



## Variable selection and corporate bankruptcy forecasts



Shaonan Tian<sup>a,\*</sup>, Yan Yu<sup>b</sup>, Hui Guo<sup>b,c</sup>

<sup>a</sup> Lucas College and Graduate School of Business, San Jose State University, San Jose, CA 95192, USA

<sup>b</sup> Carl H. Lindner College of Business, University of Cincinnati, Cincinnati, OH 45221, USA

<sup>c</sup> Economics and Management School, Wuhan University, Wuhan 430072, China

### ARTICLE INFO

#### Article history:

Received 29 August 2013

Accepted 3 December 2014

Available online 9 December 2014

#### JEL classification:

G33

G17

M41

C25

#### Keywords:

Discrete hazard model

Financial ratios

LASSO

Market information

### ABSTRACT

We investigate the relative importance of various bankruptcy predictors commonly used in the existing literature by applying a variable selection technique, the least absolute shrinkage and selection operator (LASSO), to a comprehensive bankruptcy database. Over the 1980–2009 period, LASSO admits the majority of Campbell et al. (2008) predictive variables into the bankruptcy forecast model. Interestingly, by contrast with recent studies, some financial ratios constructed from only accounting data also contain significant incremental information about future default risk, and their importance relative to that of market-based variables in bankruptcy forecasts increases with prediction horizons. Moreover, LASSO-selected variables have superior out-of-sample predictive power and outperform (1) those advocated by Campbell et al. (2008) and (2) the distance to default from Merton's (1974) structural model.

© 2014 Elsevier B.V. All rights reserved.

### 1. Introduction

Accounting and finance researchers have considered various predictive variables in the reduced-form corporate bankruptcy forecast model. Earlier studies, e.g., Beaver (1966), Altman (1968), Ohlson (1980) and Zmijewski (1984), have routinely used accounting variables, i.e., financial ratios constructed from only accounting data, as a gauge of default risk. In an attempt to improve the empirical performance of the reduced-form model, Shumway (2001) advocates for incorporating market-based variables in bankruptcy forecasts, in addition to two commonly used accounting-based variables. In a similar vein, Campbell et al. (2008; CHS thereafter) introduce new market variables and accounting variables; and they also propose a modification of the accounting variables adopted in Shumway (2001) by using the market value of assets rather than the book value. While Shumway (2001) and CHS (2008) have shown their models exhibit noticeable improvement over the models proposed in previous studies, none of existing studies has provided a formal analysis on the relative importance of a comprehensive set of bankruptcy predictors using the advanced variable selection technique. In

particular, there is no conclusive evidence on the role of accounting variables in bankruptcy forecasts. We try to fill the gap by introducing a state-of-art variable selection technique proposed by Tibshirani (1996)—the least absolute shrinkage and selection operator (LASSO).

Statisticians develop variable-selection methods to achieve two main objectives—(1) identifying relevant predictive variables and (2) improving prediction accuracy (see, e.g., Fan and Li, 2001). A formal variable-selection analysis thus allows us to shed new light on the corporate bankruptcy forecast literature in two important ways. First, it enables us to identify from an exhaustive set of bankruptcy predictors proposed in existing studies a parsimonious subset of the most relevant ones. Such identification has important implications for testing bankruptcy theories, designing regulations in credit markets, and conducting credit risk analysis. Second, as we confirm in this paper, the selected reduced-form model shows improved in-sample and out-of-sample performance, when comparing with the prominent models in the existing literature.

LASSO penalizes regression coefficients through a shrinkage method and thus provides a sparse variable-set solution. It has been widely used in variable-selection studies (see, Tibshirani, 1996; Efron et al., 2004) and is a state-of-the-art variable selection tool. LASSO enjoys the easy interpretability as the traditional subset variable selection does but has additional advantages of

\* Corresponding author. Tel.: +1 (408)924 7245; fax: +1 (408)924 3445.

E-mail address: [shaonan.tian@sjsu.edu](mailto:shaonan.tian@sjsu.edu) (S. Tian).

(1) the stability of model selection and (2) potential improvement in prediction accuracy. Compared with other commonly used variable selection methods such as the subset or stepwise selection, LASSO has several desirable statistical properties that suit particularly well for the main empirical issues that we try to address in this paper. First, given the rareness of default events, stability is a necessary requirement of variable selection techniques used for bankruptcy forecasts. LASSO is quite stable to small perturbations of data changes. Second, the shrinkage method may improve prediction accuracy. Third, LASSO produces an entire variable selection path that we can use to gauge the relative importance of the selected variables. Fourth, LASSO naturally overcomes the multicollinearity problem. Last, LASSO is computationally efficient, especially when there are a large number of candidate predictors.

We construct a comprehensive bankruptcy database by merging daily and monthly equity data from the Center for Research in Security Prices (CRSP) with annual financial information from COMPUSTAT.<sup>1</sup> A company is in default if it files for either Chapter 7 (liquidation) or Chapter 11 (reorganization) bankruptcy protection. We include an exhaustive list of 39 accounting-based variables and market-based variables that have been used in the bankruptcy literature as candidate default-risk predictors. As in Shumway (2001), Chava and Jarrow (2004), CHS (2008), and others, we model the bankruptcy risk using the discrete hazard model. Shumway (2001) emphasizes that the discrete hazard model using time-varying panel data has important advantages compared with static models using cross-sectional data (e.g., Altman, 1968; Ohlson, 1980; Zmijewski, 1984). This is because the latter ignore the fact that firms change over time and thus may produce biased and inconsistent bankruptcy probability estimates. In this paper, we adopt LASSO variable selection technique on time-varying covariates for the panel bankruptcy data.

We first focus on the forecast of one-year-ahead bankruptcy—the most commonly used forecast horizon in the existing literature, and then investigate how variable-selection results vary with forecast horizons. Over the full sample spanning the 1980–2009 period, LASSO selects seven predictive variables into the reduced-form bankruptcy forecast model. We find strong support for Shumway's (2001) argument of including market-based variables in bankruptcy forecasts. Two market variables advocated by Shumway (2001), i.e., stock return volatility and the excess stock return, and one market variable proposed by CHS (2008), i.e. stock price, enter into the LASSO-selected reduced-form model. Shumway (2001) shows that, consistent with the previous accounting studies, (1) the net income to total assets ratio and (2) the total liabilities to total assets ratio constructed from accounting data are significant predictors even when controlling for market-based variables in bankruptcy forecasts. CHS (2008), however, suggest that we should modify these two variables using the market value of assets instead of the book value. Our variable selection analysis allows us to shed light on this issue: LASSO selects CHS's modified financial ratios but not Shumway's (2001) original variables. Of CHS's eight predictive variables, five enter into our reduced-form bankruptcy forecast model, indicating that CHS have done a reasonably good job in selecting the bankruptcy predictors.<sup>2</sup> Nevertheless, the LASSO variable selection analysis differs from CHS's model in two ways. First, LASSO identifies two additional predictive variables—(1) the current liabilities to total assets ratio and (2) the total debts to total assets ratio constructed

from only accounting data. This result reaffirms the important role of accounting-based variables in bankruptcy forecasts. Second, three of CHS's predictive variables, i.e., the market capitalization, the market to book ratio, and the ratio of cash and short-term assets to the market value of assets, do not enter into the LASSO-selected reduced-form model.<sup>3</sup> In our study, the LASSO variable selection results are strikingly consistent across subsample periods: The identical sets of predictive variables are selected over the 1980–2000, 1980–2002, 1980–2005, and 1990–2009 periods.

The distance to default (DD) constructed from Merton's (1974) structural model is a popular bankruptcy risk measure among practitioners. CHS (2008) and Bharath and Shumway (2008), however, find that DD provides relatively little information about future bankruptcy beyond the variables used in their reduced-form models. When we include DD as a candidate predictor along with the other 39 predictive variables, it does not enter into the LASSO-selected reduced-form model and the set of selected predictors is identical to that obtained without DD as a candidate predictor. In the out-of-sample forecast, our limited empirical study shows the performance of the DD only model is similar to, or slightly better than, that of the CHS reduced-form model. By contrast, the LASSO-selected reduced-form model performs noticeably better than the DD only model over various out-of-sample testing periods.

CHS (2008) have advocated for constructing financial ratios using the market value of assets in default forecasts. By contrast, accounting researchers, e.g., Beaver et al. (2005), have reiterated the relevance of accounting-based variables by showing that their predictive power is strikingly consistent across time. In a similar vein, Das et al. (2009) show that accounting data provides significant supplementary information in distress risk pricing, especially for firms with limited or no trading activity. We provide support for both arguments. LASSO selects the market value of assets for the net income to total assets ratio and the total liabilities to total assets ratio but chooses the book value for the current liabilities to total assets ratio and the total debts to total assets ratio.

As CHS (2008) emphasize, a firm's market value of equity is a more accurate gauge of its prospects than is the book value of equity. Hence, the predictive power of market leverage for bankruptcy risk likely reflects the conventional wisdom that firms with lower leverage are more able to pay off its debts. In addition, a firm with higher bankruptcy costs has more incentives to lower its bankruptcy risk by taking precautionary actions, for example, adopting low target leverage or taking low-risk project (e.g., George and Hwang, 2010). The book leverage is a more reliable measure of target leverage than is market leverage because many studies (e.g., Welch, 2004; Graham and Harvey, 2001) show that firms rarely counteract to changes in their capital structure caused by fluctuations in their stock prices. Therefore, book leverage correlates negatively with bankruptcy risk possibly because it is a proxy for precautionary actions taken by firms to reduce their bankruptcy risk.

Our conjectures have an interesting implication. Because individual firms are susceptible to large idiosyncratic shocks or their fortune can change quite drastically over time, market variables are more useful in short-run forecasts than in long-run forecasts. By contrast, because ceteris paribus a prudent firm is more able to withstand a financial market storm due to its precautionary actions intended for reducing bankruptcy risk than is a reckless firm, the relative importance of accounting-based variables may increase with forecast horizons when idiosyncratic risk averages out. We find strong support for this implication. When the prediction

<sup>1</sup> Vassalou and Xing (2004) also use only CRSP and COMPUSTAT data to construct their bankruptcy database. While CHS (2008) incorporate additional proprietary data sources, our database is qualitatively similar to theirs. See Ding et al. (2012) for details.

<sup>2</sup> Of five predictive variables proposed by Shumway (2001), four variables enter into our LASSO-selected reduced-form model either directly or in a modified form.

<sup>3</sup> Using our data, we confirm CHS's finding that these three variables have statistically significant in-sample predictive power for the default risk, although these variables are not selected by LASSO.

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات