Forecasting Bankruptcy with Incomplete Information

Xu, Xin

28 May 2013

Forecasting Bankruptcy with Incomplete Information*

Xin Xu†

First draft: 28 May 2013; This draft: 31 March 2014

Abstract

We propose new specifications that explicitly account for information noise in the input data of bankruptcy hazard models. The specifications are motivated by a theory of modeling credit risk with incomplete information (Duffie and Lando [2001]). Based on over 2 million firm-months of data during 1979-2012, we demonstrate that our proposed specifications significantly improve both in-sample model fit and out-of-sample forecasting accuracy. The improvements in forecasting accuracy are persistent throughout the 10-year holdout periods. The improvements are also robust to empirical setup, and are more substantial in cases where information quality is a more serious problem. Our findings provide strong empirical support for using our proposed hazard specifications in credit risk research and industry applications. They also reconcile conflicting empirical results in the literature.

JEL Codes: C41, G17, G33.

Key Words: Credit Risk Modeling, Incomplete Information, Hazard Models, Bankruptcy Forecast, Probability of Default (PD), Forecasting Accuracy, Intensity-based Models, Reduced-form Models, Duration Analysis, Survival Analysis

*I thank Edward Altman, Darrell Duffie, Kay Giesecke, Jens Hilscher, David Lando, Francis Longstaff, Ronald Masulis, Anthony Saunders, Harald Scheule, Marti Subrahmanyam, Dragon Yongjun Tang, Ke Wang and seminar participants at 7^{th} Financial Risks International Forum (2014, France), 2013 Australasia Credit Modeling Forum for helpful comments and discussions. The SSRN link of this paper is http://ssrn.com/abstract=2347079.

†This paper was written at the University of New South Wales, Australia. Xin Xu is at Commonwealth Bank of Australia. Email: xinxu75@gmail.com. The views expressed in this paper are those of the author and do not necessarily reflect those of the employer.

1 Introduction

We propose new specifications within bankruptcy hazard functions that explicitly account for information noise in the input data. The issues on input data have been of long-standing interest in credit risk modeling, including information transparency, data integrity, data quality, and their impact on models' empirical performance. These issues have recently become increasingly more important topics, during and post the global financial crisis, in the academic literature and financial press.[1] It is thus highly desirable to develop statistical models that explicitly take into account of noise in the input data. However, to the best of our knowledge, such models are virtually non-existent in the empirical credit risk literature.

Motivated by the seminal work of Duffie and Lando [2001] on modeling credit risk with incomplete information, we propose new hazard specifications that explicitly handle noisy information, and demonstrate their empirical efficacy. Compared to the previous literature on bankruptcy hazard models, our specifications have new variables in the hazard function, which are the interaction effects between proxies for the degree of noise and time-varying covariates. Based on over two million firm-months of panel data on North American public firms during 1979–2012, within which there are more than 2,100 bankruptcies filed under Chapter 7 or Chapter 11, we show that our interaction effects significantly improve both in-sample model fit and out-of-sample forecasting accuracy. The improvements in forecasting accuracy are persistent over time, and are robust to various empirical setup. We are also able to predict the signs of the coefficients on the proposed interaction effects, which are strongly supported by the data.

When accounting information is noisy and the degree of noise is heterogeneous,[2] the theoretical results of Duffie and Lando [2001] imply that any monotonic transformation of the hazard rate is a *nonlinear* function of both the degree of noise and relevant time-varying covariates. However, such non-linearity in the hazard function is typically not modeled in the current practice of credit risk modeling.[3] By further exploring this implication, we find that one way to approximate the non-linearity is to use interaction

[1]The related news article, industry publications and academic papers include Morgenson, G., "Was There A Loan It Didn't Like?" *New York Times*, November 1, 2008; Bitner, R., *Confessions of a Subprime Lender*, Wiley, 2008; Schoolman, P., 2008, "Credit Crisis Lessons for Modelers," in *Risk Management: The Current Financial Crisis, Lessons Learned and Future Implications*; Ng and Rusticus [2013]; to cite a few.

[2]The heterogeneity in the degree of noise might be both cross-sectional and in time series.

[3]Typically, the current practice models monotonic transformations of the hazard rate as a *linear* function of time-varying covariates. Well-known examples include studies using proportional hazard models (for example, Bharath and Shumway [2008], Duffie, Saita and Wang [2007]), and those using dynamic logistic regressions (for example, Shumway [2001], Chava and Jarrow [2004], Campbell, Hilscher and Szilagyi [2008, 2011]). In particular, Cox [1972] Proportional Hazard models treat log(hazard) as a linear function of covariates (and log[baseline hazard function]). Dynamic logistic regressions treat logit(hazard) as a linear function.

effects between proxies for the degree of noise and covariates.[4] Accordingly, we develop an approach to create hazard specifications that explicitly handle information noise, which amounts to three simple steps. One, identify time-varying covariates. Two, identify a proxy for the degree of noise. Three, construct interaction effects between the identified proxy and covariates. All these variables are candidates to be selected within the bankruptcy hazard function. In particular, this paper chooses covariates from four well-known hazard models in the literature, namely, the best-performing models in Shumway [2001], Chava and Jarrow [2004], Duffie, Saita and Wang [2007], Bharath and Shumway [2008], respectively. We also choose numerous candidates as proxies for the degree of noise that are widely accepted in the finance literature, including firm size, analyst coverage and analysts' forecast variation.[5]

Our approach allows us to develop three empirically testable hypotheses, regarding our proposed interaction effects. First, we test if the signs of the coefficients on the interaction effects are consistent with theoretical predictions. Second, we test whether our interaction effects as a whole improve in-sample Goodness-of-Fit. Finally, we test if these effects improve out-of-sample forecasting accuracy.

We find strong empirical evidence consistent with the three hypotheses. First, the coefficients on our proposed interaction effects have the same signs as predicted by the first hypothesis. Second, altogether these effects significantly improve the in-sample model fit based on full-sample tests. Third, the models with our proposed effects persistently outperform those without, in out-of-sample forecasting accuracy, according to two well-accepted predictability measures, (1) Area Under ROC Curve (AUC), and (2) the captured fractions of the total number of bankruptcies within deciles ranked by model forecasts. For the first measure, the models with our effects have significantly higher year-by-year AUC in typically 6 out of the 10 holdout years, and are no worse in any other year.[6] For the second measure, the models with our effects capture more bankruptcies in top deciles and less in low-risk deciles, than the models without our effects. This implies that models achieve more accurate classification and less mis-classification by using our proposed interaction effects. The models with our effects also have predominantly higher *cumulative* captured bankruptcies in all deciles, implying an unambiguous improvements on forecasting accuracy. Finally, we conduct a variety of robustness checks. We show that our results are robust to different empirical setup, and are substantially stronger when our interaction effects are used in private firm models, where information quality is a more serious problem. Therefore, our findings provide strong empirical support for using our

[4] We note that models in Chava and Jarrow [2004] also used interaction effects, between industry groups and covariates. Nonetheless, they are not related to imperfect information.

[5] See, for example, Thomas [2002], Zhang [2006], Lin, Ma and Xuan [2011], Guo and Masulis [2012]. In particular, we use log(total assets) as a proxy for firm size in our main results, Sections 5.1 and 5.2, and adopt other proxies in robustness checks, Section 5.3.

[6] The average of the year-by-year AUC improvements are also highly significant.

proposed hazard specifications in real-world bankruptcy forecasting, where firm-specific information is likely to be noisy.

We advance the empirical literature on corporate bankruptcy prediction, or more generally corporate default prediction, which dates back at least to Altman [1968], Beaver [1966]. The state of the art in default/bankruptcy forecasting is probably represented by hazard models (also known as intensity-based models, reduced-form models, survival analysis or duration analysis). The best-known default/bankruptcy hazard models include those in, for example, Shumway [2001], Chava and Jarrow [2004], Duffie, Saita and Wang [2007], Bharath and Shumway [2008], Campbell, Hilscher and Szilagyi [2008, 2011], Chava, Stefanescu and Turnbull [2011], Duan, Sun and Wang [2012]. For comprehensive reviews on this literature, see Duffie, Saita and Wang [2007] or Giesecke, Longstaff, Schaefer and Strebulaev [2011], and references therein. In addition to introducing new hazard specifications that improve model performance, our paper also reconciles some conflicting empirical findings in the previous literature. Particularly, there have been disagreements on the statistical significance of covariates such as firm size and asset profitability in the hazard function.[7] Our paper provides plausible explanations on the discrepancies in the empirical results, thus reconciles the literature.[8]

Our approach also has a broad range of industry applications on credit risk modeling.[9] For instance, they are directly applicable to Probability of Default (PD) models that are widely used by credit rating agents, or by virtually all banking institutions (as internal rating tools), where concerns on data quality and verification quality of obligors' information are prevalent.[10]

Apart from Duffie and Lando [2001], our paper is closely related to the theoretical literature studying credit risk models with incomplete information, see, for example, Giesecke [2004, 2006], Guo, Jarrow and Zeng [2009] and the sequel. Our paper provides an empirical implementation of the theory, in justified and practical manners.[11]

Furthermore, our paper contributes to empirical studies investigating the impact of financial reporting quality on bankruptcy forecasting accuracy,[12] or the factors affecting predictability and likelihood of corporate defaults.[13] Within these types of empirical

[7]For example, Shumway [2001], Chava and Jarrow [2004] found that (relative) firm size is significant with negative signs, while Duffie, Saita and Wang [2007], Campbell, Hilscher and Szilagyi [2011] found it insignificant, or sometimes significant with positive signs. Similarly, Chava and Jarrow [2004], Bharath and Shumway [2008] found that asset profitability measures, like net income divided by assets, are significant with negative signs, but Chava, Stefanescu and Turnbull [2011] found it insignificant.

[8]See Section 3 and Section 5.3 for the explanations.

[9]This paper only investigates bankruptcy events, because bankruptcy data is publicly available and is the only data available to us.

[10]PD models are widely used in the financial industry, in areas of Basel-compliant regulatory capital measurement, economic capital management, risk management, portfolio management and pricing.

[11]Note that this paper does not consider other forms of imperfect information, for example, delayed information. They might be considered in future empirical work.

[12]See, for example, Beaver, McNichols and Rhie [2005], Beaver, Correia and McNichols [2012].

[13]See, for example, Campbell, Hilscher and Szilagyi [2011], Tang, Subrahmanyam and Wang [2012],

studies, our hazard specifications provide a potentially useful tool to account for noise in econometricians' information set. For example, our proposed interaction effects naturally serve as control variables in credit risk-related empirical tests.

Our paper also has a technical contribution. As will be demonstrated in Section 5.3, our proposed hazard specifications have a built-in mechanism to elegantly handle outliers, by automatically adjusting the responsiveness of covariates based on the outliers' degree of noise. This mechanism requires minimum (or no) distortion of the input data, and is shown to be effective in our empirical study.[14]

The remainder of the paper is organized as follows. Section 2 explores implications of the theoretical results of Duffie and Lando [2001], and develops hypotheses accordingly. Section 3 outlines the design of our empirical study. Section 4 describes the bankruptcy dataset that we construct to test hypotheses. Section 5 presents the empirical results, including evidence from full-sample tests, out-of-sample tests and robustness checks. Section 6 concludes.

2 Hazard Specifications with Imperfect Information

In this section, we propose hazard specifications that account for information noise, motivated by the theoretical results of Duffie and Lando [2001](henceforth DL). We then develop three hypotheses related to the specifications.

2.1 Theory

We explore the results of DL, who considered a filtering problem when there is noise in the observed assets of a debt issuer (henceforth "firm"). Using the notation of DL, the stock of assets of the firm, V_t, is modeled as a geometric Brownian Motion (BM)[15] with initial value of V_0. Although all parameters associated with the stochastic process of V_t are known, V_t itself is not observable to the creditors of the firm. Instead, a noisy value of assets is observed, denoted as $\hat{V}_t$. It is assumed that $\log \hat{V}_t = \log V_t + U_t$, where U_t denotes random noise that is independent of $\log V_t$, and is normally distributed with mean u and standard deviation a. Note that the standard deviation a of U_t can be interpreted as "a measure of the degree of noise" (Duffie and Lando [2001, p. 642]). We adopt this interpretation throughout this paper.

The firm will file bankruptcy when $\log V_t$ first falls to some low boundary $\underline{v}$.[16] We

Cai, Saunders and Steffen [2012], Maffett, Owens and Srinivasan [2013] as more recent studies.

[14] The capability to handle outliers is a side benefit of our approach. While Section 5.3 gives an example, we defer to future work for dedicated, full-blown empirical studies on this topic.

[15] All random variables are defined on a fixed probability space $(\Omega, \mathscr{F}, P)$.

[16] In the model of DL, $\underline{v}$ is determined by the firm owners within an optimal bankruptcy framework of Leland [1994], Leland and Toft [1996]. Note that the firm owners (or managers) have perfect information on the "true" value of assets, $\log V_t$, to decide when to file bankruptcy. Thus, only creditors' information is noisy. Problems with asymmetric information was explicitly ruled out by DL, and is not considered

denote the bankruptcy time as τ. Within this setup, Duffie and Lando [2001, Equation 26] showed that the conditional probability of bankruptcy at time t, during the period of $(s-t), s>t$, is

$$P(t,s) = Pr(\tau \leq s|\hat{V}_t, V_0, \tau > t) = 1 - \int_{\underline{v}}^{+\infty} [1 - \pi(s-t, x-\underline{v})] g(x|\hat{V}_t, V_0, t) dx, \quad (1)$$

where we refer to DL for the detailed expressions of $\pi(\cdot)$ and $g(\cdot)$.

Under the assumption that accounting report is unbiased,[17] the conditional probability of bankruptcy (henceforth, PB), $P(t,s)$, is a function of the standard deviation of noise (a), the time-t observed assets ($\hat{V}_t$), initial assets (V_0), mean and volatility of the asset growth rates, debt face value, among other parameters. See Appendix A for more detailed descriptions on these parameters.

We are interested in the joint impact on PB of a and observed (noisy) asset returns, denoted as $r_N \triangleq (\frac{\hat{V}_1}{V_0} - 1)$ at time $t=1$, assuming $V_0 = 1$ is observed with perfect information and $s=2$. Thus, we fix all other parameters at some values (see Appendix A for values of these parameters), and vary levels of a and $\hat{V}_1$. We then numerically evaluate $P(1,2)$ for various levels of a and r_N, and graphically illustrate how PB changes accordingly. This results a surface of PB as shown in Figure 1a.

There are two salient features about the shape of the PB surface in Figure 1a. First, while in general, r_N is decreasing in PB, the slope of PB with respect to r_N varies with a. With a higher level of a, the slope along the direction of r_N becomes less steep. This can be seen more clearly if we project the surface of PB onto the PB–r_N plane, resulting a contour plot in Figure 1b, i.e., PB curves with the same value of a (or "iso-a" curves). Clearly, the slope of a PB curve given a lower a is steeper than that given a higher a.

This feature implies that, when asset information becomes noisier, the observed asset returns, as a predictor, become less responsive to bankruptcy risk.[18] This feature is highly intuitive and are supported by numerous empirical studies, for example, Beaver, McNichols and Rhie [2005], Beaver, Correia and McNichols [2012].

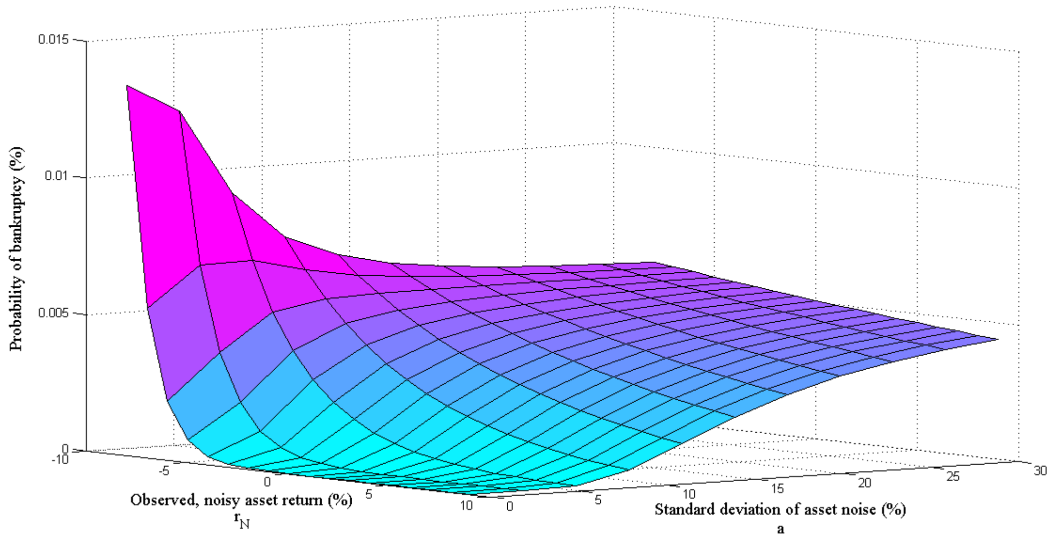
The second feature is that PB is increasing in a when r_N is above around -4%, but decreasing in a when r_N is below around -8%.[19] This feature is also intuitive, because with

here either.

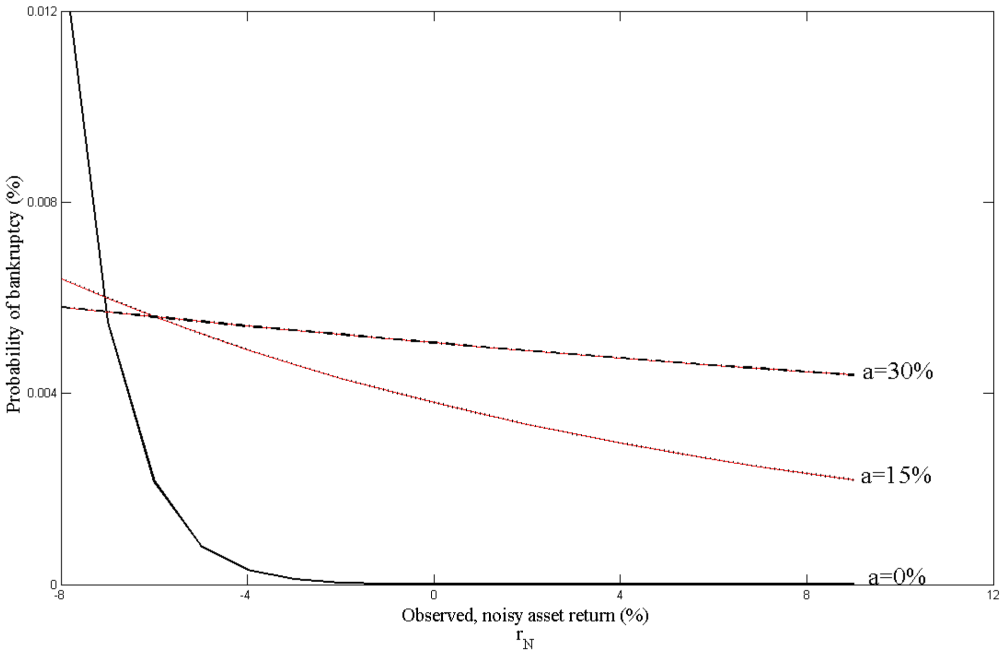
[17]Unbiased accounting report means that $u = -\frac{a^2}{2}$ so that $\mathrm{E}(U_t) = 1$. We make this assumption throughout this paper. It can be shown that the bias of accounting report does not materially impact the interaction effects between the degree of noise and covariates. Thus this assumption is not central to this paper, and will be topics of future research.

[18]In extreme cases, when the degree of noise is extremely high, the slope along the direction of r_N becomes almost flat, implying that coefficient on r_N is close to zero. In these cases, observed asset returns as a predictor will not accurately rank firms in terms of bankruptcy risk.

[19]Note that Figure 1a generalizes Duffie and Lando [2001, Figure 4], in a two-dimensional sense. Their graph corresponds to the case where r_N is zero in Figure 1a. Figure 1b is also similar to Duffie and Lando [2001, Figure 6]. We note that Figure 1 implies when r_N is near -6%, the monotonicity of PB in a is indeterminate. The threshold of (around) -6% is due to the assumptions on specific parameter values



(a) Probability of bankruptcy (PB) for various levels of standard deviation of noise, a, and observed (noisy) asset returns, $r_N \triangleq (\frac{\hat{V}_1}{V_0} - 1)$



(b) The projection of the surface of PB, in Figure 1a above, onto the PB–r_N plane

Figure 1: Theoretical probability of bankruptcy, varying the degree of noise and observed (noisy) asset returns

noisier information, one should attribute the observed high (low) asset returns more to noise rather than the true value of assets, which entails higher (lower) PB than noise-free PB. Therefore, to assess the impact of information noisiness on PB, we should consider it in context of other parameters (like r_N) as demonstrated in Figure 1.

These two features persist even if we perform any monotonic transformation of PB, such as logarithmic or logit transformations. This is because monotonic transformations do not change relative magnitude and the sign of the slope. Therefore, any monotonically transformed PB still has steeper (less steep) slope along the direction of r_N for lower (higher) a, and is increasing (decreasing) in a for higher (lower) r_N. We graphically illustrate the surface of logarithmic transformation of PB in Figure 2a. Figure 2a is identical to Figure 1a, except that the vertical axis is now log(PB). Again, we plot the projection of the surface of log(PB) onto the log(PB)–r_N plane, in Figure 2b.

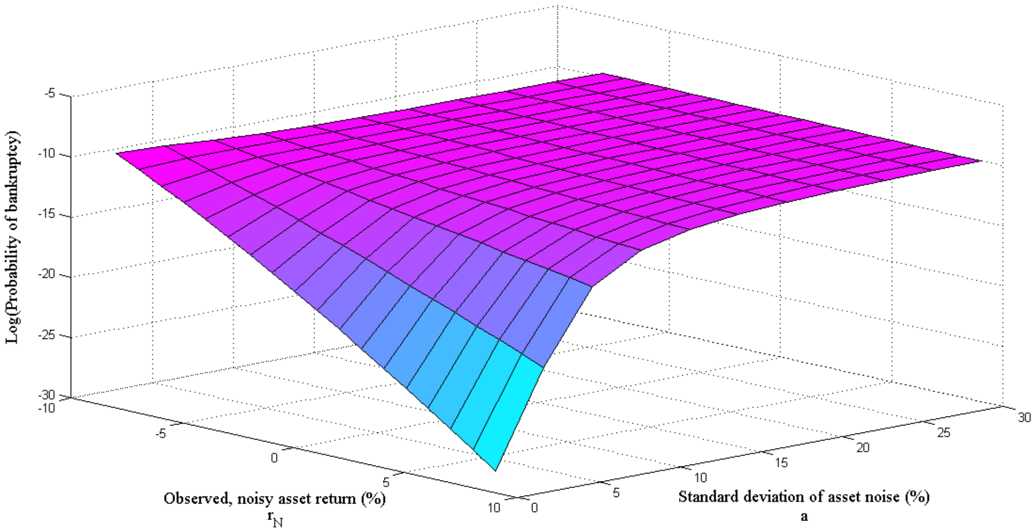
The above analysis shows that any monotonic transformation of PB is a non-linear function of both a and r_N when there exists heterogeneity in the degree of noise, either across firms or over time.[20] One obvious way to approximate such non-linearity within models of PB is to use an interaction effect between a and r_N, i.e. $(a*r_N)$. This is directly implied by Figure 2b. To provide more insights into this approximation, note that, when $(a * r_N)$ is added, the coefficient on r_N can now be viewed as a linear function of a, and thus varies depending on levels of a. This mechanism precisely models the variation of log(PB)'s slope with respect to r_N, conditional on a. The use of interaction effects, of course, also has mathematical convenience to keep (any monotonic transformation of) PB within the linear family.

Although noise is associated with firm's assets and all other parameters are assumed to be known without noise, the slope of PB with respect to other parameters might also vary with a, a feature similar to that shown in Figure 2. Such parameters include the mean and volatility of the asset growth rates, denoted as μ and σ respectively. The theoretical relationship between log(PB) and μ or σ, for various levels of a, is depicted in Figures 4 and 5, respectively, in Appendix A. Likewise, for various levels of a, we also plot the relationship between log(PB) and normalized debt face value,[21] $\frac{D}{\hat{V}_1}$, where D denotes the debt face value, in Figure 6 of Appendix A. In all graphs, log(PB) has a steeper slope (with respect to the corresponding covariate) when a is lower, and a flatter slope when a is higher. Thus, we might also incorporate interaction effects between a

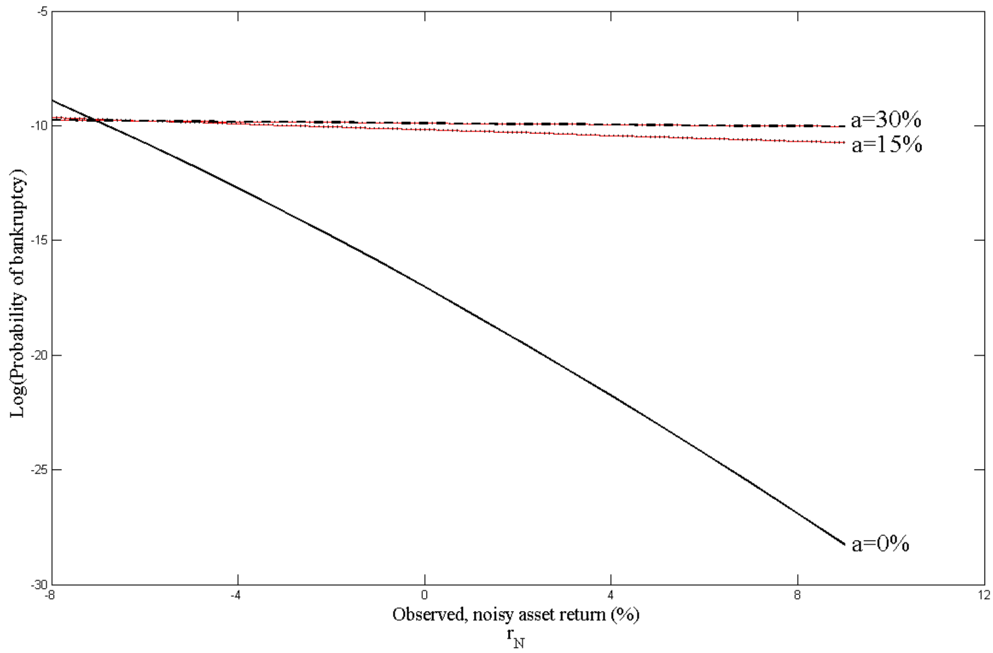
and unbiased accounting report. When we change these assumptions, the PB surface might shift along any axis (implying changes of the signs on main effects of a), but the shape of this surface, and thus the two features, remains the same.

[20] As can be seen in Figure 2b, the current practice of modeling a linear relationship, between covariates and monotonic transformation of PB, implicitly assumes that the degree of noise is the *same* across firms and over time, which is unlikely in reality.

[21] We consider normalized debt face value because it is popular in the empirical credit risk literature, typically as a proxy for leverage.



(a) log(probability of bankruptcy) (log(PB)) for various levels of standard deviation of noise, a, and observed (noisy) asset returns, $r_N \triangleq (\frac{\hat{V}_1}{V_0} - 1)$



(b) The projection of the surface of log(PB), in Figure 2a above, onto the log(PB)–r_N plane

Figure 2: Theoretical log(probability of bankruptcy), varying the degree of noise and observed (noisy) asset returns

and these parameters, i.e., $(a * \mu)$, $(a * \sigma)$ or $(a * \frac{\mathrm{D}}{\hat{V}_1})$, when modeling log(PB).[22]

We note that the empirical success of our interaction effects depends on the heterogeneity in the variables a, r_N, μ, σ and $\frac{\mathrm{D}}{\hat{V}_1}$. As is evident in Figures 2, 4, 5 and 6, if either of these variables lacks variation, the corresponding interaction effect may fail to be detected as significant in empirical estimation. Therefore, in empirical study, we try to avoid data exclusions as much as possible, in order to better exploit variation in these variables. Interestingly, this differs from the common practice of forecasting bankruptcy or default, which manipulates data to avoid extreme values (i.e., outliers) in independent variables.

2.2 Hypotheses on Proposed Hazard Specifications

It is natural to approximate PB, i.e., $P(t, s)$ in Equation (1), using hazard rate, denoted as λ_t, because λ_t is the continuous-time limit of $P(t, t+\Delta t)$, and it can be shown that this limit exists in the case of incomplete information,[23]

$$\lambda_t \triangleq \lim_{\Delta t \to 0} \frac{P(t, t + \Delta t)}{\Delta t}. \tag{2}$$

Consequently, in light of our analysis in Section 2.1, within the hazard function we can also use the interaction effects between proxies for a and time-varying covariates.

We develop three empirically testable hypotheses related to our proposed interaction effects. First, the features of PB surfaces in Figures 2, 4, 5 and 6 allow us to predict the signs of the coefficients on our proposed interaction effects.

Hypothesis 1 (Signs of the Coefficients)**.** *Supposing there is one proxy for the degree of noise, a, that is decreasing in a, then the interaction effect, between this proxy and any covariate that is decreasing (increasing) in hazard rate, will have a negative (positive) coefficient.*

For example, without any interaction effect, the observed (noisy) asset return, r_N, is decreasing in the hazard rate (as shown in Figures 2). Hypothesis 1 thus predicts any interaction effect between r_N and a proxy for a (decreasing in a) has a negative sign. It is straightforward to verify that this prediction is in accordance with the analysis in Section 2.1. We provide further intuition on Hypothesis 1 in Section 3 later (as explanations on Equation (4)).

Second, if Equation (1) indeed represents the real-world data-generating process (DGP) of bankruptcy, then we expect that our proposed hazard specifications should improve empirical performance of hazard models, including both in-sample model fit and

[22] We also find there might exist higher-order interaction effects between a, r_N and other parameters. For simplicity, we only consider first-order interaction effects in this study.

[23] This result is also due to DL, which showed that λ_t exists when there is incomplete information, and thus justified the use of hazard models, or similar statistical models, in practice.

out-of-sample forecasting accuracy. This is because, while hazard models without our proposed interaction effects try to model the relationship between monotonic transformations of PB and covariates as a hyperplane, our specifications effectively model the relationship as a surface similar to that implied by the DGP (shown in Figures 2). Better approximations of the DGP should be reflected in empirical model performance. Therefore, we develop the following two hypotheses.

Hypothesis 2 (In-Sample Goodness-of-Fit)**.** *Hazard models with the proposed interaction effects, between proxies for a and time-varying covariates, have significantly better in-sample Goodness-of-Fit than those without.*

Hypothesis 3 (Out-of-Sample Forecasting Accuracy)**.** *Hazard models with the proposed interaction effects, between proxies for a and time-varying covariates, have significantly better out-of-sample forecasting accuracy than those without.*

3 Empirical Design

We conduct empirical study in three steps. First, similar to Duffie, Saita and Wang [2007], Bharath and Shumway [2008], we specify the hazard rate as a Cox [1972] proportional hazard model (henceforth, Cox model),

$$\lambda_t = h_t \exp(\beta' X_t), \tag{3}$$

where h_t is an arbitrary and unspecified baseline hazard function common to all firms, X_t is a vector of time-varying covariates, β is a vector of coefficients.[24] β can be estimated using the partial likelihood function of Cox [1972] without requiring estimation of h_t.[25]

Next, we choose the covariates, X_t, from four well-known bankruptcy hazard models in the literature (henceforth, "reference models"), instead of identifying them by ourselves. The reasons for this design are threefold. First, the reference models are widely accepted as the state of the art in credit risk prediction, and are frequently cited.[26] Their choices of X_t also have economic interpretations that are aligned with our analysis in Section 2.1. See below the detailed descriptions of X_t within these models. Second, we use the reference models as benchmarks for model comparison purposes. Hence, their choices of X_t serve as control variables when testing the impact of our proposed interaction effects.[27]

[24]" $'$ " is the transpose operator. Cox model implies that $\log \lambda_t = \log h_t + \beta' X_t$, which fits into our analysis in Section 2.1.

[25]Because this paper only studies ranking power of models, the estimate of h_t is not required.

[26]For more recent citations of these models, see, for example, Tang, Subrahmanyam and Wang [2012], Maffett, Owens and Srinivasan [2013].

[27]We note that the reference models might not include covariates like firm liquidity (see, for example, Campbell, Hilscher and Szilagyi [2011]). We defer empirical tests on our proposed specifications using a more comprehensive set of covariates (potentially beyond those used in the extant literature) to future research.

Third, this design allows us to demonstrate the generality of our approach. Our hazard specifications impose no restriction on X_t, and thus can be used in combination with any existing choice of X_t, as long as they are properly analyzed as we demonstrate here.

Finally, we create four new models (henceforth, "augmented models") by creating interaction effects between proxies for a and covariates, and adding them into the reference models' hazard functions. Each augmented model corresponds to one reference model. Now the hazard rate of an augmented model becomes

$$\lambda_t = h_t \exp\left[\bar{\beta}' X_t + \bar{\gamma}_0 \tilde{a} + \sum_{i=1}^{I} \bar{\gamma}_i (\tilde{a} * X_t^i)\right], \tag{4}$$

where $\bar{\beta}$ denotes a vector of coefficients on the main effects of X_t, I is the number of additional interaction effects, $\tilde{a}$ denotes a proxy for a such that higher $\tilde{a}$ represents lower a, X_t^i denotes the i^{th} covariate with which $\tilde{a}$ interacts, and $\bar{\gamma}_0, \ldots, \bar{\gamma}_I$ are coefficients on $\tilde{a}$ and interaction effects respectively.

We provide more intuition, using Equation (4), why Hypothesis 1 is in accordance with our analysis in Section 2.1. First, we note that the coefficients on X_t^i are different in Equations (3) and (4). In Equation (3) where there is no interaction effect, the coefficient on X_t^i is the corresponding element within β, denoted as β_i. In Equation (4), the coefficient can be viewed as $(\bar{\beta}_i + \bar{\gamma}_i \tilde{a})$ where $\bar{\beta}_i$ is the coefficient on the main effect of X_t^i. Hypothesis 1 predicts that β_i and $\bar{\gamma}_i$ have the same sign. If this is true, then lower degree of noise (i.e., higher $\tilde{a}$) entails that $(\bar{\beta}_i + \bar{\gamma}_i \tilde{a})$ is more consistent with β_i. For example, when β_i and $\bar{\gamma}_i$ have the (same) positive sign, lower degree of noise entails that $(\bar{\beta}_i + \bar{\gamma}_i \tilde{a})$ is more positive, or less negative.[28] In other words, Hypothesis 1 implies that when the degree of noise is lower, X_t^i is more responsive to $\log \lambda_t$, which is precisely the intuition behind the analysis in Section 2.1.

The four reference models used in our study are "Model with accounting and market variables" in Shumway [2001], "Public firm model with industry effects" in Chava and Jarrow [2004], the intensity model in Duffie, Saita and Wang [2007] and "Model 7" in Bharath and Shumway [2008], respectively. These models are the best-performing one in the corresponding articles. Henceforth, we call the reference models "S01 Model", "CJ04 Model", "DSW07-S Model" and "BS08 Model", respectively. Note that "DSW07-S Model" is a simplified version of the intensity model of Duffie, Saita and Wang [2007].[29]

We recognize that some covariates chosen by the reference models can be loosely

[28] Likewise, if both β_i and $\bar{\gamma}_i$ have the negative sign, then lower degree of noise (i.e., higher $\tilde{a}$) entails that $(\bar{\beta}_i + \bar{\gamma}_i \tilde{a})$ is more negative, or less positive.

[29] We use "-S" to highlight that our implementation is a "simplified" version. There are two simplifications within our implementation. First, we use a "naïve" version of Distance-to-Default (DD) measure, developed by Bharath and Shumway [2008]. Bharath and Shumway [2008] showed that the default prediction performance of DD is robust to how it is implemented. Second, for simplicity, we do not model the time series dynamics of covariates.

interpreted as proxies for parameters analyzed in Section 2.1, namely, observed (noisy) asset return, the expected asset return, volatility of asset return and normalized debt face value.[30] Therefore, we use these covariates to construct potential interaction effects. Note that whether or not a potential interaction effect is included in our augmented models is determined by the statistical significance of its coefficient, an empirical decision that is data dependent. The covariates used in each reference model and the potential interaction effects are described as follows.[31]

S01 Model has five covariates: (1) Net Income/Total Asset (NI/TA); (2) Total Liability/Total Asset (TL/TA); (3) firm's relative size (RSIZE) defined as the difference between the logarithm of firm's equity value and the logarithm of the total NYSE & AMEX market capitalization; (4) firm's stock excess return (EXRET) defined as difference between firm's trailing one-year stock return and the value-weighted CRSP NYSE & AMEX index return; and (5) firm's stock volatility (σ_E). We use four of them to construct potential interaction effects, based on their economic interpretations. First, it is natural to (loosely) interpret NI/TA and TL/TA as proxies of observed (noisy) asset return and normalized debt face value, respectively. Second, we view σ_E as a rough proxy for volatility of asset return. Finally, although EXRET is excess return, not firm's stock return, it can be viewed as a crude approximation of the trailing one-year stock return. The trailing one-year stock return is commonly used as a "naïve" proxy for the expected asset return (see, for example, Bharath and Shumway [2008]).[32] As a result, we obtain four potential interaction effects within S01 Model.

CJ04 Model includes all the covariates used in S01 Model, with additional industry effects. Hence, the potential interaction effects in CJ04 Model are the same as in S01 Model.

In DSW07-S Model, there are four covariates: (1) a "naïve" version of Distance-to-Default measure (Naïve DD) defined as, roughly speaking, the number of standard deviations of asset growth rate by which the expected log assets exceed log debts;[33] (2) firm's trailing one-year stock return (RETURN); (3) three-month Treasury bill rate (3m T-rate); and (4) trailing one year return on the S&P500 index (SPX). We construct two potential interaction effects, with Naïve DD and RETURN respectively. This is because Naïve DD is effectively a synthesis of the expected asset return, volatility of asset return and normalized debt face value, and RETURN is commonly used as a proxy for the

[30]We also recognize that it is impossible to precisely map parameters in Section 2.1 to the covariates used in the reference models. This is because these covariates are identified based on empirical performance rather than theoretical considerations. However, we do find similarity between their economic interpretations, and thus can roughly approximate the parameters using these covariates.

[31]Also see Table 1 for a summary of covariates used in the reference models.

[32]This approximation is plausible, because, like the expected asset return, EXRET is also decreasing in PB.

[33]It is called a "naïve" version because the implementation uses naïve proxies for parameters, see Bharath and Shumway [2008] for details.

expected asset return (see, for example, Bharath and Shumway [2008]). Consequently, we have two potential interaction effects within DSW07-S Model.

There are six covariates in BS08 Model: (1) probability of bankruptcy measured using Naïve DD ($\pi_{\text{Naïve}}$), defined as $N(-\text{Naïve DD})$, where $N(\cdot)$ is the Gaussian cumulative distribution function; (2) logarithm of firm's market capitalization of equity ($\log E$), where E is defined as the product of month-end stock price and number of shares outstanding; (3) logarithm of firm's debt face value ($\log F$), where F is defined as (Compustat item "Debt in Current Liabilities")$+\frac{1}{2}$(Compustat item "Total Long-Term Debt"); and three covariates used in S01 Model, namely, (4) σ_E, (5) EXRET and (6) NI/TA. We use all of them to construct potential interaction effects. The justifications of choosing $\pi_{\text{Naïve}}$, σ_E, EXRET and NI/TA are the same as those in DSW07-S Model and S01 Model. The use of $\log E$ and $\log F$ is justified by interpreting them altogether as a proxy for normalized debt face value. As such we get six potential interaction effects within BS08 Model.

To select $\tilde{a}$ in Equation (4), i.e. proxies for the degree of noise a, there is a wide range of choices in the finance literature. In our study, we try a number of popular candidates described as follows.[34]

One natural choice of $\tilde{a}$ is firm size. Greater firm size implies less degree of noise (see, for example, Zhang [2006], Lin, Ma and Xuan [2011]). To construct interaction effects within S01 Model, CJ04 Model and DSW07-S Model, we use log(Total Asset) (log(TA)) as the proxy for firm size when reporting our main results. As a robustness check, we also try another two proxies for firm size: log(equity market value) ($\log E$), and log(Asset Rank) (log(AR)) where Asset Rank is obtained by ranking all surviving firms every month according to their total assets.[35] Note that, unlike log(TA) or $\log E$ that captures firms' absolute size effect, log(AR) measures firms' relative size cross-sectionally.[36] In BS08 Model, $\log E$ is already used as a covariate and is strongly correlated with log(TA).[37] Hence, to avoid potential multi-collinearity problems, within BS08 Model we only use

[34]We stress that we do not intend to search for the "best" proxies for the degree of noise. Our focus is to study the real-world benefits of using our proposed interaction effects, together with popular (and reasonable) candidates of $\tilde{a}$.

[35]In this study, we rank firms into 1,000 groups.

[36]We note that S01 and CJ04 models have a covariate, RSIZE, which might also be viewed as a proxy for (relative) firm size. We address potential concerns on multi-collinearity, when using RSIZE together with proxies for firm size, in several ways. First, we compute the contemporaneous correlation between RSIZE and log(TA) or log(AR), and find it is moderate within our dataset, at around 0.6–0.7. Second, we find the variance inflation factors (VIF) of RSIZE, log(TA) and log(AR) are typically around 2–5, below the standard threshold of 10. These diagnostics indicate that multi-collinearity is mild. Moreover, the out-of-sample results in Section 5.2 confirm that multi-collinearity might be less a problem when using RSIZE and log(TA) or log(AR) together. Therefore, we include RSIZE as a covariate when using log(TA) or log(AR) as proxies for firm size. Nevertheless, multi-collinearity does become a problem when we use $\log E$ as a proxy, because the contemporaneous correlation between $\log E$ and RSIZE is 0.92. Hence, we exclude RSIZE from S01 and CJ04 models when using $\log E$ as the proxy for firm size. As a robustness check (not shown here), we also take RSIZE as a proxy for firm size within the augmented S01 and CJ04 models, and obtain similarly strong results supporting our hypotheses.

[37]The contemporaneous correlation between $\log E$ and log(TA) is 0.82 within our dataset.

$\log E$ as a proxy for firm size when constructing interaction effects.

We note that interpreting firm size as a proxy for the degree of noise might reconcile the conflicting empirical findings on firm size in the extant literature. As can be shown in Figure 2, log(PB) might be increasing, decreasing, or non-monotonic, with the degree of noise, depending on the values of other covariates like observed asset returns. Therefore, in models without our proposed interaction effects, the coefficient on firm size depends on the slope of log(PB) along firm size at the average levels of other covariates,[38] and thus is data-dependent within our framework. This provides a plausible explanation why empirical studies can have different findings on statistical significance, or sometimes the sign, of the coefficient on firm size.

Apart from firm size, analyst coverage and analysts' forecast variation are also popular proxies for the degree of noise. Higher coverage and lower variation implies less degree of noise (see, for example, Thomas [2002], Guo and Masulis [2012]). Therefore, in robustness checks (Section 5.3), we also use these two proxies, namely, Analyst Coverage (AC), defined as the number of monthly analyst forecasts on EPS or NAV, and normalized variation of analysts' forecasts ($-\log(\text{CV})$), defined as $-\log(\text{Coefficient of Variation})$ where

$$\text{Coefficient of Variation} \triangleq \frac{\text{Standard deviation of analysts' forecasts}}{\text{Absolute value of mean analysts' forecasts}}. \tag{5}$$

Note that $-\log(\text{CV})$ is constructed such that it is decreasing in the degree of noise, in order to be aligned with Hypothesis 1.

4 Data

We construct a comprehensive bankruptcy dataset for North American public firms during 1979-2012,[39] including both Chapter 7 and Chapter 11 filings. We identify bankruptcies from a variety of sources, namely, New Generation Bankruptcy Database,[40] UCLA-LoPucki Bankruptcy Database, and the Fixed Income Securities Database (FISD). Following Duffie, Saita and Wang [2007], we also identify additional bankruptcies from firms with Compustat deletion reasons as "02-Bankruptcy" (Compustat items DLRSN, DLRSNI).[41] These data sources are standard in bankruptcy studies.[42] Moreover, to en-

[38]For models with our interaction effects, the coefficient on the main effect of firm size represents the slope of log(PB) along firm size when other covariates are zero.

[39]There are few bankruptcies filed in early 2013. We treat them as if they were filed in December 31, 2012.

[40]The data is publicly available at www.BankruptcyData.com.

[41]We manually verify bankruptcy date and status using a random sample of the firms with DLRSN/DLRSNI as "02", and find this indicator is highly accurate. However, we do not use DLRSN/DLRSNI value of "03-Liquidation", as we find it might be unrelated to bankruptcy.

[42]See, for example, Tang, Subrahmanyam and Wang [2012], Cai, Saunders and Steffen [2012] as more recent studies.

sure accuracy of bankruptcy dates and status, which are the response variables in our empirical study, we manually verify more than 1,000 firms that have ambiguous bankruptcy information, using SEC filings and other public information sources.[43] Finally, we link these bankruptcy events, using CIK and CUSIP, to Compustat North America Quarterly accounting data (henceforth, Compustat),[44] which is further merged with CRSP monthly stock market data (henceforth, CRSP), resulting a firm-month panel dataset.[45]

In order to properly develop independent variables, i.e., time-varying covariates and proxies for the degree of noise, within our dataset, we further require that (i) any bankrupt firm appear in both Compustat and CRSP; (ii) any bankrupt firm have bankruptcy date no later than 5 years after the last available observation in Compustat/CRSP;[46] and (iii) each firm-month observation have at least 6 months' stock returns in the previous one year, and have non-missing, nonzero equity market value in CRSP. Like Chava and Jarrow [2004], when there are multiple bankruptcies associated with a firm, we only consider the first one, and we assume uninformative left censoring.[47] Note that we try to avoid data exclusions due to data quality reasons. This is because we want to better exploit both cross-sectional and time-series variation in firms' accounting/market information, and variation in the degree of noise.[48]

After applying the above rules, we are able to obtain 2,112 bankruptcies, and 2,152,203 firm-month observations, from a total of 20,180 firms, in our final panel dataset. The total number of bankruptcies is similar to those observed in recent bankruptcy studies.[49] We plot and tabulate the bankruptcy profile of our dataset, for each year during 1979-2012, in Figure 3. Within Figure 3, Panel 3a depicts the number of bankruptcies and the bankruptcy rate by year, as blue bars and red lines, with vertical axes labeling on the left and right respectively. Panel 3b provides the detailed data used to plot Panel 3a. The general patterns, which show peaks of bankruptcies in early 1990s, early 2000s and around 2009, are consistent with those demonstrated by the previous literature. See,

[43] We search firms using CIK within SEC Filings including 8-K, Administrative Proceeding, 10-K(or 10-KSB), and so on. We also search firms by combinations of firm name, Chair/CEO name, address, phone number, IRS number, CUSIP, ticker, CIK and industry, from news, online market information, online business/company information, court documents, credit reports, and so forth.

[44] For bankrupt firms identified by UCLA-LoPucki Bankruptcy Database and Compustat DLRSN/DLRSNI, they already have GVKEY to be merged with Compustat accounting data.

[45] We carry forward Compustat quarterly observations to make it a monthly dataset.

[46] If a bankrupt firm exits Compustat/CRSP databases 5 (or more) years earlier than it files bankruptcy, we treat it as right censored at one month after the final Compustat/CRSP observation. The choice of 5 years is arbitrary. In practice, it is unlikely that creditors use information older than 5 years to make one-year ahead bankruptcy predictions.

[47] Right censoring occurs at the following three types of dates: (i) the date that a firm is deleted from Compustat (the earlier of Compustat items DLDTE and DLDTEI); (ii) if the Compustat delete date is more than 5 years later than the last available observation in Compustat/CRSP, then the firm is censored at one month after the final Compustat/CRSP observation; and (iii) otherwise, December 31, 2012.

[48] In robustness checks (Section 5.3), we follow convention in the previous literature to exclude financial firms.

[49] See, for example, Tang, Subrahmanyam and Wang [2012], Cai, Saunders and Steffen [2012].

for example, bankruptcy rate time-series in Chava and Jarrow [2004], or default rate time-series in Giesecke, Longstaff, Schaefer and Strebulaev [2011].

While most of the independent variables are developed using Compustat and CRSP data items, we construct the following variables from Datastream and IBES: three-month Treasury bill rate is from Datastream, and the two proxies for the degree of noise, namely Analyst Coverage (AC) and normalized variation of analysts' forecasts($-\log(\text{CV})$), are derived from IBES. Furthermore, we perform a number of data transformations on independent variables, as follows.

First, following Shumway [2001], Chava and Jarrow [2004], Bharath and Shumway [2008], we winsorize all Compustat-related covariates at 1^{st} and 99^{th} percentiles,[50] and impute missing values of Compustat-related covariates for any firm-month observation by carrying forward the most recent value of the relevant covariate available to that particular firm. Second, we find that only about 53% of the observations within our dataset have IBES information. For observations with no IBES information, we treat them as the noisiest observations, i.e., we set their AC to be 0, and $-\log(\text{CV})$ to be the lowest value in the data.[51] Third, where applicable, we translate all Compustat items into US Dollar using Compustat item CURUSCNQ, before deriving any independent variable.[52] Fourth, if any firm-month observation has less than 12 months' stock returns in the previous one year, we calculate the annualized trailing one year return and volatility for that particular observation.

Table 1 and Table 2 summarize the definitions and key descriptive statistics of the covariates, and of the proxies for the degree of noise, respectively, after winzorization and missing value imputation.[53] The conceptual descriptions of the independent variables are also explained in Section 3.

The summary statistics of the covariates in Table 1 are very similar to those in the previous literature (see, for example, Shumway [2001], Chava and Jarrow [2004], Bharath and Shumway [2008]), except that the standard deviations of most variables are higher. Larger variation reflects greater heterogeneity within our sample, both in cross section and in time series. This is in fact the case that we are particularly interested in.[54]

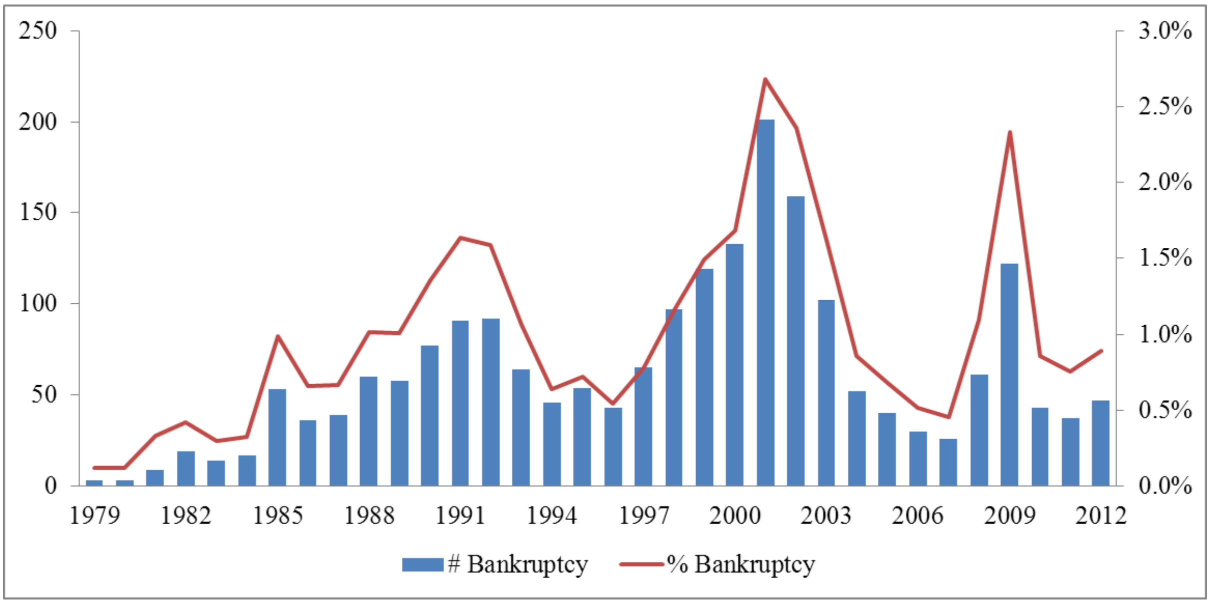
[50]The winsorization is intended to remove potential data errors within Compustat, see Shumway [2001], Chava and Jarrow [2004].

[51]We also set $-\log(\text{CV})$ to the lowest value in the sample if the standard deviation of analyst forecasts is undefined when the number of forecasts is 1, and set $-\log(\text{CV})$ to the highest value when the standard deviation of analyst forecasts is zero (i.e., complete consensus among analysts).

[52]There are more than 10% firms in our dataset are Canadian firms, with native currency of CAD.

[53]In Table 1, we only report the mean values for the three industry dummy variables, IND2–IND4, because the mean represents the proportion of observations that fall into an industry group. For example, IND4 has a mean of 0.18, indicating that 18% of the observations are in the industry group 4.

[54]The variation is larger because, in contrast to the previous literature which typically focused a subpopulation of firms (for example, industrial firms or firms listed in large stock exchanges), we try to include as many firms into our sample as possible. In particular, we notice that the covariates related to the trailing one year stock returns (e.g. RETURN, EXRET) have few extreme values. Some of the extreme returns come from "penny stock" firms and some are due to annualization of less-than-12-months



(a) The number of bankruptcies and bankruptcy rates by year, within our dataset, during 1979–2012. The bankruptcy rates are calculated as percentages of the number of surviving firms each year. The number of bankruptcies are plotted as (blue) bars with vertical axis labeling on the left, and bankruptcy rates are plotted as (red) lines with vertical axis on the right.

Year	# Bankruptcy	# Firms	% Bankruptcy	Year	# Bankruptcy	# Firms	% Bankruptcy
1979	3	2,561	0.12%	1996	43	7,892	0.54%
1980	3	2,532	0.12%	1997	65	8,323	0.78%
1981	9	2,733	0.33%	1998	97	8,401	1.15%
1982	19	4,501	0.42%	1999	119	7,983	1.49%
1983	14	4,694	0.30%	2000	133	7,891	1.69%
1984	17	5,261	0.32%	2001	201	7,504	2.68%
1985	53	5,370	0.99%	2002	159	6,748	2.36%
1986	36	5,474	0.66%	2003	102	6,299	1.62%
1987	39	5,840	0.67%	2004	52	6,063	0.86%
1988	60	5,921	1.01%	2005	40	5,891	0.68%
1989	58	5,759	1.01%	2006	30	5,836	0.51%
1990	77	5,674	1.36%	2007	26	5,747	0.45%
1991	91	5,573	1.63%	2008	61	5,562	1.10%
1992	92	5,798	1.59%	2009	122	5,231	2.33%
1993	64	6,027	1.06%	2010	43	5,020	0.86%
1994	46	7,229	0.64%	2011	37	4,907	0.75%
1995	54	7,497	0.72%	2012	47	5,262	0.89%

(b) This table reports the year, the number of bankruptcies, the number of surviving firms and the bankruptcy rate as % of the number of firms in the year, for each year within our dataset, during 1979–2012.

Figure 3: The bankruptcy profile of our dataset, for each year during 1979–2012. Our dataset is constructed by assembling bankruptcy filings of North American public firms, within the Compustat/CRSP universe, from New Generation Bankruptcy Database, UCLA-LoPucki Bankruptcy Database, the Fixed Income Securities Database and firms with Compustat deletion reasons as "02-Bankruptcy". We also require that (i) any bankrupt firm appear in both Compustat and CRSP; (ii) any bankrupt firm have bankruptcy date no later than 5 years after the last available observation from Compustat/CRSP; and (iii) any firm-month observation have at least 6 months' stock returns in the previous one year, and have non-missing, nonzero equity market value in CRSP. Applying these rules, and merging Compustat and CRSP databases, we obtain our final panel dataset that consists of 2,112 bankruptcies and 2,152,203 firm-month observations during 1979–2012.

Table 1: Summary statistics of the covariates

1979–2012, 2,112 bankruptcies, 2,152,203 firm-months, 20,180 firms in total

Variable	Definition	Mean	Median	Min	Max	Std. Dev.
NI/TA	Net Income / Total Asset (NIQ/ATQ[a])	-0.0086	0.0061	-0.4067	0.1047	0.0695
TL/TA	Total Liability / Total Asset (LTQ/ATQ)	0.5351	0.5362	0.0315	1.2523	0.2652
RSIZE	$\log \frac{\text{Market capitalization of firm equity}}{\text{NYSE \& AMEX market capitalization}}$ (CRSP)	-10.66	-10.77	-21.58	-2.48	2.09
3m T-rate	Current 3-month U.S. Treasure Bill rate (Datastream)	0.0473	0.0491	0.0001	0.1552	0.0303
SPX	Trailing one year return of S&P500 Index (CRSP)	0.0988	0.1125	-0.4476	0.5337	0.1704
σ_E	Annualized trailing one year standard deviation of monthly stock returns (CRSP)	0.50	0.41	0.00	24.09	0.38
EXRET	$r_{i(t-1)} - r_{m(t-1)}$, i.e., firm's trailing one year excess stock return over the value-weighted NYSE & AMEX return (CRSP)	0.0374	-0.0749	-1.6387	258.80 (P99: 2.89)	1.0257
RETURN	$r_{i(t-1)}$, i.e., firm's trailing one year stock return (CRSP)	0.1642	0.0466	-0.9996	259.02 (P99: 3.09)	1.0404
E	Market capitalization of firm equity, in millions of USD, (CRSP) Stock price * Number of shares outstanding (PRC*SHROUT)	1,638.41	118.13	0.00	626,550.33	9,788.28
F	Face value of debt, in millions of USD: Debt in Current Liabilities + 0.5 * Total Long-Term Debt (DLCQ+0.5*DLTTQ)	453.36	16.92	0.02	13,997.57	1,707.48
Naïve DD	A simplified version of Merton "Distance to Default" (see Bharath and Shumway [2008]): $\frac{\log[(E+F)/F]+(r_{i(t-1)}-0.5\sigma_V^2)T}{\sigma_V\sqrt{T}}$, where $\sigma_V \triangleq \frac{E}{E+F}\sigma_E + \frac{F}{E+F}(0.05+0.25\sigma_E)$ and we set $T=1$	127.41	5.93	-11.46	6,979,022	24,232.59
$\pi_{\text{Naïve}}$	$N(-\text{Naïve DD})$, where $N(\cdot)$ is the Gaussian cumulative distribution function	0.07	0.00	0.00	1.00	0.20
IND2	Industry group dummy variable, 1 if firm's CRSP SIC code in [1000, 1500) or [2000, 4000), 0 otherwise	0.46	–	–	–	–
IND3	Industry group dummy variable, 1 if firm's CRSP SIC code in [4000, 5000), 0 otherwise	0.09	–	–	–	–
IND4	Industry group dummy variable, 1 if firm's CRSP SIC code in [6000, 6800), 0 otherwise	0.18	–	–	–	–

This table reports the definitions and summary statistics of all covariates within our firm-month panel data. The definitions include, in parentheses, the relevant Compustat (Quarterly, North American) items used to calculate a particular variable (or database names if the variable is from other data sources). The sample period is 1979–2012. There are 2,112 bankruptcies, and 2,152,203 firm-month observations, from a total of 20,180 firms, in the full sample. All Compustat information is translated into US Dollar where applicable, and all Compustat-related covariates are winsorized at 1^{st} and 99^{th} percentiles. Missing values of Compustat-related covariates for any firm-month observation are imputed by carrying forward the most recent value of the relevant covariate available to that particular firm. Trailing one year returns and volatility are annualized if there are less than 12 months' stock returns in the previous one year. For the three industry dummy variables, IND2–IND4, we only report their means because they have the meaning of proportions of observations that fall into an industry group.

[a] If ATQ≤0, we use [Total Liability (LTQ) + Equity market value(E)].

Table 2: Summary statistics of proxies for the degree of noise

1979–2012, 2,112 bankruptcies, 2,152,203 firm-months, 20,180 firms in total

Proxies for a	Mean	Median	Min	Max	Std. Dev.
log(TA)	5.38	5.28	-6.91	15.17	2.38
log(AR)	5.91	6.22	0.00	6.91	0.99
AC	3.53	1.00	0.00	55.00	5.88
$-\log(\text{CV})$	-2.75	-7.25	-7.25	12.08	5.47

This table reports the summary statistics of the proxies for a, the degree of noise, used in this study. The sample period is 1979–2012. Our firm-month panel dataset has 2,112 bankruptcies, and 2,152,203 firm-month observations, from a total of 20,180 firms. log(TA), which is log(Total Asset), is defined as log(Compustat item ATQ) where ATQ is in millions of USD. log(AR) is log(Asset Rank), where Asset Rank is obtained by ranking all surviving firms every month into 1,000 groups based on their ATQ. AC, i.e., Analyst Coverage, is defined as the number of monthly analyst forecasts on EPS or NAV, obtained from IBES. $-\log(\text{CV})$ is the normalized variation of analysts' forecasts, and is defined as $-\log(\text{Coefficient of Variation})$ where Coefficient of Variation is defined as (Standard deviation of monthly analyst forecasts)/(Absolute value of mean monthly analysts forecasts), with both numerator and denominator obtained from IBES. AC is set to 0 when it has missing value. When $-\log(\text{CV})$ has missing value, we consider two cases. If it has missing value because the standard deviation of analyst forecasts is zero, it is set to the highest value (12.08). Otherwise, it is set to the lowest value (-7.25).

Table 2 provides the summary statistics of the proxies for the degree of noise, a, used in this study.[55] All proxies are constructed so that they are decreasing in a. As shown in Table 2, for AC and $-\log(\text{CV})$, more than half of the observations in our dataset have only one or two values (which are the lowest values), due to missing values. Thus, variation in these two variables are much less than other variables'.

5 Empirical Results

In this section, we report the results of our empirical study. We conduct full-sample tests to test Hypotheses 1 and 2, and conduct out-of-sample tests, on forecasting accuracy, to test Hypothesis 3. Finally, we report results on a variety of robustness checks, using both full-sample and out-of-sample tests.

5.1 Full Sample Tests

To test Hypothesis 1, we estimate four augmented models, by adding our proposed interaction effects into the reference models, using the full sample during 1979-2012. The

(yet still very large) returns. In Table 1, we report the 99^{th} percentile of these variables (in parentheses), which are close to the maximum values of these variables reported in the previous literature. We do not winsorize, however, these market-related variables in our empirical study, following convention in the previous literature. This poses no problems on our approach, because one advantage of our approach is in fact the built-in mechanism of handling such outliers. Moreover, we find that results from winsorizing these variables (not shown in this paper), at 99^{th} percentile, are almost identical to the reported results.

[55] One proxy, $\log E$, is not summarized here, because it is also a covariate in BS08 Model, and is summarized in Table 1.

four reference models are also estimated. The full-sample estimates of all models are reported in the columns labeled by the model names within Table 3.

As can be seen in Table 3, the signs and magnitudes of the coefficients within the reference models are consistent with those in the previous literature.[56] For augmented models, we select interaction effects by initially including all the potential ones described in Section 3 (i.e., four interaction effects in S01 Model, four effects in CJ04 Model, two effects in DSW07-S Model, six effects in BS08 Model), and then eliminating any interaction effect whose coefficient is not significantly different from zero at 10% level.[57] As shown in Table 3, we select three, four, one and three interaction effects in the final augmented S01, CJ04, DSW07-S and BS08 Model, respectively. We report, in parentheses, standard errors that robust to model mis-specification (see, for example, Lin and Wei [1989], Allison [2010]),[58] which are typically larger than conventional model-based standard errors.[59]

In Table 3, the signs of the coefficients on the interaction effects, in the augmented models, are consistent with the predictions of Hypothesis 1. For example, the coefficient on NI/TA has a negative sign in CJ04 Model. In the augmented CJ04 Model, the coefficient on interaction effect between NI/TA and log(TA) also has a negative sign, as predicted by Hypothesis 1. Note that the coefficient on NI/TA is now $[-1.57 - 0.26\log(\text{TA})]$. Hence, firms with higher total assets, i.e., with lower degree of noise, will have greater coefficients on NI/TA in magnitude (with a negative sign). This implies that the slope along the direction of NI/TA becomes steeper for lower degree of noise, which is precisely what Figure 2b illustrates. Likewise, the coefficient on TL/TA is $[2.09 + 0.15\log(\text{TA})]$, and thus is greater in magnitude (with a positive sign) when the degree of noise is lower. Again this is consistent with Hypothesis 1. As will be shown in robustness checks (Section 5.3), this conclusion is robust to the choices of data constructions and alternative proxies for the degree of noise.[60]

Therefore, we find strong empirical evidence consistent with Hypothesis 1. We are

[56] The only exception is the coefficient on "3m T-rate" in DSW07-S Model, which has a different sign from that in Duffie, Saita and Wang [2007]. However, when we use a dynamic logistic regression (which implies non-proportional hazard rates), we find that the sign is consistent with that in Duffie, Saita and Wang [2007]. In spite of this sensitivity of "3m T-rate" to different hazard functional forms, we find that DSW07-S Model used here has high out-of-sample predictive power, as shown in Section 5.2 below, which shows its validity. Also note that the magnitude of the coefficient on "3m T-rate" is about 100 times of that in Duffie, Saita and Wang [2007] because they used values in percentage whereas we use decimal values.

[57] This variable selection method is called "backward elimination". The selection of interaction effects is, of course, not unique. We also try other combinations of interaction effects within the augmented models, and find similar improvements on model performance.

[58] Typical model mis-specification includes omission of other covariates or non-linear terms, which are relevant to reduced-form models studied here.

[59] The robust standard errors are obtained from the "sandwich" variance estimator. It is unclear whether the robust standard errors were used in the previous literature.

[60] We stress that Hypothesis 1 says nothing about the main effect of the proxy for the degree of noise. We include the main effects in order to facilitate inclusion of interaction effects.

Table 3: Full-sample estimates of four bankruptcy hazard models, with and without our proposed interaction effects

Dependent Variable: Time to Bankruptcy
1979–2012, 2,152,203 firm-months, 2,112 bankruptcies
(Robust standard errors in parentheses)

Variable	S01 Model		Augmented S01 Model		CJ04 Model		Augmented CJ04 Model	
NI/TA	-0.40**	(0.20)	-0.45	(0.31)	-0.38	(0.29)	-1.57***	(0.38)
TL/TA	2.98***	(0.09)	2.76***	(0.09)	2.85***	(0.13)	2.09***	(0.19)
EXRET	-1.77***	(0.13)	-0.20**	(0.10)	-1.64***	(0.13)	-0.14	(0.10)
RSIZE	-0.18***	(0.02)	-0.31***	(0.02)	-0.18***	(0.02)	-0.35***	(0.02)
σ_E	0.21***	(0.02)	0.14***	(0.04)	0.20***	(0.02)	0.12***	(0.04)
IND2					-0.50***	(0.14)	-0.53***	(0.15)
IND3					-0.14	(0.26)	0.22	(0.26)
IND4					-0.64	(0.49)	-0.24	(0.47)
NI/TA*IND2					0.89**	(0.39)	0.77**	(0.37)
TL/TA*IND2					0.51***	(0.18)	0.56***	(0.18)
NI/TA*IND3					0.13	(0.60)	0.69	(0.58)
TL/TA*IND3					0.29	(0.30)	-0.39	(0.30)
NI/TA*IND4					-3.30***	(0.72)	-2.40***	(0.78)
TL/TA*IND4					-0.24	(0.56)	-1.13**	(0.54)
log(TA)			-0.02	(0.02)			-0.04	(0.04)
EXRET*log(TA)			-0.32***	(0.02)			-0.30***	(0.02)
NI/TA*log(TA)			-0.51***	(0.06)			-0.26***	(0.08)
σ_E*log(TA)			0.04**	(0.02)			0.04**	(0.02)
TL/TA*log(TA)							0.15***	(0.04)
	DSW07-S Model		Augmented DSW07-S Model		BS08 Model		Augmented BS08 Model	
Naïve DD	-0.40***	(0.02)	-0.32***	(0.03)				
RETURN	-1.10***	(0.15)	-0.92***	(0.15)				
3m T-rate	21.70***	(2.90)	20.45***	(2.87)				
SPX	1.62***	(0.29)	1.57***	(0.28)				
log(TA)			-0.06***	(0.01)				
(Naïve DD)*log(TA)			-0.02***	(0.00)				
$\pi_{\text{Naïve}}$					1.47***	(0.14)	1.42***	(0.14)
$\log E$					-0.24***	(0.02)	-0.08**	(0.04)
$\log F$					0.25***	(0.02)	0.22***	(0.02)
$1/\sigma_E$					-0.59***	(0.04)	-0.52***	(0.04)
EXRET					-0.79***	(0.12)	-0.74***	(0.12)
NI/TA					-3.37***	(0.18)	-1.80***	(0.27)
NI/TA*($\log E$)							-0.77***	(0.09)
$(\log E) * (\log E)$							-0.05***	(0.01)
$(\log F) * (\log E)$							0.02***	(0.01)

This table reports on the parameter estimates of the four reference models, and the corresponding augmented models which have additional effects related to information noise. The reference models are labeled as "S01", "CJ04", "DSW07-S" and "BS08" models respectively, in the corresponding columns. The augmented models are next to the corresponding reference models. The sample period is 1979–2012. There are 2,112 bankruptcies, and 2,152,203 firm-month observations in the full sample. All models are Cox [1972] proportional hazard models with time-varying covariates. All explanatory variables are described in Table 1, and all logarithms are natural logarithms. A positive coefficient on a particular variable in the reference models implies that the hazard rate is increasing in that variable, and vice versa. In the augmented models, the coefficient on a covariate (if it also has an interaction effect) is a linear function of the proxy for the degree of noise. Robust standard errors are are reported in parentheses (see Lin and Wei [1989] or Allison [2010]). (*** significant at 1% level, ** significant at 5% level, * significant at 10% level).

unaware of, apart from our study, any prediction, or alternative explanation, about the signs of the coefficients on these interaction effects, in the empirical literature of credit risk.

To test Hypothesis 2, following Chava and Jarrow [2004], we conduct a likelihood ratio test to test if our proposed effects as a whole, i.e., interaction effects and proxies for the degree of noise, significantly improve the in-sample Goodness-of-Fit. The results are reported in Panel A of Table 4. Here we treat the reference models as the constrained versions of the augmented models, by constraining the coefficients on our proposed effects to zero. The unconstrained versions are simply augmented models. The difference in (the minimized) -2log-likelihood ($-2\log L$) of the two versions provides a likelihood ratio test (χ^2-test), under the null hypothesis that the constrained model is the true model. As can be seen from the column labeled as "χ^2 Statistics" in Panel A of Table 4 (which is the difference in $-2\log L$ of two models), the likelihood of the augmented models is much, and significantly, larger than that of the corresponding reference models, implying that we can easily reject the null hypothesis and conclude that the augmented models are significantly better in terms of in-sample Goodness-of-Fit. This is consistent with Hypothesis 2, and again strongly supports the validity of our proposed hazard specifications in bankruptcy forecasting.

We also report, in Panel B of Table 4, alternative Goodness-of-Fit measures that are popular in the literature, namely McFadden's pseudo-R^2 (pseudo-R^2) and Akaike Information Criteria (AIC).[61] Higher pseudo-R^2 or lower AIC implies better in-sample model fit. As shown in Panel B of Table 4, the augmented models have better Goodness-of-Fit by either measure, which are again consistent with Hypothesis 2.[62]

5.2 Out-of-Sample Tests of Forecasting Accuracy

With respect to Hypothesis 3, we conduct out-of-sample tests, in three steps, to assess forecasting accuracy of the reference and augmented models.

First, similar to Chava and Jarrow [2004], Duffie, Saita and Wang [2007], we build ten holdout samples, one for each year during 2003–2012.[63] Any firm whose information is available at the beginning of a holdout year is included as one observation into the holdout sample for that particular year.[64] Every observation also has an indicator variable,

[61] I thank Jens Hilscher for suggesting these measures. Examples in the previous literature that used pseudo-R^2 include Campbell, Hilscher and Szilagyi [2008] or Campbell, Hilscher and Szilagyi [2011].

[62] Like the results in Panel A of Table 4, measures in Panel B are also based on log-likelihood. Therefore, it might be unsurprising that we find similar conclusions using likelihood ratio tests and alternative measures.

[63] Like Chava and Jarrow [2004], we also generate 1,000 bootstrapped holdout samples over 2003-2012. The results on the bootstrapped samples are very similar to those reported in this paper, and are available upon request.

[64] Note that a firm is included only if it has information available at the beginning of a holdout year. Thus, for example, firms entering the database in the middle of a holdout year will not be included into the holdout sample of that year.

Table 4: In-sample Goodness-of-Fit on our proposed hazard specifications

Panel A: Likelihood Ratio test

	$-2\log L$ Unconstrained	$-2\log L$ Constrained	χ^2 Statistics	Degree of Freedom	p-value
Augmented S01 Model	30301.24	31199.65	898.40	4	< 0.001
Augmented CJ04 Model	30047.78	31048.98	1001.20	5	< 0.001
Augmented DSW07-S Model	31297.20	31364.65	67.45	2	< 0.001
Augmented BS08 Model	30203.37	30418.94	215.57	3	< 0.001

Panel B: Alternative Goodness-of-Fit measures

	Pseudo-R^2	AIC
S01 Model	0.1511	31209.65
Augmented S01 Model	0.1755	30319.24
CJ04 Model	0.1552	31076.98
Augmented CJ04 Model	0.1824	30085.78
DSW07-S Model	0.1466	31372.65
Augmented DSW07-S Model	0.1484	31309.20
BS08 Model	0.1723	30430.94
Augmented BS08 Model	0.1782	30221.37

This table reports on a Likelihood Ratio test and alternative Goodness-of-Fit measures, in Panel A and B respectively. The Likelihood Ratio test tests the statistical significance of all effects related to information noise as a whole (including both main and interaction effects). The augmented models, with our proposed effects, are considered as unconstrained models, and the reference models, without these effects, are considered as constrained models (by constraining the coefficients on our proposed effects to zero). The difference in $-2\log L$ between the constrained and unconstrained models are listed in the column labeled as "χ^2 Statistics". The alternative Goodness-of-Fit measures include McFadden's pseudo-R^2 (pseudo-R^2) and Akaike Information Criteria (AIC), which are listed in the corresponding columns for the reference and augmented models respectively. Higher pseudo-R^2 or lower AIC implies better in-sample model fit. The sample period is 1979–2012. There are 2,112 bankruptcies, and 2,152,203 firm-month observations in the full sample.

indicating whether it files bankruptcy within the relevant holdout year. In total, within the entire ten-year holdout period, we have 53,636 observations to be predicted, and 558 bankruptcy events, from 8,905 (distinct) firms.

Second, we produce one-year-ahead forecasts for all models. We use the same set of independent variables as those in Table 3. At the beginning of each holdout year, we re-estimate coefficients on all independent variables based on the information available at that time. We then forecast bankruptcies within the particular holdout year, using the re-estimated models and available information. For example, to predict bankruptcies within the holdout sample of 2005, we re-estimate all models using the firm-month panel data from January 1, 1979 to December 31, 2004, and then use information available at December 31, 2004, together with the re-estimated models, to produce forecasts of bankruptcy in 2005. Like most studies in the previous literature, we consider rank ordering, not the exact probability of bankruptcy. Hence, the model forecasts are in fact a score. For example, the scores for the augmented models are[65]

$$Score = \bar{\beta}' X_t + \bar{\gamma}_0 \tilde{a} + \sum_{i=1}^{I} \bar{\gamma}_i (\tilde{a} * X_t^i), \tag{6}$$

where the notations are the same as in Equation (4), and $\bar{\beta}$, $\bar{\gamma}_0, \ldots, \bar{\gamma}_I$ are estimated from data.

Finally, we rank observations within each holdout year according to the model scores, and assess forecasting accuracy of the rankings. The standard measures of bankruptcy forecasting accuracy are Receiver Operating Characteristic (ROC) curve together with its summary statistics, Area Under ROC Curve (henceforth, AUC), and Cumulative Accuracy Profile (CAP) curve together with the associated summary statistics, Gini coefficient (GINI). Because AUC and GINI are equivalent,[66] we only report AUC as our first measure of forecasting accuracy, both year-by-year and on average. AUC is also used in, for example, Chava and Jarrow [2004], Duffie, Saita and Wang [2007]. The second measure we employ is the one adopted in Shumway [2001], Chava and Jarrow [2004], Bharath and Shumway [2008], which is the captured proportions of the total number of bankruptcies within deciles. To calculate this measure, for each holdout year, we count the number of captured bankruptcies, within each decile that is formed by sorting and grouping model scores. Then we aggregate the number of bankruptcies in each decile across the entire 10-year holdout period, and calculate them as percentages of the total number of bankruptcies. We recognize that this measure is equivalent to a bankruptcy-

[65] In our case, the one-year-ahead survival probability produced by the Cox model, assuming all firms have updated information at the beginning of a holdout year, is a monotonic transformation of the score in Equation (6). Consequently, the rank orderings produced by survival probability and the score are the same.

[66] $GINI = 2AUC - 1$. For more details of these, and other, measures of discriminative power of rating systems, see, for example, Engelmann, Hayden and Tasche [2003].

weighted aggregate CAP curve over the entire 10-year holdout period. We adopt this measure because it provides another perspective and more details on models' predictive performance.

We report the out-of-sample AUC of all models, in corresponding columns of Table 5 (labeled by model names). Higher AUC implies better forecasting accuracy. To compare AUC of an augmented model with AUC of its corresponding reference model in each holdout year, we conduct a χ^2-test of DeLong, DeLong and Clarke-Pearson [1988], which tests the difference of two correlated ROC curves.[67] The test results are reported in the four columns labeled as "AUC Difference", with χ^2 statistics in parentheses.[68] We also report the average AUC difference over the 10 holdout years. In spirit of the Lyapounov Central Limit Theorem, we conduct a χ^2-test to test if the average AUC difference is significantly different from zero.[69]

In Table 5, the reference models have notably very high forecasting accuracy, all with average out-of-sample AUC above 0.9. This is consistent with the results from the previous literature, and might be attributed to proper variable selection, high frequency (monthly frequency) in observations, availability of market information and high-quality accounting information in Compustat. On top of such predictive models, Table 5 shows that our interaction effects still achieve statistically significant and persistent improvements on forecasting accuracy. Typically, in more than half of the 10 holdout years, the augmented models, with our interaction effects, significantly *outperform* their corresponding reference models, at <10% level, but *never* perform significantly worse.[70] In fact, out of the 10 holdout years, there are very few years in which the augmented models' AUC is less than that of their corresponding reference models.[71] Not surprising, the average improvements in AUC of the augmented models are highly significant.[72]

To the best of our knowledge, the above out-of-sample results, especially the year-by-year AUC improvements, are one of the strongest pieces of evidence in the literature.[73]

[67]Because we calculate AUC of two models within the same holdout year, a test on correlated ROC curves is deemed appropriate.

[68]We only focus on *out-of-sample* results in Table 5, for each holdout year and averaged over the entire holdout period. The *in-sample* Goodness-of-Fit statistics (based on data before any holdout year) are omitted here for brevity and are available upon request.

[69]Assuming each AUC difference is drawn independently from different normal distributions, the Lyapounov Central Limit Theorem gives the sampling distribution of the average AUC difference. We can then conduct a χ^2-test in a standard way, on the null hypothesis that the mean of the sampling distribution is zero.

[70]We also find similar results on the 1,000 bootstrapped holdout samples. The augmented models are typically significantly better than the reference models in around 60% of the 1,000 holdout samples, and insignificantly better in 25%–30%, insignificantly worse in about 10%, significantly worse in less than 5%.

[71]Note that the χ^2-test has the null hypothesis of "AUC Difference= 0". Had it been an one-tailed t-test with the null hypothesis of "AUC Difference< 0", we would have seen more holdout years with significant improvements.

[72]Similar to Chava and Jarrow [2004], we also conduct (paired) t-tests and nonparametric tests on the 1,000 bootstrapped holdout samples, which confirm the results.

[73]We note that Chava and Jarrow [2004], Duffie, Saita and Wang [2007] conducted similar out-of-

Table 5: Out-of-sample forecasting accuracy: Area Under ROC Curve (AUC)

2003–2012, 53,636 one-year-ahead forecasts for each model, and 558 bankruptcies, from a total of 8,905 firms
(χ^2 statistics in parentheses)

Holdout Period	# of Firms	# of Bank-ruptcy	S01 Model (1)	Augmented S01 Model (2)	AUC Difference (2)-(1)	(χ^2)	CJ04 Model (3)	Augmented CJ04 Model (4)	AUC Difference (4)-(3)	(χ^2)
2003	6172	101	0.9244	0.9353	1.09×10^{-2} **	(5.29)	0.9244	0.9350	1.06×10^{-2} **	(4.92)
2004	5845	52	0.9058	0.9296	2.38×10^{-2} ***	(11.99)	0.9218	0.9391	1.73×10^{-2} ***	(8.56)
2005	5593	40	0.9259	0.9357	9.80×10^{-3} ***	(7.52)	0.9340	0.9426	8.60×10^{-3} ***	(14.09)
2006	5531	30	0.8841	0.9075	2.34×10^{-2} **	(4.94)	0.8884	0.9162	2.78×10^{-2} **	(4.07)
2007	5471	26	0.9145	0.9011	-1.34×10^{-2}	(1.25)	0.8993	0.8907	-8.60×10^{-3}	(1.28)
2008	5275	61	0.8857	0.8942	8.50×10^{-3} *	(3.58)	0.8882	0.9051	1.69×10^{-2} ***	(8.81)
2009	5150	122	0.8823	0.8797	-2.60×10^{-3}	(0.21)	0.8734	0.8791	5.70×10^{-3}	(0.85)
2010	4839	43	0.9205	0.9296	9.10×10^{-3}	(1.79)	0.9292	0.9387	9.50×10^{-3} *	(3.47)
2011	4704	37	0.8942	0.9081	1.39×10^{-2}	(1.75)	0.9019	0.9139	1.20×10^{-2}	(1.41)
2012	5056	46	0.8993	0.9115	1.22×10^{-2} *	(2.82)	0.9091	0.9215	1.24×10^{-2}	(2.25)
Average			0.9037	0.9132	9.57×10^{-3} ***	(15.26)	0.9070	0.9182	1.12×10^{-2} ***	(21.79)

Holdout Period	# of Firms	# of Bank-ruptcy	DSW07-S Model (5)	Augmented DSW07-S (6)	AUC Difference (6)-(5)	(χ^2)	BS08 Model (8)	Augmented BS08 Model (7)	AUC Difference (8)-(7)	(χ^2)
2003	6172	101	0.9245	0.9277	3.20×10^{-3} **	(6.47)	0.9273	0.9317	4.40×10^{-3} **	(4.34)
2004	5845	52	0.9230	0.9279	4.90×10^{-3}	(1.80)	0.9467	0.9510	4.30×10^{-3} *	(2.72)
2005	5593	40	0.9390	0.9372	-1.80×10^{-3}	(0.35)	0.9440	0.9481	4.10×10^{-3}	(0.75)
2006	5531	30	0.8777	0.8946	1.69×10^{-2} ***	(7.56)	0.9230	0.9283	5.30×10^{-3}	(0.80)
2007	5471	26	0.9089	0.9133	4.40×10^{-3}	(0.32)	0.9235	0.9156	-7.90×10^{-3}	(0.26)
2008	5275	61	0.8786	0.8860	7.40×10^{-3} **	(4.17)	0.9188	0.9154	-3.40×10^{-3}	(0.19)
2009	5150	122	0.8654	0.8666	1.20×10^{-3}	(0.38)	0.8573	0.8619	4.60×10^{-3}	(1.96)
2010	4839	43	0.8921	0.9030	1.09×10^{-2} ***	(6.75)	0.9204	0.9377	1.73×10^{-2} ***	(7.65)
2011	4704	37	0.9284	0.9366	8.20×10^{-3} **	(5.18)	0.9268	0.9328	6.00×10^{-3}	(0.79)
2012	5056	46	0.8742	0.8863	1.21×10^{-2} ***	(10.12)	0.8976	0.9108	1.32×10^{-2} **	(5.31)
Average			0.9012	0.9079	6.74×10^{-3} ***	(24.57)	0.9185	0.9233	4.81×10^{-3} **	(4.70)

This table reports on Area Under ROC Curve (AUC) of all models for every holdout year during 2003–2012. There are 558 bankruptcies, and 53,636 firm-year observations to be predicted in the entire 10-year holdout period. At the beginning of each year, all models are estimated using data available at that time, and produce one-year-ahead forecasts. The accuracy of a model, for a particular year, is measured by AUC, using the forecasts and actual bankruptcies within that year. The AUCs of the reference and the corresponding augmented models, together with their differences, are reported in columns labeled by their model names. A positive (negative) AUC difference, in a particular year, implies that a model with the proposed interaction effects predicts bankruptcies more (less) accurately than the model without, in that year. A χ^2-test, which compares two correlated ROC curves (see DeLong, DeLong and Clarke-Pearson [1988]), is conducted to test the statistical significance of the AUC differences being different from zero, with the χ^2 statistics reported in parentheses. Finally, the average AUC and average AUC difference, over the entire holdout period, are calculated. A χ^2-test, in spirit of the Lyapounov Central Limit Theorem, is conducted to test if the average AUC differences are significantly different from zero (*** significantly different from zero at 1% level, ** significant at 5% level, * significant at 10% level).

The high statistical significance of the AUC increments manifests the robustness and persistence of the improvements in forecasting accuracy.[74] By controlling the main effects of the covariates from the reference models, we ensure that the improvements come purely from our proposed interaction effects. Therefore, the tests on out-of-sample AUC provide strong evidence that our interaction effects markedly improve hazard model specifications.

We examine the second measure of forecasting accuracy, the fractions of the total number of bankruptcies captured within deciles, aggregated over the entire 10 holdout years, in Panel A of Table 6. Panel B is similar to Panel A, except that, in each decile, we calculate the cumulative fractions (over the previous deciles) of the total number of bankruptcies.

To interpret the results in Panel A of Table 6, a model is deemed better than another if it captures more bankruptcies in the first few deciles. Apparently, the augmented models capture more bankruptcies, than their corresponding reference models, in all of the top two or three deciles.[75] For example, Augmented BS08 Model captures 75.45% of the total number of bankruptcies in the first decile, higher than 74.74% captured by BS08 Model. Likewise, in the second and third deciles, Augmented BS08 Model also captures more bankruptcies (13.26% vs 12.72%, and 4.3% vs 4.12%, respectively). We also note that, within the low-risk deciles (deciles 6-10), the augmented models capture *less* bankruptcies, and thus has less misclassification, than the corresponding reference models. Therefore, Panel A of Table 6 demonstrates that our proposed interaction effects indeed improve the discriminative power of the reference models.

In Panel B of Table 6, we report the cumulative captured proportions of total bankruptcies, which are effectively CAP curves. As a standard criterion to compare two CAP curves, one model unambiguously outperforms another if its CAP curve is no lower than the CAP curve of another model in *all* deciles. Clearly, this is the case in Panel B of Table 6, where the CAP curves of the augmented models are higher than or equal to, in all deciles, those of the corresponding reference models. The evidence is unambiguous, and further confirms the superior predictive performance of the augmented models.

Altogether, Table 5 and Table 6 provide strong empirical evidence that our proposed interaction effects significantly improve the out-of-sample forecasting accuracy of hazard

sample tests. However, Chava and Jarrow [2004] reported the average AUC difference but no year-by-year results. Duffie, Saita and Wang [2007] reported year-by-year AUC but without comparing them to any benchmark using statistical tests.

[74] The absolute magnitude in the AUC differences depends on how noisy a specific sample is. In robustness checks (Section 5.3), we show that, when data is noisier and less frequently updated, the absolute magnitude in AUC improvements can be substantial.

[75] We note that the augmented models not only outperform in the first decile, as usually seen in the previous literature, but are also better in the second, and sometimes third, deciles. This is expected, because our proposed interaction effects impact the rankings of firms if firms have different degrees of noise. Such impacts are not necessarily associated with the predicted probability of bankruptcy. Thus, the improvements on captured bankruptcies might occur in multiple deciles.

Table 6: Out-of-sample forecasting accuracy: captured bankruptcies within deciles ranked by model forecasts

2003–2012, 53,636 one-year-ahead forecasts for each model, and 558 bankruptcies, from a total of 8,905 firms

Panel A: Fractions of bankruptcies captured within deciles ranked by model forecasts (%)

Decile	S01 Model	Augmented S01 Model	CJ04 Model	Augmented CJ04 Model	DSW07-S Model	Augmented DSW07-S Model	BS08 Model	Augmented BS08 Model
1	74.01	74.91	75.63	76.34	71.15	72.22	74.73	75.45
2	11.47	11.83	11.29	11.29	14.34	14.87	12.72	13.26
3	6.09	5.56	5.20	5.02	4.66	5.02	4.12	4.30
4	3.23	2.15	2.51	2.51	3.41	1.97	3.05	2.69
5	1.43	1.79	1.43	1.25	1.43	1.43	1.25	1.25
6-10	3.76	3.76	3.94	3.58	5.02	4.48	4.12	3.05

Panel B: Cumulative fractions of bankruptcies captured within deciles ranked by model forecasts (%)

Decile	S01 Model	Augmented S01 Model	CJ04 Model	Augmented CJ04 Model	DSW07-S Model	Augmented DSW07-S Model	BS08 Model	Augmented BS08 Model
1	74.01	74.91	75.63	76.34	71.15	72.22	74.73	75.45
2	85.48	86.74	86.92	87.63	85.48	87.10	87.46	88.71
3	91.58	92.29	92.11	92.65	90.14	92.11	91.58	93.01
4	94.80	94.44	94.62	95.16	93.55	94.09	94.62	95.70
5	96.24	96.24	96.06	96.42	94.98	95.52	95.88	96.95
6-10	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

This table reports on the forecasting accuracy of all models during the holdout period of 2003–2012, by first sorting firms each year into deciles according to model forecasts, and then counting and aggregating the number of bankruptcies captured within each decile. There are 558 bankruptcies, and 53,636 firm-year observations to be predicted in the entire holdout sample. Panel A reports the fractions of total number of bankruptcies captured within each decile. Panel B reports the cumulative fractions of bankruptcies (i.e., captured bankruptcies within a decile and its previous deciles). Both panels report the results of the reference and corresponding augmented models in the columns labeled by the model names. There are three key indicators that one model predicts bankruptcies more accurately than another model: (i) it captures more bankruptcies in the first few deciles (deciles 1 and 2, or 3); (ii) it captures less bankruptcies within the low-risk deciles (deciles 6-10); (iii) the cumulative fractions of bankruptcy that it captures are no lower in all deciles.

models. Hence, these results strongly support the use of our hazard specifications in real-world bankruptcy forecasting, where firm-specific information is likely to be noisy.

5.3 Robustness Check

In this section, we test the robustness of our proposed hazard specifications in three aspects. First, we check if our empirical results are robust to the choices of proxy for the degree of noise. Second, following convention in the literature, we remove financial firms from our sample and check if the results persist. Third, instead of using the empirical setup in the previous sections, we test the impact of our hazard specifications in an environment where information quality is a more serious problem. Within this environment, we have no market information, less frequently updated financial reports, fewer explanatory covariates, and many outliers. This environment is typical for creditors (or rating agents) to predict bankruptcy/default of private firms.[76] In the following, we present all results using robust standard errors, although the standard errors are not reported here for brevity.

As the first robustness check, we report, in Table 7, the combined results of the full-sample estimates and out-of-sample forecasting accuracy, using a number of alternative proxies for the degree of noise (the second column of Table 7), including firm's asset rank (log(AR)), equity market value ($\log E$),[77] analyst coverage (AC) and normalized variation in analyst forecasts ($-\log(\text{CV})$), as described in Section 3. The datasets (including 10 holdout samples) and reference models we use here are the same as those described before, in Sections 3 and 4. Using these alternative proxies, we construct new augmented models by adding the interaction effects (the third column of Table 7) into the corresponding reference models (the first column of Table 7). For brevity, we only report the full-sample estimates of the interaction effects in the augmented models (the fourth column of Table 7), and the average out-of-sample AUC differences, between the augmented and corresponding reference models (the fifth column of Table 7).[78]

As shown in Table 7, the coefficients on all interaction effects have the same signs as those predicted by Hypothesis 1, irrespective of what proxies for the degree of noise are adopted, and what reference models are used. This result is striking, and shows that the predictions of Hypothesis 1 are indeed general rules rather than special cases. As for out-of-sample forecasting accuracy, compared to results using log(TA) as a proxy, log(AR)

[76] Such creditors include, for example, banking institutions, or investors in private firm debts.

[77] Recall that when adding $\log E$, as a proxy for the degree of noise, to S01 Model and CJ04 Model, we exclude RSIZE as a covariate to prevent multi-collinearity problems, due to strong contemporaneous correlation between them (0.92). As another robustness check, we also use RSIZE as a proxy, and get similarly favorable results on our proposed interaction effects. The results are not shown for brevity, and are available upon request.

[78] The details of other in-sample and out-of-sample tests, omitted here for brevity, are available upon request.

Table 7: Robustness check: alternative proxies for the degree of noise

Full sample: 1979–2012, 2,152,203 firm-months, 2,112 bankruptcies
Holdout samples: 2003–2012, 53,636 one-year-ahead forecasts for each model, 558 bankruptcies

Reference Model	Proxy for the Degree of Noise	Our Proposed Interaction Effect	Full Sample Estimate	Average Uplift of Out-of-Sample AUC
S01 Model	log(AR)	EXRET*log(AR)	-0.38***	1.03×10^{-2} ***
		TL/TA*log(AR)	0.30***	
		σ_E*log(AR)	0.11***	
	$\log E$	EXRET*$\log E$	-0.15***	7.50×10^{-3} ***
		NI/TA*$\log E$	-0.80***	
		TL/TA*$\log E$	0.09**	
		$\sigma_E * \log E$	0.10***	
	$-\log(\text{CV})$	EXRET*$(-\log(\text{CV}))$	-0.15***	3.98×10^{-3} **
		σ_E*$(-\log(\text{CV}))$	0.03***	
CJ04 Model	log(AR)	EXRET*log(AR)	-0.35***	1.19×10^{-2} ***
		TL/TA*log(AR)	0.37***	
		σ_E*log(AR)	0.09***	
	$\log E$	EXRET*$\log E$	-0.13**	7.51×10^{-3} ***
		NI/TA*$\log E$	-0.79***	
		TL/TA*$\log E$	0.12***	
		$\sigma_E * \log E$	0.09***	
	$-\log(\text{CV})$	EXRET*$(-\log(\text{CV}))$	-0.15***	2.98×10^{-3} *
		σ_E*$(-\log(\text{CV}))$	0.03**	
DSW07-S Model	log(AR)	(Naïve DD)*log(AR)	-0.06***	7.68×10^{-3} ***
	$-\log(\text{CV})$	RETURN*$(-\log(\text{CV}))$	-0.10***	3.71×10^{-3} ***
	AC	RETURN*AC	-0.09**	2.06×10^{-3} ***
BS08 Model	$-\log(\text{CV})$	RETURN*$(-\log(\text{CV}))$	-0.11***	3.40×10^{-3} ***
	AC	RETURN*AC	-0.12***	2.18×10^{-3} *

This table provides a summary of both full-sample estimates, and average improvements in out-of-sample forecasting accuracy, of the augmented models when alternative proxies for the degree of noise are adopted. The full-sample period is 1979–2012. There are 2,112 bankruptcies, and 2,152,203 firm-month observations in our dataset. The holdout sample period is 2003–2012. There are 558 bankruptcies, and 53,636 firm-year observations to be predicted within the entire holdout sample. We construct the augmented models by adding the interaction effects (the third column) into the corresponding reference models (the first column). The interaction effects are created using alternative proxies for the degree of noise (the second column). When using $\log E$ as a proxy within S01 and CJ04 Models, we exclude RSIZE as a covariate to prevent multi-collinearity problems, due to strong contemporaneous correlation between them (0.92). For brevity, we only report the full-sample estimates of the coefficients on the interaction effects (the fourth column), and the average out-of-sample AUC differences between the augmented and corresponding reference models (the fifth column). The statistical significance of the full-sample estimates is tested using robust standard errors (not reported here). (*** significantly different from zero at 1% level, ** significant at 5% level, * significant at 10% level) and $\log E$ give similar improvements in AUC.[79] This confirms the validity of using firm size to create our interaction effects. For AC and $-\log(\text{CV})$, the AUC improvements are also significant on average, at <10% level, although the absolute magnitude of the improvements is marginal. Given the lack of variation in these two variables due to the high proportion of missing values,[80] this out-of-sample predictive performance is in fact surprisingly good. It indicates that they are appropriate proxies for the degree of noise, and that our hazard specifications work reasonably well even when data availability of good proxies is low.

In the second robustness check, we follow convention in the previous literature to exclude financial firms in our sample, i.e. firms with CRSP SIC between 6000 and 6800. We then re-run all the models specified in Table 3.[81] We report the combined results of the full-sample estimates and out-of-sample forecasting accuracy in Table 8, which has a similar format to that of Table 7.

We find that both in-sample and out-of-sample results, in Table 8, are actually slightly better than those using data without exclusions (i.e. results reported in Sections 5.1 and 5.2). Our interaction effects now have even higher in-sample statistical significance, again with the expected signs. For example, the variable "σ_E*log(TA)" in the augmented S01 Model now has a *p*-value of <0.0001, as opposed to 0.04 in Table 3. The improvements in out-of-sample forecasting accuracy brought by our interaction effects are also greater. For example, now the augmented BS08 Model has an average uplift of AUC 0.0074 (significant at 1% level), increasing from 0.0048 (significant at 5% level) reported in Table 5. Similar results are observed for all augmented models in Table 8. Therefore, we demonstrate that our hazard specifications are robust to the choices of firm subpopulation.

In the third robustness check, we consider a different empirical setup, where firms' market information is unavailable, firms' accounting reports are updated annually, and there are fewer explanatory covariates available, potentially with many outliers. Such environment might be more realistic for practical default prediction, for instance, on bank loans within an internal rating system. While the setup poses more difficulties for modeling, our proposed hazard specifications are expected to bring substantially more benefits in this case where imperfect information becomes a more severe problem. We show this is indeed the case, in the following investigation.

Now we use Compustat Annual (North America) accounting data to construct independent variables. Without the requirements of joining CRSP data, we are able to include more bankruptcy events in the new dataset, 2,537 in total. The full sample now has 290,811 firm-year observations, from a total of 27,443 firms. As before, we perform

[79] In the case of log(AR), results are even better.

[80] Recall that, as shown in Section 4, these two variables have more than half of their values with only one or two values, due to missing value.

[81] We also use asset rank as a proxy for the degree of noise on this dataset, and obtain similar results (not reported here).

Table 8: Robustness check: non-financial firms

Full sample: 1979–2012, 1,769,316 firm-month observations, 1,895 bankruptcies
Holdout samples: 2003–2012, 41,955 one-year-ahead forecasts for each model, 470 bankruptcies

Reference Model	Our Proposed Interaction Effect	Full Sample Estimate	Average Uplift of Out-of-Sample AUC
S01 Model	EXRET*log(TA)	-0.25***	1.12×10^{-2} ***
	NI/TA*log(TA)	-0.36***	
	σ_E*log(TA)	0.10***	
CJ04 Model	EXRET*log(TA)	-0.24***	1.24×10^{-2} ***
	NI/TA*log(TA)	-0.26***	
	TL/TA*log(TA)	0.19***	
	σ_E*log(TA)	0.09***	
DSW07-S Model	(Naïve DD)*log(TA)	-0.03***	6.69×10^{-3} ***
BS08 Model	NI/TA*$\log E$	-0.72***	7.42×10^{-3} ***
	$\log E * \log E$	-0.05***	
	$\log F * \log E$	0.02***	

This table provides a summary of both full-sample estimates, and average improvements in out-of-sample forecasting accuracy, of the augmented models when financial firms (with CRSP SIC between 6000 and 6800) are excluded. The full-sample period is 1979–2012. There are 1,895 bankruptcies, and 1,769,316 firm-month observations in our dataset. The holdout sample period is 2003–2012. There are 470 bankruptcies, and 41,955 firm-year observations to be predicted within the entire holdout period. Based on this dataset, we test all the models specified in Table 3. For brevity, we only report the full-sample estimates of the coefficients on the interaction effects in the augmented models (the third column), and the average out-of-sample AUC differences between the augmented and corresponding reference models (the fourth column). The statistical significance of the full-sample estimates is tested using robust standard errors (not reported here). (*** significantly different from zero at 1% level, ** significant at 5% level, * significant at 10% level) winsorization (at 1^{st} and 99^{th} percentiles), missing value imputation (by carrying forward) and currency conversion (CAD to USD). To facilitate this "private firm" modeling environment, we choose the "Private Firm Model" in Chava and Jarrow [2004] as our reference model, which uses the two financial ratios, NI/TA and TL/TA, from S01 Model.[82] The augmented model is developed by using total assets (log(TA)) as the proxy for the degree of noise, and interacting it with NI/TA and TL/TA respectively.[83] We report the combined results on the full-sample estimates and out-of-sample forecasting accuracy in Table 9.

We make a number of notes on Table 9. Within the reference model, interestingly the full-sample estimate of coefficient on NI/TA is insignificant. Further investigation of the data reveals that the insignificance might be caused by many (more than 1%) extreme values in this variable, or outliers.[84] These outliers turn out to be less correlated with bankruptcy events, and thus cause the main effect of NI/TA to be insignificant. We note that these outliers typically have very low total assets, and, as small firms, their accounting information might have high degree of noise. A potential solution is to further winsorize NI/TA at, for example, 5^{th} percentile. Nevertheless, although further winsorization can make this variable statistically significant *in sample*, the *out-of-sample* forecasting accuracy deteriorates significantly.[85] This is probably because over-winsorization distorts the data too much and produces unintended consequences (for example, biased estimates), which poses a dilemma for the winsorization approach to handle outliers.

On the contrary, our approach does not excessively distort the data. We augment the data with a new variable, i.e., the degree of noise, and use our interaction effects to take into account of noise in the data. As shown in Table 9, this approach remarkably improves the hazard specification. First, the full-sample estimates of our interaction effects (-0.09 and 0.14, respectively) are highly significant, at <1% level. Second, with our interaction effects, the coefficient on NI/TA becomes $[-0.05 - 0.09 \log(\text{TA})]$, which is now statistically significant because both -0.05 and -0.09 are significantly different from zero.[86] Third, the column labeled as "$-2 \log L$" shows that the log-likelihood, as an in-sample Goodness-of-Fit measure, of the augmented model is significantly higher than that of the reference model, as demonstrated by the difference of $-2 \log L$ and its p-value, in the last two rows

[82] The reference model adopted here is, of course, a simplistic one. It is not central to us how to build a full-fledged model for private firms, or bank loans. Our intention is to show the potential benefits of our hazard specifications when they are used within a similar setup.

[83] We also use asset rank as the proxy to create interaction effects. The results, not reported here, are similar.

[84] Obviously, winsorization at 1^{st} percentile does not solve this problem, because there are more than 1% outliers, and NI/TA still has extreme values after winsorization.

[85] The average out-of-sample AUC of the model built on the 5-percentile-winsorized data is 0.7618, significantly *worse* than that of the reference model (0.7882, as shown in Table 9), which is built on the 1-percentile-winsorized data but has insignificant NI/TA.

[86] In this case, we might say that the coefficient on NI/TA is not significantly different from zero if neither -0.05 nor -0.09 are significantly different from zero, a hypothesis that we can easily reject here.

Table 9: Robustness check: private firm models

Full sample: 1979–2012, 290,811 firm-year observations, 2,537 bankruptcies
Holdout samples: 2003–2012, 90,407 one-year-ahead forecasts for each model, 758 bankruptcies

Model	Variable	Full Sample Estimate	$-2\log L$	Average of Out-of-Sample AUC
Private Firm Model	NI/TA	-0.02	45,933	0.7882
	TL/TA	0.24***		
Augmented Private Firm Model	NI/TA	-0.05**	44,596	0.8304
	TL/TA	0.17***		
	log(TA)	-0.12***		
	NI/TA*log(TA)	-0.09***		
	TL/TA*log(TA)	0.14***		
Difference (= Private Firm Model - Augmented Model)			1,337	-0.0422
p-value from a χ^2-test			< 0.0001	< 0.0001

This table provides a summary of full-sample estimates, in-sample Goodness-of-Fit and average out-of-sample forecasting accuracy, of the reference and augmented models within a "private firm"-style empirical setup. We use Compustat Annual (North America) accounting data, without joining CRSP data. The full-sample period is 1979–2012. There are 2,537 bankruptcies, and 290,811 firm-year observations in our dataset. The holdout sample period is 2003–2012. There are 758 bankruptcies, and 90,407 firm-year observations to be predicted within the entire holdout period. We perform winsorization (at 1^{st} and 99^{th} percentiles), missing value imputation (by carrying forward) and currency conversion (CAD to USD). We adopt the "Private Firm Model" from Chava and Jarrow [2004] as our reference model. We develop augmented models using total assets (log(TA)) as the proxy for the degree of noise, and interact it with NI/TA and TL/TA respectively. The full-sample estimates of the coefficients on independent variables of both models are reported in the third column. The statistical significance of the full-sample estimates is tested using robust standard errors (not reported here). In the the fourth and fifth columns, respectively, we report -2log-likelihood ($-2\log L$), as an in-sample Goodness-of-Fit measure, and the average out-of-sample AUC, of both models, together with their differences and p-values from χ^2-tests. (*** significant at 1% level, ** significant at 5% level, * significant at 10% level) in Table 9.

Furthermore, to shed lights on how our approach improves the in-sample model fit, we note that the coefficient on NI/TA is now dynamic, depending on TA. When TA is small, which implies that the degree of noise is high, the coefficient automatically decreases in magnitude, reflecting that NI/TA becomes less responsive to bankruptcy risk when information is noisier. Therefore, our hazard specifications, represented by the "Augmented Private Firm Model" in Table 9, have a built-in mechanism to handle outliers driven by information noise.

Finally, our specifications dramatically improve the out-of-sample forecasting accuracy. As can be seen from the last column of Table 9, the average improvement in out-of-sample AUC (over 10 holdout samples) is 0.0422, which is highly significant, both statistically and in magnitude. If we translate AUC into GINI, which is a popular measure used in the industry, GINI of the reference and augmented models are 0.57 and 0.66, respectively. They are in different magnitude from a practical perspective. Thus, this investigation demonstrates that our proposed hazard specifications are robust to empirical setup, and their potential benefits might be substantial in real-world applications.

The third robustness check also reconciles the conflicting empirical findings on NI/TA in the literature. The statistical insignificance of the coefficient on NI/TA in the reference model is the same phenomenon documented in the previous literature (see, for example, Chava, Stefanescu and Turnbull [2011]). Our approach provides a plausible explanation why the coefficient on NI/TA can be insignificant, and how to handle these situations.

6 Conclusions

We introduce new hazard specifications that explicitly handle information noise in the input data, and empirically show their efficacy, using full-sample tests, out-of-sample tests, and a variety of robustness checks.

Our paper advances the literature in a number of ways. First, our specifications improve the empirical performance of popular hazard models, on both in-sample Goodness-of-Fit and out-of-sample forecasting accuracy. Second, we provide an empirical implementation of a theory of modeling credit risk with incomplete information (Duffie and Lando [2001]). Third, we highlight the importance, and provide a tool, to take into account of information noise within credit risk-related studies. Fourth, our specifications have a built-in mechanism to elegantly handle outliers, without excessively distorting data.

We also expect our proposed hazard specifications have a broad range of real-world applications in the financial industry. Our approach is theoretically justified, and practically easy to implement. We demonstrate the empirical success of our specifications, and potentially substantial benefits of using them in cases where data quality is a more serious problem. These advantages of our specifications are particularly appealing to the industry.

There are several possibilities for future work. First, while this paper predicts bankruptcy events for public firms, if data is available, it might also be interesting to test our proposed hazard specifications using default events, and for private firms. Second, we might explore alternative ways to construct proxies for the degree of noise. Potential candidates include corporate governance quality, expert judgments, or syntheses of multiple information quality measures. Third, we can conduct empirical study comparing different ways to handle outliers in default prediction. Finally, we might study other forms of incomplete information, like biased accounting reports or delayed information, or other ways to handle information quality issues.

Appendices

A Theoretical Probability of Bankruptcy with Incomplete Information

The conditional probability of bankruptcy (PB) of the debt issuer (firm), in Equation (1), is a function of the following parameters (see Duffie and Lando [2001]), when we set $t = 1, s = 2$,

1. V_0: the initial value of firm assets. We normalize it to be 1 throughout this paper, so that all other parameters can be expressed as a multiple of V_0;

2. μ: the expected growth rate of firm assets;

3. σ: the volatility of firm assets' growth rate;

4. $\delta > 0$: total cash flow generated by firm, expressed as a fraction of assets. In other words, the amount of cash flow at time t is δV_t;

5. r: the constant discount rate used to discount future cash flows by both firm and creditors. Because the DL model assumes all economic agents are risk-neutral, r is thus the riskless interest rate determined by the market;

6. $\theta > 0$: tax rate of firm;

7. α: a fraction of assets, $\in [0, 1]$, representing the loss due to friction costs in event of default/bankruptcy (i.e., bankruptcy costs);

8. $D > 0$: the face value of debt issued by firm, modeled as a consol bond;

9. $C > 0$: a constant coupon paid by the firm debt;

10. $\hat{V}_1$: the observed (noisy) assets;

11. $a > 0$: the standard deviation of noise associated with log(assets);

12. $u = -\frac{a^2}{2}$: the mean of noise associated with log(assets), assuming accounting report is unbiased.

Note that the bankruptcy threshold, $\underline{v}$, in Equation (1) is determined endogenously by other parameters.

To plot Figures 2 and 6, we vary $\hat{V}_1$ and a, and then calculate $r_N = (\frac{\hat{V}_1}{V_0} - 1)$ and $\frac{D}{\hat{V}_1}$, in Figure 2 and 6 respectively. We fix other parameters at the following set of values,

$$V_0 = 1; \quad \mu = 1.125\%; \quad \sigma = 5\%; \quad \delta = 5\%; \quad r = 6\%; \quad \theta = 35\%; \quad \alpha = 30\%;$$
$$D = 1.28; \quad C = 0.0787. \tag{7}$$

Note that Equation (7) implies that the coupon rate of the debt is $C/D = 6.15\%$ per annum.

Likewise, to plot Figure 4, we vary μ and a, fixing other parameters at the following set of values,

$$V_0 = 1; \quad \hat{V}_1 = 1; \quad \sigma = 5\%; \quad \delta = 5\%; \quad r = 6\%; \quad \theta = 35\%; \quad \alpha = 30\%;$$
$$D = 1.28; \quad C = 0.0787. \tag{8}$$

To plot Figure 5, we vary σ and a, fixing other parameters at the following set of values,

$$V_0 = 1; \quad \hat{V}_1 = 1; \quad \mu = 1.125\%; \quad \delta = 5\%; \quad r = 6\%; \quad \theta = 35\%; \quad \alpha = 30\%;$$
$$D = 1.28; \quad C = 0.0787. \tag{9}$$

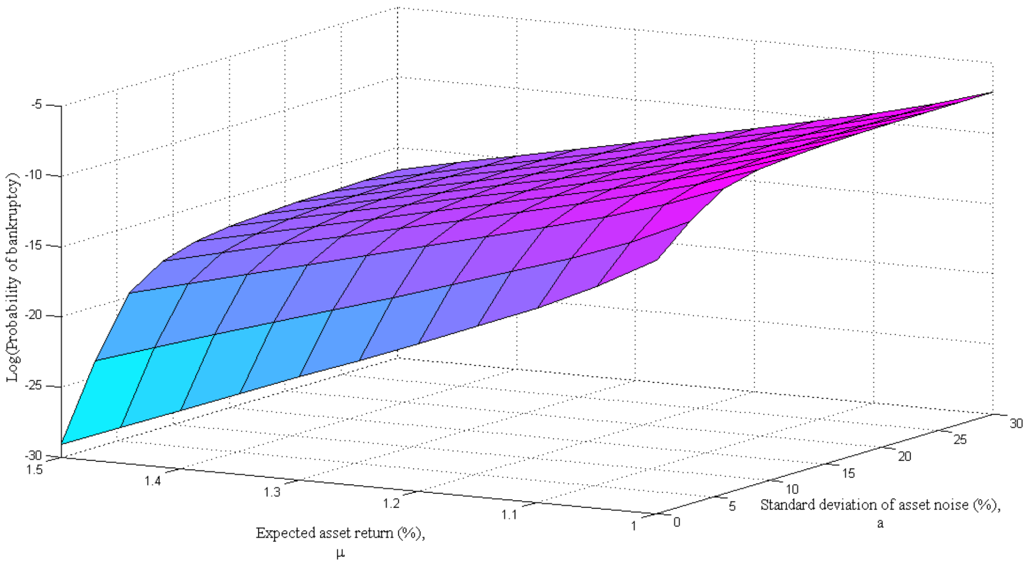


Figure 4: Theoretical log(probability of bankruptcy), varying the degree of noise and expected growth rate of assets

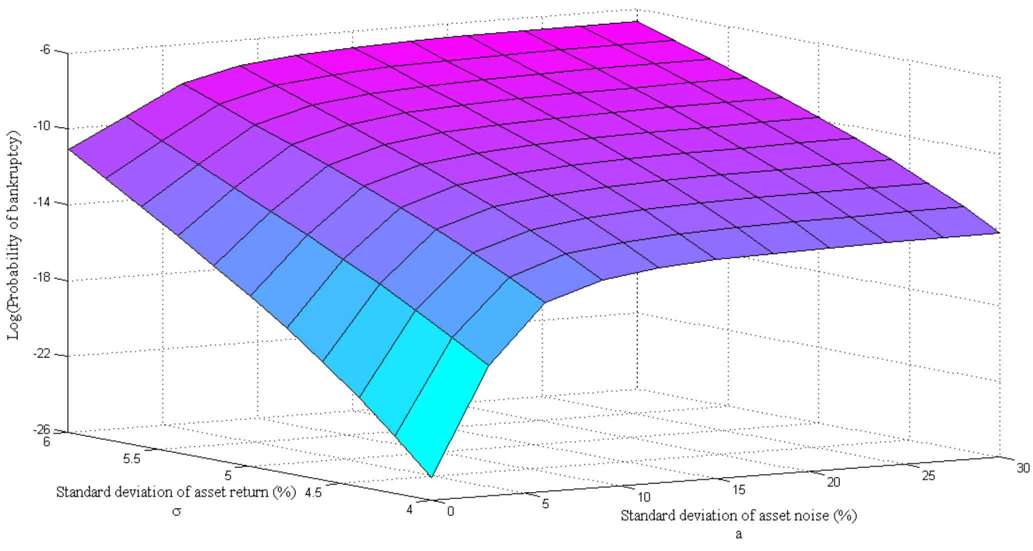


Figure 5: Theoretical log(probability of bankruptcy), varying the degree of noise and volatility of asset growth rate

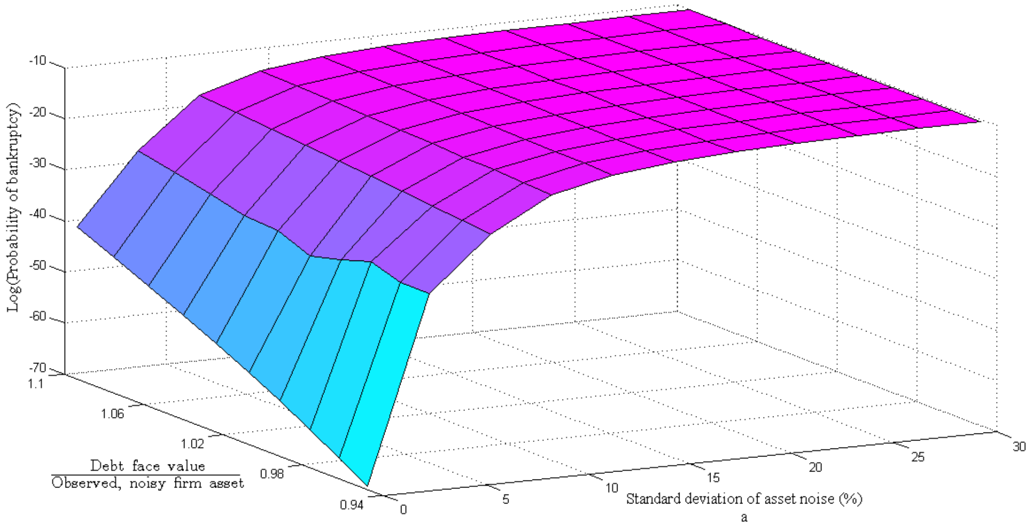


Figure 6: Theoretical log(probability of bankruptcy), varying the degree of noise and normalized debt face value (normalized by observed, noisy assets)

References

Allison, P.D., 2010, *Survival Analysis Using SAS®: A Practical Guide*, Second Edition, SAS Institute Inc.

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

Beaver, W.H., 1966, Financial ratios as predictors of failure, *Journal of Accounting Research*, 4, 71–111.

Beaver, W.H., McNichols, M.F. and Rhie, J-W, 2005, Have financial statements become less informative? evidence from the ability of financial ratios to predict bankruptcy, *Review of Accounting Studies*, 10, 93–122.

Beaver, W.H., Correia, M. and McNichols, M.F., 2012, Do differences in financial reporting attributes impair the predictive ability of financial ratios for bankruptcy?, *Review of Accounting Studies*, 17, 969–1010.

Bharath, S.T. and Shumway, T., 2008, Forecasting default with the Merton Distance to Default model, *The Review of Financial Studies*, 21, 1339–1369.

Cai, J., Saunders, A. and Steffen, S., 2012, Syndication, interconnectedness, and systemic risk, Working paper, SSRN.

Campbell, J.Y., Hilscher, J. and Szilagyi, J., 2008, In search of distress risk, *The Journal of Finance*, 63, 2899–2939.

Campbell, J.Y., Hilscher, J. and Szilagyi, J., 2011, Predicting financial distress and the performance of distressed stocks, *Journal of Investment Management*, 9, 1–21.

Chava, S. and Jarrow, R.A., 2004, Bankruptcy prediction with industry effects, *Review of Finance*, 8, 537–569.

Chava, S., Stefanescu, C. and Turnbull, S.M., 2011, Modeling the loss distribution, *Management Science*, 57, 1267–1287.

Cox, D.R., 1972, Regression models and life tables (with discussion), *Journal of the Royal Statistical Society*, Series B 34, 187–220.

DeLong, E.R., DeLong, D.M. and Clarke-Pearson, D.L., 1988, Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach, *Biometrics*, 44, 837–845.

Duan, J-C, Sun, J. and Wang, T., 2012, Multiperiod corporate default prediction – a forward intensity approach, *Journal of Econometrics*, 170, 191–209.

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

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

Engelmann, B., Hayden, E. and Tasche, D., 2003, Measuring the discriminative power of rating systems, Discussion paper, Deutsche Bundesbank.

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

Giesecke, K., 2006, Default and information, *Journal of Economic Dynamics and Control*, 30, 2281–2303.

Giesecke, K., Longstaff, F.A., Schaefer, S. and Strebulaev, I., 2011, Corporate bond default risk: A 150-year perspective, *Journal of Financial Economics*, 102, 233–250.

Guo, L. and Masulis, R., 2012, Information quality and CEO turnover, Working paper, UNSW.

Guo, X., Jarrow, R.A. and Zeng, Y., 2009, Credit risk models with incomplete information, *Mathematics of Operations Research*, 34, 320–332.

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

Leland, H.E. and Toft, K.B., 1996, Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads, *The Journal of Finance*, 51, 987–1019.

Lin, C., Ma, Y. and Xuan, Y., 2011, Ownership structure and financial constraints: evidence from a structural estimation, *Journal of Financial Economics*, 102, 416–431.

Lin, D.Y. and Wei, L.J., 1989, The robust inference for the cox proportional hazards model, *Journal of the American Statistical Association*, 84, 1074–1078.

Maffett, M., Owens, E. and Srinivasan, A., 2013, Default prediction around the world: The effect of constraints on pessimistic trading, Working paper, SSRN.

Ng, J. and Rusticus, T.O., 2013, Banks' survival during the financial crisis: The role of financial reporting transparency, Working paper, SSRN.

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

Tang, Y., Subrahmanyam, M. and Wang, Q., 2012, Does the tail wag the dog? the effect of credit default swaps on credit risk, Working paper, NYU.

Thomas, S., 2002, Firm diversification and asymmetric information: evidence from analysts' forecasts and earnings announcements, *Journal of Financial Economics*, 64, 373–396.

Zhang, X., 2006, Information uncertainty and stock returns, *The Journal of Finance*, 61, 105–137.