A Discrete Transformation Survival Model with Application to Default Probability Prediction

A. Adam Ding, Shaonan Tian, Yan Yu and Hui Guo *

Abstract

Corporate bankruptcy prediction has received paramount interest in academic research, business practice and government regulation. Consequently, accurate default probability prediction is extremely important. We propose to apply a discrete transformation family of survival analysis to corporate default risk predictions. The new transformation family is shown to include the popular Shumway’s model and Cox proportional hazards model. We show that a transformation parameter different from those two models is needed for default prediction using the bankruptcy data. In addition, out-of-sample validation statistics show improved performance. The estimated default probability is further used to examine a popular asset pricing question that whether the default risk has carried a premium. Due to some distinct features of bankruptcy application, the proposed discrete transformation survival model with time-varying covariates is fundamentally different from the continuous survival models in the literature. Their links and differences are also discussed.

Key Words: Corporate Bankruptcy Prediction; Credit Risk; Logistic regression; Proportional Hazards; Survival analysis.

* A. Adam Ding is an Associate Professor of Mathematics at Northeastern University; email: a.ding@neu.edu. Shaonan Tian is a PhD student of and Yan Yu is an Associate Professor of Business Analytics at the University of Cincinnati; email: tiansn@mail.uc.edu; Yan.Yu@uc.edu. Hui Guo is an Associate Professor of Finance at the University of Cincinnati, Cincinnati, OH, 45221.
1 Introduction

Corporate bankruptcy has long been one of the most significant threats for many businesses. It not only increases the financial loss to its creditors but also has a negative impact to the society and the aggregate economy. More alarmingly, the data released by the Administrative Office of the U.S. courts shows that in the recent decades business failures have occurred at higher rates than at any time since the early 1930’s. The default loss has also maintained at an startling level of trillions of dollars.

Accurate default probability prediction is of great interest for all academics, practitioners and regulators. Corporate default forecasting models are used by regulators to monitor the financial health of banks, funds, and other institutions. Practitioners use probability default forecasts in conjunction with models to price corporate debt and for internal rating based approach (Schönbucher 2003; Lando 2004). Academics use bankruptcy forecasts to test various conjectures such as the hypothesis that default risk is priced in stock return (Campbell, Hilscher and Szilagyi 2008). Given recent economic condition, the importance of accurate default predictive model validation is even more substantially promoted by the Basel Committee on Banking Supervision under the current framework of Basel II 1.

Despite vast literature on bankruptcy prediction (see Altman 1993 for a survey), most research prior to past decade are concentrated on static modeling using cross-sectional data (e.g. Altman 1968; Ohlson 1980; Zmijewski 1984). Though multi-period firm characteristics are observed, prior researchers only choose to use one period observation with a single-period logistic regression or discriminant analysis.

On the other hand, the event of default can be considered as a terminal event for the company. This is mathematically equivalent to the death event in the survival analysis which has generated a huge body of literature. From the survival analysis point of view, predicting the time to default based on the various measurable financial and market variables at current time naturally corresponds to analyze covariate effects on the survival time. In continuous time survival analysis, the proportional hazards model (Cox 1972) is the most

---

1Basel II is an international business standard that requires financial institutions to maintain enough cash reserves to cover risks incurred by operations.

While there is a natural correspondence between survival analysis and the default risk modeling, such a link was not explored in the literature until Shumway (2001) proposed a discrete hazard model. The conditional default probability $\pi_{i,k}$ that the $i$th firm files bankruptcy at time $t_k$ given it survives past time $t_{k-1}$ is modeled through a multi-period logistic regression by the $i$th firm’s specific characteristics $Z_i(t_k)$ at time $t_k$,

$$\pi_{i,k} = \frac{1}{1 + \exp(-\alpha - \beta^T Z_i(t_k))}.$$  

Here the time-varying covariate values $Z_i(t_k)$ are usually firm’s financial ratios obtained from accounting statements and firm’s market variables from public trading record; $\beta$ is the covariate effect parameter and $\alpha$ is a scalar parameter. Not surprisingly, Shumway (2001) also shows that hazard, or survival modeling, is advantageous by coping with time-varying panel data, while static model ignores the fact that firms change through time and may produce biased and inconsistent bankruptcy probability estimates. This discrete hazard model quickly gains popularity in corporate bankruptcy prediction and is used in Chava and Jarrow (2004); Bharath and Shumway (2008); Campbell et al. (2008). The Cox proportional hazards model was also used for bankruptcy prediction in Duffie, Saita and Wang (2007); Duffie, Eckner, Horel, and Saita (2009).
In this paper, we propose a new discrete-time transformation model family to bankruptcy probability prediction with time-varying covariates

\[
\pi_{i,k} = \begin{cases} 
1 - \frac{1}{[1+c \exp(\alpha_k + \beta Z_i(t_k))]^{1/c}}, & c > 0; \\
1 - \exp[-\exp(\alpha_k + \beta Z_i(t_k))], & c = 0.
\end{cases}
\]  

(1)

or

\[
\pi_{i,k} = \begin{cases} 
1 - \frac{1}{1+\rho \exp(\alpha_k + \beta Z_i(t_k))}, & \rho > 0; \\
\frac{1}{1+\exp(\alpha_k + \beta Z_i(t_k))}, & \rho = 0.
\end{cases}
\]  

(2)

Here \(c\) (or \(\rho\)) are scalar transformation parameters. The proposed discrete-time transformation models are derived formally in Section 3 by applying a monotonic transformation on difference of the minus log survival functions. This transformation model family contains the Shumway (2001) model when \(c = 1\) or \(\rho = 0\) and Cox proportional hazards model (Cox 1972) when \(c = 0\) or \(\rho = 1\) as special cases.

The estimated covariate effect parameter \(\beta\) can be used to rank companies’ default risk based on their covariate values \(Z_i(t_k)\): companies with higher \(\beta Z_i(t_k)\) values would have higher default risk at given time \(t_k\). Thus \(\beta Z_i(t_k)\) can be considered as a credit score like those given out by the rating agencies such as Moody’s and S&P. Abundant literature and ongoing research are dedicated to obtain good credit scores. However, actual default probability is needed to assess the portfolio risk for calculate banking reserves as in Basel II. Actual default probability is also essential to combine with the loss given default measure (Schuermann 2005).

Figure 1 shows the default probability curves for different transformation parameter values. We can see that same scores would correspond to different default probabilities under different transformation parameters. In our empirical analysis, we apply the proposed discrete-time transformation model family on a comprehensive bankruptcy data set spanning from 1981 to 2006. Log-likelihood plot of the fit on quarterly firm observations show that the optimal transformation parameter resides near the point \(c = 10\) which is neither Shumway’s model nor Cox proportional hazards model.

[Insert Figure 1]

[Insert Figure 2]
In addition, out-of-sample prediction with withholding period 2002-2006 shows improved accuracy ratio and model goodness-of-fit. We also investigate the asset pricing implication conjectured by Fama and French (1996) that investors require a positive return premium for holding stocks with high default probabilities. We sort stocks into portfolios by the predicted default probability using the proposed discrete transformation survival model. We find that stocks with higher default probabilities deliver anomalously lower returns, which challenges the original Fama and French’s (1996) conjecture. Our findings, however, are consistent to those documented by some recent literature such as Campbell et al. (2008).

The original motivation of our proposed discrete transformation models comes from the continuous time generalized odds-rate model of Dabrowska and Doskum (1988) with time-invarying covariate and Zeng and Lin (2006, 2007) for time-varying covariate. However our proposed models are fundamentally different from those models. In Section 3 we show in detail that our transformation is on the difference of the minus log survival functions, while transformation of Zeng and Lin (2006) depends on the entire history of covariate values. Simple discrete extension of Zeng and Lin (2006) is ill-defined, due to some unique features for bankruptcy prediction application. In particular: Firstly, actual calendar time needs to be used. This is because same firm specific characteristics at different calendar time may expose to different default probability due to different macroeconomic conditions. Secondly, differing from the classical survival model set-up, companies from the bankruptcy database do not share a common starting point. This is due to the use of calendar time (see discussion in Section 3.4). Data for most firms start from the beginning of the sample observation period despite that they have prior accounting statements and trading activities. On the other hand, a number of firms enter in the middle of the sample period since they have just started public trading. Hence, the transformation model of Zeng and Lin (2006), depending on the entire history of covariates from the same starting point, is not well-defined. Thirdly, the proposed discrete-time transformation model enjoys the same appealing “memoryless feature” as Shumway (2001). That is, the conditional default probability only depends on the last available observation, instead of the whole path of covariates as in Zeng and Lin (2006). The first two features of bankruptcy prediction imply that a practical model needs
to be memoryless. Finally, by nature, the accounting and market information used in the default prediction study are collected only at discrete time period over a fixed time window. For example, accounting data from sample period 1981-2006 are obtained for our study from the quarter end balance sheet, income statement and cash flow report.

The rest of paper is organized as following. Section 2 describes the bankruptcy data we use in the study. Section 3 presents the discrete time transformation survival model and links to continuous time survival analysis. Empirical results of the corporate bankruptcy application are given in Section 4. The Appendixes give detailed mathematical derivations.

2 Bankruptcy Data

In our study, we develop a comprehensive bankruptcy database by merging The Center for Research in Security Prices (CRSP) with Compustat from Standard & Poor’s (COMPUS-TAT) database through Wharton Research Data Services (WRDS)\(^2\). The CRSP database provides a complete collection of security data including price, return, and volume data for the three major stock exchange markets: NYSE, AMEX, and NASDAQ. COMPUSTAT maintains quarterly accounting information for companies including reports of Income Statement, Balance Sheet, and Statement of Cash Flows etc. Our bankruptcy database includes all the publicly traded companies in the United States between 1981 and 2006. To measure the probability of default using our proposed transformation model, we need a set of exploratory variables and an event indicator of bankruptcy for default companies. In our empirical study, we define a firm as default if it files under either Chapter 7 or Chapter 11 bankruptcy code. Because it usually takes a long time to settle bankruptcy disputes, in some cases, the COMPUSTAT updates the default status with a substantial delay. The delay makes it difficult to identify accurately the corporate default in the most recent period. To address this issue, we end our sampling period in the year 2006 and restrict our sampling time window from 1981 to 2006. In addition, eight covariate measures are constructed for the exploratory variables: profitability, leverage, short-term liquidity, the market-to-book ratio (MB), volatility and excess return over the S&P 500 index return, as well as the firm’s

\(^2\)website: http://wrds-web.wharton.upenn.edu/wrds/
relative size to the S&P 500 index value and the price. The formation of this set of covariates follows Campbell et al. (2008) closely.

Note that a firm may exit the database anytime due to its financial healthy status, it may also enter the database in different time periods. For most healthy firms, their Initial Public Offering (IPO) date was prior to the year of 1981. In this case, the firms enter our database coincidentally with the start of our sampling period. On the other hand, there may be firms with IPO date after 1981. This case is particularly common during the “dot-com bubble” period in the late 1990s and early 2000. Under such situation, we record the first observation of the firm at the time of its first trading date. For example, the IPO date for the IT company “Microsoft Corporation” was March 1986. Therefore, the firm “Microsoft Corporation” enters our database in the year of 1986, instead of the starting time of our sampling period, 1981. Such property of having “different starting time” is unique for bankruptcy data, and therefore, needs special treatment when linking with the survival models.

3 Model

3.1 A Discrete Time Transformation Survival Model

Financial data are discrete in nature. For example, commonly used predictors such as firms’ financial ratios are obtained through accounting statements quarterly. Hence, a discrete time model is needed for corporate bankruptcy prediction.

Suppose there are $K$ fixed observation time $t = t_1, t_2, ..., t_K$ for the whole observation period. For example, these are the quarter end date for quarterly data. Total $i = 1, 2, ..., n$ public firms are in the data base during the sample period, each with observed data $(B_i, X_i, \Delta_i, Z_{i,k}), k = 1, ..., K$. Here $B_i$ denotes the starting time – the first time the firm is publicly traded during the observation period. If a firm in the data base is traded prior to $t_1$, then $B_i = t_1$. Here $X_i$ denotes the last time the firm is observed during the observation period. It is subjected to right censoring by the end of observational period. If a firm files bankruptcy after $t_K$, then $X_i = t_K$. $Z_{i,k} = Z_i(t_k)$ is the $d$-dimensional covariate vectors for firm $i$ at time $t = t_k$. $\Delta_i$ is the so called censoring indicator in survival analysis:
\( \Delta_i = 1 \) if the \( i \)th firm enters bankruptcy filing process at \( t = X_i \); \( \Delta_i = 0 \) otherwise. A healthy firm may experience early exit from the data base, such as merger or acquisition. In those cases, \( X_i < t_K \) but \( \Delta_i = 0 \).

Denote \( T \) the survival time and \( S_Z(t) = Pr(T > t \mid Z = z) \) the survival function given \( Z \). Here \( Z \) is the covariate process over the whole time period. And denote

\[
\pi_{i,k} = Pr(T = t_k \mid X \geq t_k, Z(t_k) = Z_{i,k})
\]

the conditional probability that the firm files bankruptcy at time \( t_k \) given it is at risk at time \( t_k \) (survival past time \( t_{k-1} \)).

Formally, let

\[
G[- \log \frac{S_Z(t_k)}{S_Z(t_{k-1})}] = \exp[\beta^T Z(t_k)]G[- \log \frac{S_0(t_k)}{S_0(t_{k-1})}],
\]

where \( G \) is a strictly increasing transformation function with \( G(0) = 0 \) and \( G(\infty) = \infty \); \( \beta \) is a \( d \)-dimensional covariate effect parameter; \( S_0(\cdot) \) is the (baseline) survival function when \( Z \equiv 0 \).

Our discrete time transformation survival model then takes the form in (1)

\[
\pi_{i,k} = \begin{cases} 1 - \frac{1}{1+c \exp(\alpha_k + \beta^T Z_{i,k})^{1/c}}, & c > 0; \\ 1 - \exp[- \exp(\alpha_k + \beta^T Z_{i,k})], & c = 0. \end{cases}
\]

for transformation function \( G_c \) belongs to family

\[
G_c(x) = \begin{cases} \frac{1}{c}[\exp(cx) - 1], & c > 0; \\ x, & c = 0. \end{cases}
\]

Or it takes the form in (2)

\[
\pi_{i,k} = \begin{cases} 1 - \frac{1}{1+\rho \exp(\alpha_k + \beta^T Z_{i,k})^{1/\rho}}, & \rho > 0; \\ 1+\exp(\alpha_k + \beta^T Z_{i,k}), & \rho = 0. \end{cases}
\]

for transformation function \( G_\rho \) belongs to family

\[
G_\rho(x) = \begin{cases} \frac{1}{\rho}\log[1 + \rho(\exp(x) - 1)], & \rho > 0; \\ \exp(x) - 1, & \rho = 0. \end{cases}
\]

Appendix 1 gives detailed derivations of the discrete time transformation survival models (1) and (2) based on (3) and monotonic transformations (4) and (5). Here \( G_c \) and \( G_\rho \) are
common monotonic transformations used in the survival analysis literature. Inverse of \( G_c \) is equivalent to a class of logarithmic transformations considered in Chen et al. (2002). Inverse of \( G_\rho \) is similar to the class of Box-Cox transformations.

Note when \( c = 0 \) or \( \rho = 1 \), the proposed discrete time transformation survival models are equivalent to the classical Cox proportional hazard model. When \( c = 1 \) or \( \rho = 0 \), these are the popular so-called discrete hazard model proposed in Shumway (2001) and is then followed by most bankruptcy prediction literature such as Chava and Jarrow (2004), Campbell et al. (2008) etc. Our transformation model would estimate the transformation parameter \( c \) (or \( \rho \)) in addition to parameters \( \alpha_k \) and \( \beta \).

### 3.2 Estimation and Algorithm

The likelihood function for the proposed discrete transformation survival model is

\[
L = \prod_{i=1}^{n} \prod_{k: B_i < t_k \leq X_i} \pi_{i,k}^{\Delta_{i,k}} (1 - \pi_{i,k})^{1 - \Delta_{i,k}},
\]

where \( \Delta_{i,k} = \Delta_i I\{X_i = t_k\} \) and \( \pi_{i,k} = Pr(T = t_k | X \geq t_k, Z(t_k) = Z_{i,k}) \) is modeled by (1) and (2). The log-likelihood function is then

\[
l = \sum_{i=1}^{n} \sum_{k: B_i < t_k \leq X_i} \Delta_{i,k} \log(\pi_{i,k}) + (1 - \Delta_{i,k}) \log(1 - \pi_{i,k}).
\]

Therefore, for a fixed \( c \) (or \( \rho \)) value, the fitting of \( \alpha_k \)'s and \( \beta \) can be implemented mathematically using logistic regression with a specified link function on independent \( \Delta_{i,k} \)'s even though \( \Delta_{i,k} \)'s are dependent in the data set. When the model assumption of Cox proportional hazards model or Shumway (2001) model holds, our estimates would give about the same results. Our transformation model would provide better fit when the transformation parameters are not close to 0 or 1.

For estimation, we could maximize (7) with (1) or (2) over the transformation parameter \( c \) (or \( \rho \)), covariate effect parameter \( \beta \) and \( \alpha_k \)'s simultaneously on the data set. Instead, in our work, we compute the parametric maximum likelihood estimator for the covariate effect parameter \( \beta \) and \( \alpha_k \) on a grid search over the transformation parameter \( c \) (or \( \rho \)). At each point value of \( c \) (or \( \rho \)) values, likelihood given by (7) is maximized over the parameter \( \beta \) and
This is a standard nonlinear optimization procedure. Many scientific optimization packages can conduct the maximization computation. In our numerical work, we use `lsqnonlin()` function in MATLAB optimization toolbox. Other numerical optimization methods such as Newton-Raphson algorithm can also be applied.

When the size of the parameter set expands, the standard nonlinear optimization procedure may not be computationally efficient. Alternatively, we may use profile-likelihood method (Murphy and van der Vaart, 2000) to estimate. Our algorithm is as below. For a given $c$ (or $\rho$) over a fixed grid window,

- **Step 0.** Initialize parameter $\hat{\beta}^{(0)}$ and $\hat{\alpha}^{(0)}_k$ for $k = 1, 2, ..., K$. Sensible initial values, such as a vector of 0 or 1 can be used. Alternatively, we may use the parameter estimated through Shumway’s multi-period logistic regression as an initial estimate.

- **Step 1.** Given estimated $\hat{\beta}^{(j)}$, obtain $\hat{\alpha}^{(j+1)}_k$ for $k = 1, 2, ..., K$. Each $\alpha_k$ is estimated only on the set of firm observations at time $t_k$ for $k = 1, 2, ..., K$. Hence, the likelihood given by equation (7) is maximized over one parameter for each time period. This can be done almost instantaneously.

- **Step 2.** Given estimated $\hat{\alpha}^{(j)}_k$ for $k = 1, 2, ..., K$, estimate $\hat{\beta}^{(j+1)}$ on the entire data set. This step involves the entire dataset to conduct the optimization procedure. However, the dimension of covariate parameter $\beta$ is relatively small so that the convergence is quite fast.

Then iterate step 1 and step 2 until convergence.

Note that the optimization algorithm does not guarantee a global maximum value on likelihood. To hedge the local minimum situation, we experiment on different initial values of $\beta$ and $\alpha_k$s. Our experience shows that the sensible initial values, such as a vector of 0 or 1 may easily lead to a satisfactory performance.

**Remark 1:** The proposed transformation model on the bankruptcy prediction has three parametric components: $d$-dimensional covariate effects parameter $\beta$, $K$-dimensional baseline parameter $\alpha_k = \alpha(t_k)$ and the scalar transformation parameter $c$ (or $\rho$). This is a usual
parametric problem set-up. Hence, under some mild conditions, large sample properties for estimates of our proposed discrete transformation models (1) or (2) can be easily established following the standard parametric maximum likelihood estimators with total fixed $d + K + 1$ parameters to estimate. Root-n consistency and asymptotic normality are readily available for further inference.

### 3.3 Link to Continuous Time Survival Models

The proposed discrete transformation models are naturally motivated from the continuous generalized odds-rate model of Dabrowska and Doskum (1988a) with time-invarying covariate and Zeng and Lin (2006) for time-varying covariate $Z(t)$. However, the proposed models are fundamentally different due to the nature of bankruptcy application.

Most importantly, our transformation $G$ is on $\left[ -\log \frac{S_Z(t_k)}{S_Z(t_{k-1})} \right]$ or equivalently on the difference of minus log of survival functions. Equation (3) can be rewritten as

$$ G[-\log S_Z(t_k) - (-\log S_Z(t_{k-1}))] = \exp[\beta^T Z(t_k)]G[-\log S_Z(t_k) - (-\log S_Z(t_{k-1}))]. \tag{8} $$

Instead, the transformation of Zeng and Lin (2006) is on the cumulative hazard function $\Lambda_Z(t) = -\log S_Z(t)$

$$ G(\Lambda_Z(t)) = \int_0^t e^{\beta^T Z(s)}d\Lambda(s), $$

which depends on the entire history of covariate values $\{Z(s)\}_{s=0}^t$. Here $\Lambda(.)$ is an unspecified increasing function.

After some calculation (see Appendix 2), Zeng and Lin (2006)’s model can be reexpressed by

$$ \frac{d}{dt} G[\Lambda_Z(t)] = \exp[\beta^T Z(t)] \frac{d}{dt} G[\Lambda_0(t)]. \tag{9} $$

Equation (8) can be regarded as taking the difference of $-\log S_Z(t_k)$ first and then take transformation $G$ on the difference. Equation (9) applies the transformation $G$ on $\Lambda_Z(t) = -\log S_Z(t_k)$ first before taking difference. Again note that (8) and (9) are fundamentally different transformation models.

Zeng and Lin (2006) can not be extended similarly to the bankruptcy prediction application. When the covariate $Z(t)$ is time-varying, their likelihood function depends on the
values of the covariate process \( Z(s) \) for all time \( s \) between \( s = 0 \) and \( s = X_i \). Therefore their nonparametric maximum likelihood estimator can not be found unless we observe the whole covariate process \( Z(s) \) between \( s = 0 \) and \( s = X_i \). For a new public company starts at time \( B_i > 0 \), certainly the covariate process \( Z(s) \) does not exist for the time interval between \( s = 0 \) and \( s = B_i \). That is, \( \Lambda_Z(B_i) \) is unknown as long as there is covariate effect and the time-varying covariate is not known for \( t < B_i \). Hence, simple discrete extension of Zeng and Lin (2006) is ill-defined for bankruptcy prediction application. By taking difference first as in equation (8), our approach yields measures that only depend on covariate values at one last period but not on the whole past history of covariates.

Note when \( Z \) is time invariant covariates, the generalized odds-rate model of Dabrowska and Doskum (1988a) has similar transformation identity as equation (9) since

\[
G[\Lambda_Z(t)] = \exp[\beta^T Z] G[\Lambda_0(t)],
\]

where \( \Lambda_Z(t) \) is their cumulative hazard function in continuous time.

Transformation \( G_c(\cdot) \) of (4) is commonly used in continuous time transformation survival models (e.g. Dabrowska and Doskum, 1988; Chen et al. 2002; Zeng and Lin, 2006). The generalized odds-rate model family include the classical Cox proportional hazard model \( (c = 0) \) and proportional odds model \( (c = 1) \) as special cases. The Cox proportional hazard model corresponds to \( c = 0 \), a boundary point on the parameter space, for the family \( G_c \). Transformation \( G_\rho(\cdot) \) of (5) yields alternative families where the Cox proportional hazards model corresponds to an interior point of the parameter space with Cox proportional hazards model \( (\rho = 1) \) and proportional odds model \( (\rho = 0) \) as special cases.

3.4 Discussion

Doksum and Gasko (1990) pointed out a correspondence between logistic regression models in binary regression analysis and the generalized odds-ratio model in survival analysis with time-invariant covariate. The binary regression is done for the indicator variable on survival past a fixed time point. In section 2.1, we similarly show that the discrete transformation (3) with time-varying covariate in survival analysis corresponds to logistic regressions (1) or (2)
on indicators for survival in each discrete time period as if those indicators were independent. This new correspondence links the survival analysis model techniques to the application of bankruptcy probability modeling.

Literature on survival analysis were extensive in biomedical fields. Although it is natural to model the bankruptcy of a company as a survival event, the advanced survival analysis model results have rarely been applied to the bankruptcy prediction. There is a technical reason for this. In most survival analysis theory, the focus is on the survival probability curves from a common starting time \( t = 0 \), generally a clinic event such as the beginning of a medical treatment or diagnose of the disease. That is, the time used is not the actual calendar time. The literature on modeling of bankruptcy prediction, in contrast, generally are done with calendar time. It is important to understand the reason for the difference. In the biomedical applications, it is reasonable to expect that individuals with same covariate values (physical characteristics, medical treatments, etc.) will follow the same biological process after, say, a surgical operation. Therefore, those patients should have same probability to survive after one year, regardless the operation was done in calendar year 1985 or year 1990. However, we would not expect companies with same covariate values (financial characteristics) to have same one year bankruptcy probability in calendar year 1985 as in calendar year 1990. This is due to the different macroeconomics environment in 1985 versus 1990. With the actual calendar time, we can no longer ensure a common starting time \( t = 0 \) for all individuals as companies are not all started at the same time. The application of survival analysis models to the bankruptcy prediction needs to take this difference into account. For example, the transformation model (9) is not well-defined unless covariate process \( Z(t) \) exists from the common starting time. So it can not be applied directly to the bankruptcy prediction problem. We propose the alternative transformation model (3) instead.

4 Empirical Results

4.1 The Data

To measure the distress risk using our proposed transformation survival model, we need an event indicator of bankruptcy for the distressed firms and a set of exploratory variables.
In our work, we classify a firm as in distress if it files under either Chapter 7 or Chapter 11 bankruptcy protection code. To obtain the list of distressed firms, we take the firms as default if its reported deletion reason is liquidation or bankruptcy by COMPUSTAT or it was delisted from CRSP due to the same reason. As such, our bankruptcy database includes 1565 firms that file bankruptcy from January 1981 to December 2006. The default indicator equals to one in the month that firm was delisted due to Chapter 7 or Chapter 11 bankruptcy filing. All the other exiting reasons such as merger or acquisition would set the default indicators to zero.

Table 1 reports the properties of firm defaults by year in more details. We note that the default rate exhibits substantial variation across time. The pattern reflects mainly the fact that firms have more difficulties in fulfilling their financial obligations during business recessions than during business expansions. Specifically, in our sample, the default rate peaks in the year 1991, when the economy was in a recession accompanied by a severe credit crunch due to monetary tightening (e.g., Bernanke, Lown, and Friedman, 1991). Similarly, the default rate is at an elevated level in the year 2001, when the economic fell into recession following the burst of the technology bubble. Such time-varying default rate highlights the importance of taking into account the distinct macro-economic effect featured by different calendar time on default probability, as in our proposed survival model.

Table 1 reports the properties of firm defaults by year in more details. We note that the default rate exhibits substantial variation across time. The pattern reflects mainly the fact that firms have more difficulties in fulfilling their financial obligations during business recessions than during business expansions. Specifically, in our sample, the default rate peaks in the year 1991, when the economy was in a recession accompanied by a severe credit crunch due to monetary tightening (e.g., Bernanke, Lown, and Friedman, 1991). Similarly, the default rate is at an elevated level in the year 2001, when the economic fell into recession following the burst of the technology bubble. Such time-varying default rate highlights the importance of taking into account the distinct macro-economic effect featured by different calendar time on default probability, as in our proposed survival model.

Table 1 reports the properties of firm defaults by year in more details. We note that the default rate exhibits substantial variation across time. The pattern reflects mainly the fact that firms have more difficulties in fulfilling their financial obligations during business recessions than during business expansions. Specifically, in our sample, the default rate peaks in the year 1991, when the economy was in a recession accompanied by a severe credit crunch due to monetary tightening (e.g., Bernanke, Lown, and Friedman, 1991). Similarly, the default rate is at an elevated level in the year 2001, when the economic fell into recession following the burst of the technology bubble. Such time-varying default rate highlights the importance of taking into account the distinct macro-economic effect featured by different calendar time on default probability, as in our proposed survival model.

To construct the exploratory variables for our model, we merge the daily and monthly equity data from CRSP with the quarterly and annually updated accounting data from COMPUSTAT. We adopt eight covariates as in Campbell et al. (2008), which has been considered to be the state-of-art model in the bankruptcy literature. Furthermore, in a separate working paper on dynamic variable selection (XXX, 2011), we also confirm that these variables remain significant in explaining default probability. Among these eight covariates, three are accounting ratio measures. Profitability (NIMTA) is calculated by dividing the net income by the market value of the total asset. Leverage (TLMTA) is calculated by dividing total liability by the market value of the total asset. Short-term liquidity (CASHMTA) is
calculated by dividing the cash and short-term asset by the market value of the total asset. Here the market value of the total asset is the sum of the firm market equity and its book liability.

We follow Daniel and Titman (2006) to determine the firm book equity in market-to-book ratio (MB) calculation. To avoid the outliers of the book equity, we adjust the book equity by adding 10% of the difference between the book equity and the market equity. Four market variables based on the equity information are excess return (EXRET), firm’s market capitalization or relative size (RSIZE), volatility (SIGMA) and stock price (PRICE). We use the standard deviation of the daily stock price over the previous three month to estimate the volatility. Both the excess return and the firm’s relative size are evaluated over the S&P 500 index as the market value. Except for volatility, the other three market measures including the excess return, the firm’s relative size, and the price enter the estimation model in its log scale.

To be consistent with previous literature, we further modify our data set in the following ways. Companies report their accounting data with a delay. To ensure we use the accounting information that is available at the time of the forecast, we lag all the annually updated accounting measures, including net income, total liability and the cash and short-term assets by four months and quarterly updated accounting data by two months. All the covariate measures are constructed after such lagging on the accounting data. Great care has also been taken in aligning the fiscal time with the calendar time.

Table 2 reports the summary statistics of the eight covariate measures. Panel A gives the summary statistics on the entire 1,812,730 firm-month observations. Panel B summarizes the eight covariates’ statistical property on the default group only, a total of 1,565 observations. Comparison between Panel A and Panel B demonstrates strong distribution differences among the eight covariates between the entire data set and the default group.

[Insert Table 2]
4.2 Results

We apply the proposed discrete transformation survival models (1) and (2) to the quarterly updated bankruptcy data from 1981 to 2006 using the two-step profile likelihood algorithm described in Section 3.2. Figure 2 gives the model maximum log-likelihood values over a grid of transformation parameters $c$ and $\rho$. From Figure 2, we observe that the log-likelihood decreases monotonically over $\rho$ so that the optimal model resides in the $G_c$ transformation. The log-likelihood increases sharply from $c = 0$ and stabilizes around the interval from $c = 10$ to $c = 12$ with little variation. The maximum is achieved around the point of $c = 11.5$. Clearly neither Shumway’s model ($c = 0$ or $\rho = 1$) nor Cox proportional hazards ($c = 1$ or $\rho = 0$) is optimal on the bankruptcy probability prediction for this data set.

Similarly, Figure 3 plots the observed values for the log-likelihood function of the fit on the annually updated data set under the transformation functions $G_c$ in (4) and $G_\rho$ in (5). Again, we observe that the log-likelihood value is maximized neither close to the $c = 1$ for Shumway’s model nor $c = 0$ for Cox proportional hazards model, but at around $c = 10$.

[Insert Figure 3]

We further withhold data from 2002 till 2006 for validation purpose and estimate the default probability via expanding window approach. To that end, we predict the probability of default for each validation year on an estimation window from the start of the sampling period up to the forecasting period to eliminate look-ahead bias. For the reason of computational efficiency, we use annually updated firm data. We investigate accuracy ratio\(^3\) (see definition in Duffie et al. 2007) and the Hosmer-Lemeshow goodness-of-fit test (Hosmer and Lemeshow 2000, page 147) statistics on the validation data set by comparing three different models: the optimal transformation model with $c = 10$, Shumway’s model $c = 1$ and Cox proportional hazards model $c = 0$.

Table 3 displays our out-of-sample performance evaluation results. We find our optimal transformation model with $c = 10$ yields a slightly better accuracy ratio. The high p-value for the Hosmer-Lemeshow test attests the out-of-sample improvement of our selected model.

\(^3\)Accuracy Ratio is a similar measure as Area Under a ROC Curve (AUC).
in the overall significance.

[Insert Table 3]

**Remark 2:** The prediction ability of a binary response prediction model has two components: (a) discrimination, that is, the ability to discriminate between those subjects experiencing the event of interest and those not; (b) calibration, that is, providing correct prediction probability for event occurrence (Hosmer and Lemeshow 2000). Higher accuracy ratio reflects higher discrimination. The Hosmer-Lemeshow test checks proper calibration. For our default prediction model at time $t_k$, the discrimination is achieved through the scores $\beta^* Z_i(t_k)$. Proper calibration requires good estimation of the transformation parameter $c$ (or $\rho$) and baseline $\alpha_k = \alpha(t_k)$ in addition to good estimation of covariate parameter $\beta$. When calculating the out-of-sample Hosmer-Lemeshow test, we ignored the prediction of $\alpha(t_k)$ here by using its data estimation. The Hosmer-Lemeshow tests show that the standard models of $c = 1$ or $c = 0$ are not properly calibrated for out-of-sample predictions while the selected transformation model of $c = 10$ significantly improves the calibration. Using correct transformation parameter also leads to better estimation of $\beta$ thus improving the discrimination as well, although the improvement in this component is minimal. The out-of-sample predictions of $\alpha(t_k)$ requires further modeling and will be a topic for future research.

### 4.3 Asset Pricing Implication

We further investigate Fama and French’s (1996) conjecture that investors require a positive return premium for holding distressed stocks. As in Campbell et al. (2008), we measure the distress premium by sorting stocks according to their predicted default probabilities, estimated from the selected optimal discrete transformation survival model. In particular, at the beginning of each year from 1985 to 2006, we update the firm’s default probability only using historically available data. We then form 10 portfolios according to their default risk distribution and hold each portfolio for one year. Detailed specification of the cut-off percentile points for such ten portfolios is listed in Table 4. Note that those percentile cut-off points are not equally spaced, but provide finer grids to the tail of the distribution. Table 4 summarizes our findings.
We first investigate the average of simple return\textsuperscript{4} in excess of the S&P 500 index return for each portfolio. From Panel A, we observe stocks with high conditional default probabilities have substantially lower returns than do stocks with low conditional default probabilities. For example, the portfolio with the lowest default probability has an annualized average excess return of 3.29%, compared with the $-12.03\%$ for the portfolio with the highest default probability. Such a finding poses a significant challenge to Fama and French’s (1996) conjecture that distressed stocks have a higher expected return.

We then examine the question “can stock return anomalies be explained by the three-factor model” (Fama and French 1996) by regressing each portfolio’s value-weighted return on the standard Fama and French three factors\textsuperscript{5}. Panel B shows portfolio with high distress risk tends to have high loadings on the size factor (Small Minus Big or SMB) and value risk factor (High Minus Low or HML), yet an anomalously negative alpha. Alpha is the estimated intercept after fitting the value-weighted return on the three factors. A significant alpha suggests that the risk premium is not fully priced by the three factors. In the first row of Panel A, we observe that the reported alpha for the highest default risk portfolio in the 99th to the 100th percentile is $-18.96\%$ with a t-statistics of $-19.62$. The alpha deviates significantly from 0. Therefore, it shows that distress risk estimated from the selected optimal transformation model cannot be fully explained by the commonly used Fama and French’s (1996) three-factor model. Additional risk factor may be needed in order to explain the anomalous stock return. These results confirm the findings in Campbell et al. (2008), where a discrete hazard model with $c = 1$ is used.

Panel C shows the average of the default probabilities of stocks ($\hat{p}$), market capitalization (rSize) and market-to-book equity ratio (MB) within each portfolio. We can see that as the average default probability increases, the portfolio tends to have a monotonically decreased market capitalization. By contrast, the market-to-book equity ratio first decreases and then increases when the default risk is elevated. In summary, our asset pricing implication

\textsuperscript{4}equal-weighted return
\textsuperscript{5}Data available at Professor Kenneth R. French’s website.
demonstrated by Table 4 imposes challenges to the standard Fama and French’s (1996) conjecture, but is consistent with Campbell et al. (2008)’s findings among others.

4.4 Conclusion

We applied our proposed discrete transformation survival model to the bankruptcy data from 1981 to 2006. An optimal model with c around 10 was selected using two-step profile likelihood estimation. This model is recommended over Shumway’s model (c = 1) or Cox Proportional Hazards model (c = 0), when accurate default probability prediction is needed. Out-of-sample performance is examined through annual rolling window approach for withholding sample from 2002 to 2006 and is shown with improved accuracy ratio as well as model goodness-of-fit. Further asset pricing implication challenges the famous Fama and French’s (1996) conjecture on investors demand risk premiums for distress companies. However, the findings are consistent to some recent literature such as Campbell et al. (2008). Finally, we need to re-emphasize that this proposed discrete transformation survival model is fundamentally different from the continuous survival models in the literature because of some unique features of bankruptcy application. The proposed discrete transformation survival model may be potentially used in other applications with similar characteristics.

References


Appendix

Appendix 1: Derivations of the discrete transformation models (1) or (2).

For our discrete time model where $T$ can only take values at $t_1, ..., t_K$, the survival function $S(t) = P(T > t)$, $t = t_1, ..., t_K$.

$$G[- \log \frac{S_Z(t_k)}{S_Z(t_{k-1})}] = \exp[\beta^T Z(t_k)]G[- \log \frac{S_0(t_k)}{S_0(t_{k-1})}], \quad (A.1)$$

Assuming independent censoring,

$$\pi_{i,k} = Pr(T = t_k | X \geq t_k, Z(t_k) = Z_{i,k}) = Pr(T = t_k | T \geq t_k, Z(t_k) = Z_{i,k}).$$

Hence it can be represented as

$$\pi_{i,k} = Pr(T = t_k | T \geq t_k, Z(t_k) = Z_{i,k})$$
$$= 1 - \frac{S_Z(t_k)}{S_Z(t_{k-1})}$$
$$= 1 - \exp\{-G^{-1}[G(- \log \frac{S_0(t_k)}{S_0(t_{k-1})}) \exp(\beta^T Z_{i,k})]\}$$
$$= 1 - \exp\{-G^{-1}\exp(\alpha_k + \beta^T Z_{i,k})\}$$

where $\alpha_k = \log[G(- \log \frac{S_0(t_k)}{S_0(t_{k-1})})]$.

When $G$ belongs to the family (4)

$$G_c^{-1}(x) = \left\{ \begin{array}{ll} \frac{1}{c} \log(1 + cx), & c > 0; \\ x, & c = 0. \end{array} \right.$$ \hspace{1cm}

Hence

$$\exp[-G_c^{-1}(x)] = \left\{ \begin{array}{ll} \frac{1}{(1+cx)^{1/c}}, & c > 0; \\ \exp(-x), & c = 0. \end{array} \right.$$ \hspace{1cm}

This gives

$$\pi_{i,k} = \left\{ \begin{array}{ll} 1 - \frac{1}{(1+c\exp(\alpha_k+\beta^T Z_{i,k}))^{1/c}}, & c > 0; \\ 1 - \exp(-\exp(\alpha_k+\beta^T Z_{i,k})), & c = 0. \end{array} \right. \quad (A.2)$$

When $c > 0$, note left hand side of equation (A.1) indicates

$$G[- \log \frac{S_Z(t_k)}{S_Z(t_{k-1})}] = \frac{1}{c} [\exp(c(- \log \frac{S_Z(t_k)}{S_Z(t_{k-1})})) - 1] = \frac{1}{c} [\exp(\log \frac{S_Z(t_{k-1})}{S_Z(t_k)}) - 1] = \frac{1}{c} \frac{S_Z(t_{k-1}) - S_Z(t_k)}{S_Z(t_k)}.$$ \hspace{1cm}

Right hand side of equation (A.1) indicates

$$\exp[\beta^T Z(t_k)]G[- \log \frac{S_0(t_k)}{S_0(t_{k-1})}] = \exp[\beta^T Z(t_k)]\frac{1}{c} \frac{S_0(t_{k-1}) - S_0(t_k)}{S_0(t_k)}.$$ \hspace{1cm}
Equation (A.1) indicates
\[ \frac{S_Z(t_{k-1}) - S_Z(t_k)}{S_Z(t_k)} = \exp[\beta \tau Z(t_k)] \frac{S_0(t_{k-1}) - S_0(t_k)}{S_0(t_k)}. \]

In particular, when \( c = 1 \) this leads to
\[ \frac{S_Z(t_{k-1}) - S_Z(t_k)}{S_Z(t_k)} = \exp[\beta \tau Z(t_k)] \frac{S_0(t_{k-1}) - S_0(t_k)}{S_0(t_k)}, \]
equivalently
\[ \frac{\pi_{i,k}}{1 - \pi_{i,k}} = \exp[\beta \tau Z(t_k)] \frac{\pi_{i,0}}{1 - \pi_{i,0}}. \]

This is equivalent to binary regression model with logit link with panel data as defined in equation (A.2), which is the popular so-called discrete hazard model in bankruptcy literature following Shumway (2001).

When \( c = 0 \), Equation (A.1) indicates
\[ -\log \frac{S_Z(t_k)}{S_Z(t_{k-1})} = \exp[\beta \tau Z(t_k)](-\log \frac{S_0(t_k)}{S_0(t_{k-1})}). \]
or
\[ -\log S_Z(t_k) - (-\log S_Z(t_{k-1})) \]
\[ \frac{-\log S_0(t_k) - (-\log S_0(t_{k-1}))} = \exp[\beta \tau Z(t_k)]. \]

This is equivalent to proportional hazard model in continuous case when regarding \(- \log[S(t)]\), the cumulative hazard function in continuous case as step functions with jumps at \( t = t_1, \ldots, t_K \) and the difference of \(- \log[S(t)]\) is equivalent to hazard rate function in continuous sense.

For \( G \) in the family (5),
\[ G_{\rho}^{-1}(x) = \begin{cases} \log\left\{ 1 + \frac{1}{\rho}[\exp(\rho x) - 1] \right\}, & \rho > 0; \\ \log(1 + x), & \rho = 0. \end{cases} \]

Hence
\[ \exp[-G_{\rho}^{-1}(x)] = \begin{cases} \frac{1}{1 + \frac{1}{\rho} \exp(\rho x - 1)}, & \rho > 0; \\ \frac{1}{1 + x}, & \rho = 0. \end{cases} \]

This gives
\[ \pi_{i,k} = \begin{cases} \frac{1}{1 + \frac{1}{\rho} \exp(\rho \exp(\alpha_k + \beta \tau Z_{i,k}) - 1)}, & \rho > 0; \\ \frac{1}{1 + \exp(\alpha_k + \beta \tau Z_{i,k})}, & \rho = 0. \end{cases} \]
Appendix 2: Derivation of transformation equation (9).

Zeng and Lin (2006) assumes that

$$\Lambda_Z(t) = G^{-1}\left\{ \int_0^t e^{\beta T Z(s)} d\Lambda(s) \right\},$$

where $\Lambda(\cdot)$ is an unspecified increasing function. The baseline hazard is

$$\Lambda_0(t) = G^{-1}\left\{ \int_0^t 1 d\Lambda(s) \right\} = G^{-1}\left\{ \Lambda(t) \right\}.$$

Notice that their $\Lambda(t)$ is not the baseline hazard, $G^{-1}\{\Lambda(t)\}$ is. So their $\Lambda(t) = G\{\Lambda_0(t)\}$.

We have $d\Lambda(t) = dG\{\Lambda_0(t)\}$. Therefore, $dG\{\Lambda_Z(t)\} = e^{\beta T Z(t)} d\Lambda(t) = e^{\beta T Z(t)} dG\{\Lambda_0(t)\}$. 

<table>
<thead>
<tr>
<th>Year</th>
<th>Num of Bankruptcy</th>
<th>Number of Firms</th>
<th>Default Rate (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1981</td>
<td>32</td>
<td>4085</td>
<td>0.78</td>
</tr>
<tr>
<td>1982</td>
<td>45</td>
<td>4345</td>
<td>1.04</td>
</tr>
<tr>
<td>1983</td>
<td>36</td>
<td>4474</td>
<td>0.81</td>
</tr>
<tr>
<td>1984</td>
<td>49</td>
<td>4785</td>
<td>1.02</td>
</tr>
<tr>
<td>1985</td>
<td>53</td>
<td>4944</td>
<td>1.07</td>
</tr>
<tr>
<td>1986</td>
<td>80</td>
<td>4971</td>
<td>1.61</td>
</tr>
<tr>
<td>1987</td>
<td>39</td>
<td>5203</td>
<td>0.75</td>
</tr>
<tr>
<td>1988</td>
<td>68</td>
<td>5444</td>
<td>1.25</td>
</tr>
<tr>
<td>1989</td>
<td>73</td>
<td>5404</td>
<td>1.35</td>
</tr>
<tr>
<td>1990</td>
<td>81</td>
<td>5362</td>
<td>1.51</td>
</tr>
<tr>
<td>1991</td>
<td>125</td>
<td>5345</td>
<td>2.34</td>
</tr>
<tr>
<td>1992</td>
<td>93</td>
<td>5357</td>
<td>1.74</td>
</tr>
<tr>
<td>1993</td>
<td>47</td>
<td>5615</td>
<td>0.84</td>
</tr>
<tr>
<td>1994</td>
<td>50</td>
<td>6508</td>
<td>0.77</td>
</tr>
<tr>
<td>1995</td>
<td>51</td>
<td>7003</td>
<td>0.73</td>
</tr>
<tr>
<td>1996</td>
<td>57</td>
<td>7226</td>
<td>0.79</td>
</tr>
<tr>
<td>1997</td>
<td>71</td>
<td>7587</td>
<td>0.94</td>
</tr>
<tr>
<td>1998</td>
<td>95</td>
<td>7554</td>
<td>1.26</td>
</tr>
<tr>
<td>1999</td>
<td>63</td>
<td>7188</td>
<td>0.88</td>
</tr>
<tr>
<td>2000</td>
<td>69</td>
<td>7008</td>
<td>0.99</td>
</tr>
<tr>
<td>2001</td>
<td>79</td>
<td>6738</td>
<td>1.17</td>
</tr>
<tr>
<td>2002</td>
<td>79</td>
<td>6310</td>
<td>1.25</td>
</tr>
<tr>
<td>2003</td>
<td>44</td>
<td>5881</td>
<td>0.75</td>
</tr>
<tr>
<td>2004</td>
<td>25</td>
<td>5634</td>
<td>0.44</td>
</tr>
<tr>
<td>2005</td>
<td>20</td>
<td>5572</td>
<td>0.36</td>
</tr>
<tr>
<td>2006</td>
<td>10</td>
<td>5523</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table 1: **Firm data:** This table lists the number of defaults and number of active firms each year in our sampling period. The number of active firm is calculated by averaging over the number of active firms across all months of the year.
Table 2: Summary Statistics for quarterly updated firm-month observations

<table>
<thead>
<tr>
<th>Variable</th>
<th>NIMTA</th>
<th>TLMTA</th>
<th>EXRET</th>
<th>RSIZE</th>
<th>SIGMA</th>
<th>MB</th>
<th>PRICE</th>
<th>CASHMTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>−0.0073</td>
<td>0.4285</td>
<td>−0.0131</td>
<td>−10.6007</td>
<td>0.5996</td>
<td>2.1824</td>
<td>2.1981</td>
<td>0.1209</td>
</tr>
<tr>
<td>Median</td>
<td>0.0045</td>
<td>0.3970</td>
<td>−0.0096</td>
<td>−10.7276</td>
<td>0.4685</td>
<td>1.5392</td>
<td>2.4532</td>
<td>0.0460</td>
</tr>
<tr>
<td>Std. Dev</td>
<td>0.1070</td>
<td>0.2868</td>
<td>0.1701</td>
<td>2.0776</td>
<td>0.4952</td>
<td>1.7278</td>
<td>1.3281</td>
<td>0.8102</td>
</tr>
<tr>
<td>observations: 1,823,241</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Mean     | −0.1137| 0.7164 | −0.2373| −13.4158| 1.6761 | −0.3841| −0.3678| 0.1911  |
| Median   | −0.0444| 0.8386 | −0.1110| −13.6183| 1.4030 | 0.0171 | −0.4700| 0.0292  |
| Std. Dev | 0.4151 | 0.2905 | 0.5180 | 1.7834 | 1.4362 | 1.5942 | 1.5809 | 1.6681  |
| observations: 1,562 |
Table 3: Out-of-Sample Accuracy Ratio and Hosmer Lemeshow Test statistics and its p-value for $c = 0$ (Cox proportional hazards model), $c = 1$ (Shumway’s model) and $c = 10$ (the optimal selected transformation model) under the transformation function $G_c$ on annually updated bankruptcy data.

<table>
<thead>
<tr>
<th>$c$</th>
<th>Accuracy Ratio</th>
<th>Hosmer Lemeshow Test $\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c = 0$</td>
<td>0.7956</td>
<td>22.8885</td>
<td>0.0035</td>
</tr>
<tr>
<td>$c = 1$</td>
<td>0.8006</td>
<td>19.4592</td>
<td>0.0126</td>
</tr>
<tr>
<td>$c = 10$</td>
<td>0.8095</td>
<td>8.7790</td>
<td>0.3613</td>
</tr>
</tbody>
</table>

Figure 1: This plots how the same scores are translated into different probabilities under different transformation parameter $c$ or $\rho$ values. The solid line gives the probabilities converted according to Shumway’s model $c = 1$ or $\rho = 0$. The optimal parameter value from the fit on the real bankruptcy data is around $c = 10$. 
### Table 4: Asset Pricing Results

We sort all the stocks according to their predicted default probabilities, obtained from the selected optimal discrete transformation survival model. Ten portfolios are constructed based on percentile cut-off. For example, first column “0005” refers to the portfolio of stocks with the default probability below the 5th percentile and the second column displays portfolio of stocks in 5th to 10th percentile. Panel A reports the average of simple excess return for each portfolio. Panel B summarizes the results of estimate on constant alpha and three factor loadings with their corresponding t-statistics obtained by regressing the value-weighted return on the Fama-French three (RM, HML,SMB) factors. Panel C reports the average fitted default probability ($\hat{p}$), market capitalization (rSize) and market-to-book equity ratio (MB). * denotes significant at 5%, ** denotes significant at 1%.

<table>
<thead>
<tr>
<th>Portfolios</th>
<th>0005</th>
<th>0510</th>
<th>1020</th>
<th>2040</th>
<th>4060</th>
<th>6080</th>
<th>8090</th>
<th>9095</th>
<th>9599</th>
<th>9900</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Portfolio Return</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Return(%)</td>
<td>3.2875</td>
<td>2.1037</td>
<td>1.3872</td>
<td>0.5437</td>
<td>-0.3416</td>
<td>-1.3635</td>
<td>-2.8139</td>
<td>-4.5521</td>
<td>-6.6981</td>
<td>-12.0232</td>
</tr>
<tr>
<td><strong>Panel B: Fama-French Three-Factor Regression Coefficient</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha(%)</td>
<td>0.7998</td>
<td>-0.3106</td>
<td>-0.6905</td>
<td>-1.2457</td>
<td>-2.4045</td>
<td>-4.2112</td>
<td>-6.7159</td>
<td>-9.4513</td>
<td>-12.1528</td>
<td>-18.9572</td>
</tr>
<tr>
<td>RM</td>
<td>-0.1295</td>
<td>-0.0119</td>
<td>0.0767</td>
<td>0.1239</td>
<td>0.1839</td>
<td>0.3275</td>
<td>0.4276</td>
<td>0.6669</td>
<td>0.6641</td>
<td>0.6746</td>
</tr>
<tr>
<td>(−7.11)**</td>
<td>(-0.50)</td>
<td>(3.50)**</td>
<td>(5.16)**</td>
<td>(4.05)**</td>
<td>(5.63)**</td>
<td>(4.82)**</td>
<td>(6.25)**</td>
<td>(4.66)**</td>
<td>(2.74)**</td>
<td></td>
</tr>
<tr>
<td>HML</td>
<td>-0.2178</td>
<td>0.0476</td>
<td>0.2256</td>
<td>0.3332</td>
<td>0.1819</td>
<td>0.2373</td>
<td>0.1442</td>
<td>0.4827</td>
<td>0.5235</td>
<td>0.7920</td>
</tr>
<tr>
<td>(−7.95)**</td>
<td>(1.34)</td>
<td>(6.84)**</td>
<td>(9.23)**</td>
<td>(2.66)**</td>
<td>(2.71)**</td>
<td>(1.08)</td>
<td>(3.01)**</td>
<td>(2.44)</td>
<td>(2.13)**</td>
<td></td>
</tr>
<tr>
<td>SMB</td>
<td>0.0289</td>
<td>0.1767</td>
<td>0.1905</td>
<td>0.3195</td>
<td>0.5177</td>
<td>0.8115</td>
<td>1.0051</td>
<td>1.3279</td>
<td>1.2795</td>
<td>1.5402</td>
</tr>
<tr>
<td>(1.30)</td>
<td>(6.16)**</td>
<td>(7.12)**</td>
<td>(10.91)**</td>
<td>(9.34)**</td>
<td>(11.43)**</td>
<td>(9.29)**</td>
<td>(10.20)**</td>
<td>(7.36)**</td>
<td>(5.12)**</td>
<td></td>
</tr>
<tr>
<td><strong>Panel C: Portfolio Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{p}$(%)</td>
<td>0.0270</td>
<td>0.0448</td>
<td>0.0671</td>
<td>0.1205</td>
<td>0.2321</td>
<td>0.4599</td>
<td>0.8844</td>
<td>1.5032</td>
<td>2.7754</td>
<td>6.5600</td>
</tr>
<tr>
<td>MB</td>
<td>3.6068</td>
<td>2.9953</td>
<td>2.5172</td>
<td>2.0419</td>
<td>1.9462</td>
<td>1.9039</td>
<td>1.6422</td>
<td>2.8665</td>
<td>0.7333</td>
<td>1.1215</td>
</tr>
</tbody>
</table>
Figure 2: This plots the observed values of the log-likelihood functions for the quarterly bankruptcy data from 1981 to 2006: (a) pertains the transformation function $G_c$; (b) pertains the transformation function $G_\rho$. 

Figure 3: This plots the observed values of the log-likelihood functions for the annual bankruptcy data from 1981 to 2006: (a) pertains the transformation function $G_c$; (b) pertains the transformation function $G_\rho$. 