross validation type score.
Table 3: MLEs of regime-switching intensity models

<table>
<thead>
<tr>
<th></th>
<th>$M_D$</th>
<th>$M_L$</th>
<th>$RS_I$</th>
<th>$RS_{DtD}$</th>
<th>$RS_{I,DtD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_01$</td>
<td>-3.113</td>
<td>-5.714</td>
<td>-4.979</td>
<td>-5.867</td>
<td>-5.448</td>
</tr>
<tr>
<td></td>
<td>(0.062)**</td>
<td>(0.123)**</td>
<td>(0.177)**</td>
<td>(0.128)**</td>
<td>(0.133)**</td>
</tr>
<tr>
<td>$\mu_02$</td>
<td>-5.899</td>
<td></td>
<td></td>
<td>-6.985</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.178)**</td>
<td></td>
<td></td>
<td>(0.186)**</td>
<td></td>
</tr>
<tr>
<td>RET</td>
<td>-1.256</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.061)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DtD_1$</td>
<td>-1.072</td>
<td>-0.966</td>
<td>-0.972</td>
<td>-0.420</td>
<td>-0.625</td>
</tr>
<tr>
<td></td>
<td>(0.023)**</td>
<td>(0.019)**</td>
<td>(0.019)**</td>
<td>(0.031)**</td>
<td>(0.024)**</td>
</tr>
<tr>
<td>$DtD_2$</td>
<td></td>
<td></td>
<td></td>
<td>-2.106</td>
<td>-3.901</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.079)**</td>
<td>(0.269)**</td>
</tr>
<tr>
<td>NITA</td>
<td>-8.094</td>
<td>-7.763</td>
<td>-7.908</td>
<td>-8.081</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.378)**</td>
<td>(0.399)**</td>
<td>(0.376)**</td>
<td>(0.378)**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.122)**</td>
<td>(0.124)**</td>
<td>(0.127)**</td>
<td>(0.126)**</td>
<td></td>
</tr>
<tr>
<td>STD</td>
<td>0.620</td>
<td>0.655</td>
<td>0.600</td>
<td>0.609</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.076)**</td>
<td>(0.083)**</td>
<td>(0.084)**</td>
<td>(0.077)**</td>
<td></td>
</tr>
<tr>
<td>G3M</td>
<td>-0.076</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.011)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S&amp;P500</td>
<td>3.433</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.479)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIX</td>
<td>0.008</td>
<td>-0.015</td>
<td>0.005</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)**</td>
<td>(0.005)**</td>
<td>(0.003)**</td>
<td>(0.003)**</td>
<td></td>
</tr>
<tr>
<td>$p_{11}$</td>
<td>0.995</td>
<td>0.590</td>
<td>0.676</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.404)**</td>
<td>(0.292)**</td>
<td>(0.053)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.995</td>
<td>0.702</td>
<td>0.649</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.485)**</td>
<td>(0.254)**</td>
<td>(0.054)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>loglik</td>
<td>-7827.874</td>
<td>-7313.331</td>
<td>-7233.233</td>
<td>-7176.601</td>
<td>-7154.689</td>
</tr>
<tr>
<td>AIC</td>
<td>15665.748</td>
<td>14614.662</td>
<td>14484.465</td>
<td>14371.201</td>
<td>14329.378</td>
</tr>
<tr>
<td>BIC</td>
<td>15682.895</td>
<td>14594.086</td>
<td>14515.329</td>
<td>14402.065</td>
<td>14363.671</td>
</tr>
</tbody>
</table>

1. $M_D$ is Duffie, Saita, and Wang (2007)’s model and $M_L$ is the model chosen using LASSO method without regime-switching effect. $RS_I$ and $RS_{DtD}$ refer to RS model in intercept and DtD factor loading respectively, and $RS_{I,DtD}$ is the model of RS parameters in both intercept and DtD.

2. Numbers in parentheses are standard errors. Significant levels at 95%, and 99% are denoted by one, two and three stars, respectively.
Table 4: Out-of-sample performance

<table>
<thead>
<tr>
<th></th>
<th>$M_D$</th>
<th>$M_L$</th>
<th>$RS_I$</th>
<th>$RS_{Dtd}$</th>
<th>$RS_{L,Dtd}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS</td>
<td>0.415</td>
<td>0.356</td>
<td>0.237</td>
<td>0.337</td>
<td>0.339</td>
</tr>
<tr>
<td>AE</td>
<td>0.103</td>
<td>0.102</td>
<td>0.101</td>
<td>0.101</td>
<td>0.101</td>
</tr>
<tr>
<td>AUC</td>
<td>93.142</td>
<td>94.528</td>
<td>94.542</td>
<td>94.462</td>
<td>94.832</td>
</tr>
</tbody>
</table>

1 Numbers are performance measures of one-period ahead default prediction during January 2009 to December 2009. BS is Brier score and AE is absolute error. AUC is area under ROC curve.

Figure 3: GCV-style score as a function of LASSO constraint parameters.
<table>
<thead>
<tr>
<th>Name</th>
<th>Definitions / Variables Included</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Firmspecific</strong></td>
<td></td>
</tr>
<tr>
<td>ASSTE*</td>
<td>log of total asset adjusted (TA) deflated to 2005 dollars using GDP deflator</td>
</tr>
<tr>
<td>CASH*</td>
<td>cash and equivalence to TA</td>
</tr>
<tr>
<td>DtD*</td>
<td>distance to default measure</td>
</tr>
<tr>
<td>METL*</td>
<td>market value of asset to total liability</td>
</tr>
<tr>
<td>MKTBE*</td>
<td>market to book ratio</td>
</tr>
<tr>
<td>NITA*</td>
<td>net income to TA</td>
</tr>
<tr>
<td>PROFIT*</td>
<td>operating income before depreciation to TA</td>
</tr>
<tr>
<td>RATING*</td>
<td>debt rating dummy</td>
</tr>
<tr>
<td>RET</td>
<td>log($1 + R_{it}$) - log($1 + R_{S&amp;P500,t}$)</td>
</tr>
<tr>
<td>RSIZE*</td>
<td>log of market to S&amp;P500 market value</td>
</tr>
<tr>
<td>SALES*</td>
<td>sales to TA</td>
</tr>
<tr>
<td>STD</td>
<td>standard deviation of RET for one year</td>
</tr>
<tr>
<td>TLTA*</td>
<td>total liability to TA</td>
</tr>
<tr>
<td><strong>Macro</strong></td>
<td></td>
</tr>
<tr>
<td>SR Rate</td>
<td>Treasury constant maturity rate / G3M, G6M, G1</td>
</tr>
<tr>
<td>LR Rate</td>
<td>Treasury constant maturity rate / G3, G5, G7, G10</td>
</tr>
<tr>
<td>Term Spread</td>
<td>G3-G3M, G3-G6M, G3-G1, G5-G3M, G5-G6M, G5-G1</td>
</tr>
<tr>
<td>Bond Rate</td>
<td>Moody’s seasoned corporate bond yield / Aaa and Baa</td>
</tr>
<tr>
<td>Credit Spread</td>
<td>Baa-Aaa</td>
</tr>
<tr>
<td>VIX</td>
<td>Chicago board options exchange market volatility index</td>
</tr>
<tr>
<td>S&amp;P500</td>
<td>one year trailing S&amp;P500 index return</td>
</tr>
<tr>
<td>CF3*</td>
<td>Chicago Fed national activity index’s 3-month moving average</td>
</tr>
<tr>
<td>CPgro*</td>
<td>growth rate of corporate profits after tax</td>
</tr>
<tr>
<td>GDPgro*</td>
<td>growth rate of gross domestic product</td>
</tr>
<tr>
<td>NFCPATA Xgro*</td>
<td>growth rate of nonfinancial corporate business profits after tax</td>
</tr>
<tr>
<td>INDPROgro*</td>
<td>growth rate of industrial production index</td>
</tr>
</tbody>
</table>

1. Variables in * are three months lag values when estimation parameters in LASSO in order to mimic the real default prediction exercise.

2. We follow Campbell, Hilscher, and Szilagyi’s [2008] transformed method to book equity values to prevent mis-measured the values of total asset. TA := ATQ + 0.1*(ME-BE), where BE := SE - PSTKQ + TXDITCQ. SE := SEQQ or CEQQ + PSTKQ. ME := PRCCQ * CSHPRQ, S&P500MV := S&P500 market value. MVA := ME + LTQ. We use COMPSTAT quarterly data items NIQ (#69), OIBDPQ (#21), DLCQ (#45), DLTTQ (#51), PSTKQ (#55), TXDITCQ (#52), PRCCQ (#14), CSHPRQ (#15), SEQQ (#60), CEQQ (#59), ATQ (#44), LTQ (#54), SALEQ (#2), CHEQ (#36) to construct firmspecific accounting variables. After firmspecific variables are constructed, each of variables is winsorized using a 5/95 percentile interval to prevent outliers, e.g., Shumway [2001], Duffie, Saita, and Wang [2007] and Campbell, Hilscher, and Szilagyi [2008].

3. Rating is an indicator which equals to 1 if, within a year, a firm has rating greater than subordinated debt rating: "BBB-" (or senior debt rating: "A-3") in S&P’s ratings definitions.
Figure 4: Monthly number of defaults during January 1990 through December 2008 (bars), and the total across firms of estimated regime-specific default intensities of $M_L$ model. (dotted line) $RS_{L,DiD}$ model implied high-default-rate regime aggregate intensity is dashed line and solid line is low-default-rate regime aggregate intensity. Gray zone: smoothed probabilities of high-default-rate regime are greater than 0.5.
Figure 6: In-sample area under ROC curve (AUC). Aggregate one year in-sample AUC for one-period ahead default predictions during 1998 to 2008. Dashed black line: $M_D$ model. Dashed dotted blue line: $M_L$ model. Dotted red line: $RS_I$ model, and solid red curve: $RS_{I,DtD}$ model.
Figure 7: Out-of-sample receiver operating characteristic (ROC) curve. Aggregate one year out-of-sample ROC curve for one-period ahead default prediction during January 2009 to December 2009. Solid curve: $RS_{I,DtD}$ model. Dashed black curve: Duffie, Saita, and Wang (2007) model $M_D$. Area under ROC are 0.95 and 0.93 for $RS_{I,DtD}$ and $M_D$ model respectively.