Essays on corporate default risk and equity return

Gang Liu
University at Albany, State University of New York, gangliu@hotmail.com

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etal

Part of the Economics Commons

Recommended Citation
https://scholarsarchive.library.albany.edu/legacy-etal/681

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive.
Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
Essays on Corporate Default Risk
and Equity Return

by

Gang Liu

A Dissertation
Submitted to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

College of Arts & Sciences
Department of Economics
2012
To My Family:

My parents,

Sisters,

Husband and Son
Acknowledgement

Although only my name appears on the cover of this dissertation, many people have contributed to it. Therefore, I would like to express my gratitude towards all those people who have made this dissertation possible.

My deepest gratitude is to my advisor, Professor Bruce C. Dieffenbach for his encouragement and expert advice. I have been very lucky to have an advisor who gave me freedom to explore on my own, and at the same time the guidance to overcome the setbacks in my work. Without his constructive criticism and insightful comments, the dissertation won’t be in its current form.

I’m deeply grateful to Professor David M. Smith for carefully reading my dissertation, and giving detailed comments for every edition of my dissertation. He pinpointed all inconsistencies and encouraged the use of correct grammar in my English writings. His encouraging words and optimistic spirit always empower me when my steps faltered.

I am also thankful to Professor Terrence W. Kinal for his continuous support and guidance during my graduate studies. And I benefit a lot from suggestions and comments from him while writing my third-year paper and dissertation.

Next, I would like to thank my professors and friends at University at Albany. Particularly, I would like to acknowledge Professor Kwan K. Yun for his care about students. He encourages his students and guides them to meet high standards.

Last but not least, I’m deeply grateful to my family for their unfailing love and support all these years. They granted me line of credit without asking for equity held against it. I deeply appreciate their belief in me. My immediate family, my parents, sisters, husband, and son, to whom this dissertation is dedicated to, has been a constant source of love, support and strength to me. Thank you for being there for me.
Abstract

The theme of this dissertation is to predict firm’s default risk from different approaches, and evaluate the effect of default risk on cross-sectional stock returns.

Currently, among the empirical work that uses structural models to predict probability of bankruptcy (PD), no one has dealt with the noise in stock prices. Yet, ignoring the noise could lead to biased estimates of PD. Chapter 1 tackles this issue by applying a filtering technique (Duan and Fulop, 2007) on observed stock prices. The proposed structural model of credit risk is a down-and-out barrier call (DOC) option model, developed by Black and Cox (1976) and Brockman and Turtle (2003). We implement the proposed method on U.S. transportation firms from year 1999 to year 2008.

Default data show clear evidence of cycles. In chapter 2, we include macro factors in reduced form hazard models to capture the time-variation in default rate and the impact of business cycle on firms’ creditworthiness. Other covariates used include firm-specific accounting ratios and market return variables. We construct macro factors from fifty macroeconomic time series data using principal component analysis. Our proposed hazard model delivers a noticeable improvement over Shumway’s (2001) model and show that macroeconomic conditions affect firm’s creditworthiness differently. Unlike Sueyoshi (1995), our specification tests suggest the choice of a hazard model specification is innocuous in our context of firm’s default prediction.

Chapter 3 explores the effect of default risk on stock returns using default probability estimated from our one-year ahead hazard model in chapter 2. Our study captures a pattern different from that of current empirical literature. To uncover this credit risk
puzzle (i.e., the negative risk-return relation for the distressed portfolios), we adjust the portfolio returns by several risk-based asset-pricing models. We also conduct tests for competing asset pricing models by generalized method of moments (GMM) and find evidence that support default risk as a factor that should be considered in explaining stock returns.
## Contents

### Chapter 1

**Predicting Corporate Bankruptcy when Stock Prices are Contaminated by Trading Noises---A Structural Model of Default Risk with Particle Filtering MLE Method**

1. Introduction and Literature Review ......................................................... 2
   1.1 Estimation Approaches ................................................................. 7
   1.2 Motivation ............................................................................... 10

2. Structural Models and Default Probabilities ........................................... 12
   2.1 Standard Call Option Approach: Merton/KMV ............................... 13
   2.2 Down-and-out Barrier Option (DOC) approach .............................. 15

3. Estimation of Structural Models ............................................................. 18
   3.1 Case 1: DOC barrier option model without noise in equity prices ...... 18
   3.2 Case 2: DOC barrier option model with possible noise in equity prices... 22

4. Estimation using the Particle Filter ......................................................... 24
   4.1 Particle filter (PF) ...................................................................... 24
   4.2 Firm value filtering scheme for DOC barrier model ....................... 26
   4.3 Maximum likelihood estimation ................................................... 27

5. Estimation with empirical data .............................................................. 28
   5.1 Description of the data .................................................................. 28
   5.2 Variable specification .................................................................... 30
   5.3 Empirical results ......................................................................... 32
   5.3.1 Parameter estimation ............................................................... 32
Chapter 2

Reduced Form Hazard Models of Corporate Default Prediction including Macro Factors as Covariates

1. Introduction and related literature ................................................. 63

2. Econometric framework .............................................................. 67
   2.1 General setup: an approximate dynamic factor model ................. 67
   2.2 Estimation Methodology: a two-step procedure ......................... 69

3. Data Description
   3.1. Macro economic time series used for constructing macro factors .... 72
   3.2. Data for hazard model regression ............................................. 75

4. Empirical results ........................................................................... 79

5. Robustness check ................................................................. 92

6. Conclusion and extension ........................................................... 102

Appendix ...................................................................................... 104

References .................................................................................. 106
Chapter 3

The Effect of Default Risk on Cross-section Stock Returns

1. Introduction .................................................................112

2. Default risk and cross section equity return ......................115
  2.1. Raw return pattern ...................................................115
  2.2. Risk-adjusted return pattern ......................................119

3. Pricing of default risk---asset pricing model test ...............124
  3.1 Econometric specification and tests of linear factor pricing models ....125
  3.2 Empirical analysis....................................................129
    3.2.1 Test models.......................................................129
    3.2.2 Test assets.......................................................131
    3.2.3 Data..............................................................133
    3.2.4 Results...........................................................133

4. Conclusions ..............................................................137

References ........................................................................139
Chapter 1: Predicting Corporate Bankruptcy when Stock Prices are Contaminated by Trading Noises---A Structural Model of Default Risk with Particle Filtering Maximum Likelihood Estimation

Abstract

Currently, among the empirical work that uses structural models to predict probability of bankruptcy (PD), no one has dealt with the noise in stock prices. Yet, ignoring the noise could lead to biased estimates of PD. Our paper attempts to make the implementation of structural models more feasible by applying a filtering technique on observed stock prices. The proposed structural model of credit risk is a down-and-out barrier call (DOC) option model, developed by Black and Cox (1976) and Brockman and Turtle (2003). DOC model recognizes that a bankruptcy or default may happen anytime rather than only on the days when debt matures, and this flexibility would not be possible with a standard call option model by Black and Scholes (1973) and Merton (1974). The estimation method used is based on “smoothed auxiliary particle filtering” techniques developed by Duan and Fulop (2009). We implement the proposed method on U.S. transportation firms from year 1999 to year 2008. Unobserved asset values, and its drift and volatility, as well as firm-specific implied barriers are backed out simultaneously from the observed equity prices, and probabilities of default are computed 1-year, 2-year, and 3-year ahead respectively. Our bankruptcy probabilities measures display better discriminatory power than the ones implied in other commonly adopted bankruptcy models, such as Altman (1968), Merton (1974)/ KMV and Shumway (2001).

JEL classification: G13, G33, C22
1. Introduction and literature review

Due to the rapid growth of the markets for corporate bonds and credit derivatives, academics and practitioners have recently shown a marked interest in credit risk models for derivative pricing and risk management purposes. In addition, the new regulatory requirements of Basel Accord II provide strong incentive for financial institutions to assess the risk inherent in their credit portfolios and determine capital requirements.

The objective of this paper is to model noise in stock price, apply a new estimation method and investigate the empirical performance of a down-and-out barrier option (DOC) model to forecast corporate defaults. We estimate default probabilities for U.S. firms in the transportation industry in the period 2000-2008. We choose transportation industry because these firms are strongly affected by the business cycle. We also provide a comprehensive comparison of the discriminatory power among several distinct bankruptcy\textsuperscript{1} models.

Academics in the fields of accounting and finance have been forecasting bankruptcy for decades. The traditional models adopt fundamental analysis; the philosophy of these models, such as Altman (1968) and Ohlson (1980) is to find which accounting-based measures are important in assessing the credit risk of a firm. Modern approaches to probability of default use current market data about equity and/or debt to back out a market measure of the creditworthiness of a potential borrower. Basically, there are two different conceptual approaches for default risk modeling: structural models and reduced

\textsuperscript{1} We use bankruptcy and default interchangeably in this paper. For the definition of bankruptcy, and included types of bankrupt firms in empirical investigation, please refer to the data description part in section 6.
form models\textsuperscript{2}. This chapter examines the default probabilities (PDs) that are generated by a “structural” model of credit risk. Specifically, we adopt a down-and-out call (DOC) option framework (Black and Cox, 1976; Brockman and Turtle, 2003).

The family of structural models, pioneered by Black-Scholes (1973) and Merton (1974) (hereafter, “BSM”), adopt the contingent claim approach, i.e., all securities of a firm are treated as a contingent claim on the firm’s assets. In Black-Scholes (1973), a firm’s equity is viewed as a standard European call option on the firm’s fundamental value or asset value with strike price equal to the promised payment of corporate liabilities, and a time-to-maturity the same as that of the total debt. The seminal work of Merton (1974) extends the Black-Scholes (1973) option pricing framework to the evaluation of corporate debt, and shows that by put-call parity, the payoff to liability holders of a firm is equal to the value of a risk-free discount bond minus the value of a put option.

One byproduct of Merton (1974) is the probability that a firm will default. In this setup, the risk of a firm’s default process is linked to the variability in the firm’s underlying asset value. It’s assumed that the value of the firm moves randomly over time with a given expected return and volatility. At option maturity, if the asset value hits the default boundary (or barrier\textsuperscript{3}), then the firm is in default.

\textsuperscript{2} Reduced models are pioneered by Jarrow and Turnbull (1995), where bankruptcy is modeled as a hazard rate process, rather than as a microeconomic model of the firm’s capital structure. Models along this line include Jarrow, Lando, and Turnbull (1997), and Duffie and Singleton (1999). They use market information about bond spreads to extract default probabilities (because lenders require a spread over the risk-free interest rate to compensate for the default risk they bear), but give no economic explanation of default causality.

\textsuperscript{3} A default boundary, or barrier, is a level such that when the firm’s asset value falls beneath this level, it will cause the firm to default. For simplicity, we consider in our paper only the case of a constant barrier. In BSM model with zero-coupon bond assumption, the default boundary is zero until the bond matures, and then set to the principal value of debt at maturity.
One possible weakness of Merton’s approach is the assumption that default only occurs at the maturity of the debt. Under this assumption, the value of a firm’s underlying asset could go anywhere prior to maturity. However there may exist a barrier below which a firm’s fundamental value is not allowed to fall. One argument for the early bankruptcy barrier is debt covenants that contain positive net worth provision or require the firm to maintain certain financial ratios above pre-specified levels (e.g debt-to-equity ratios) as depicted in Brockman and Turtle (2003).\(^4\) Another limitation of viewing common stock of a firm as standard call option is the implication that a shareholder-aligned manager will take infinite risk when choosing a project. As first pointed out by Galai and Masulis (1976), the value of a standard call option is strictly increasing in the volatility of the underlying assets. Hence the manager may reject a more profitable project in favor of a project with negative NPV but higher risk (or larger variance of return), a problem described as “asset substitution” and “under-investment” in Jensen and Meckling (1976) and Myers (1977).

Therefore, we need a different model that keeps the option analogy but also provides a limit on the risk choices of managers. Such a requirement is satisfied if we treat the firm’s equity as a down-and-out (knocked-out) barrier option.\(^5\) There are two general types of specification of the asset value that triggers bankruptcy:


\(^5\) Indeed, Hao (2005) shows the sign of Greek Vega, i.e., the first derivative for an option value with regard to the volatility of the underlying asset value, is not definite for a DOC option. This feature prevents managers from taking infinite risk in project choice. On the contrary, Vega is always positive for a standard call option.
Exogenous vs. Endogenous Barrier

Black and Cox (1976) first modeled early default feature by imposing an exogenous default barrier and viewing a firm’s equity as a down-and-out call option (DOC) on the firm’s asset value. In their model, a firm is in default when its asset value falls beneath a pre-specified firm-specific barrier, which can be either a constant or a time varying variable. Corporate equity will be knocked out by bankruptcy, and asset ownership is then transferred from shareholders to liability holders. Longstaff and Schwartz (1995, LS) extend the Black and Cox model to introduce a stochastic risk-free interest rate that follows the Vasicek (1977) process. The LS model considers coupon-paying bonds with an exogenous default barrier that equals the principal value of debt; however, as Huang and Huang (2002) and others argued, it is more reasonable to specify a default barrier that is some fraction of debt principal.

Leland and Toft (1996, LT) endogenizes the default barrier by introducing taxes and bankruptcy costs as factors in determining the optimal asset value which triggers bankruptcy. In the LT model, to determine the equilibrium default barrier endogenously, the equity pricing formula should satisfy the smooth-pasting condition. Under this condition, the equity holders choose the optimal default barrier that maximizes equity value. In other words, at the default-triggering asset level, the partial derivative of equity with respect to the asset value is zero. As pointed out by Simone (2009), the equity value is still positive and this value could be the liquidation’s price. The optimal barrier represents the level of guarantees that a company must have in order to borrow money. Of course, it affects the payoff of equity holders.

---

We choose an exogenous rather than endogenous barrier in this paper’s DOC option pricing setting. The reason is that the F.O.C. equation for the DOC model is highly nonlinear in the barrier level term. Hence, it will be extremely difficult to derive a closed form solution for this endogenous barrier level, and then substitute it into the joint likelihood function of observed equity values to obtain the maximum likelihood estimates of the DOC model parameters. Indeed, the cost of a more realistic model is its tractability. In these cases we need to rely on numerical methods.

Our paper adopts a DOC option framework with an exogenous default barrier set to some fraction of the principal value of debt. Rather than specifying default barriers subjectively, we propose a statistical method to estimate the barrier level of a sample of industrial firms through a closed-form DOC valuation model of Merton (1973). There are many different ways to implement structural credit risk models. Our Particle Filter-Maximum Likelihood Estimation (PF-MLE) method is different from the “proxy approach” in Brockman and Turtle (2003, hereafter, “BT”), or the “volatility restriction approach” in Reisz and Perlich (2007). And this difference in implementation surely affects the performance of the structural models. We will explain in more detail below.

The major empirical challenge in all structural credit risk models is that the firm’s asset value is an unobserved latent factor. Since the asset value of a firm is not typically traded, market prices cannot be observed. The key distinguishing feature of different estimation approaches is how the unobserved asset value and its volatility are estimated. And in our DOC option framework, there is one additional parameter i.e., the default barrier ($H$), which acts as the absorbing barrier for the firm’s asset value.
Next section 1.1 presents a review of the literature on default probability estimation methods. In section 1.2, we propose applying particle filtering technique to MLE method to deal with noise in stock price.

1.1 Estimation approaches

1.1.1 The proxy approach

In the proxy approach, the market value of a firm’s assets is approximated by the sum of the market value of equity and the book value of liabilities. The volatility of the asset value is then computed directly from the returns of the proxy firm value. Empirical studies using the proxy approach include Jones, Mason and Rosenfeld (1984), Brockman and Turtle (2003), and others. Though it is easy to implement, the proxy approach would lead to an upward biased estimate of the asset values since it replaces the market value of debt with its higher book value. Also, Wong and Choi (2006) showed theoretically and empirically that this approximation leads to a significant overestimation of the default barrier in DOC model of Brockman and Turtle (2003). To get rid of the bias, they propose a maximum likelihood method (MLE) to estimate the asset values, asset volatilities, and default barriers; and their simulation proves the preciseness of the MLE estimators in DOC option framework. Following Wong and Choi (2006), we will adopt the MLE method in our paper, but combine it with a particle filtering technique to account for the noise in equity price data.
1.1.2 The volatility restriction approach (VR)

The volatility restriction (VR) method is the most popular way of implementing structural models. First proposed by Ronn and Verma (1986), and widely used in academia (Lyden and Saraniti, 2000; Delianedis and Geske, 2003) and the commercial world\(^7\), the VR method is to solve a system of two nonlinear equations that relate the observed stock value and estimated stock volatility with model outputs (i.e., asset value and asset value volatility). The first equation is an equity pricing formula, describing the equity value \(S\) as an option on the underlying asset value \(V\); the second equation restricting the equity volatility \(\sigma_S\) to match the asset volatility \(\sigma_V\) is derived from the equity-pricing equation via Ito’s lemma.

\[
S = C(V, \sigma_V) \quad \text{and} \quad \sigma_S = \sigma_V \frac{V}{S} \frac{\partial S}{\partial V}
\]

The above equation system could be solved with Newton-Raphson iterative algorithm implemented via computer, and the speed is fast if \(C(V, \sigma_V)\) is a standard call option pricing formula. However when it was applied to a DOC option framework, Reisz and Perlich (2007) found that “a Newton-Raphson-type search does not always converge due to the very flat slope of the DOC formula at certain points”. For this reason, we are not in favor of the VR method in estimating our DOC model. Another major drawback of this VR approach, as Duan (1994) states, is that it forces the Ito lemma to hold at each time point, and this could be too restrictive. When the market value of equity changes abruptly at a certain point, we may not get a solution for the equations. Even when the

---

\(^7\) Moody’s KMV apply this approach to generate an estimate of asset volatility, which is used as an input in their estimation process.
results are obtained, it would be biased, e.g. a jump in the market value of equity could lead to an overestimation of asset volatility. (See Crouhy (1997) and KMV (2002))

1.1.3 Iterative procedure /KMV method

Moody’s KMV uses a complex iterative procedure to solve for asset volatility. First, it applies the above volatility restriction approach to obtain the initial value of asset volatility, which is used to determine the inverted asset values corresponding to the observed equity prices through the equity pricing formula. The volatility of the resulting asset return is then used as the input for the next iteration step, and so on. Iteration is continued until convergence. Duan, Gauthier and Simonato (2004) shows that the KMV method turns out to produce a point estimate identical to the maximum likelihood estimates for Merton’s (1974) model. However, it cannot provide a sampling distribution for statistical inference. For the DOC option model we adopt in this paper, the limitation of the KMV method is more obvious in that it cannot estimate the capital structure parameter, i.e. the default barrier ($H$).

1.1.4 Maximum likelihood estimation

The transformed-data ML estimation proposed by Duan (1994) has been largely ignored in the credit risk literature. Even fewer empirical studies of structural models of credit risk have used the ML approach in forecasting corporate defaults, with the exception of Hao (2005). Until recently, Wong and Choi (2006) and Ericsson and Reneby

---

8 KMV is a trademark of KMV Corporation. Stephen Kealhofer, John McQuown and Oldrich Vasicek founded KMV Corporation in 1989. On February 11, 2002, Moody’s announced that it was acquiring KMV for more than $200 million in cash. “Expected Default Frequency” (EDF), a measure similar to default probability, is reported by Moody’s-KMV monthly.
(2005) showed, by a series of simulations, that the ML approach outperforms the VR and Proxy approach in terms of both efficiency and biasness. In our DOC model of default prediction, we will apply the MLE method while combining it with a particle filtering technique to account for the effects of noise in equity data. The implementation procedures of this PF-MLE method are given in detail in Section 3 and Section 4. We give the justification for this new method below.

1.2 Motivation---modeling noise in stock price

The stock market provides a rich source of information regarding bankruptcy. If the market is efficient, then equity prices should contain information about a firm’s asset value and reflect its financial health well. However, equity prices are probably influenced by market microstructure (Hasbrouck (1993) and Madhavan et al. (1997)).

In this paper, we rely on equity price information to extract parameters of our DOC model. Ignoring trading noise could lead to biased estimates of asset volatility, which is an essential input for our measure of default probabilities. So, we need to apply a filtering technique on observed stock prices. Conventional methods described above in section 1.1, such as the proxy approach by Brockman and Turtle (2003), VR approach by Reisz and Perlich (2007), or MLE approach by Ericsson and Reneby (2002), all assume a one-to-one relationship between equity value and firm values, failing to account for the noise in equity prices.

To account for the above fact and consequences, we adopt a smoothed particle filtering-MLE approach proposed by Duan and Fulop (2009) in our paper. This approach is an extension of the transformed-data maximum likelihood method developed by Duan...
It treats the equity as noisy signal about the asset value. The particle filtering estimation is a complicated nonlinear filtering problem on a state-space model. Duan and Fulop (2009) apply this approach to the estimation of the standard call option model of Merton (1974). In our paper, we’ll apply this method in a down-and-out call option model to predict a firm’s probability of default (PD). The proposed method makes implementation of a DOC barrier option model more feasible.

To our knowledge, no empirical study of the structural models of bankruptcy prediction has dealt with the noise in equity prices. We are the first to apply the MLE approach with a smoothed particle filter algorithm in the framework of the barrier option when equity prices are contaminated by trading noise. This paper contributes to the literature by empirically examining the significance of trading noise in equity prices, by proposing a statistical framework to capture the effect of trading noise on default probability in DOC model, and by providing a comprehensive comparison of bankruptcy prediction performance of the barrier option model with other structural models (Merton/KMV) and accounting-based measures.

The rest of this paper is organized as follows: Section 2 summarizes the analytical framework for standard call option and barrier options to provide the necessary background for subsequent sections. A general description of the estimation problem is then presented in section 3, followed by an exposition of the PF-MLE estimation technique for DOC model in section 4. Section 5 discusses our data set, and presents empirical analysis for firms in transportation industry. Section 6 provides some concluding comments and indicates possible directions for future research.
2. Structural Models and Default Probabilities

The structural models of credit risk view a firm’s equity and debt as contingent claims written on its fundamental value. In both standard call option and down-and-out barrier option approaches, the firm’s underlying asset value is assumed to follow a Geometric Brownian Motion (GBM) in the following form:

\[ dV = \mu_v V dt + \sigma_v V dW_t \]

where \( V \) is unobserved asset value, \( \mu_v \) is expected asset return (or asset drift), \( \sigma_v \) is the volatility of the firm’s asset return, and \( W_t \) is a standard Brownian motion (or Wiener process).

Applying Ito’s lemma with \( f(V) = \log(V) \), we can derive the log asset value process as

\[
\begin{align*}
    d \ln V &= f'(V) dV + \frac{1}{2} f''(V) V^2 \sigma_v^2 dt \\
    &= \frac{1}{V} (\mu_v V dt + \sigma_v V dW_t) \cdot \frac{1}{2} \sigma_v^2 dt \\
    &= (\mu_v - 0.5 \sigma_v^2) dt + \sigma_v dW_t \\
\end{align*}
\]

It follows that

\[ \ln V_t = \ln V_0 + (\mu_v - 0.5 \sigma_v^2) t + \sigma_v W_t \]  \hspace{1cm} (2)

Exponentiation gives the expression for \( V \), \( V_t = V_0 \exp[(\mu_v - 0.5 \sigma_v^2) t + \sigma_v W_t] \). In most cases, to work with normal distributions of the logarithm of \( V_t \) is easier than with the lognormal distribution itself. So in this paper, we adopt equation (2). \( \ln V_t \) is normally distributed with mean \( \ln V_0 + (\mu_v - 0.5 \sigma_v^2) t \) and variance \( \sigma_v^2 t \)
Next, we describe the analytical framework for standard call (SC) option and DOC barrier options below. SC option model is provided for reason to compare the result with that from the DOC model in Section 6.

2.1 Standard Call Option Approach: Merton/KMV

In the seminal work of Merton (1974), a firm’s equity $S$ is viewed as a European call option written on the firm’s asset value $V_t$. It is assumed that firm’s debt has the form of a pure discount bond of promised payment $(F)$ that matures at time $T$. Hence the firm does not default prior to the debt maturity date. On maturity date $T$, if $V_T > F$, then the firm pays off its debt and the value of equity at this time is $V_T - F$. If $V_T < F$, then firm will default on its debt obligations. The value of equity is hence zero. Therefore, the payoff of the firm’s equity at maturity time $T$ is

$$S(T) = \max(V_T - F, 0)$$ (3-1)

This shows the equity value is a call option on the assets of the firm, with strike price equal to the promised payments to debt holders $F$ on maturity day.

The market value of a firm’s equity, $S_t$, can be priced by the standard Black-Scholes (1973) option pricing formula and written as,

$$S_{Merton}(t) \equiv C_{BS}(V_t; \sigma, r, F, r, T-t) = V_t \Phi(d_1) - F e^{-r(T-t)} \Phi(d_1 - \sigma \sqrt{T-t})$$ (3)

where

$$d_1 = \ln\left(\frac{V_t}{F}\right) + \left(\frac{r + \sigma^2}{2}\right)(T-t)$$

and $\Phi(\cdot)$ is the standard normal distribution function, $r$ is the risk-free interest rate, $\tau \equiv T-t$ is the maturity of the option (or equivalent debt contract), $F$ is the strike price, equal to the sum of liabilities due within $\tau$.
years. The first term in Eq. (3) reflects the expected benefit from obtaining a stock outright. The second term measures the present value of paying the exercise price on the expiration date. The difference of the two terms is the payoff of holding the equity.

In standard call option approach, the firm is in default when the asset value falls below an exercise boundary\(^9\) at the maturity date. Under the real probability measure\(^10\), the default probability for Merton (1974) can be derived as:

\[
P_{\text{Merton}}(t) = P(V_t < F_t) = \Phi\left(-\frac{\ln(V_t / F) + (\mu_v - 0.5\sigma_v^2)(T-t)}{\sigma_v\sqrt{T-t}}\right)
\]  

(4)

where \(\frac{\ln(V_t / F) + (\mu_v - 0.5\sigma_v^2)(T-t)}{\sigma_v\sqrt{T-t}}\) is referred to as the ‘Distance-to-Default’ (DD) by Moody’s KMV. DD is the difference between the assets’ value and the total liabilities scaled by the standard deviation of the firm’s asset value, i.e. a measure of asset volatility adjusted leverage ratio. The larger the DD, the less likely a firm will default. It is obvious that firms with higher asset variability or higher financial leverage have smaller distance to default, hence a higher probability of default than firms without such characteristics.

To use \(P_{\text{Merton}}(t)\) to forecast bankruptcy, we assume that a firm has a discount bond outstanding with a face value equal to the sum of short-term debt plus half of long-term debt just as KMV does. But the Merton/KMV model which we’ll test in the empirical part later is different from that actually used by Moody’s KMV. The major difference is that we assume the asset value follows a lognormal distribution, and use a standard

---

\(^9\) The boundary is equal to the principal value of total debt in Merton (1974). The Moody’s KMV approach sets an exogenous barrier that is equal to the sum of short term debt plus half of long term debt (see Crosbie and Bohn, 2002), and argues that this choice captures adequately the financing constraints of firms. Vassalou and Xing (2004) do the same.

\(^10\) Alternative ‘risk neutral probability’ use risk-free interest rate \(r\) instead of firm’s expected asset return as the drift of its underlying asset. In general, risk neutral default probability is greater than or equal to real default probability if investors are assumed to be risk-averse.
normal distribution to convert DD into default probabilities. However, Moody’s does not need the GBM assumption of asset value, and constructed a proprietary Expected Default Frequency (EDF) database mapping DD into default probabilities. Since we do not have access to Moody’s proprietary database, we cannot perfectly replicate the methods of Moody’s KMV. Notice that if we rank firms by their relative default probability, the assumption of distribution of DD is innocuous. Therefore our results will “emphasize the model’s ability to rank firms by default risk rather than its ability to calculate accurate probabilities” as Bharath and Shumway (2003) does.

2.2 Down-and-Out Barrier Option (DOC) approach

Black and Cox (1976) introduce an early bankruptcy and view a firm’s equity as a down-and-out call option (DOC) written on the firm’s assets \( V \), with default barrier or knock-out value of the firm \( H \), asset volatility \( \sigma_v \) debt \( F \) with maturity \( \tau \equiv T - t \) as strike price, and risk-free interest rate. Via the closed-form formula for a DOC option valuation by Merton (1973), the equity value can be written as:

\[
S_{\text{barrier}}(t) \equiv C_{\text{DOC}}(V_t; \sigma_v, H, F, r, T - t) = V\Phi(a) - Fe^{-r(T-t)}\Phi(a - \sigma_v\sqrt{T-t})
\]

\[
- \left( \frac{H}{V} \right)^{2\eta} \left[ V\Phi(b) - F(V / H)^2 e^{-r(T-t)}\Phi(b - \sigma_v\sqrt{T-t}) \right]
\]

where \( \Phi(\bullet) \) is the standard normal distribution function, and \( a, b \) is path-dependent terms, relying on the value of \( F \) and \( H \):

\[
a = \left[ \ln(\frac{V_t}{\text{Max}(F, H)}) + (r + 0.5\sigma_v^2)(T - t) \right] \frac{\sigma_v\sqrt{T-t}}{\sigma_v\sqrt{T-t}}
\]
Following Brockman and Turtle (2003), we simplify our discussion in this paper by assuming a zero payment to the equity holders upon failure. Also when default barrier is set to zero $H=0$, the DOC option will collapse to a standard call option (SC) in Merton (1974). It implies that the conventional SC model is fully nested in our proposed DOC structural model.

The first two terms in Eq. (5) are expressions for the standard call option with strike $F$, volatility $\sigma_v$, and time to expiration $\tau \equiv T - t$. The last two terms are terms involving the barrier $H$, which guarantee that option value given by Eq. (5) is zero at the knockout value of the firm, i.e., when $V = H$. When the asset value hits the default point, the firm is assumed to default.

Empirical evidence (Wong and Choi (2004), Brockman and Turtle (2003)) supports the claim that equity behaves as a barrier option. Wong and Choi (2004), Hao (2005) and our paper have found that the barrier for industrial firms is much lower than the face value of debt, i.e. $H < F$. In this case, Eq. (5) can be rewritten as a compact form as below

$$C_{\text{DOC}}(V_t; \sigma_v, H, F, r, T - t) =$$

$$C_{\text{BS}}(V_t; \sigma_v, F, r, T - t) - \left(\frac{H}{V}\right)^{2q} C_{\text{BS}}(V_t; \sigma_v, FV^2 / H^2, r, T - t)$$

(5-1)
In the special case when a firm’s asset value is much larger than the barrier \((V >> H)\),
the option is unlikely to terminate early before expiration. The effect of the second term
in Eq.(5-1) is negligible, so the value of the call is approximately equal to that priced by
Black-Scholes formula (Derman and Kani, 1997). Investors of the firm are expected to
receive that payoff at expiration.

Given the low level of the default barrier, the estimated probability of bankruptcy in
the DOC option framework should take two possible situations into consideration: one is
that the firm’s asset value \(V\) hits the barrier \(H\) before debt maturity date (early
bankruptcy); the other is that conditional on no default (i.e. \(V \geq H\) before maturity), the
firm cannot pay off its maturing liabilities at the end of the period (late bankruptcy).\(^\text{11}\)

Based on the above logic, the implied closed-form solution for the failure
probability in a DOC option framework can be written as

\[
PD_{\text{barrier}}(t) = \Phi\left( \frac{\ln\left(\frac{F}{V}\right) - (\mu_V - \sigma_V^2 / 2)(T-t)}{\sigma_V \sqrt{T-t}} \right) + \exp\left( \frac{2(\mu_V - \sigma_V^2 / 2)\ln\left(\frac{H}{V}\right)}{\sigma_V^2} \right) \Phi\left( \frac{\ln\left(\frac{H^2}{V^2F}\right) + (\mu_V - \sigma_V^2 / 2)(T-t)}{\sigma_V \sqrt{T-t}} \right)
\]

when \(H < F\). For derivation of \(PD_{\text{barrier}}(t)\) and similar probability measures, see

As mentioned in Section 1, we use a DOC model with exogenous barrier level
depending on debt principal \(F\). That is, the barrier level is some fraction of debt principal.
So, we rewrite the default barrier in the following form:

\[
H = \alpha F \quad \text{where} \quad \alpha \text{ may be equal to, greater or less than one}
\]

\(^{11}\) In Brockman and Turtle’s framework, the firm-specific barrier levels are significantly large, mostly
exceeding debt level. So the second situation of late bankruptcy we discussed here is not applicable in BT’s
model.
\( \alpha \) is a parameter that needs to be estimated in our DOC model.

### 3. Estimation of structural models

In our paper, we use information from the stock market as well as balance sheet information for our empirical estimation. The major challenges in implementation are the estimation of the unobserved asset value \( V \) and the treatment of noise in equity prices. As we discussed in Section 1, among many alternatives of approach (such as proxy approach, volatility restriction approach, MLE), we will adopt Particle Filtering-MLE as a feasible estimation procedure for our DOC barrier option model specified in section 2. In this section, we first outline the transformed-data ML estimation procedures for DOC model assuming stock price is an accurate signal of the firm’s asset value; and then we introduce PF-MLE as a feasible method for estimating DOC model using noisy equity data.

#### 3.1 Case 1: DOC barrier option model without noise in equity prices

##### 3.1.1 Parameter estimation

In order to compute the default probabilities from this model, we need to estimate the parameters \( \theta = (\sigma_v, \mu_v, \alpha) \) using historic data, where \( \sigma_v \) is volatility of asset value, \( \mu_v \) is expected return of asset (or asset drift), and \( \alpha \) is a ratio related to default barrier, set as some fraction of the principal value of the debt, i.e. \( \alpha = H / F \), where \( H \) is default boundary, and \( F \) is the repayment on debt at maturity date.
3.1.2 Transformed-data maximum likelihood estimation

We apply the transformed-data MLE method of Duan (1994) to obtain the log-likelihood function of discretely sampled equity values on a firm that survives the entire sample period \( \{S_1, S_2, \ldots, S_n\} \) based on the assumption that the firm value is log-normally distributed and the equity value is an option on the firm’s assets. Then the parameters are obtained by maximizing the log-likelihood function for the equity values.

In our barrier option model, suppose an equity pricing formula \( S_i = DOC(V_i) \) is pre-specified as in Equation (5), and the logarithm of the asset value \( \ln V_i \) is latent and assumed to follow a normal distribution. Therefore, the density function of \( \ln V_i \) can be derived through Equation (2) as:

\[
f \left( \ln V_i \mid \ln V_{i-1} \right) = \frac{1}{\sigma \sqrt{2\pi \Delta t}} \exp \left\{ -\frac{\left[ \ln V_i - \ln V_{i-1} - \left( \mu_V - 0.5\sigma^2 \right) \Delta t \right]^2}{2\sigma^2 \Delta t} \right\}
\]

(8)

The Jacobian term can be written as:

\[
\frac{\partial S_i}{\partial \ln V_i} = \frac{\partial S_i}{\partial V_i} \times \frac{\partial V_i}{\partial V_i} = j'(V_i) \times V_i
\]

(9)

where \( j'(V_i) \) is called the option delta (i.e. the first derivative of the equity value with respect to the asset value). Note that the form of \( j'(V_i) \) depends on the equity pricing formula in a different model.\(^{12}\) Applying Jacobian transformation technique yields the following conditional likelihood function of \( S_i \):

\(^{12}\) In Merton’s model, delta of the standard option equals \( V_i \Phi(d_1) \), derived from Equation (3). We apply MLE method to estimate Merton’s model as well. The procedures are similar as that for DOC model described here.
\[ g(S_t | S_{t-1}) = f(\ln V_t | \ln V_{t-1}) \left( \frac{\partial S_t}{\partial \ln V_t} \right)^{-1} \]

\[ = \frac{1}{\sigma_v \sqrt{2\pi \Delta t}} \times \exp \left\{ -\frac{\left[ \ln V_t - \ln V_{t-1} - \left( \mu_v - 0.5\sigma_v^2 \right) \Delta t \right]^2}{2\sigma_v^2 \Delta t} \right\} \left( j'(V_t) \times V_t \right)^{-1} \] (10)

It turns out to be the log-likelihood function of the firm’s asset value plus a term related to the Jacobian of the transformation. The log-likelihood function for joint equity values \( \{S_1, S_2, \ldots, S_n\} \) in the DOC barrier option model can be written as,

\[ L(\mu_v, \sigma_v; S_1, S_2, S_3, \ldots, S_n) = \sum_{i=2}^{n} \left\{ \ln f(\ln V_i | \ln V_{i-1}) - \ln \left( \frac{\partial S_i}{\partial \ln V_i} \right) \right\} \]

(11)

\[ = -\frac{n-1}{2} \ln (2\pi \sigma_v^2 \Delta t) - \frac{1}{2} \sum_{i=2}^{n} \frac{\left[ \ln V_i - \ln V_{i-1} - \left( \mu_v - 0.5\sigma_v^2 \right) \Delta t \right]^2}{2\sigma_v^2 \Delta t} - \sum_{i=2}^{n} \ln(V_i) - \sum_{i=2}^{n} \ln(\Delta V_i) \]

where \( V_i = DOC^{-1}(S_i; \mu_v, \sigma_v, \alpha) \), \( j'(V_i) \) is the delta of barrier option, derived from Equation (5), taking the following form.

\[ j'(V_i) \equiv \frac{\partial S_i}{\partial V_i} = \Phi(a_i) + \varphi(a_i) \frac{1}{\sigma_v \sqrt{T-t}} - F \varphi(b_i - \sigma_v \sqrt{T-t}) \frac{1}{\sigma_v \sqrt{T-t} V_i} \]

\[ - \left( \frac{H}{V_i} \right)^{2\eta} (1 - 2\eta) \Phi(b_i) + \left( \frac{H}{V_i} \right)^{2\eta} \varphi(b_i) \frac{1}{\sigma_v \sqrt{T-t}} \]

\[ + F \left( \frac{H}{V_i} \right)^{2\eta} (2 - 2\eta) \frac{1}{V_i} \Phi(b_i - \sigma_v \sqrt{T-t}) - F \left( \frac{H}{V_i} \right)^{2\eta-2} \varphi(b_i - \sigma_v \sqrt{T-t}) \frac{1}{\sigma_v \sqrt{T-t} V_i} \] (12)

where \( \varphi(\bullet) \) is density function of standard normal distribution, early default barrier is some fraction of debt principal value \( H = \alpha F \).

### 3.1.3 Survivorship consideration

Duan et al.(2004) adopts another joint density function to take into account the fact that for a surviving company, its asset value must have stayed above the barrier for the
entire sample period. The original approach of Duan (1994) leads to an upward bias in
the asset drift, thus a biased probability of default measure. Following Duan et al. (2004),
we use the following log-likelihood function in our MLE implementation, which corrects
the survivorship bias.

\[
L' = L + \sum_{\tau=2}^{n} \ln \left( 1 - \exp \left( -\frac{2}{\sigma_y^2 \Delta t} \ln \frac{V_{\tau-1} V}{\alpha F} \ln \frac{V_{\tau}}{\alpha F} \right) \right) \\
- \ln \left( \Phi \left( \frac{\mu_y - 0.5 \sigma_y^2 (n-1) \Delta t - \ln \frac{\alpha F}{V_0}}{\sqrt{(n-1) \Delta t \sigma_y}} \right) \right) \\
- \exp \left( \frac{2}{\sigma_y^2} (\mu_y - 0.5 \sigma_y^2) \left( \ln \frac{\alpha F}{V_0} \right) \right) \Phi \left( \frac{\mu_y - 0.5 \sigma_y^2 (n-1) \Delta t + \ln \frac{\alpha F}{V_0}}{\sqrt{(n-1) \Delta t \sigma_y}} \right)
\]

ML estimators are parameters that maximize the likelihood function \( L' \) in equation (13),
subject to the constraints that the market values of equities are equal to the DOC barrier
option pricing that is,

\[
\max L'(\mu_y, \sigma_y, \alpha) \quad \text{s.t.} \quad S_t = DOC(V_t; \sigma_y, \alpha, F, r, T - t), \forall t = 1, 2, ..., n. \quad (14)
\]

### 3.1.4 Implementation procedures of MLE in DOC model

The implementation steps of the MLE approach are as follows:

1. Given an initial guess of \( \mu_y, \sigma_y, \alpha \), \(^{13}\) calculate the firm’s asset value \( V_0 \) by

   solving equation (5), i.e., the DOC option pricing formula.

---

\(^{13}\) We thank H.Y.Wong for suggestion on solving the convergence problem in DOC model. As the
likelihood function is not smooth at \( \alpha = 0 \) or \( \alpha = 1 \), thus, we may obtain a singular Hessian matrix. So
we vary the initial value for alpha through a grid search method from zero to two. After computing the
likelihood function for a series of initial alpha, we pick the one with the largest value, and get
corresponding estimates. We also thank H. Hao for providing part of the Matlab codes used in this paper.
2. Plug $V^0$ into the likelihood equation (13). Maximize it with respect to $\mu_V, \sigma_V, \alpha$ respectively, and obtain new estimates of asset drift ($\mu_V$), asset volatility ($\sigma_V$) and barrier level ratio ($\alpha$)

3. Given the new estimates from the previous step, calculate $V'$

4. Repeat step 2 and step 3 until $L'$ the likelihood converges (i.e. smaller than a pre-specified tolerance level).

The optimization software we use is Matlab’s medium-scale line search.\(^\text{14}\)

3.2 Case 2: DOC barrier option model with possible noise in equity prices

When the observed equity prices $S$ contain trading noise, the one-to-one relationship between equity value and firm values no longer holds. Equity prices are best viewed as noisy signals about the asset value.\(^\text{15}\) So we cannot back out the unobserved asset value $V$ from the DOC equity pricing formula (Equation (5)) via conventional methods such as Newton-Ralphson algorithm. Therefore, the estimation of the unobserved asset value $V$ becomes a complicated nonlinear filtering problem. In this subsection, we set up a brief analytical framework for Case 2; the procedures of MLE estimation via smoothed localized Sampling/Importance Resampling particle filter are described in Section 4.

A General Setup for Estimation

Assume we observe the equity prices $S$ at a fixed frequency $z$ (for daily data, $z=1/250$). Our econometric model consists of a transition (or state) equation for the latent

\(^{14}\) It’s computation-intensive in processing, e.g. on a Windows version of Matlab running on a PC with a 3.0 GHz Pentium 4 processor and 1.0G of RAM, the estimation of parameters in Merton model for a single firm with 252 observations (the number of trading days in a year) takes around 10 minutes; or equivalently, it takes 30 days running for one PC to estimate 4320 firm-year observations.

\(^{15}\) See Max (2005).
variable $V$, describing the dynamics of asset values, and a measurement (or observation) equation that maps asset values $V$ into the observed equity price $S$.

### 3.2.1 The transition equation

The asset value $V$ follows a geometric Brownian motion as in Equation (1), and $\log V$ is normally distributed. We can write the discrete time form of Equation (2) as

$$\ln V_{t+1} = \ln V_t + \left(\mu_V - 0.5\sigma_V^2\right) t + \sigma_V \sqrt{t} \epsilon_t \quad \text{where} \quad \epsilon_t \sim N(0,1) \quad (15)$$

### 3.2.2 The observation equation

Since the size of trading noise is expected to be proportional to the stock price, we take a multiplicative log-normal observation error as a natural choice for noise,

$$S_t = DOC(V_t; \sigma_V, \alpha, F, r, T-t) \cdot \exp(\delta \zeta_t) \quad \text{where} \quad \zeta_t \sim N(0,1) \quad (16)$$

$DOC(V_t; \sigma_V, \alpha, F, r, T-t)$ is the nonlinear option pricing function for equity given in Equation (5) and (7). The logarithmic of equity prices are as follows:

$$\ln S_t = \ln DOC(V_t; \sigma_V, \alpha, F, r, T-t) + \delta \zeta_t \quad \text{where} \quad \zeta_t \sim N(0,1) \quad (17)$$

### 3.2.3 The state space model

The state space model we will estimate consists of Equation (15) and (16), and $\epsilon_t \perp \zeta_t$.\textsuperscript{16} It contains four parameters, denoted by $\theta = (\sigma_V, \delta, \mu_V, \alpha)$. We implement the model using daily stock data, but all parameters will be expressed in annual units below. The equations in the system specify how the unobserved state of asset value $V$ is linked to the observed stock price. This relationship enables us to infer the firm asset value path using the stock data. Note that specification issues can arise, as our estimates for the magnitude of trading noise, $\delta$, depend on the use of DOC model of Black and

\textsuperscript{16}It’s the most easy to implement. Extensions could be made by taking the correlations of $\epsilon_t$ and $\zeta_t$ into account, or by using a distribution with fatter tails for the noise structure.
Cox (1976), and the multiplicative error structure assumption. Following Duan et al. (2009), we will conduct a specification check in Section 6. Next we describe the particle filter based MLE method and its application in DOC barrier model.

4 Estimation using the Particle Filter

For the more general case (Section 3.2, case 2 above) of DOC barrier option model with possible noise in equity prices, we propose the method of MLE via particle filter.

4.1 Particle filter (PF)

Particle filters, also known as sequential Monte Carlo methods, are sophisticated model estimation techniques based on simulation. It has been used extensively in the engineering field, and recently in financial applications.\(^\text{17}\) The PF provides a convenient filter for non-linear models such as the structural credit risk models we consider here. The technique relies on a set of discrete points (or particles) that are updated iteratively to approximate the true distribution of the unobserved state variable at different stage. Because the PF procedure is relatively new in finance, we provide the details of this method in the Appendix.

For the state space model we construct in Section 3.2, by Bayes’ theorem, the filtering density \( f(V_{t+1} \mid S_{t+1}, \theta) \) can be approximated by:

\[
\begin{align*}
\hat{f}(V_{t+1} \mid S_{t+1}, \theta) &\propto f(S_{t+1} \mid V_{t+1}, \theta) \hat{f}(V_{t+1} \mid S_t, \theta), \\
&= \frac{1}{M} \sum_{i=1}^{M} f(V_{t+1} \mid V_t^i, \theta)
\end{align*}
\]

where \( \hat{f}(V_{t+1} \mid S_t, \theta) \).

\(^\text{17}\) See Gordon, Salmond and Smith(1993) and Doucet et al.(2001) for information on the theory and application of particle filtering.
The Sampling/Importance Resampling (SIR) method by Rubin(1987) is a very commonly used particle filtering method. One simple algorithm is as follows:

1) Sample $M$ particles $V_{t+1}^i$ from $f(V_{t+1} | V_t^i, \theta)$, via transition equation (15),

2) Give each particle a filtering weight of $W_{t+1}^i = \frac{\omega_{t+1}^i}{\sum_{i=1}^{M} \omega_{t+1}^i}$, where

$$\omega_{t+1}^i = f(S_{t+1} | V_{t+1}^i, \theta)$$

3) Resample from the weighted sample $\{V_{t+1}^{(i)}, W_{t+1}^{(i)}, i = 1, ..., M\}$ to obtain a new equal weight sample of size $M$.

4) Repeat step 1), 2), and 3) for $t = 1, ..., T$.

The above SIR particle filter is not efficient, as it has not taken into account the information in $S_{t+1}$. Following Duan (2009), we utilize the “auxiliary filtering” idea of Pitt and Shephard (1999) to design our particle filter for the DOC barrier option model. In first step, we sample a pair $(V_t^i, V_{t+1}^i)$ instead of a point $V_{t+1}^i$, and the weight assigned to each particle needs to be adjusted accordingly. Our sampler takes advantage of the knowledge of $S_{t+1}$, in that the sampling of $V_{t+1}$ is around the implied asset value by $S_{t+1}$ under no trading noise.

Another key challenge in the use of the likelihood function evaluated via particle filter is that it is not generally smooth in the underlying parameters. For this reason, we use the “smooth bootstrap resampling” method proposed by Pitt(2002). The smoothing enables gradient based optimization and computation of parameter standard errors using conventional first order techniques.
4.2 Firm value filtering scheme for DOC barrier model

In this section, we describe our smoothed auxiliary particle filter sampler for DOC barrier option model.\(^{18}\)

**Step 1: Sampling**

Suppose a set of equal weight particles \(\{V^i_t\}_{i=1}^M\) are known at time \(t\). Draw \(\zeta_{t+1}^i\) from standard normal distribution, and compute \(V^i_{t+1} = DOC^{-1}\left(S_{t+1} \exp\left(-\delta_{t+1}^i\right)\right)\) to generate \((V^i_t, V^i_{t+1})\).

**Step 2: Assign to each sample point \(V_{t+1}^i\) a normalized importance weight \(^{19}\) of**

\[
W_{t+1}^i = \frac{\omega_{t+1}^i}{\sum_{i=1}^M \omega_{t+1}^i}, \quad \text{where} \quad \omega_{t+1}^i = \frac{f(V_{t+1}^i \mid V_t^i, \theta)}{j'(V_{t+1}^i) \exp(\delta_{t+1}^i)}
\]

\(\theta = (\sigma, \delta, \mu, \alpha)\), and \(j'(V_{t+1}^i)\) is the delta of barrier option, as defined in (12) being evaluated at \(V_{t+1}^i\).

**Step 3: Resampling**

The motivation for this step is to draw particles from the current particle set with probabilities proportional to their weights. That is, we want to sample more times from particles that have higher likelihood and eliminate the low probability particles. This will improve the statistical efficiency. We use the resampling method proposed by Pitt(2002):

\(^{18}\) The particle filter algorithm is coded and implemented in Microsoft Visual C++ 6.0 environment. A dll function file is outputted from Visual C++, and acts as an interface connecting two software environments(Matlab and Visual C++) for fulfillment of particle filtering estimation. We thank Professor J.C.Duan for providing its original codes. We modified the codes and extend it to the situation of a DOC model.

\(^{19}\) Derivation available upon request.
construct a smoothed empirical distribution using the weighted sample \[\{V_{i+1}^{(i)}, W_{i+1}^{(i)}, i = 1, ..., M\}\], then resample a new equal weight sample of size \(M\).

Step 1, step 2 and step 3 are repeated for \(t = 1, \ldots, T\). Once the particles and weights have been computed for each date, we are ready to construct the filtered path of \(V_t\) by

\[\bar{V}_t = \sum_{i=1}^{M} W^i_t V^i_t\] for each \(t\). The PF thus delivers a time series of filtered state variable \(V\).

4.3 Maximum likelihood estimation

We use one year of daily stock returns to estimate the state space model specified in section 3.2.3 by maximizing the likelihood. To this end, we need a method to derive a likelihood function for the observed equity values in a model with a latent variable. Pitt (2002) shows that one may simply average the importance weights across particles, take logs, and sum over time to create a log likelihood function as follows:

\[L^* = \sum_{t=1}^{T} \ln \left( \frac{1}{M} \sum_{i=1}^{M} W^i_t \right) \tag{20}\]

The parameters of the state space model \(\theta = (\sigma_r, \delta, \mu_r, \alpha)\) can then be estimated by maximizing equation (20). The estimates will then be plugged into Equation (6) to calculate the probability of default \(PD_{noise}^{DOC}\).

\[\text{After step 2, we drop the first entry of the pair } (V^i_t, W^i_{i+1}) \text{, and get weighted sample for } V^{i}_{i+1}.\]
5. Estimation with Empirical Data

5.1 Description of the Data

Our sample consists of U.S. industrial corporations, with average total assets of more than $50 million between calendar year 1999 and 2008. At this stage, our analysis is focused on transportation industry, because these firms are strongly affected by the business cycle. Specifically, we included firms with SIC code 37 (transportation equipment), 40 (Railroads), and 41, 42, 44, 45, 47 (other transportation). The accounting data and SIC codes are from Compustat Industrial Database, while market information of equity are from CRSP Daily Stock Return File.

Our data of firm bankruptcies are drawn from three sources:

(1) The Altman-NYU Salomon Center Bankruptcy List, which includes detailed information of all Chapter 7 and Chapter 11 bankruptcies filed during 2000 and 2008 by public firms. (2) Compustat’s research file and footnotes, with footnote 35 having code 02 (filing for Chapter11) or 03 (chapter 7). (3) CRSP’s bankruptcy delisting code (572 or 574). Not all bankruptcies can be used in our analysis, as many of them are related to either financial or foreign firms. To compare out-of-sample bankruptcy forecast accuracy, we also need a control group of non-bankrupt firms for each calendar year. We follow Dichev (1998) and Hao (2005) to use a broad definition of financial distress to ensure that the sample of non-bankrupt firms are relatively distress-free: any firm with CRSP

---

21 We restrict our sample to non-financial firms, (SIC codes 6000 ~ 6999 are excluded), so that the leverage ratios are more relatively comparable across firms.
22 We thank D.F. He for providing us access to some portion of the database.
23 In DOC model, Bankruptcy is defined as either the underlying asset value crosses the barrier from above at any time before maturity; or conditional on no default (i.e. having always stayed above the barrier), the underlying asset value is greater than barrier, but less than the amount of debt owed at maturity.
24 We thank Professor Altman for granting us usage of this data set.
delisting codes 400, 550-585 (bankruptcy, liquidation or poor performance) for a calendar year or within the next two years is disqualified as a non-bankrupt firm for that calendar year.

Since we need both share price data and accounting data in MLE estimation via particle filter, our samples include only firms that are listed in the intersection of Compustat and CRSP database between year 1999 and 2008. After cleaning the data, our samples consist of 1731 firm-year observations with 411,916 daily observations \(^{25}\), representing 272 individual transportation firms and 30 initial corporate bankruptcies. The distribution of bankrupt and non-bankrupt firms is shown in table 1.

### Table 1: Distribution of Bankrupt and Non-bankrupt Firms

The second column of the table reports the number of firms for which our PF-MLE could be applied each year. The third column reports the number of firms that filed for bankruptcy (Chapter 7 or Chapter 11) for each year in our sample period.

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of Firms in sample</th>
<th>Bankrupt Firms</th>
<th>Nonbankrupt Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>213</td>
<td>3</td>
<td>210</td>
</tr>
<tr>
<td>2000</td>
<td>189</td>
<td>4</td>
<td>185</td>
</tr>
<tr>
<td>2001</td>
<td>173</td>
<td>3</td>
<td>170</td>
</tr>
<tr>
<td>2002</td>
<td>162</td>
<td>6</td>
<td>156</td>
</tr>
<tr>
<td>2003</td>
<td>153</td>
<td>3</td>
<td>150</td>
</tr>
<tr>
<td>2004</td>
<td>161</td>
<td>2</td>
<td>159</td>
</tr>
<tr>
<td>2005</td>
<td>159</td>
<td>5</td>
<td>154</td>
</tr>
<tr>
<td>2006</td>
<td>157</td>
<td>1</td>
<td>156</td>
</tr>
<tr>
<td>2007</td>
<td>156</td>
<td>1</td>
<td>155</td>
</tr>
<tr>
<td>2008</td>
<td>153</td>
<td>2</td>
<td>151</td>
</tr>
</tbody>
</table>

\(^{25}\) Direct data exchange between SAS and Matlab is ruled out. Yet Excel’s involvement imposes limits on the size of a dataset that can be moved in a single pass. e.g. Size of an Excel 2002 spreadsheet is limited to 65,536 rows and 256 columns. We use MySQL software for high-volume data manipulation in Matlab, and as the “middleman” for data exchange between SAS and Matlab.
5.2 Variable Specification

Table 2 presents the summary statistics for the variables used.

The inputs to DOC option model include the face value of debt \((F)\), the risk free rate\((r)\), the maturity of option life\((\tau \equiv T - t)\), and the market value of firm’s equity \((E, \text{ in millions of dollars})\). For \(r\) the risk free rate, we use 1-Year Treasury Constant Maturity Rate obtained from the Board of Governors of the Federal Reserve System, available at [http://research.stlouisfed.org/fred/data/irates/gs1](http://research.stlouisfed.org/fred/data/irates/gs1). The equity values are calculated from CRSP as the product of daily closing share price and the number of shares outstanding.

Following Brockman and Turtle (2003), we always take the option’s life \(\tau \equiv T - t\) to be 10 years for backing out structural parameters \(\theta = (\sigma, \delta, \mu, \alpha)\), which allows us to make a fair comparison with other empirical studies. Accordingly, the strike price \(F\) is set to total liabilities of the firm that are due within 10 years (Compustat data item 54). This measure for strike price \(F\) is different from that in Moody’s KMV(2002) or Vassalou and Xing(2003), which use current debt (Compustat data item 45) plus one half of long term debt (Compustat data item 51). In their work, the maturity of option life \((\tau \equiv T - t)\) is set to 1, or 2, or 3…years, depending on the default prediction horizon. As Reisz and Perlich(2007) pointed out, it is unrealistic to assume that the firm has a life of 1, 2, or 3 years and back out market asset value and its volatility again.

We use Compustat Industrial Quarterly Database to get the firm’s book value of debt. We also collect from Compustat Annual Database each firm’s accounting ratios of working capital to total assets, retained earnings to total assets, sales to total assets, market value of equity to total liabilities, etc. These variables are used in the alternative bankruptcy prediction models (such as Shumway’s hazard model, or Altman’s Z-score).
Table 2: Summary Statistics for the sample

Table 2 reports summary statistics for the transportation firms’ data we used. $S$ is in millions of dollars, taken from CRSP as the total number of shares outstanding times the stock price. $F$ is face value of debt in millions of dollars (Compustat item 54) in our DOC model estimation. Risk-free interest rate is measured as 1-Year Treasury Constant Maturity Rate. WC/TA is the firm's ratio of working capital to total assets, RE/TA is ratio of retained earnings to total assets, EBIT/TA is earnings before interest and taxes divided by total assets, ME/TL is the ratio of market value of equity to book value of total liabilities, Sale/TA is the sales/total assets ratio. NI/TA is the firm's ratio of net income to total assets. TL/TA is total liabilities divided by total assets. Relative return is firm's prior year return relative to the CRSP value weighted index return in the same period. Relative size is the logarithm difference between firm's market equity capitalization and the total market value from the CRSP file.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std.</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market value of equity ($S$)</td>
<td>3164</td>
<td>8307</td>
<td>453.038</td>
<td>0.19861</td>
<td>84433</td>
</tr>
<tr>
<td>Debt value ($F$)</td>
<td>6085</td>
<td>33447</td>
<td>370.61</td>
<td>0.335</td>
<td>460442</td>
</tr>
<tr>
<td>Debt to equity ratio ($F/S$)</td>
<td>2.695</td>
<td>8.8028</td>
<td>0.9092</td>
<td>0.0046</td>
<td>532.33</td>
</tr>
<tr>
<td>Risk free interest rate ($r$)</td>
<td>3.571</td>
<td>1.6709</td>
<td>3.64</td>
<td>0.49</td>
<td>6.33</td>
</tr>
<tr>
<td>WC/TA (item 179/item6)</td>
<td>0.142</td>
<td>0.1813</td>
<td>0.1111</td>
<td>-0.562</td>
<td>0.616</td>
</tr>
<tr>
<td>RE/TA (item 36/item 6)</td>
<td>0.144</td>
<td>0.4146</td>
<td>0.2019</td>
<td>-2.063</td>
<td>0.892</td>
</tr>
<tr>
<td>EBIT/TA (item 178/item6)</td>
<td>0.077</td>
<td>0.0896</td>
<td>0.0789</td>
<td>-0.318</td>
<td>0.316</td>
</tr>
<tr>
<td>ME/TL (item 199*item25/item 181)</td>
<td>2.361</td>
<td>3.7312</td>
<td>1.1345</td>
<td>0.006</td>
<td>24.457</td>
</tr>
<tr>
<td>Sale/TA (item 12/ item6)</td>
<td>1.324</td>
<td>0.7912</td>
<td>1.1478</td>
<td>0.204</td>
<td>3.979</td>
</tr>
<tr>
<td>NI/TA (item 172/item 6)</td>
<td>0.0315</td>
<td>0.0902</td>
<td>0.0393</td>
<td>-0.434</td>
<td>0.216</td>
</tr>
<tr>
<td>TL/TA (item 181/ item6)</td>
<td>0.6055</td>
<td>0.2145</td>
<td>0.606</td>
<td>0.105</td>
<td>1.246</td>
</tr>
<tr>
<td>Relative return</td>
<td>0.0558</td>
<td>0.5096</td>
<td>-0.0154</td>
<td>-0.817</td>
<td>2.184</td>
</tr>
<tr>
<td>Relative Size</td>
<td>-10.129</td>
<td>1.9603</td>
<td>-10.15</td>
<td>-14.48</td>
<td>-5.653</td>
</tr>
</tbody>
</table>

Note:
1. In KMV/Merton model estimation, $F$, liability, is computed as Compustat item 45+0.5*Compustat item 51, i.e. debt in current liabilities plus half of the long-term debt.
2. All items with numbers (such as item 36, item 6, etc.) are from Compustat.
model) we consider later in the paper. To ensure that statistical results are not heavily influenced by extreme values, we winsorize all the financial ratio variables constructed from Compusat data. We set all observations higher (lower) than the 99th (1st) percentile of each variable to that value.

In order to make sure that all the information used to calculate default probability is observable at the time of prediction\textsuperscript{26}, we lag the quarterly (annual) data to ensure that each firm’s fiscal quarter (year) ends at least two (six) months before the beginning of the new fiscal quarter (year).

5.3 Empirical Results

For our DOC model with possible trading noise in equities, we run the estimation using the 1000-particle smoothed auxiliary SIR filter. Since a firm’s leverage will experience cyclical variation, we run the estimation for each firm for each year, so the asset volatility and exogenous bankruptcy barrier varies over the years.

5.3.1 Parameter Estimation

5.3.1.1. Implied default barrier vs. Debt ($\alpha$)

Table 3 reports the statistics of the estimates for the barrier-to-debt ratio ($\alpha$) for firms in transportation related industries from 1999-2008. The estimation is done for each firm for each year, using daily data. Consequently the number of estimations is enormous. We can see from Panel A that over 75% of the whole sample has a non-zero default barrier. The median firm has a default barrier at around 19% of its liabilities. The firms in

\textsuperscript{26} The federal securities laws require publicly traded companies to disclose information on an ongoing basis. Form 10-K (annual report) had to be filed with the SEC within 90 days after the end of the company's fiscal year. The deadline for filing their quarterly reports is within 45 days. Some firms have exceptions.
Table 3. MLE estimates of Barrier-to-debt ratio ($H/F$)

The table reports MLE estimates of the barrier to debt ratio. Sample includes 1676 firm-year data for 30 unique bankrupt firms and 1544 non-bankrupt firm-year data from 1999 to 2008. Industries are classified by SIC code. Estimation is done for each firm for each year, using daily data.

<table>
<thead>
<tr>
<th>Number of firm-year observation</th>
<th>Percentile</th>
<th>Proportion of $H &lt; F$</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
<td>25%</td>
<td>50%</td>
</tr>
<tr>
<td>Pooled 1676</td>
<td>0</td>
<td>0.029</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Panel A: Barrier-to-debt ratio estimates

<table>
<thead>
<tr>
<th>Year</th>
<th>Pooled</th>
<th>Number of firm-year observation</th>
<th>Percentile 5%</th>
<th>Percentile 25%</th>
<th>Percentile 50%</th>
<th>Percentile 75%</th>
<th>Percentile 95%</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>213</td>
<td>0</td>
<td>0.008</td>
<td>0.061</td>
<td>0.4</td>
<td>2.035</td>
<td>0.374</td>
<td>0.911</td>
</tr>
<tr>
<td>2000</td>
<td>189</td>
<td>0</td>
<td>0.007</td>
<td>0.0589</td>
<td>0.367</td>
<td>1.718</td>
<td>0.393</td>
<td>0.926</td>
</tr>
<tr>
<td>2001</td>
<td>173</td>
<td>0</td>
<td>0.032</td>
<td>0.244</td>
<td>0.809</td>
<td>2.243</td>
<td>0.571</td>
<td>0.809</td>
</tr>
<tr>
<td>2002</td>
<td>162</td>
<td>0</td>
<td>0.035</td>
<td>0.256</td>
<td>0.711</td>
<td>1.608</td>
<td>0.525</td>
<td>0.864</td>
</tr>
<tr>
<td>2003</td>
<td>153</td>
<td>0</td>
<td>0.054</td>
<td>0.261</td>
<td>0.55</td>
<td>1.343</td>
<td>0.44</td>
<td>0.876</td>
</tr>
<tr>
<td>2004</td>
<td>161</td>
<td>0</td>
<td>0.044</td>
<td>0.169</td>
<td>0.43</td>
<td>1.055</td>
<td>0.382</td>
<td>0.944</td>
</tr>
<tr>
<td>2005</td>
<td>159</td>
<td>0.001</td>
<td>0.028</td>
<td>0.16</td>
<td>0.462</td>
<td>1.7</td>
<td>0.4</td>
<td>0.887</td>
</tr>
<tr>
<td>2006</td>
<td>157</td>
<td>0</td>
<td>0.051</td>
<td>0.149</td>
<td>0.279</td>
<td>1.43</td>
<td>0.35</td>
<td>0.905</td>
</tr>
<tr>
<td>2007</td>
<td>156</td>
<td>0</td>
<td>0.041</td>
<td>0.3</td>
<td>0.833</td>
<td>3.671</td>
<td>0.73</td>
<td>0.801</td>
</tr>
<tr>
<td>2008</td>
<td>153</td>
<td>0.006</td>
<td>0.285</td>
<td>0.812</td>
<td>1.657</td>
<td>4.246</td>
<td>1.21</td>
<td>0.549</td>
</tr>
</tbody>
</table>

Panel B: Barrier-to-debt ratio by year

Panel C: Barrier-to-debt ratio by industry

<table>
<thead>
<tr>
<th>Industry</th>
<th>Pooled</th>
<th>Number of firm-year observation</th>
<th>Percentile 5%</th>
<th>Percentile 25%</th>
<th>Percentile 50%</th>
<th>Percentile 75%</th>
<th>Percentile 95%</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transportation Equipment (37)</td>
<td>736</td>
<td>0</td>
<td>0.031</td>
<td>0.204</td>
<td>0.598</td>
<td>1.94</td>
<td>0.516</td>
<td>0.834</td>
</tr>
<tr>
<td>Railroad (40)</td>
<td>82</td>
<td>0</td>
<td>0.055</td>
<td>0.389</td>
<td>0.874</td>
<td>1.263</td>
<td>0.531</td>
<td>0.829</td>
</tr>
<tr>
<td>Other transportation (41,42,44,45,47)</td>
<td>858</td>
<td>0</td>
<td>0.026</td>
<td>0.169</td>
<td>0.569</td>
<td>2.318</td>
<td>0.538</td>
<td>0.864</td>
</tr>
</tbody>
</table>
the top quartile of the whole sample have a barrier greater than 60% of total liabilities. Our analysis also indicates that roughly 34% of firms have a barrier that is statistically different from zero at the 5% level. All these support the inclusion of a default barrier in model specification. Furthermore, 85.1% of the whole sample has a ratio of less than 1, implying most firms have a barrier below their total liability. We then disaggregate the sample by years, and report the results in Panel B. We observe that the medians of the barrier-to-debt ratios exhibit a pattern of increase in the range of year 2001-2003 and 2006-2008 (Figure 1). It is interesting to notice that recessions hit the economy in both ranges. One possible explanation is that investors require more protection during times of financial stability, thus pricing equities with a higher default barrier level. Panel C presents \( \alpha \) estimates for firms in transportation related industries identified by two digits SIC code. The three subgroups show similar distribution pattern of barrier level. As in Wong and Choi (2004) and Hui (2005), our estimates are much lower than that of

![Figure 1. Barrier-to-debt Ratio for transportation industry.](image)

The Figure shows the median of the barrier-to-debt ratios estimates for transportation industry from 1999 to 2008. The ratios are derived from estimation for each firm for each year, using daily data.
Brockman and Turtle (2004). The reason is that they use a proxy approach for estimation, which leads to upward bias in $\alpha$.

5.3.1.2 Trading noise ($\delta$)

Table 4 reports the estimates of parameters $\theta = (\sigma_y, \delta, \mu_y, \alpha)$ from MLE via particle filtering. The mean and median values of the noise measure are 0.369 and 0.296 respectively. Recall the noise structure $\delta \xi_t$ in Equation (17), where $\xi_t \sim N(0,1)$. So for $\delta = 0.3$, it implies that if the trading noise is 1 standard deviation, then the trading noise amounts to 0.3% of the stock price. Some firms face negligible trading noise ($\delta = 0$). Though the estimates for noise are not big in magnitude, the omission of trading noise may have a real impact on the volatility estimates. This point is shown clearly in summary result of asset volatility ($\sigma_y$) estimates.

Table 4 Particle Filter-MLE estimates for transportation industry

The table presents the particle-filter MLE estimates for transportation industry from 1999 to 2008. An estimation is done for each firm for each year, using daily data. $\sigma_{wo} / \sigma$ is the ratio of asset volatility estimates without trading noise over the one with. $\sigma_{wo}$ is obtained through MLE method, $\sigma$ is the estimates from particle filter method.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>5%</th>
<th>25%</th>
<th>75%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>0.369</td>
<td>0.296</td>
<td>0.128</td>
<td>0.215</td>
<td>0.424</td>
<td>0.817</td>
</tr>
<tr>
<td>$100*\delta$</td>
<td>1.422</td>
<td>0.3</td>
<td>0</td>
<td>0.187</td>
<td>0.594</td>
<td>4.252</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.08</td>
<td>0.084</td>
<td>-0.55</td>
<td>-0.114</td>
<td>0.268</td>
<td>0.69</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.528</td>
<td>0.19</td>
<td>0</td>
<td>0.03</td>
<td>0.605</td>
<td>2.109</td>
</tr>
<tr>
<td>$\sigma_{wo} / \sigma$</td>
<td>1.0556</td>
<td>1.0121</td>
<td>0.999</td>
<td>1</td>
<td>1.042</td>
<td>1.252</td>
</tr>
</tbody>
</table>
Check the Nature of Noise Estimates

According to market microstructure theory, the more liquid a firm, the smaller trading noise $\delta$ it should have. To ascertain the nature of our empirical noise estimates $\hat{\delta}$ (i.e. make sure it is in line with our belief of microstructure noise, not simply due to mis-specification error of DOC model), we conduct a series of rank correlation tests between our noise estimates ($\hat{\delta}$) and commonly adopted measure of market liquidity across firms in our sample. We follow Duan and Fulop (2009) and use their proxy measures for market liquidity, namely, percentage bid-ask spread, firm size and trading volume. Firm size is measured as the product of stock price and number of shares outstanding. For each firm year observation in our sample, we take average of the daily values for each liquidity measure over that specific year. It’s expected that firms with larger market equity capitalization or trading volume, or with smaller closing ask-bid difference, will have a smaller trading noise. Table 5 presents our result. As expected, both the firm size and trading volume are negatively and significantly related to the noise estimate $\delta$ in all three tests (Spearman, Pearson, Kendall’s rank correlation$^{27}$) at 99% confidence interval. Also, the expected positive relationship between $\delta$ and trading volume is indicated in all tests except Pearson test. Even so, the unexpected negative estimate in Pearson test is insignificant both economically and statistically. All these results support our estimates of noise as intuitively plausible.

\footnote{\textsuperscript{27} For details of Spearman, Pearson, Kendall’s test, refer to Appendix.}
Table 5: Relationship of the noise estimates with alternative measures of liquidity

The table presents the correlation coefficient between our noise estimate and alternative measures of liquidity. Our noise estimates are derived from estimation for each firm for each year using daily data. The analysis is implemented on transportation industry from 1999 to 2008. For each firm year observation in our sample, we take average of the daily values for each liquidity measure over that specific year to derive corresponding liquidity measure. Cross-section sample size is 1676. p- values are in parenthesis. Firm size is calculated as stock price times shares outstanding from CRSP.

<table>
<thead>
<tr>
<th></th>
<th>Pearson Correlation Coefficients</th>
<th>Spearman Correlation Coefficient</th>
<th>Kendall Tau b Correlation Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated Noise</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Percentage Bid-Ask</td>
<td>-0.00033</td>
<td>0.27</td>
<td>0.1843</td>
</tr>
<tr>
<td>Spread</td>
<td>(0.9894)</td>
<td>(&lt;0.0002)</td>
<td>(&lt;.0001)</td>
</tr>
<tr>
<td>log(Firm Size)</td>
<td>-0.0726</td>
<td>-0.269</td>
<td>-0.1825</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(&lt;0.0004)</td>
<td>(&lt;.0001)</td>
</tr>
<tr>
<td>log(Trading Volume)</td>
<td>-0.0657</td>
<td>-0.21858</td>
<td>-0.15</td>
</tr>
<tr>
<td></td>
<td>(0.0071)</td>
<td>(&lt;0.0003)</td>
<td>(&lt;.0001)</td>
</tr>
</tbody>
</table>

LR ratio test of significance of noise

The maximum likelihood estimation allows us to construct directly a likelihood ratio (LR) test for null hypothesis of no trading noise. The test is

\[ H_0 : \delta = 0 \text{ vs. } H_1 : \delta > 0 \]

Let \( \log L_0 \equiv \) the logarithm of joint likelihood from estimation under \( H_0 : \delta = 0 , \)

\[ \log L_1 \equiv \] the logarithm of joint likelihood from Particle filter estimation with noise.

Under \( H_0 : \delta = 0 , \) the test statistics \( LR = -2(\log L_0 - \log L_1) \) is no longer subject to a traditional \( \chi^2 \) distribution with 1 degree of freedom, as \( \delta \) is the lower bound of parameter set. Gourieroux and Monfort (1995) pointed out a simple solution to compute the critical
value of this new distribution: divide a significant level by 2, and find corresponding critical value from $\chi^2$ distribution table. The likelihood-ratio test rejects the null hypothesis if the value of the test statistic $LR$ is smaller than this corrected critical value. Our analysis suggests that 83.48% firms in transportation industry face a statistically significant trading noise at 5% level.

5.3.1.3 Asset volatility ($\sigma_y$)

The magnitude of estimated asset volatility is consistent with that of prior empirical work. In Ericsson and Reneby (2005), business risk is deemed high if $\sigma_y = 40\%$ and low if $\sigma_y = 20\%$. Our estimate $\sigma_y$ has a median value of 30\% in between. Also, 88.16\% of our estimates $\sigma_y$ are significant at 5\% level, showing it is a fairly efficient measure. The omission of trading noise at stock exchange will lead to upward bias in asset volatility, as trading noise is embodied in genuine asset volatility. Hence asset value will be underestimated, leading to overestimation of default probability. In our sample of transportation industry, the upward bias is in the order of 5.56\% on average with the maximum bias at 25.2\% as shown in Table 4.

5.3.1.4 Asset drift ($\mu_y$)

As for the asset drift estimate $\mu_y$, the estimation result is not desirable due to large standard errors. This is not surprising, as it is well known in the literature that it is hard to estimate the expected return of an asset governed by a diffusion process. Nonetheless, our
estimate $\mu_Y$, is still in consistent range.\textsuperscript{28} If we calculate firm-specific $\mu_Y$, as the sum of risk free rate ($r =3.6\%$, the median value in table 2) and the market price of risk ($\lambda = 0.15$, the stable estimate from Huang and Huang (2003)) multiplied by the estimated firm-specific asset risk $\sigma_Y = 30\%$, the computed value of $\mu_Y$, is roughly 8.1%, in the same order of magnitude as the estimated $\mu_Y, 0.08$ in Table 4.

5.3.2 Bankruptcy prediction

In this subsection, we compare how different models (Merton/KMV, DOC option, Altman’s Z-score, Shumway’s discrete hazard model) perform when trying to predict 1-, 2-, and 3-year-ahead bankruptcies.

5.3.2.1 Estimating implied default probability in structural credit risk models

Figures 2-4 give graphical presentation of the derived probability default measures from Merton/KMV model and DOC option framework for 1-year-ahead forecast. Aggregate default probability is computed as the simple average of the default likelihood indicators of all firms. Figure 2 shows that PD derived from particle filter is less (or equal to) that derived from traditional MLE method. As we mentioned earlier, ignoring trading noise would lead to upward biased asset volatility measure, hence overestimated PD. Figure 3 presents the comparison of PD from Merton/KMV and DOC option model. It is not surprising to observe that the latter exceeds the former one. In principle, $PD_{\text{call}} < PD_{\text{DOC}}$, because DOC option model recognize the possibility that firms may go bankrupt prior to debt maturity, i.e., the early bankruptcy.

\textsuperscript{28} We need to estimate the actual drift of the asset process, as bankruptcy prediction should use actual probabilities, not risk neutral ones.
Figure 2. Aggregate Default Probability Measure: PD(DOC, MLE) vs. PD(DOC, Particle filter) This figure shows default probability derived from particle filter is lower than that derived from traditional MLE method. Ignoring trading noise would lead to upward bias in asset volatility measure, hence overestimated PD. Note: During 1999–2008, there are two recession periods as defined by NBER, One is March 2001-November 2001, the second starts from December 2007.

Figure 3. Aggregate Default Probability Measure: Merton/KMV vs. PD (DOC model, Particle filter) The figure presents the comparison of PD from Merton/KMV and DOC option model. We observe \( PD_{\text{call}} < PD_{\text{DOC}} \) because DOC option model recognize the possibility of early bankruptcy, that is, firms may go bankrupt prior to debt maturity. Note: During 1999–2008, there are two recession periods as defined by NBER, One is March 2001-November 2001, the second starts from December 2007.
Figure 4. Average Default Probability (DOC model, Particle filter) The Figure shows the forecasted average default probability for bankrupt firms and non-bankrupt firms for transportation industry from 1999 to 2008. The default probability is derived from DOC model with particle filtering techniques applied to MLE method. Note: During 1999–2008, there are two recession periods as defined by NBER, One is March 2001–November 2001, the second starts from December 2007.

We also compute the average default probability for bankrupt firm and non-bankrupt firm respectively. Figure 4 presents the results. On average, the default probability for bankrupt firms is much higher than that for non-bankrupt firms. This suggests that our default probability measures do capture default risk. For both group, the default probability varies with the business cycle. While for bankrupt firms, the degree of volatility in default probability is much larger.

5.3.2.2 Altman’s Z-score model

Altman (1968) constructs a function that combines financial ratios into a single indicator variable, often called the Z-score, which discriminates non-default firms from those that default. The parameters in its model are achieved via multiple discriminant
analysis. Following Altman (2000), we will adopt a convenient form of the Z-score model as below in our analysis:

\[ ALT(t) = 0.012 \times \frac{WC}{TA} + 0.014 \times \frac{RE}{TA} + 0.033 \times \frac{EBIT}{TA} + 0.006 \times \frac{ME}{TL} + 0.999 \times \frac{SALE}{TA} \] \quad (21)

Where \( \frac{WC}{TA} \) is the firm’s ratio of working capital to total assets, \( \frac{RE}{TA} \) is ratio of retained earnings to total assets, \( \frac{EBIT}{TA} \) is earnings before interest and taxes divided by total assets, \( \frac{ME}{TL} \) is the ratio of market value of equity to book value of total liabilities, \( \frac{SALE}{TA} \) is the sales/total assets ratio. The lower a firm’s distress potential, the larger its Z-score. A conventional definition for zones of discrimination is as follows: for \( Z > 2.99 \), the firm lies in a “safe” zone; for \( 1.8 < Z < 2.99 \), “Grey” Zone; \( Z < 1.80 \), “Distress” Zone.

5.3.2.3 Shumway’s (2001) discrete hazard model

Realizing that single-period models ignore the fact that firms change through time, Shumway (2001) proposes a hazard model that uses all available information, i.e. data through the life of a firm to determine each firm’s bankruptcy risk. The hazard function \( \lambda(t) \) in our paper, takes the logistic function form as below:

\[
\log \frac{\lambda_i(t)}{1 - \lambda_i(t)} = \beta_0 + \beta_iX_i(t) + \varepsilon
\] \quad (22)

where \( X_i(t) \) is the vector of explanatory variables related to firm’s financial information, which are used to forecast failure for firm at time \( t \). Notice that the model incorporate time-varying covariates by making \( X \) depend on time \( t \). Following Shumway (2001), \( X_i(t) \) in our paper include five terms, i.e. NI /TA, TL/TA, Relative return, Relative Size
as defined in Table 2, and \( \sigma \) (the volatility of firm’s equity return). Namely, NI/TA is the firm's ratio of net income to total assets. TL/TA is total liabilities divided by total liabilities divided by total assets. Relative return is firm's prior year return relative to the CRSP value weighted index return (vwretd). Firm’s relative size is the logarithm difference between firm's market equity capitalization and the total market value from the CRSP file. \( \sigma \) (the volatility of firm’s equity return) is obtained through regressing monthly return of stock on value weighted index return and then taking the standard deviation of the residual of this regression, i.e. variance in stock return which cannot be explained by average market.

The hazard function \( \lambda(t) \), gives the probability of failure at \( t \) conditional on surviving up to \( t \).

Let \( \tilde{\lambda}_i(t) \equiv \text{prob ( firm } i \text{ file for bankruptcy at time } t | \text{ surviving up to time } t \) \)

\( Y_{it} \equiv \text{indicator for default}, 1 \text{ if firm } i \text{ default at time } t, 0 \text{ otherwise.} \)

\[ t_i \equiv \text{the last time period when firm } i \text{ is observed} \]

Suppose there are \( N \) firms in the sample. The parameter \( \beta \) of the discrete hazard model could be estimated by maximizing the joint density of firm observations below:

\[
\log L = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \left\{ Y_{it} \ln \tilde{\lambda}_i(t) + (1 - Y_{it}) \left[ 1 - \ln \tilde{\lambda}_i(t) \right] \right\}
\]

It is a conventional binary choice model for panel data. Due to the nonlinear likelihood functions and complicated forms, the hazard models might be difficult to estimate. Shumway (2001) demonstrated that a multiperiod logit model is equivalent to a discrete-time hazard model in that they have the same likelihood function. So in practice, discrete-time hazard models can be estimated with a computer program that estimates logit
models. It's easy to do. We treat each firm year as an independent observation. The dependent variable in a logit model is set equal to one only in the year when the firm files bankruptcy. However, the $\chi^2$ test statistics produced by the logit programs need to be corrected by adjusting the sample size to account for the lack of independence between observations, i.e. the degree of freedom is number of unique firms, instead of the total firm-year observations in the estimation.

Given $\hat{\beta}$, the estimate of $\beta$ obtained from above logit program, the probability of default within one year is:

$$SH(1) = \frac{1}{1 + \exp(- (\beta_0 + \hat{\beta}_i, X))}$$

where $X$ are data from one year prior to bankruptcy. Probability to default within $n$ years is derived as:

$$SH(n) = (1 - SH(1))^{n-1} * SH(1)$$

where $X$ are data from $n$ year prior to bankruptcy.

### 5.3.3 Evaluation of default prediction performances

In evaluating the performance of different bankruptcy forecast models, we require the end of the forecast horizon to still be within our sample, i.e. before December 2008. So, 3-year-ahead prediction can be made only up to December 2005. As the forecast horizon increases, the number of firm-years we used decreases.

We use the concept of Receiver Operating Characteristic (ROC) curve to compare forecast accuracy. A ROC curve can be created by varying from zero to one the threshold that maps estimated default probability to class predictions (i.e. predicting default if the
probability is larger than the threshold value, and no default otherwise). For each threshold, the ROC curve could tell for a certain level of type I error or false positive rate (percentage of non-defaults that are mistakenly classified as defaults on the $x$-axis), what percentage of true defaults is correctly detected as defaults (True positive rate on the $y$-axis) and hence the type II error rate (what percentage of true default cannot be detected, given by the vertical distance from the chosen point to the top of the figure). Thus ROC curve reflects the quality of a ranking. Given two models, the one with a ROC curve further to the top left than the other, has a better prediction power (ranking). The AUC curve (area under ROC curve) also captures this idea. A perfect model would assign all defaulting firms a greater probability of default than any surviving firm, so its ROC curve will go from $(0,0)$ to $(0,1)$, and then to $(1,1)$, and AUC will be 1. A model without any power has a ROC curve being the main diagonal (i.e. it increases True Positive rate equally likely as it does on False Positive rate), with an AUC of 0.5. For example, the AUC is computed to be 0.9 for DOC model in one-year ahead prediction using the trapezoidal rule (Rosner, 2000). This means that the DOC model has an 90% chance of correctly distinguishing a bankrupt form an insolvent firm based on the ordering the default probability.

So we can compare the AUC scores of the default probability inferred from Merton/KMV, DOC, Shumway’s, ALT ‘s Z score for 1-, 2- 3- year-ahead default prediction to rank the models’ performance. Figures 5-7 and Table 6 present the results.
Table 6: Comparison of Forecast Accuracy (Common Firms)

Table 6 compares the ability of default probability estimates from different models to predict default within one year, two years, and three years. We report two measures to evaluate the forecast performance: AUC (area under curve) score from ROC (receiver operating characteristic) curve, and percent concordant. Note: “within two years” means that 2004 data can predict 2006 default, for example.

<table>
<thead>
<tr>
<th></th>
<th>DOC (PF-MLE)</th>
<th>DOC (MLE)</th>
<th>Merton/KMV</th>
<th>Shumway</th>
<th>Altman</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Concordant (%)</td>
<td>50.4</td>
<td>81.1</td>
<td>75.1</td>
<td>90.7</td>
<td>87.2</td>
</tr>
<tr>
<td>AUC score</td>
<td>0.724</td>
<td>0.873</td>
<td>0.846</td>
<td>0.926</td>
<td>0.903</td>
</tr>
</tbody>
</table>

Panel A: Default within one year (27 bankruptcies)

| Percent Concordant (%) | 42.9 | 71.6 | 50.2 | 72 | 72.5 |
| AUC score              | 0.666 | 0.78 | 0.7 | 0.8 | 0.806 |

Panel B: Default in second year (23 bankruptcies)

| Percent Concordant (%) | 40 | 71.9 | 38.6 | 66.3 | 73.5 |
| AUC score              | 0.64 | 0.78 | 0.634 | 0.76 | 0.796 |

With data one year prior to bankruptcy, Panel A of Table 6 and Figure 5 shows that DOC model is uniformly more powerful than Merton/KMV model and other statistical models, no matter whether the model is estimated with transformed data MLE method, or MLE via the particle filter approach.

We repeat the same analysis with data two years prior to bankruptcy (in Panel B of Table 6 and Figure 6). As the forecast horizon increases, all the models lose prediction accuracy by various degrees. AUC statistics still favor the barrier option framework over
Merton/KMV, but the superiority of the DOC model over Shumway’s model becomes less.

![ROC Curve](image)

**Figure 5. ROC curve for One-year-ahead default prediction**

The figure presents the forecast performance comparison of default probabilities measures from Merton/KMV, DOC, Shumway’s and ALT ‘s Z score model. The forecast horizon is one- year ahead. For example, 2004 data can predict 2005 default. ROC (receiver operating characteristic) curve reflects the quality of a ranking. Given two models, the one with a ROC curve further to the top left than the other, has a better prediction power (ranking).

Results with data three years prior to bankruptcy are shown in Panel C of Table 6 and Figure 7. The AUC score from the DOC model with traditional MLE method is less than that from one estimated with particle-filtering technique. So we see the necessity of taking account of trading noise in structural bankruptcy prediction model as the forecast horizon increase. Also, Altman’s default measure is expected to win over Merton’s probability measure in medium- to long- term forecast horizon.
Figure 6. ROC curve for two-year-ahead default prediction
The figure presents the forecast performance comparison of default probabilities measures from Merton/KMV, DOC, Shumway’s and ALT ‘s Z score model. The forecast horizon is two-year ahead. For example, 2004 data can predict 2006 default. ROC (receiver operating characteristic) curve reflects the quality of a ranking. Given two models, the one with a ROC curve further to the top left than the other, has a better prediction power (ranking).

Figure 7. ROC curve for three-year-ahead default prediction
The figure presents the forecast performance comparison of default probabilities measures from Merton/KMV, DOC, Shumway’s and ALT ‘s Z score model. The forecast horizon is three-year ahead. For example, 2004 data can predict 2007 default. ROC (receiver operating characteristic) curve reflects the quality of a ranking. Given two models, the one with a ROC curve further to the top left than the other, has a better prediction power (ranking).
Percent Concordant

‘Percent Concordant’, is another measure commonly used to evaluate the default forecast performance. A pair of bivariate observations $\{X_1, Y_1\}$ and $\{X_2, Y_2\}$ is said to be concordant if $\text{sign}(X_1 - X_2) = \text{sign}(Y_1 - Y_2)$. If we let $X$ denote observed response value (1 if default, 0 otherwise), $Y$ denote the predicted probabilities of default from model estimation, then a concordant pair is a pair of observation where the observation with response value of 0 has a lower predicted mean score than the observation with response value of 1. Percent concordant is defined as $n_c/n$, where $n_c$ is the number of pairs that are concordant, and $n$ is the total number of distinct pairs. The percent concordant measures from Table 6 support our conclusion made from ROC curve analysis.

Rank Percentile

We also conduct rank correlation analysis for all the default likelihood measure. Table 7 presents the result. As expected, Altman’s Z score is negatively correlated with all other measures. All correlation coefficients are statistically significantly at the 1% level.

Table 7: Kendall’s Rank Correlation of different PD measures
The table presents the correlation coefficient between different default probability measures. The estimation is done for each firm for each year. The sample includes transportation industry from 1999 to 2008. Cross-section sample size is 1676. p-values are in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Altman's Z score</th>
<th>Shumway</th>
<th>PD (Merton)</th>
<th>PD (DOC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altman's Z score</td>
<td>1</td>
<td>-0.09601</td>
<td>-0.22714</td>
<td>-0.15502</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(&lt;.0001)</td>
<td>(&lt;.0001)</td>
<td>(&lt;.0001)</td>
</tr>
<tr>
<td>Shumway</td>
<td>1</td>
<td>0.42584</td>
<td>0.49685</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(&lt;.0001)</td>
<td>(&lt;.0001)</td>
<td></td>
</tr>
<tr>
<td>PD (Merton)</td>
<td>1</td>
<td></td>
<td>0.70853</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(&lt;.0001)</td>
<td></td>
</tr>
<tr>
<td>PD (DOC)</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
6. Conclusions and Extension

In this paper, we apply a new estimation method to a default forecasting model (DOC), and investigate its empirical performance using noisy equity prices data. We provide empirical validation of the model by showing that the implied barriers are statistically significant for a sample of U.S. transportation firms from 1999 to 2008. Our noise estimate derived from the MLE method via smoothed auxiliary particle filtering technique suggests that ignoring trading noise would lead to biased estimates of asset volatility. We then apply the model and the method to bankruptcy prediction for 1-, 2- 3-years ahead horizons, and finds that the newly estimated model dominates the commonly used models, such as Altman (1968), Merton(1974)/ KMV and Shumway(2001) in terms of accuracy and AUC ranking.

Our default probability measure is higher than the actual or expected default frequency, suggesting a need of recalibration. One way is to apply a model such as a logistic type that provides a monotonic transformation of current probability measure, and combines it with accounting ratios/variable. Other extensions could be made to introduce jump in asset values, or a stochastic process for interest rates, or adopt a new error structure by assuming innovation in asset return correlated with noise in the stock price. Case studies and quarterly updated estimates of PD might provide additional rich and vivid information. It is also interesting to investigate the estimated default risk on stock returns, or to extend our analysis to more industries/sectors, and account for the correlation between the probabilities of default. These lines of inquiry await further investigation.
Appendix

A.1 Particle Filter

The particle filters (PF), known variously as bootstrap filtering, condensation algorithm (Isard and Blake(1996)), are usually used to estimate Bayesian models and are the sequential analogue of Markov chain Monte Carlo (MCMC) batch methods. It uses simulations to sequentially approximate the filtering density.

Following Pitt and Shephard (1999), we model unobserved state $x_t$, $t=1, \ldots, n$, as a first order Markov process satisfying $x_t | x_{t-1} \sim f(x_t | x_{t-1})$ with initial density $f(x_0)$, where $f(x_t | x_{t-1})$ is often termed as “transition density” of the state. And the time series of observations $y_t$, $t=1, \ldots, n$, are assumed conditionally independent given the unobserved state $x_t$. In other words, each $y_t$ only depends on $x_t$, $y_t | x_t \sim f(y_t | x_t)$, where $f(y_t | x_t)$ is called “measurement density”. One form of this scenario is a state space model consisting of two equations as follows:

Transition Equation: $x_t = g(x_{t-1}, \varepsilon_t), t = 1, \ldots, n$ (A-1)

Measurement Equation: $y_t = h(x_t, \nu_t)$ (A-2)

Where $h(\cdot)$ and $g(\cdot)$ are known functions, both $\varepsilon_t$ and $\nu_t$ are independent and identically distributed sequences with known PDF and $\varepsilon_t \perp \nu_t$.

The goal of filtering problem is to estimate the state $x_t$ based on all observed data $y_1, \ldots, y_n$, $t=1, \ldots, n$. Specifically, we want to compute posterior probability density
function $f(x_t | y_1, \ldots, y_t) = f(x_t | Y_t), t = 1, \ldots, n$.\footnote{MCMC would model the full posterior $f(x_1, \ldots, x_t | y_1, \ldots, y_t)$. If well designed, particle filters can be much faster than MCMC.} If posterior PDF $f(x_{t-1} | Y_{t-1})$ is known at time $t-1$, then the following two-stage Bayesian filtering technique could be applied recursively to solve the above problem:

Stage 1: Derive the “prediction density” from the state density of previous period via transition density $f(x_t | x_{t-1})$,

$$f(x_t | Y_{t-1}) = \int f(x_t | x_{t-1}) f(x_{t-1} | Y_{t-1}) dx_{t-1} \quad (A-3)$$

Stage 2: According to Bayesian theorem, the prediction density $f(x_t | Y_{t-1})$ is then combined with the measurement density $f(y_t | x_t)$ to produce the following “posterior filtering density” up to some proportionality,

$$f(x_t | Y_t) \propto f(y_t | x_t) f(x_t | Y_{t-1}) dx_t \quad (A-4)$$

However, usually no closed-form solutions to the integrals in stage 1 are available for many natural dynamic models. If $h(\cdot)$ and $g(\cdot)$ are non-linear functions, and/or if $\epsilon_t$ or $\nu_t$ are non-Gaussian processes in the above state space models, numerical methods must be used to approximate the filtering densities and solve the above Bayesian filtering problems. The endeavors along this line include Extended Kalman filter, Unscented Kalman filter, Gaussian sum methods, grid-based methods, etc. As an alternative to the above methods, PF performs better with the advantage that, with sufficient particles/samples, they can be much more accurate.

**PF** sequentially approximates the filtering distribution of the unobserved state by a weighted set of particles. The basic algorithm is as follows:
Assume a set of M state particles/samples from \( f(x_{t-1} \mid Y_{t-1}) \) and corresponding weights \( \{x_{t-1}^{(i)}, W_{t-1}^{(i)}, i = 1,\ldots,M \} \) represents the posterior density \( f(x_{t-1} \mid Y_{t-1}) \) at time \( t-1 \) \((t=1,\ldots,n)\) with \( \sum_{i=1}^{N} W_{t-1}^{(i)} = 1 \), and \( M \) is taken to be very large. Then, following the 2-stage Bayesian filtering procedure above, the particle set will be updated to represent the posterior density \( f(x_t \mid Y_t) \) at current period \( t \). And this algorithm will be iterated sequentially to advance the system to next period and so on up to \( t=n \).

There are several ways to sample from the empirical prediction density. Among them, the Sampling/Importance Resampling (SIR) method by Rubin (1987) is a very commonly used particle filtering method. This implementation proceeds in three steps as follows:

**Step 1:** Simulating the state forward: Sampling

This is done by computing \( x_t^i \) from the original set of particles \( \{x_{t-1}^i\}_{i=1}^{M} \) known at time \( t-1 \) via transition equation (A-1). Here, we assume equal weighted sample as in most literature, i.e., \( W_{t-1}^{(i)} = \frac{1}{M} \). Then, with the \( M \) equal weighted samples, we can produce an approximation to the prediction density, i.e.

\[
\hat{f}(x_t \mid Y_{t-1}) = \frac{1}{M} \sum_{i=1}^{M} f(x_t \mid x_{t-1}^i)
\]

**Step 2:** Computing and normalizing the weights: Importance Sampling

30 Other methods include Rejection sampling, MCMC sampling. They will involve evaluating density functions in very large order and computationally infeasible.
At this point, we have a vector of $M$ possible values of $x^i_t$, and equation (A-2) offers a simple way to evaluate the likelihood that the observations $y_t$ has been generated by $x^i_t$. Hence, we are able to assign to each particle a filtering weight of

$$\omega^i_t = f(y_t | x^i_t),$$

Finally, we have to normalize the weights by

$$W^i_t = \frac{\omega^i_t}{\sum_{j=1}^{M} \omega^j_t}$$

**Step 3: Resampling**

The motivation for this step is to avoid the degeneracy of the algorithm, i.e. to avoid the situation that all but one of the importance weights is close to zero.

Resample from the weighted sample $\{x^{(i)}_{t-1}, W^{(i)}_{t-1}, i = 1, \ldots, M\}$ to obtain a new equal weight sample of size $M$.

Step 1, step 2 and step 3 are repeated for $t=1, \ldots, T$. Once the particles and weights have been computed for each date, we are ready to construct the filtered path of $\overline{x}_t$ by

$$\overline{x}_t = \sum_{i=1}^{N} W^i_t x^i_t \text{ for each } t.$$ The PF thus delivers a time series of filtered state variable $x$.

**A.2. Correlation Coefficients**

Pearson’s correlation coefficient (often denoted by $r$) is a measure of the linear dependence between two variables. It’s defined as the covariance of the two variables
divided by the product of their standard deviations. Based on a samples, the sample correlation coefficient \( r \) is given as :

\[
r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}}
\]

The correlation coefficient is inside the interval \([-1,1]\). A value of 0 implies that there is no linear correlation between the two variables. A value of 1 (-1) implies that a linear equation fits the relationship between the two variables perfectly, with all data lying on a upward- (downward-) sloping line.

A rank correlation coefficient is a statistics used to measure the correspondence between rankings of two measured quantities \( \{X_i, Y_i\} \) and to assess the significance of that correspondence. Spearman’s \( \rho \) and Kendall’s \( \tau \) are two of the most popular rank correlation statistics:

1. **Spearman’s rank correlation coefficient** \( \rho \) is calculated with the following procedure: convert the \( n \) original observations \( \{X_i, Y_i\} \) into ranks \( \{x_i, y_i\} \), then for each observation calculate the difference between the ranks of the two variables, \( d_i = x_i - y_i \). If there are no tied ranks, then

\[
\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}
\]

For tied ranks, take the mean of their positions in the ascending order of the values and assign to each of the equal values this mean rank.

2. **Kendall’s rank correlation coefficient** \( \tau \) is defined as

\[
\tau = \frac{n_c - n_d}{\frac{1}{2} n(n-1)}
\]
where \( n_c \) is the number of pairs that are concordant, \( n_d \) is the number of discordant pairs and the denominator is the total number of pairs in the data set.

The rank coefficient has a value between +1 and −1. If the two rankings are the same, the coefficient has value 1; if one ranking is the inverse of the other, the coefficient has value of −1; if the rankings are completely independent, the coefficient has value of 0.

### A.3 Receiver operating characteristic (ROC) curve

A receiver operating characteristic (ROC) is a graphical plot of true positives rate (TPR) vs. false positives rate (FPR), for a binary classifier system as its discrimination threshold is varied.

In our paper, ROC curve serves as a tool to select optimal models among alternative default models. Default risk factors are identified in the sample and a logistic regression model is fit to the data. We treat each firm year as an independent observation. The dependent variable in the model is set equal to one only in the year when the firm files bankruptcy. Suppose we have a sample of \( n \) observations, let \( n1 \) represent the number of defaults. We denote this default group as \( C1 \), and the remaining non-defaulting group as \( C2 \), which has \( n2 = n-n1 \) observations. An estimated default probability, \( \hat{\pi}_i \), is calculated for the \( i \)-th observation.

Now, using the \( n \) firm-year observations, we test the performance of the model in default prediction based on the estimated probability of default. The higher values of the estimated probability are expected to relate with default. ROC curve can be constructed by shifting the threshold at which the estimated probability is considered a predictor of
default, i.e., predicting default if the probability is larger than the threshold value, and no default otherwise. For each threshold \( z \), the following measures can be computed, where \( I(\cdot) \) is the indicator function:

Number of defaults correctly predicted: \( POS(z) = \sum_{i \in C_1} I(\hat{\pi}_i \geq z) \)

Number of non-defaults correctly predicted: \( NEG(z) = \sum_{i \in C_2} I(\hat{\pi}_i < z) \)

Number of non-defaults incorrectly predicted as defaults: \( FALPOS(z) = \sum_{i \in C_2} I(\hat{\pi}_i \geq z) \)

Number of defaults incorrectly predicted as non-defaults: \( FALNEG(z) = \sum_{i \in C_1} I(\hat{\pi}_i < z) \)

Then, we obtain the values for:

\[
\text{Sensitivity} = \frac{POS(z)}{n_1} \quad \text{and} \quad 1 - \text{Specificity} = \frac{FALPOS(z)}{n_2}.
\]

Where sensitivity is defined as the ability of a model to detect the bankruptcy when it is truly present, i.e., true positive rate; specificity is the ability of a model to exclude the status of insolvency in firms who did not go bankrupt, i.e., true negative rate. A ROC curve can be constructed by plot Sensitivity (on the y-axis) against 1-Specificity (on the x-axis). The different points on the curve correspond to different cut points used to determine if the firm is bankrupt. Of course, by lowering the cutoff, we should be able to predict more bankruptcies, but we will also increase our false positives.
References:


Bharath, S.T., Shumway, T. (2007). Forecasting Default with the Merton Distance to Default Model. Review of Financial Studies (Forthcoming)


Quarterly, 3(3) (Spring 1997) pp. 73-80;


Chapter 2: Reduced Form Hazard Models of Corporate Default Prediction

including Macro Factors as Covariates

Abstract

Default data show clear evidence of cycles. In this chapter, we include macro factors in reduced form hazard models to capture the time-variation in default rate and the impact of business cycle on firms’ creditworthiness. Other covariates used include firm-specific accounting ratios and market return variables. We construct macro factors from fifty macroeconomic time series data using principal component analysis. Our proposed hazard model delivers a noticeable improvement over Shumway’s (2001) model in terms of model fit and predictive power. Macroeconomic conditions affect firm’s creditworthiness differently. In our paper, the latent macro factors load more on financially healthier firm with bigger size. This is consistent with findings in McNeil and Wendi (2007) and Koopman and Lucas (2008). Unlike Sueyoshi (1995), our specification tests suggest the choice of a hazard model specification is innocuous in our context of firm’s default prediction. The effects of the covariates on default rate are nearly proportional, and do not differ much across specifications. For ease of computation, we adopt a simple dynamic logistic hazard model. The forecasted default probability with horizon of one year will be used as a measure of default risk in Chapter 3.
1. Introduction and Related Literature

For long, a common approach to assess firm’s default probability has been the structural model based on distance to default. The structural approach is appealing in that it captures the important aspects of the process determining corporate failure. While in recent decades, reduced form model is gaining more popularity. The reduced form approach chooses the inputs by estimating their relative importance in fitting historical defaults (Hilscher, Jarrow and Deventer, 2008).

Intuitively, reduced form models can never be less accurate than the structural models of default. Because the reduced form approach can incorporate many different inputs in addition to the inputs from structural approach. Our empirical studies in this chapter verify this point. We extend the previous literature by considering a wider range of explanatory variables.

The latest generation of modeling default is dominated by duration analysis. ¹ It probably represents the state of the art in default forecasting with reduced form models. Empirical duration analysis of corporate bankruptcy prediction includes works of Shumway (2001) and Hillegeist, Keating, Cram and Lundstedt (2004) who uses a discrete duration model, and Shumway and Bharath (2008) who use a continuous Cox proportional hazard model in predicting one-year default. They find that in the presence of market leverage and volatility information, among other covariates, distance to default, a probability measure derived from Black-Scholes-Merton models adds relatively little information.

¹ Many researchers have questioned the traditional Altman (1968) type accounting ratio-based statistical models, because they are based on discrete book value accounting data, and do not allow non-linear effects among the different credit risk factors.
Our estimation methodology in this chapter is along the line of duration analysis on reduced form models of default.\(^2\)

We broaden our definition for firm’s default as failure to meet its financial obligation, which includes bankruptcies, delisting for financial reasons, or D (default) ratings issued by a leading credit rating agency S&P. We measure the duration as the time a firm has been in a state of non-default, and assess the importance of various factors in predicting that duration. Here, we focus on predicting the first case of a firm entering default, and disregard what happens afterwards. Thus, a statistical model that takes into account of multiple default, or revival of a firm is beyond the scope of this study.

Default data show clear evidence of cycles. As seen from Figure 1 and Table 1, the default rate reaches peak around year 2001’s recession. It also increases dramatically as the time approaches year 2008. So, we include macro factors as additional covariates in hazard models to capture the time-variation in default rate and the impact of business cycle on firms’ creditworthiness. We construct macro factors from fifty macroeconomic time series data using principal component analysis. The constructed factors are believed to summarize the co-movement among high-dimensional macroeconomic time series (such as interest rate, US industrial growth rate, employment, stock market return, etc.).

Our study contributes to the current literature on duration analysis of corporate default in the following ways:

*Information Content:* Our prediction model exploits all available information at a

---

\(^2\) In our study, most predictors of our model are time-varying, such as accounting variables, aggregate economic variables. And for many firms we do not observe a complete duration, because they have not filed for bankruptcy or default at the end of our sampling period, or possibly they exit due to reasons other than default. So we adopt duration analysis in this chapter because of its usefulness in dealing with censoring data and time-varying covariates.
Table 1. Actual Default Frequency Table
This table lists the total number of active firms, defaults and default frequency for every year of our sample period. Default is defined as failure to meet its financial obligation, including bankruptcies, delisting for financial reasons, or D (default) ratings issued by a leading credit rating agency S&P.

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of active firms</th>
<th>Number of defaults</th>
<th>Default frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>2871</td>
<td>26</td>
<td>0.009056</td>
</tr>
<tr>
<td>2000</td>
<td>2707</td>
<td>66</td>
<td>0.024381</td>
</tr>
<tr>
<td>2001</td>
<td>2471</td>
<td>85</td>
<td>0.034399</td>
</tr>
<tr>
<td>2002</td>
<td>2312</td>
<td>68</td>
<td>0.029412</td>
</tr>
<tr>
<td>2003</td>
<td>2277</td>
<td>48</td>
<td>0.02108</td>
</tr>
<tr>
<td>2004</td>
<td>2256</td>
<td>20</td>
<td>0.008865</td>
</tr>
<tr>
<td>2005</td>
<td>2246</td>
<td>21</td>
<td>0.00935</td>
</tr>
<tr>
<td>2006</td>
<td>2211</td>
<td>12</td>
<td>0.005427</td>
</tr>
<tr>
<td>2007</td>
<td>2114</td>
<td>10</td>
<td>0.00473</td>
</tr>
<tr>
<td>2008</td>
<td>1972</td>
<td>30</td>
<td>0.015213</td>
</tr>
</tbody>
</table>

Figure 1. Predicted (one-month-ahead) Versus Actual Default Rate
The figure plots the actual and predicted default rates from 1999 to 2008. Default is defined as failure to meet its financial obligation, including bankruptcies, delisting for financial reasons, or D (default) ratings issued by a leading credit rating agency S&P. Red solid line shows the actual annual default rate. Green slashed line shows the fitted annual average default rate derived from model 5 of table 3 with 1-month ahead forecast horizon. Blue slashed line on the bottom is the part of default frequency not captured by the model (i.e., actual value minus fitted value).
given point in time. It integrates firm-specific accounting data, market variables and constructed macroeconomic factors. Firm-specific accounting information is not limited to only leverage and volatility of asset as used in structural model; it considers, among others, ratios of working capital to total assets, retained earnings to total assets, etc. The inclusion of changing aggregate economic factors and its interaction with firm-specific characteristic variables allows our model to capture the effect of adverse business cycle on different firm’s creditworthiness, which is absent in the empirical default duration analysis mentioned above.  

Robustness check for hazard model: We compare the performance of a few conventional hazard models in our context of corporate default prediction. We show that the choice of distributional form for hazard models is innocuous in our paper.

From analysis of over 3000 publicly traded U.S. firms from 1999 to 2008, several important facts emerge from the results. Macroeconomic conditions affect firm’s creditworthiness differently. In our paper, the latent macro factors load more on financially healthier firm with bigger size. This is consistent with findings in McNeil and Wendi (2007) and Koopman and Lucas (2008). Our proposed hazard model delivers a noticeable improvement over Shumway’s (2001) model in terms of model fit and predictive power. Unlike Sueyoshi (1995), our specification tests suggest the choice of a hazard model specification (logistic regression, Complementary Loglog model or Cox proportional hazard model) is innocuous in our context of firm’s default prediction. The effects of the covariates on default rate are nearly proportional, and do not differ much across specifications. For ease of computation, we adopt a simple dynamic logistic hazard

---

3 D. Duffie et al. (2007) adopts a relatively small set of firm-specific and macroeconomic covariates due to the computational complexity in their model. They did not consider the interaction of macroeconomic state with firms of different characteristics, either.
model. The forecasted default probability with a horizon of one year will be used as a measure of default risk in Chapter 3.

The remainder of this chapter is organized as follows. In section 2, econometric framework for corporate default prediction is presented. Section 3 provides data description. Empirical results are shown in section 4. Section 5 presents specification test for hazard models. Concluding remarks are given in section 6.

2. Econometric Framework

2.1. General Setup: An Approximate Dynamic Factor Model

We consider forecasting corporate default with predictors including macro factors constructed from a large number of macroeconomic series. Following Stock and Watson (2002), we work with a static representation of the dynamic factor model. To be specific, let \( X_t = (X_{1t}, X_{2t}, \ldots, X_{Mt})' \) be an \( N \)-dimensional multiple macroeconomic time series\(^4\), let \( P_{i,t+j} \equiv \text{Prob}(Y_{i,t+j} = 1 \mid Y_{i,t+j-1} = 0) \) be firm \( i \)'s conditional probability of default in \( j \) period, given that it has survived for \( j-1 \) period, where \( Y_{i,t} \) is an indicator that equals one if the firm defaults in period \( t \). \( P_{i,t+j} \) is the \( j \)-step-ahead variable to be forecast.

Assume that \((X_t, P_{i,t+j})\) admit a factor model representation with \( r \) common latent factors \( F_t \),

\[
X_t = \Lambda F_t + \epsilon_t, \tag{1}
\]

and

\(^4\) We transformed original macroeconomic and financial market time series to make them satisfy the I(0) process requirement for \( X_t \), so each series is a zero mean stationary process. See data description section 3.1. for details.
\[ P_{i,t+j} = h(F_t, z_u, \varepsilon_{i,t+j}) \]  \hspace{1cm} (2)

where \( \Lambda = (\lambda_1, \lambda_2, ..., \lambda_N)' \) is an \( N \times k \) matrix of factor loadings, \( \varepsilon_t = (e_{1t}, e_{2t}, ..., e_{zt})' \) is an \( N \times 1 \) vector of idiosyncratic disturbances, \( j \) is the forecast horizon. \( z_u \) is a \( m \times 1 \) vector of observed variables for firm \( i \) (e.g., firm’s accounting data, market return variables), that together with \( F_t \) are useful indicators to forecast \( P_{i,t+j} \), and \( \varepsilon_{i,t+j} \) is the resulting forecast error.

Note that \( h(\cdot) \) is a form of hazard function. For now, we assume \( P_{i,t+j} \) follows a logistic distribution, and (2) can be rewritten as

\[ P_{i,t+j} = \frac{1}{1 + \exp(-\alpha_j - \beta_{jF}'F_t - \beta_{jz}'z_u - \beta_{jz}'z_u \times z_u - \varepsilon_{i,t+j})} \]  \hspace{1cm} (3)

We relax this assumption in section 5 specification test. We allow the coefficients in (3) to vary with the horizon of the prediction. We also include the interaction terms between macro factors and some firm-specific variables in (3), to capture the different impact of macro state on firms with different characteristics.

Note that \( k \), the dimension of macro factors is smaller than \( N \), the dimension of macroeconomic series, i.e. there is a reduction in dimensionality. The underlying assumption is that the series \( X_t \) may be driven by a small number of unobservable common factors \( F_t \), or in other words, \( F_t \) summarize the co-movement of the observed variables.

For identification purposes it is assumed that the \( k \) factors are independent, factor loadings are normalized as \( \Lambda'\Lambda / N = I_k \). These assumptions are made in standard factor models. For details, see Stock and Watson (2002), Bai and Ng (2002).
In classic factor analysis model, the errors $e_t$ were assumed to be cross-sectionally and temporally $i.i.d$. However, these assumptions are not appropriate in our corporate default forecast application where macroeconomic variables are used. These macro variables are serially correlated, and may be cross-correlated (e.g. various measures of treasury rate) even after controlling for the common factors. Here, we follow Stock and Watson (2002) in allowing the error terms to be both serially correlated and (weakly) cross-sectionally correlated. Stock and Watson (2002) showed that the estimated factors are consistent in such a serially correlated version of the approximate factor model introduced by Chamberlain and Rothschild (1983).

The static representation of the dynamic factor model in (1) and (3) could easily be adapted to allow lags of the factors to enter the equations for $X_t$ and $P_{i,t+j}$. For details, see Stock and Watson (2002). Then we can simply estimate the dynamic factors $F_t$ using the method of principal component. That is the main advantage of a dynamic factor model in the form of (1)-(3), and the estimators are readily calculated even for large $N$.\(^5\)

2.2 Estimation Methodology: a two-step procedure

Note that the factors $F_t$, their loadings $\Lambda$, as well as the idiosyncratic disturbances $e_t$ are not observable. However, we observe $X_{it}, i = 1,2,\cdots,N$. So we construct forecasts of $P_{i,t+j}$ based on (1) and (3) using a two-step procedure.

\(^5\) Koopman (2010) use Gaussian Maximum Likelihood and numeric sampling method to estimate the parameters for the joint stochastic process of $\{\log(P/(1-P), X_t\}$ and then extract the unknown values of factors using signal extraction algorithms and iterative nonlinear methods which can be computationally prohibitive when $N$ is large.
In first step, we estimate a time series of factors \( \{ \hat{F}_j \} \) from (1) using the method of principal component (See details below). In second step, the estimators \( (\hat{\alpha}_j, \hat{\beta}_{jF}, \hat{\beta}_{jz}, \text{ and} \hat{\beta}_{jFz}) \) are typically estimated by maximum likelihood method. We can also fit mode (3) by running a logit regression on a set of observations where each firm period is treated as an independent observation. The dependent variable takes the value one if firm default at that time period and zero otherwise. A standard logit software can do the work. Then, a forecast default probability \( P_{i,t+j} \) could be formed as

\[
\frac{1}{1 + \exp(-\hat{\alpha}_j - \hat{\beta}_{jF} \hat{F}_i - \hat{\beta}_{jz} z_{it} - \hat{\beta}_{jFz} \hat{F}_i \times z_{it})}.
\]

Estimation of the common factors---principal components analysis

Estimates of the factors \( F_i \) and factor loadings \( \lambda_i \) are obtained by solving the following optimization problem

\[
V(k) = \min_{\Lambda, F^k} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \lambda_k F^k_i)^2
\]

If we concentrate out \( F_i \) and use the normalization that \( \Lambda'\Lambda / N = I_k \), then the optimization problem is identical to maximizing \( \text{tr}(\Lambda'X'\Lambda) \), where \( X \) is the \( T \times N \) matrix with \( t \)-th row \( X_{it}' \), and \( \text{tr}(\cdot) \) denotes the matrix trace. The matrix of factor loadings is obtained by setting \( \Lambda \) equal to the eigenvectors corresponding to the \( k \) largest eigenvalues of the \( N \times N \) matrix \( X'X \). The resulting factor matrix (principal components estimator of \( F \)) is then \( \hat{F}^k = X'\hat{\Lambda} / N \), which is a weighted average of the transformed
observables. We do a further step in renormalizing the factor vectors to have zero mean and standard deviation of one.

Estimating the number of factors (k)

We are left with an important question, that is, the determination of the number of factors (k) underlying X_{it}. Theoretically, a model with more factors can fit no worse than a model with fewer factors, but it comes at the cost of efficiency since more factor loadings are to be estimated. Traditional AIC or BIC rule will no longer be able to consistently estimate the number of factors (For proof, see Bai and Ng (2002)).

Bai and Ng (2002) proposed a consistent estimator for the number of static factors in a large N and T approximate factor models based on a penalized least squares objective function. We will follow Bai and Ng (2002) in estimating the number of factors (k) in our paper.

Specifically, let \( \hat{F}^k \) be a matrix of k factors estimated using the method of principal components, and let \( V(\hat{k}, \hat{F}^k) = \min_{\lambda} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \lambda_i^k \hat{F}_{it}^k)^2 \) denotes the average total sum of squares from time series regressions of \( X_{it} \) on the \( k \) factors for all \( i = 1, 2, \ldots, N \). Then a loss function can be used to determine the optimal number of factors \( k \). In our paper, we use the following readily applicable criteria proposed by Bai and Ng (2002):

\[
IC_{p1}(k) = \ln(V(\hat{k}, \hat{F}^k)) + k \left( \frac{N + T}{NT} \right) \ln \left( \frac{NT}{N + T} \right) \tag{5}
\]

\[
IC_{p2}(k) = \ln(V(\hat{k}, \hat{F}^k)) + k \left( \frac{N + T}{NT} \right) \ln C_{NT}^2 \tag{6}
\]
\[ IC_{p3}(k) = \ln\left( \nu(k, \hat{F}^k) \right) + k \left( \frac{\ln C_{NT}^2}{C_{NT}^2} \right) \]  

(7)

where \( C_{NT} = \min\{\sqrt{N}, \sqrt{T}\} \) and the second term in each criterion is the penalty for overfitting, specified as a function of the sample size in both dimension \((N, T)\). The smaller the criterion, the better captured the variations in \( X_{it} \).

3. Data Description

3.1. Macro economic time series used for constructing macro factors

The observable macro economic variables used for constructing macro factors are drawn from several broad categories of data: (I) Production, income and macro Indicators; (II) Employment, unemployment and hours (III) Housing and sales (IV) Sales, orders and production (V) Interest rate (VI) Money and bank lending (VII) Price index (VIII) Average hourly earnings (IX) Exchange rates (X) Stock market returns. All of the data are from Federal Reserve Bank of St. Louis website, except for two series, S&P 500 stock price index annualized quarterly return and return volatility, which were computed from S&P 500 stock price index by the author. A complete list of the variable names appears in Table 2. The sample range for the macro time series is January 1978 - December 2008.

Prior to the estimation of macro factors \( F_t \), each individual data series is transformed from its release values in two ways. First, each series is subject to a stationary-inducing transformation, possibly by taking first differencing (e.g., the unemployment rate), log-difference (e.g., the industrial production data), etc. In some cases, like the Institute for
Table 2: Macroeconomic Time Series Data

The table present the time series that were used to construct the macro factors in section 3.1. The transformation codes are 1=no transformation, 2=first difference, 3=logarithm, 4= first difference of logarithm, 5=second difference of logarithm. The following abbreviations appear in the data description: SA=seasonally adjusted, SAAR=seasonally adjusted at an annual rate. The series were taken from the St. Louis Fed online database, http://research.stlouisfed.org/fred2/

<table>
<thead>
<tr>
<th>Series number.</th>
<th>Transformation</th>
<th>Total</th>
<th>No. of mnemonic</th>
<th>code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I) Production, Income and Macro Indicators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.indpro</td>
<td>4</td>
<td>Industrial production index (2007=100, sa)</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.ipman</td>
<td>4</td>
<td>Industrial Production: Manufacturing (2007=100, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.dspic</td>
<td>4</td>
<td>Real Disposable Personal Income (bil of Chained 2005 $, saar)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.napm</td>
<td>1</td>
<td>ISM Manufacturing: PMI composite index (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.napmpi</td>
<td>1</td>
<td>ISM Manufacturing: production index (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.mcumfn</td>
<td>1</td>
<td>Capacity Utilization: Manufacturing (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.tcu</td>
<td>1</td>
<td>Capacity Utilization: Total Industry (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.umcsent</td>
<td>1</td>
<td>University of Michigan: Consumer Sentiment Index (1966 Q1=100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(II) Employment, Unemployment and Hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.napmei</td>
<td>1</td>
<td>ISM Manufacturing: employment index (%)</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.payems</td>
<td>4</td>
<td>All Employees: Total nonfarm (thous., sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.unrate</td>
<td>1</td>
<td>Civilian Unemployment Rate (%), sa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.ic4wsa</td>
<td>3</td>
<td>4-Week Moving Average of Initial Claims (sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.uempmean</td>
<td>1</td>
<td>Average (Mean) Duration of Unemployment (weeks, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.awhnonag</td>
<td>1</td>
<td>Average Weekly Hours Of Production And Nonsupervisory Employees: Total private</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.awhman</td>
<td>1</td>
<td>Average Weekly Hours of Production and Nonsupervisory workers: Manufacturing (sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.awotman</td>
<td>1</td>
<td>Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(III) Housing and Sales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.houst</td>
<td>3</td>
<td>Housing Starts: Total (thous, saar)</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.permit</td>
<td>3</td>
<td>New Private Housing Units Authorized by Building Permits (thous, saar)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.hsn1f</td>
<td>3</td>
<td>New One Family Houses Sold: United States (thous, saar)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.hnfsepussa</td>
<td>3</td>
<td>New One Family Homes For Sale at end of month (thous, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(IV) Sales, Orders and Inventories</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.napmnoi</td>
<td>1</td>
<td>ISM Manufacturing: New order index (%)</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.napmiii</td>
<td>1</td>
<td>ISM Manufacturing: Inventories Index (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.napmsdi</td>
<td>1</td>
<td>ISM Manufacturing: Supplier deliveries index (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.rrsfs</td>
<td>4</td>
<td>Real Retail and Food Services Sales (mil $, sa, 1982-84=100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.usasartmism</td>
<td>4</td>
<td>Total Retail Trade in United States (sa, 2005=100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Series number.</td>
<td>Transformation</td>
<td>Total No. of series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>----------------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mnemonic</td>
<td>code</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(V) Interest Rate</td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. fedfunds</td>
<td>2</td>
<td>Effective Federal Funds Rate (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. gs1</td>
<td>2</td>
<td>1-Year Treasury Constant Maturity Rate (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. gs10</td>
<td>2</td>
<td>10-Year Treasury Constant Maturity Rate (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. mortg</td>
<td>2</td>
<td>30-Year Conventional Mortgage Rate (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. baa</td>
<td>2</td>
<td>Moody's Seasoned Baa Corporate Bond Yield (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31. dp</td>
<td>1</td>
<td>default premium, baa-gs10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32. tp</td>
<td>1</td>
<td>term premium, gs10-gs1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33. yc</td>
<td>1</td>
<td>yield curve slope, gs10-fedfunds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(VI) Money and Bank Lending</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34. m1sl</td>
<td>5</td>
<td>M1 Money Stock (bil $, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35. m2sl</td>
<td>5</td>
<td>M2 Money Stock (bil $, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36. busloans</td>
<td>4</td>
<td>Commercial and industrial loans at all commercial banks (bil $, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37. realln</td>
<td>4</td>
<td>Real Estate Loans at All Commercial Banks (bil $, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38. totalsl</td>
<td>4</td>
<td>Total Consumer Credit Outstanding (bil $, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(VII) Price Index</td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40. ppilfe</td>
<td>5</td>
<td>Producer Price Index: Finished Goods (1982=100, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41. cpiaucsl</td>
<td>5</td>
<td>Consumer Price Index for All Urban Consumers: All Items (1982-1984=100, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consumer Price Index for All Urban Consumers: All Items Less</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42. cpilfesl</td>
<td>5</td>
<td>Food &amp; Energy (1982-1984=100, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Personal Consumption Expenditures: Chain-type Price Index</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43. pcepi</td>
<td>4</td>
<td>Personal Consumption Expenditures: Chain-Type Price Index Less (2005=100, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44. pcepilfe</td>
<td>4</td>
<td>Food and Energy (2005=100, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(VIII) Average Hourly Earnings</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45. ahetpi</td>
<td>5</td>
<td>Avg Hr Earnings of Production Workers: total private ($, sa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46. aheman</td>
<td>5</td>
<td>Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(IX) Exchange Rates</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47. twexbxml</td>
<td>4</td>
<td>Trade Weighted US dollar Exchange Index: Broad (1997=100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48. twexxml</td>
<td>4</td>
<td>Trade Weighted US dollar Exchange Index: Broad (1997=100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(X) Stock Market Returns</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49. sp500</td>
<td>4</td>
<td>S&amp;P 500 stock price index annualized quarterly return</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50. sp500va</td>
<td>4</td>
<td>S&amp;P 500 stock price index return volatility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total no. of series</td>
<td></td>
<td></td>
<td>50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supply Management’s Employment Index, the data requires no transformation, as the data is in percent already. Table 2 also lists the transformation method used for each variable. The decision to make which transformation was made judgmentally after preliminary unit root tests and inspection of the data. Second, after the transformation, each series is demeaned and standardized to have zero mean and unit variance.

3.2. Data for Hazard Model Regression
3.2.1 Sample and Covariates

We are particularly interested in the sensitivity of default hazard rates to firm-specific, market variables and macroeconomic factors. In this chapter, we broaden our definition for firm’s default as failure to meet its financial obligation, which includes bankruptcies, delisting for financial reasons, or D (default) ratings issued by a leading credit rating agency S&P. We obtained corporate default data from three sources, namely, Altman-NYU Salomon Center Bankruptcy List, Compustat and CRSP.

Our sample period for hazard model regression in (3) is January 1999 ~ December 2008. It consists of 3404 unique firms, covering 265792 firm-months of monthly data. Our dataset includes 386 defaults.

The firm-specific covariates we have consider and tested in the hazard rate models include: a firm’s relative size (RLSIZE) measured as the logarithm difference between firm's market equity capitalization and the total market value from the CRSP file; a firm’s trailing one-year stock return relative to the CRSP value weighted index return.
Following Shumway (2001), we also include the volatility of a firm’s equity return (SIGMARET) as a covariate of interest.\(^6\)

We also collect from Compustat Database, and compute each firm’s several accounting ratios as predictors. Namely, the firm's ratio of net income to total assets (NI/TA), total liabilities to total assets (TL/TA), working capital to total assets (WC/TA), retained earnings to total assets (RE/TA), sales to total assets (SALE/TA), market value of equity to book value of total liabilities (ME/TL), etc. To ensure statistical results are not heavily influenced by extreme values, we winsorized all above financial ratios in the same way as we did in previous chapter.

Since Compustat provides only quarterly and yearly data, for each month we take the firm’s accounting ratios to be the value derived from corresponding quarter. We have lagged the accounting ratio data, for one thing, to make sure that all the information used to compute default hazard rate is observable at the time of prediction; for another, to tackle with the problem of state-dependence /rate dependence. That is, the values of these predictors at time \( u \) can be potentially influenced either by a firm’s bankruptcy status or its hazard rate at time \( u \) (Singer and Willett (2003), p.441).

3.2.2.Person-period data

In our survival analysis of firm’s default event, we use person-period type data instead of person-record data. A person-period data set has multiple lines of data for each

\(^6\) This volatility is obtained through regressing monthly return of stock on value weighted index return and then taking the standard deviation of the residual of this regression, i.e. variance in stock return which cannot be explained by average market
subject under study.\(^7\) So, it allows each firm to contribute data whenever it is at risk, i.e., one for each period when it was at risk of event occurrence. Although the multiple records for each firm do not appear to be obtained independently of each other, it shouldn’t be a concern for yielding appropriate estimation results considering that we are modeling a hazard rate for each firm in each of its periods of risk. In other words, we can assume that all records in the person-period data set are conditionally independent (Singer and Willett (2003), p.384), because we are modeling the conditional probability of default given the firm surviving up to each particular time period and some predictors in each period.

3.2.3 Data Sampling—left-truncation and right-censoring in Duration Analysis

For consistency of exposition, we will assume for now that \( T \) is a discrete random variable. The idea applies to a continuous case as well. Let \( t_i \) be a duration variable of interest with probability mass \( f(t_i) = \Pr(T = t_i) \), and we define the survivor function at time \( t_i \) as the probability that the survival time \( T \) is at least \( t_i \):

\[
S(t_i) = \Pr(T \geq t_i) = \sum_{u=t_i}^\infty f(u) \tag{8}
\]

Then the hazard rate below defines the firm’s conditional probability of default given that it has survived to that time.

\[
h(t_i) = \Pr(T = t_i \mid T \geq t_i) = \frac{f(t_i)}{S(t_i)} \tag{9}
\]

\(^7\) For details and examples of person-period data set, see Singer and Willett (2003), section 10.5.1 and section 11.3.1.
Ideally, we would like to observe a random sample of completed durations from each subject. Suppose we observe \( \{t_1, t_2, \ldots, t_N, \} \). The log likelihood function of the sample is simply

\[
\log L = \sum_{i=1}^{N} \ln f(t_i)
\]

(10)

But in practice we are more likely to have censored duration data. In our study, for many firms we do not observe a complete duration \( t_i \), because they have not default at the end of our sampling period, or possibly they exit due to reasons other than default. In this case, the log likelihood function of the sample can be written as

\[
\log L = \sum_{i=1}^{N} \left[ d_i \ln f(t_i) + (1 - d_i) \ln S(C_i) \right]
\]

(11)

where indicator \( d_i \) is used to distinguish between censored and complete spells, which takes value 1 if the firm defaults by year \( t_i \), zero otherwise; \( C_i \) is the number of years we observe the firm before it is right censored.

In addition to the right censoring problem mentioned above, our likelihood specification will also need to take into account of the special sampling design we adopt. Namely, it is a mix of stock sampling and inflow sampling. We include firms whose stocks are traded in one of the three major US markets in 1999, and we also include all firms that begin trading after 1999. We do not adopt a simple flow sampling method, because it would lead to the problem of missing a significant portion of firms as we observe samples in which durations are long.

However, a stock sampling would result in length-biased non-random samples (Lancaster, 1979). Let \( Z_i \) be the length of time firms have been trading in the market.
before the starting time of observation. In a stock sample, firms with a smaller \( Z_i \) are less likely to be sampled. Therefore, for both completed and censored spells, we have to condition on the fact that we already know the firm has survived in the state up to the sampling time to be at risk. Thus, contributions to the likelihood are conditioned on \( T > Z_i \).

So, the log likelihood function for our sample with left truncation and right censoring can be written as following:

\[
\log L = \sum_{i=1}^{N} [d_i \ln f(t_i) + (1 - d_i) \ln S(C_i) - \ln S(Z_i)]
\]

(12)

We estimate the parameters of our model using a maximum likelihood approach.

4. Empirical Results

4.1. Construction of macroeconomic factors

To determine the optimal number of factors \((k)\) behind the 50 macroeconomic time series, we use the information criteria proposed by Bai and Ng (2002) as shown in (5) –(7). All criteria point to \( k = 5 \) factors. We estimate the factors using principal components analysis method specified in section 2.2.

Figure 2 displays the \( R^2 \) of the regressions of the 50 individual macro time series against each of the first five empirical factors estimated over the full sample period (1978-2008). These \( R^2 \) are plotted as bar charts with one chart for each series. The series are ordered numerically using the ordering in Table 2. Figure 3 indicates the shares of variation in each time series that can be attributed to each common macroeconomic factor.
Figure 2 $R^2$ of the regressions of the individual macro time series against each factor

Figure 2 displays the $R^2$ of the regressions of the 50 individual macro time series against each of the first five empirical factors estimated over the full sample period (1978-2008). These $R^2$ are plotted as bar charts with one chart for each factor. The series are ordered numerically using the ordering in Table 2. The broad categories are: (I) Production, income and macro Indicators; (II) Employment, unemployment and hours (III) Housing and sales (IV) Sales, orders and production (V) Interest rate (VI) Money and bank lending (VII) Price index (VIII) Average hourly earnings (IX) Exchange rates (X) Stock market returns
Figure 2 (cont’d): $R^2$ of the regressions of the individual macro time series against each factor. Figure 2 above displays the $R^2$ of the regressions of the 50 individual macro time series against each of the first five empirical factors estimated over the full sample period (1978-2008). These $R^2$ are plotted as bar charts with one chart for each factor. The series are ordered numerically using the ordering in Table 2. The broad categories are: (I) Production, income and macro Indicators; (II) Employment, unemployment and hours (III) Housing and sales (IV) Sales, orders and production (V) Interest rate (VI) Money and bank lending (VII) Price index (VIII) Average hourly earnings (IX) Exchange rates (X) Stock market returns

Figure 3 Shares of explained variation in macroeconomic time series data Figure 3 above indicates the shares of variation in each time series that can be attributed to each common macroeconomic factor. The 50 individual macro time series are ordered numerically using the ordering in Table 2.

81
Broadly speaking, the first factor loads primarily on production, employment and sales; the second factor on housing and interest rate; the third, on employment, interest rate and stock market return; the fourth, on employment and inflation; the fifth, on inflation. Taken together, these five factors account for 56% of the variance of the 50 time series in the dataset (24.67%, 12.92%, 7.18%, 6.38%, and 4.79% respectively). The shares explained variation in macro series ranges from about 15% (real disposable personal income, real estate loans) to about 90% (ISM manufacturing index, total nonfarm employees) except for 3 series (producer price index, average hourly earnings of workers for total private, and for manufacturing industry), which have explained shares below 10%. Removing the three poorly explained macroeconomic time series do not alter the patterns of our constructed factors, so we keep the 50 time series.

4.2. Findings from hazard model analysis

Table 3 reports dynamic logistic regression results for various alternative specifications in (3). The forecast horizon is one month ahead. We start with Shumway’s (2001) bankruptcy model. We note from model 1 that all the included variables in Shumway (2001) enter significantly and with expected signs, except for NI/TA, which is statistically insignificant. In model 2, we include significant variables (LT/AT, RLSIZE, EXCESSRET, SIGMARET) from Shumway (2001), balance sheet ratios (WC/TA, RE/TA, SALE/TA) from Altman (1968), as well as firm’s book-to-market ratio (BM)\(^8\). In model 3, we remove from model 2 the insignificant variable and variables with

\(^8\) Firms with high book-to-market (BM) ratio of its equity are perceived to be more risky, hence investors will require a premium for holding stocks of such firms. Fama and French (1993) used HML (high minus low, the return differential between firms with high BM and low BM) as a risk factor in pricing cross-sectional stock return.
Table 3  Logistic Hazard Model Estimates
This table presents results from logistic regressions of the default indicators on predictor variables. Model 1 is Shumway’s (2001) model. Model 5 is our proposed best model. See section 3.2.1 for definitions of all covariates. All models are estimated at one-month-ahead horizon. The sample range is 1999-2008. t-statistics are in parentheses. *** (**, *) indicates significant at 1% (5%, 10%) level.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-22.93</td>
<td>-20.47</td>
<td>-20.24</td>
<td>-20.45</td>
<td>-25.23</td>
</tr>
<tr>
<td></td>
<td>(-32.63)**</td>
<td>(-26.71)**</td>
<td>(-27.61)**</td>
<td>(-28.24)**</td>
<td>(-23.74)**</td>
</tr>
<tr>
<td>WC/TA</td>
<td>-1.50</td>
<td>-1.45</td>
<td>-1.33</td>
<td>-1.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-10.23)**</td>
<td>(-10.55)**</td>
<td>(-11.35)**</td>
<td>(-11.10)**</td>
<td></td>
</tr>
<tr>
<td>RE/TA</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.23)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SALE/TA</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.99)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NI/TA</td>
<td>-0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LT/TA</td>
<td>0.04</td>
<td>0.29</td>
<td>-0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.71)**</td>
<td>(5.76)**</td>
<td>(-1.21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SigmaRET</td>
<td>2.01</td>
<td>1.78</td>
<td>1.71</td>
<td>1.73</td>
<td>1.79</td>
</tr>
<tr>
<td></td>
<td>(8.60)**</td>
<td>(6.39)**</td>
<td>(6.59)**</td>
<td>(6.72)**</td>
<td>(6.64)**</td>
</tr>
<tr>
<td>RLSIZE</td>
<td>-0.83</td>
<td>-0.70</td>
<td>-0.71</td>
<td>-0.72</td>
<td>-0.96</td>
</tr>
<tr>
<td></td>
<td>(-23.0)**</td>
<td>(-17.65)**</td>
<td>(-18.9)**</td>
<td>(-19.27)**</td>
<td>(-18.01)**</td>
</tr>
<tr>
<td>EXCESSRET</td>
<td>-0.82</td>
<td>-0.90</td>
<td>-0.83</td>
<td>-0.83</td>
<td>-0.88</td>
</tr>
<tr>
<td></td>
<td>(-20.71)**</td>
<td>(-18.24)**</td>
<td>(-20.03)**</td>
<td>(-20.09)**</td>
<td>(-19.89)**</td>
</tr>
<tr>
<td>Book-to-Market (BM)</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.52)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor1</td>
<td></td>
<td></td>
<td></td>
<td>-3.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-3.55)**</td>
<td></td>
</tr>
<tr>
<td>Factor2</td>
<td></td>
<td></td>
<td></td>
<td>3.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4.37)**</td>
<td></td>
</tr>
<tr>
<td>Factor3</td>
<td></td>
<td></td>
<td></td>
<td>-2.37</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(3.39)**</td>
<td></td>
</tr>
<tr>
<td>Factor1*size</td>
<td></td>
<td></td>
<td></td>
<td>-0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-3.77)**</td>
<td></td>
</tr>
<tr>
<td>Factor2*size</td>
<td></td>
<td></td>
<td></td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4.35)**</td>
<td></td>
</tr>
<tr>
<td>Factor3*size</td>
<td></td>
<td></td>
<td></td>
<td>-0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-3.32)**</td>
<td></td>
</tr>
<tr>
<td>c score (area under ROC curve)</td>
<td>0.90</td>
<td>0.924</td>
<td>0.913</td>
<td>0.913</td>
<td>0.925</td>
</tr>
<tr>
<td>Percent concordant</td>
<td>83.6</td>
<td>86.8</td>
<td>84.9</td>
<td>84.9</td>
<td>87.2</td>
</tr>
<tr>
<td>Total obs.</td>
<td>249043</td>
<td>235296</td>
<td>246382</td>
<td>246382</td>
<td>246382</td>
</tr>
<tr>
<td>Default events</td>
<td>379</td>
<td>357</td>
<td>377</td>
<td>377</td>
<td>377</td>
</tr>
<tr>
<td>Pseudo-R2</td>
<td>0.3114</td>
<td>0.3463</td>
<td>0.3295</td>
<td>0.3289</td>
<td>0.3412</td>
</tr>
<tr>
<td>Adjusted Pseudo-R2</td>
<td>0.3097</td>
<td>0.3433</td>
<td>0.3277</td>
<td>0.3275</td>
<td>0.3377</td>
</tr>
</tbody>
</table>
counterintuitive signs (RE/TA, SALE/TA). Comparison of model 1 and model 3 shows that in the presence of liquidity measure (WC/TA), the leverage ratio (LT/TA) changes sign and is no longer a significant predictor. Therefore, in model 4, we propose using WC/TA instead of LT/TA among other Shumway (2001) explanatory variables.

We report MaFadden’s pseudo-$R^2$ and adjusted pseudo-$R^2$ coefficient for each specification, calculated as $1 - \frac{\ln L_1}{\ln L_0}$, and $1 - \frac{\ln L_1 - k}{\ln L_0}$ respectively, where $L_1$ is the likelihood of the full model, $L_0$ is the likelihood of a null model with only a constant, $k$ is the number of predictors. A pseudo-$R^2$ coefficient in range of 0.2~0.4 generally indicates a good enough model.

Model 4 is a better model than Shumway model 1 in terms of both model fitness and forecast performance. We note that the adjusted pseudo-$R^2$ coefficient increases significantly from 0.3097 in model 1 to 0.3275 in model 4. The $c$ score (i.e. area under ROC curve in our Chapter 1, a measure of forecast performance) increases from 0.90 to 0.913, indicating model 4 has a higher predictive power than model 1. The score select option in logistic regression also favors WC/TA over LT/TA as a better predictor.\(^9\)

To get the relative importance of changes in the different predictor variables, we calculate the proportional change in the default probability arising from one-standard-deviation increase in each predictor for a typical firm that initially has sample mean values of the predictors. Such an increase in working capital ratio (WC/TA) reduces the probability of default by 26.9% of its initial value, the corresponding effects are a 15.8% decrease in leverage.

\(^9\) Stepwise and backward option in logistic or cox regression is not recommended. It may unwisely eliminate variables that should be included if the subsequent inclusion of other variables causes the significance of a variable to drop below a particular statistical limit. This can be a problem for main effects components of an interaction. See Yaffee and Austin (1995).
increase for volatility, a 81.4% reduction for firm’s relative size, and a 35.6% reduction for firm’s excess return.

Forecast including macro factors as predictors

Table 1 presents the annual counts of active firms and default firms over the time. The solid line in Figure 1 also shows the annual realized average default rate. We note that there is considerable variation in the corporate default rate over time. The default rate climbed to peak when the economy is in recession. To capture the time effect, we include macroeconomic factors constructed from section 4.1 as additional predictors in hazard model regression.\footnote{In preliminary analysis, we also run regressions on a set of year dummies and our explanatory variables. The result indicates that time effects cannot be omitted from the model. Our proposed model with macro factors delivers better explanatory power and forecast performance than the model using year dummies. We also constructed business cycle dummies, and found them not significant predictor in our corporate default model. Given a higher proportion of firm defaults occurring in telecom industry around 2001, and in financial industry around 2008, we also consider industry dummies in our model. The industry dummies, turned out to add little improvement in explanatory power over our proposed model, and even worse in terms of forecast performance. All these results, not reported here, are available upon request.} We also consider the interaction terms between the macro factors and firm size to capture possible cross-sectional effects in default, which may be useful for explaining the cross-sectional stock return in chapter 3. The results are illustrated in Table 3, Model 5. The first three factors, as well as their interaction terms enter the model significantly at 1% level. We note that the adjusted pseudo-$R^2$ coefficient increases further from 0.3275 in model 4 to 0.3376 in model 5. The c score increases from 0.913 to 0.924, indicating model 5 is a best model so far.

Our proposed hazard model 5 delivers a noticeable improvement over Shumway’s (2001) in terms of model fit and predictive power. The $R^2$ increases by nearly 3%. The magnitude is comparable to that in Campbell, Hilscher and Szilagyi (2008). It does not mean the impact of business cycle on firms’ default risk is minimal. As we include firm-
specific variables as well, the impact of business cycle not captured by our macro factors may has already been absorbed in these variables.

Macroeconomic conditions affect firm’s creditworthiness differently. We note that all common factors load more on larger sized firms, because all the interaction terms enter with the same sign as its main effect. e.g., -3.08 for macro factor 1, and -0.16 for interaction term between factor 1 and firm size. If we perceive bigger firms as safer firms as Fama and Frech (1993) do, then it implies that common factors tend to load more on default probabilities of financially healthier firms. This is consistent with findings in McNeil and Wendi (2007) and Koopman and Lucas (2008)). The Basel II framework also set lower asset correlations for financially weaker firms, assuming these firms have lower systematic risk. See Basel Committee on Banking Supervision (2004).

Is the impact of business cycle on firm’s creditworthiness larger for financially healthy firms than for weak firms? We should be cautious in making such conclusions. We may say, for firms of a higher credit quality, a larger share of its total default risk is due to business cycle risk. But overall, the total default risk is still lower for large sized firm.

The interaction terms are statistically significant at 1% level based on \( t \)-statistics.\(^{11}\) We can also test the significance of the interaction terms through a likelihood ratio (\( LR \)) test between two models (one is full model with interaction terms included, the other not). Under the null hypothesis (restricted model without interaction terms), the test statistics is chi-square distributed with \( k \) degree of freedom, calculated as 
\[-2(ln L_0 – ln L_1),\]
where \( k \) is the number of restrictions, \( L_0 \) and \( L_1 \) are the likelihood for

\(^{11}\) The generalized Wald chi-square test also supports the significance of the interaction terms. However, it becomes somewhat unreliable as the ratio exceeds four. We do not report the results here.
the restricted model and full model, respectively. In our paper, the test statistic takes a value of $-2 \times (-1889.62 - (-1864.75)) = 49.75$, which exceeds $x_{0.01}^2(3) = 11.3$. So we reject the hypothesis that macro factors do not interact with firms’ size to predict corporate default.

Similarly, a LR test for the significance of both macro factors and interaction terms supports the inclusion of the first 3 factors and 3 interaction terms in our proposed default prediction model ($-2 \times (-1864.75 - (-1892.78)) = 56.06$, larger than $x_{0.01}^2(6) = 16.8$).

Annual average predicted default rate is shown as the dashed line in Figure 1. We first computed the fitted probability of default for each firm using the coefficients from our proposed model, and then take average across all the firms. As seen from Figure 1, our model captures much of the variation in firms’ default rate over the time, including the cyclical spikes in the early 2000s and the strong increase around 2008.

Macro factors versus macro variables

Are our constructed macro factors better predictors than individual macro variables for corporate default forecast? The comparison is illustrated in table 4. Overall, the answer to above question is positive. We choose the macro variables based on previous findings that identified the stance of monetary policy, inflation, and real economy activity as important state variables. Specifically, we use three time series, namely, federal funds rate (FEDFUNDS), first-difference of natural logarithms of CPI (all items less food and energy, DLCPILFE), and first-difference of natural logarithms of industrial production (DLINDPRO). We interact these series with firm’s relative size (RLSIZE) accordingly.
Table 4 Macro Factors vs. Macro Variables in Default Forecast
This table compares the predictive power of Macro factors in forecasting defaults with that of Macro variables. All other explanatory variables are from model 5 of Table 3. The forecast horizon is one-month ahead. Sample range is 1999-2008 for all U.S. firms. Logistic regression model is used. t-statistics are in parentheses. *** (**, *) indicates significant at 1% (5%, 10%) level.

<table>
<thead>
<tr>
<th></th>
<th>Model (i)</th>
<th>Model (ii)</th>
<th>Model (iii)</th>
<th>Model (iv)</th>
<th>Model (v)</th>
<th>Model (vi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-25.23</td>
<td>-12.90</td>
<td>-6.76</td>
<td>-5.17</td>
<td>-6.67</td>
<td>-6.34</td>
</tr>
<tr>
<td></td>
<td>(-23.74)***</td>
<td>(-8.09)***</td>
<td>(-100.61)***</td>
<td>(-40.24)***</td>
<td>(-99.98)***</td>
<td>(-47.86)***</td>
</tr>
<tr>
<td>Factor1</td>
<td>-3.08</td>
<td>2.62</td>
<td>-0.49</td>
<td>(-3.55)***</td>
<td>(4.7)***</td>
<td>(7.75)***</td>
</tr>
<tr>
<td></td>
<td>(4.37)***</td>
<td>(-10.23)***</td>
<td>(5.26)***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor2</td>
<td>3.99</td>
<td>-6.18</td>
<td>0.37</td>
<td>(4.37)***</td>
<td>(11.53)***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-3.39)***</td>
<td>(10.62)***</td>
<td>(2.00)***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor3</td>
<td>-2.37</td>
<td>4.95</td>
<td>0.11</td>
<td>(4.35)***</td>
<td>(11.53)***</td>
<td></td>
</tr>
<tr>
<td>Factor1*size</td>
<td>-0.16</td>
<td>0.17</td>
<td>(-3.77)***</td>
<td>(5.54)***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor2*size</td>
<td>0.20</td>
<td>-0.36</td>
<td>(4.35)***</td>
<td>(11.53)***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor3*size</td>
<td>-0.11</td>
<td>0.27</td>
<td>(-3.32)***</td>
<td>(10.67)***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inflation</td>
<td>-1.94</td>
<td>-0.73</td>
<td>-0.17</td>
<td>(-1.02)</td>
<td>(-9.27)***</td>
<td>(-1.16)</td>
</tr>
<tr>
<td>(DLCPI_LFE)</td>
<td>(DLCPILFE)</td>
<td></td>
<td></td>
<td></td>
<td>(-10.62)***</td>
<td>(-0.68)</td>
</tr>
<tr>
<td>interest rate</td>
<td>-2.00</td>
<td>-0.18</td>
<td>-0.02</td>
<td>(-5.33)***</td>
<td>(-10.62)***</td>
<td>(-0.68)</td>
</tr>
<tr>
<td>(FEDFUNDS)</td>
<td>(FEDFUNDS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>real production</td>
<td>-0.47</td>
<td>-0.02</td>
<td>-0.04</td>
<td>(-2.64)***</td>
<td>(-1.92)*</td>
<td>(-2.56)**</td>
</tr>
<tr>
<td>(DLINDPRO)</td>
<td>(DLINDPRO)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inflation*size</td>
<td>-0.10</td>
<td>14.46</td>
<td>-0.10</td>
<td>(1.00)</td>
<td>(-9.44)***</td>
<td></td>
</tr>
<tr>
<td>interest rate*size</td>
<td>-0.10</td>
<td>-3.59</td>
<td>(5.31)***</td>
<td>(10.84)***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>real production*size</td>
<td>-0.03</td>
<td>-0.34</td>
<td>(-2.47)**</td>
<td>(1.86)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>other explanatory</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>variables used</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c score (area under</td>
<td>0.925</td>
<td>0.923</td>
<td>0.877</td>
<td>0.722</td>
<td>0.584</td>
<td>0.504</td>
</tr>
<tr>
<td>ROC curve)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent concordant</td>
<td>87.2</td>
<td>86.7</td>
<td>80.2</td>
<td>52.1</td>
<td>29.9</td>
<td>3</td>
</tr>
<tr>
<td>Total obs.</td>
<td>246382</td>
<td>246382</td>
<td>253684</td>
<td>253684</td>
<td>265792</td>
<td>265792</td>
</tr>
<tr>
<td>Default events</td>
<td>377</td>
<td>377</td>
<td>385</td>
<td>385</td>
<td>386</td>
<td>386</td>
</tr>
<tr>
<td>Pseudo-R2</td>
<td>0.3412</td>
<td>0.3372</td>
<td>0.2211</td>
<td>0.0840</td>
<td>0.0115</td>
<td>0.0015</td>
</tr>
<tr>
<td>adjusted Pseudo-R2</td>
<td>0.3377</td>
<td>0.3337</td>
<td>0.2190</td>
<td>0.0819</td>
<td>0.0105</td>
<td>-0.0006</td>
</tr>
</tbody>
</table>

88
Model \((a)\) and \((b)\) include explanatory variables from model 5 of table 3. The comparisons of model \((a)\) with \((b)\), or model \((c)\) with \((d)\), or model \((e)\) with \((f)\), indicate unanimously that a model using macro factors surpasses its counterpart model using macro variables, in terms of model fit (larger adjusted \(pseudo-R^2\)) and predictive power (larger \(c\) score). The superiority of model \((a)\) over model \((b)\) may not seem impressive at first glance. But considering that the three variables we used in model \((b)\) have been chosen through a lot of work in trials and errors, usage of model \((a)\) with macro factors might bring us a piece of mind.

The better performance of macro factors over macro variables in predicting default may be attributable to its high-dimension-deduction feature. The constructed factors are believed to summarize the co-movement among high-dimensional macroeconomic time series (such as interest rate, US industrial growth rate, employment, stock market return, etc.). If we instead choose judgmentally several variables from large number of macro time series as predictors, plentiful information about the state of the economy may have been left out. Moreover, the macro factors are made orthogonal in construction, i.e. each factor is independent of another factor, making them good predictor candidates. Yet, the macro variables are themselves highly interrelated. e.g., Taylor’s rule set target federal funds rate as a function of inflation rate and real production growth rate.

\textit{Comparison with Distance to default measure}

We find in the presence of market leverage and volatility information, among other covariates, distance to default, a probability measure derived from Black-Scholes-Merton models adds relatively little information (results not shown here). This conclusion is
consistent with Hillegeist, Keating, Cram and Lundstedt (2004) and Shumway and Bharath (2008). So in our proposed model, we did not include the distance to default as a predictor, making our analysis readily accessible.

**Forecast at different horizons**

Table 5 and Figure 4 show hazard model results for forecast at longer horizons. We use our best model specification from model 5 of Table 3. As the forecast horizon increase from one month to 12 month, the performance of the predictors deteriorates as expected. Adjusted pseudo-\(R^2\) coefficient decreases from 0.34 to 0.12. The c score

![Figure 4. Predicted (one-year-ahead) Versus Actual Default Rate](image)

The figure plots the actual and predicted default rates from 1999 to 2008. Default is defined as failure to meet its financial obligation, including bankruptcies, delisting for financial reasons, or D (default) ratings issued by a leading credit rating agency S&P. Red solid line shows the actual annual default rate. Green slashed line shows the fitted annual average default rate derived from model 5 of table 3 with 1-year ahead forecast horizon. Blue slashed line on the bottom is the part of default frequency not captured by the model (i.e., actual value minus fitted value).
Table 5  Logistic Hazard Model Estimates at Different Forecast Horizon
Table 5 presents 1-month, 6-month and 1-year ahead logistic hazard model estimation results. The model used is from model 5 of Table d3. See section 3.2.1 for definitions of the covariates. Sample range is 1999-2008 for all U.S. firms. t-statistics are in parentheses.
*** (**, *) indicates significant at 1% (5%, 10%) level

<table>
<thead>
<tr>
<th></th>
<th>1-month ahead</th>
<th>6-month ahead</th>
<th>1-year ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-25.23</td>
<td>-19.61</td>
<td>-16.29</td>
</tr>
<tr>
<td></td>
<td>(-23.74)***</td>
<td>(-20.86)***</td>
<td>(-18.05)***</td>
</tr>
<tr>
<td>WC/TA</td>
<td>-1.31</td>
<td>-1.28</td>
<td>-1.31</td>
</tr>
<tr>
<td></td>
<td>(-11.10)***</td>
<td>(-9.68)***</td>
<td>(-9.56)***</td>
</tr>
<tr>
<td>SigmaRET</td>
<td>1.79</td>
<td>1.80</td>
<td>2.06</td>
</tr>
<tr>
<td></td>
<td>(6.64)***</td>
<td>(7.11)***</td>
<td>(9.31)***</td>
</tr>
<tr>
<td>RLSIZE</td>
<td>-0.96</td>
<td>-0.70</td>
<td>-0.53</td>
</tr>
<tr>
<td></td>
<td>(-18.01)***</td>
<td>(-14.31)***</td>
<td>(-11.09)***</td>
</tr>
<tr>
<td>EXCESSRET</td>
<td>-0.88</td>
<td>-0.72</td>
<td>-0.63</td>
</tr>
<tr>
<td></td>
<td>(-19.89)***</td>
<td>(-16.45)***</td>
<td>(-13.61)***</td>
</tr>
<tr>
<td>Factor1</td>
<td>-3.08</td>
<td>-2.54</td>
<td>-1.93</td>
</tr>
<tr>
<td></td>
<td>(-3.55)***</td>
<td>(-3.31)***</td>
<td>(-2.62)***</td>
</tr>
<tr>
<td>Factor2</td>
<td>3.99</td>
<td>3.05</td>
<td>2.61</td>
</tr>
<tr>
<td></td>
<td>(4.37)***</td>
<td>(3.74)***</td>
<td>(3.33)***</td>
</tr>
<tr>
<td>Factor3</td>
<td>-2.37</td>
<td>-1.14</td>
<td>-0.64</td>
</tr>
<tr>
<td></td>
<td>(-3.39)***</td>
<td>(-1.70)*</td>
<td>(-1.01)</td>
</tr>
<tr>
<td>Factor1*size</td>
<td>-0.16</td>
<td>-0.12</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>(-3.77)***</td>
<td>(-3.09)***</td>
<td>(-2.14)***</td>
</tr>
<tr>
<td>Factor2*size</td>
<td>0.20</td>
<td>0.16</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>(4.35)***</td>
<td>(3.72)***</td>
<td>(3.31)***</td>
</tr>
<tr>
<td>Factor3*size</td>
<td>-0.11</td>
<td>-0.06</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>(-3.32)***</td>
<td>(-1.86)*</td>
<td>(-1.24)</td>
</tr>
<tr>
<td>c score (area under ROC curve)</td>
<td>0.925</td>
<td>0.868</td>
<td>0.812</td>
</tr>
<tr>
<td>percent concordant</td>
<td>87.2</td>
<td>78.5</td>
<td>69.9</td>
</tr>
<tr>
<td>total obs.</td>
<td>246382</td>
<td>230299</td>
<td>211418</td>
</tr>
<tr>
<td>default events</td>
<td>377</td>
<td>371</td>
<td>353</td>
</tr>
<tr>
<td>Pseudo-R2</td>
<td>0.3412</td>
<td>0.1884</td>
<td>0.1276</td>
</tr>
<tr>
<td>adjusted Pseudo-R2</td>
<td>0.3377</td>
<td>0.1848</td>
<td>0.1238</td>
</tr>
</tbody>
</table>
decreases from 0.92 to 0.81. The coefficients and significance levels for most predictors also decline. However, almost all the variables remain statistically significant, except for factor 3 and its interaction terms, which are no longer significant predictors in 1-year-ahead forecast. Two predictors are particularly important at long horizons: the magnitude of WC/TA coefficients is roughly the same across different horizons; the coefficient on the volatility term (SIGMARET) increase with the horizon.

In chapter 3, we will analyze the effect of default risk on cross-sectional stock returns. The default risk measure we used there is the forecasted default probability derived from our best model 5 in table 3 at a horizon of 1-year.

5. Robustness Check

So far we are assuming that the hazard models take a discrete logistic specification as in Equation (3). Is this assumption innocuous in our context of firms’ default prediction? In this section, we will relax this assumption, and look at two conventional models utilized in event history analysis, namely complementary log-log (Cloglog) model and Cox proportional hazard model. We will compare their performance in predicting firms’ default in section 5.1. In section 5.2, we address specification issues through some tests. Overall, the choice of a hazard model specification for easiness of computation is innocuous in our context of corporate default forecast.

5.1 Alternative hazard models estimation

The proportional hazard assumption is empirically questionable as shown in McCall (1994). Narendranathan and Stewwart (1993) and Sueyoshi (1995) argue that the choice of distributional form for the binomial models is not innocuous in a duration context, as even if the overall model fit in likelihood terms does not allow them to discriminate between models, the effects of the covariates on the exit probability may differ largely across specifications.
5.1.1. Cox’s proportional hazard model (1972)

If time is truly continuous and one wishes to estimate the effects of the covariates without making any assumption about the baseline hazard, then Cox’s (1972) model is an appropriate model. According to Cox (1972), the survival time of each firm, denoted as \( t \), follows its own hazard function \( h(t) \),

\[
h_i(t) = h(t; x_{it}) = h_0(t)
\]  
\[
\exp(x'_{it} \beta) \tag{13}
\]

where \( h_0(t) \) is a baseline hazard rate common to all firms.\(^{13}\) \( x_{it} \) is the vector of explanatory variables for the \( i \)th firm. In our study, it consists of firm-specific variables, macro factors, and interaction terms. The predictor \( x_{it} \), is time-varying in that its value for any given firm can vary over time during a firm’s life. \( \beta \) is the vector of regression parameters associated with \( x_{it} \), assumed to be the same for all firms. Taking logs for both sides of equation (13), we derive the log hazard rate as following:

\[
\log h_i(t) = \log h_0(t) + x'_{it} \beta \tag{14}
\]

Note that the derivative of the log hazard rate remains constant over time. For a continuous covariate \( x \), it implies that each unit increase in levels of the predictor will result in same proportional change in the hazard rate, independent of time.

The parameters \( \beta \) can be estimated via partial likelihood function introduced by Cox (1972, 1975). The baseline survivor function can be computed using the Breslow

\(^{13}\) No assumptions need to be made about the nature or shape of \( \lambda_0(t) \), so it allows more flexibility in the specification of the hazard function.
(1972) estimate. For details about estimating the proportional hazard model, one may refer to Cox and Oakes (1984) and our appendix A.

5.1.2. C-Log-Log Model

Another popular extension of the proportional hazards model to discrete time starts from the survival function.

\[ 1 - h(t; x_i) = [1 - h_0(t)]^{\exp(x_i'\beta)} \]  

(15)

Applying the complementary log-log transformation of the hazard rate, we obtain the following generalized linear model

\[ \log(-\log(1 - h(h; x_i))) = \alpha_i + x_i'\beta \]  

(16)

where \( \alpha_i = \log(-\log(1 - h_0(t))) \). This model can be fitted a generalized linear model with complementary log-log links.

5.1.3. Hazard model estimates of alternative specifications

Table 6 reports the results of Cox proportional hazard model, C-Log-Log Model, as well as logistic model. Each model is estimated at a forecast horizon of one month. There are 246382 firms and 377 defaults in the sample over 1999-2008.

Overall, the different model specifications yield quite similar results. All coefficients enter significantly at 1% level, with expected signs. The coefficient values also have similar magnitude across the model specifications, even closer for logistic model and c-log-log model. Logistic model is slightly better than c-log-log model and cox proportional model in terms of model fit.
Table 6 reports the results of Cox proportional hazard model, C-Log-Log Model and logistic model estimation. Each model is estimated at a forecast horizon of one month for sample range 1999-2008. \( t \)-statistics are in parentheses. *** (**, *) denotes significance at 1% (5%, 10%) level. See section 3.2.1 for definitions of the covariates.

<table>
<thead>
<tr>
<th></th>
<th>Logistic Model</th>
<th>C-log Model</th>
<th>Cox Proportional Hazard Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-25.23</td>
<td>-24.96</td>
<td>-24.61</td>
</tr>
<tr>
<td></td>
<td>(-23.74)*****</td>
<td>(-24.80)*****</td>
<td>(-1.41)</td>
</tr>
<tr>
<td>WC/TA</td>
<td>-1.31</td>
<td>-1.09</td>
<td>-1.41</td>
</tr>
<tr>
<td></td>
<td>(-11.10)*****</td>
<td>(-10.25)*****</td>
<td>(-13.36)*****</td>
</tr>
<tr>
<td>SigmaRET</td>
<td>1.79</td>
<td>1.78</td>
<td>1.60</td>
</tr>
<tr>
<td></td>
<td>(6.64)*****</td>
<td>(7.11)*****</td>
<td>(6.11)*****</td>
</tr>
<tr>
<td>RLSIZE</td>
<td>-0.96</td>
<td>-0.94</td>
<td>-1.12</td>
</tr>
<tr>
<td></td>
<td>(-18.01)*****</td>
<td>(-18.78)*****</td>
<td>(-18.94)*****</td>
</tr>
<tr>
<td>EXCESSRET</td>
<td>-0.88</td>
<td>-0.80</td>
<td>-0.97</td>
</tr>
<tr>
<td></td>
<td>(-19.89)*****</td>
<td>(-19.71)*****</td>
<td>(-23.41)*****</td>
</tr>
<tr>
<td>Factor1</td>
<td>-3.08</td>
<td>-3.06</td>
<td>-2.46</td>
</tr>
<tr>
<td></td>
<td>(-3.55)*****</td>
<td>(-3.73)*****</td>
<td>(-2.66)*****</td>
</tr>
<tr>
<td>Factor2</td>
<td>3.99</td>
<td>3.92</td>
<td>4.48</td>
</tr>
<tr>
<td></td>
<td>(4.37)*****</td>
<td>(4.53)*****</td>
<td>(4.40)*****</td>
</tr>
<tr>
<td>Factor3</td>
<td>-2.37</td>
<td>-2.25</td>
<td>-2.65</td>
</tr>
<tr>
<td></td>
<td>(-3.39)*****</td>
<td>(-3.46)*****</td>
<td>(-3.40)*****</td>
</tr>
<tr>
<td>Factor1*size</td>
<td>-0.16</td>
<td>-0.16</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>(-3.77)*****</td>
<td>(-3.93)*****</td>
<td>(-2.86)*****</td>
</tr>
<tr>
<td>Factor2*size</td>
<td>0.20</td>
<td>0.19</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(4.35)*****</td>
<td>(4.53)*****</td>
<td>(4.44)*****</td>
</tr>
<tr>
<td>Factor3*size</td>
<td>-0.11</td>
<td>-0.11</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>(-3.32)*****</td>
<td>(-3.42)*****</td>
<td>(-3.43)*****</td>
</tr>
<tr>
<td>Model Log-likelihood</td>
<td>-1864.75</td>
<td>-1877.75</td>
<td>-1912.13</td>
</tr>
<tr>
<td>Pseudo-R2</td>
<td>0.3412</td>
<td>0.3343</td>
<td>0.3219</td>
</tr>
<tr>
<td>obs</td>
<td>246382</td>
<td>246382</td>
<td>246382</td>
</tr>
<tr>
<td>default</td>
<td>377</td>
<td>377</td>
<td>377</td>
</tr>
</tbody>
</table>

Note that the intercept is not identifiable in Cox proportional model due to partial likelihood estimation method used. For details of Cox model, see appendix A.

5.2 Hazard model specification tests

5.2.1. Is the choice of hazard model specification innocuous?
In section 5.1.3., we find that the covariates in our models (logistic, c-loglog, cox proportional hazard) have coefficient values of similar magnitude. Then do the effects of the covariates on the default probability differ a lot across specifications so that model choice would be a concern? Our answer is no; the choice of hazard model specification is innocuous in our context of firm’s default prediction. We show it next.

We measure the effects of the covariates on default rate by examining the first derivative of the log hazard rate with respect to the explanatory variable,

\[
\frac{\partial \ln(\alpha_i + x'_i \beta)}{\partial x_m} = \beta_m \frac{h'(\alpha_i + x'_i \beta)}{h(\alpha_i + x'_i \beta)}
\] (17)

where \(h'(\cdot)\) denotes the first derivative of \(h(\cdot)\). This expression reflects the proportional change in the hazard rate arising from changes in the explanatory variables.

For Cox model, relative change is \(\frac{h'(\alpha_i + x'_i \beta)}{h(\alpha_i + x'_i \beta)} = 1\). This is the perfect proportionality feature of the model.

For a logistic model, \(\frac{h'(\alpha_i + x'_i \beta)}{h(\alpha_i + x'_i \beta)} = \frac{\exp[-(\alpha_i + x'_i \beta)]}{1 + \exp[-(\alpha_i + x'_i \beta)]}\).

For C-loglog model, \(\frac{h'(\alpha_i + x'_i \beta)}{h(\alpha_i + x'_i \beta)} = \frac{\exp\{-\exp(\alpha_i + x'_i \beta)\} \times \exp(\alpha_i + x'_i \beta)}{1 - \exp\{-\exp(\alpha_i + x'_i \beta)\}}\).

Define \(Z = \alpha_i + x'_i \beta\). Figure 5 plots the relative change \(\frac{h'(z)}{h(z)}\) for various model specifications. The degree of non-proportionality inherent in each model is captured by the absolute value of its slope. As we see, logistic, c-loglog and cox proportional hazard model do not differ much for very low values of \(z\), exhibiting almost perfectly proportional hazards.
Figure 5 Specification Test: Relative change in hazard rate Figure 5 plots relative change in hazard rate, \( \frac{h'(z)}{h(z)} \), for various hazard model specifications. Blue curve is for logistic model, red curve is for C_Loglog model, and yellow curve is for Cox’s proportional hazard model. 

In our corporate default prediction context, the hazard (default) rate is very low, close to 1% (0.1%) annually (monthly). See table 1. This implies a value as low as about –4.6 (-6.9) for \( z \) in Figure 5. Therefore, the change in the hazard rate arising from changes in the explanatory variables is almost perfectly proportional, irrespective of the model used. We believe the choice of hazard model specification is innocuous in our context of firm’s default prediction. For ease of computation, we can choose a logistic or C-loglog model.

5.2.2. Proportionality test in Cox’s model

In this section, we check the proportional hazard feature of our data through a proportionality test in Cox’s model. If the proportional hazards assumption holds, then
for large samples, the Schoenfeld residuals should be zero for each predictor and will be approximately uncorrelated (See Singer and Willett, 2003 and our appendix B).

As Schoenfeld (1982) suggested, we plotted the residuals against the ranked duration time for each predictor in Figure 6. We fail to find a noticeable non-horizontal trend for all the residual patterns. This indicates the effect of the predictor is constant over time. We also compute the simple correlation between Schoenfeld residuals and rank of duration time in Table 7. Most of the correlations are statistically insignificant. For some with small \( p \)-values, the degree of correlation is also low. In sum, our data are consistent with the proportionality assumption.

Table 7: Pearson correlation coefficient between rank of duration and schoenfeld residual

Table 7 shows the correlation coefficients between rank of duration and schoenfeld residual of each predictor (WC/TA, sigmaRET, RLSIZE, RETEXCESS, Factor1, Factor2, Factor3, Factor1*size, Factor2*size, Factor3*size). \( p \)-values are in parentheses. There are 377 defaults from 1999 to 2008.

<table>
<thead>
<tr>
<th>WC/TA</th>
<th>SigmaRET</th>
<th>RLSIZE</th>
<th>EXCESSRET</th>
<th>Factor1</th>
<th>Factor2</th>
<th>Factor3</th>
<th>Factor1*size</th>
<th>Factor2*size</th>
<th>Factor3*size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank of Duration Time</td>
<td>0.006</td>
<td>0.099</td>
<td>-0.176</td>
<td>-0.141</td>
<td>-0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.912)</td>
<td>(0.056)</td>
<td>(0.001)</td>
<td>(0.006)</td>
<td>(0.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor1*size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor2*size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor3*size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank of Duration Time</td>
<td>-0.008</td>
<td>-0.037</td>
<td>0.015</td>
<td>0.007</td>
<td>0.057</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.872)</td>
<td>(0.475)</td>
<td>(0.773)</td>
<td>(0.886)</td>
<td>(0.266)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6. Schoenfeld residuals from one-month-ahead Cox’s model estimation

Figure 6 plot schoenfeld residual for each predictor (WC/TA, sigmaRET, RLSIZE, RETEXCESS, Factor1, Factor2, Factor3, Factor1*size, Factor2*size, Factor3*size) against the rank of duration time. See section 3.2.1 for definition of the predictors. The schoenfeld residuals are obtained from Cox’s proportional model regression in section 5.2.2. There are 377 defaults from 1999 to 2008. If the proportional hazards assumption holds, then for large samples, the Schoenfeld residuals should be zero for each predictor and will be approximately uncorrelated. We will not find a noticeable non-horizontal trend for all the residual patterns.
Figure 6 (cont’d). Schoenfeld residuals from one-month-ahead Cox’s model estimation

Figure 6 plot schoenfeld residual for each predictor (WC/TA, sigmaRET, RLSIZE, RETEXCESS, Factor1, Factor2, Factor3, Factor1*size, Factor2*size, Factor3*size) against the rank of duration time. See section 3.2.1 for definition of the predictors. The schoenfeld residuals are obtained from Cox’s proportional model regression in section 5.2.2. There are 377 defaults from 1999 to 2008. If the proportional hazards assumption holds, then for large samples, the Schoenfeld residuals should be zero for each predictor and will be approximately uncorrelated. We will not find a noticeable non-horizontal trend for all the residual patterns.
Figure 6 (cont’d). Schoenfeld residuals from one-month-ahead Cox’s model estimation

Figure 6 plot schoenfeld residual for each predictor (WC/TA, sigmaRET, RLSIZE, RETEXCESS, Factor1, Factor2, Factor3, Factor1*Size, Factor2*Size, Factor3*Size) against the rank of duration time. See section 3.2.1 for definition of the predictors. The schoenfeld residuals are obtained from Cox’s proportional model regression in section 5.2.2. There are 377 defaults from 1999 to 2008. If the proportional hazards assumption holds, then for large samples, the Schoenfeld residuals should be zero for each predictor and will be approximately uncorrelated. We will not find a noticeable non-horizontal trend for all the residual patterns.
6. Conclusions and Extension

Default data show clear evidence of cycles. In this chapter, we include macro factors as additional covariates in hazard models to capture the time-variation in default rate and the impact of business cycle on firms’ creditworthiness. We construct macro factors from fifty macroeconomic time series data using principal component analysis. Our proposed hazard model delivers a noticeable improvement over Shumway’s (2001) model in terms of model fit and predictive power. Macroeconomic conditions affect firm’s creditworthiness differently. In our paper, the latent macro factors load more on financially healthier firm with bigger size. This is consistent with findings in McNeil and Wendi (2007) and Koopman and Lucas (2008)).
Unlike Sueyoshi (1995), our specification tests suggest the choice of a hazard model specification is innocuous in our context of firm’s default prediction. The effects of the covariates on default rate are nearly proportional, and do not differ much across specifications. For ease of computation, we adopt a simple dynamic logistic hazard model. The forecasted default probability with a horizon of one year will be used as a measure of default risk in Chapter 3.

Throughout the paper, we assumed no unobserved heterogeneity in the hazard. Lancaster (1979) argued that in the specification of the hazard there might be omitted explanatory variables. Singer and Willett (1992) note that "omission of an important independent variable amounts to pooling of heterogeneous populations defined by the different values of the omitted predictor" (p. 38). It’s well known that the neglect of unobserved heterogeneity may cause spurious state dependence of the estimated hazard (See Allison, 1982, Lancaster, 1979, Nickell, 1979). Heckman and Singer (1984) proposed a semiparametric method to make the correction. Jain and Vilcassim (1991) illustrate the use of an unobserved heterogeneity component in their investigation of household purchase timings. We believe accounting for unobserved heterogeneity in our duration analysis of a firm’s default risk to be a promising direction for future research.
Appendix A. Partial Likelihood Estimation

Let \( t_i \) denote the observed time (either censoring time or event time) for subject \( i \). Based on the hazard function (13), a partial likelihood can be constructed from the datasets as

\[
L(\beta) = \prod_{t_i \text{ uncensored}} \frac{\exp(X'_i \beta)}{\sum_{t_j \geq t_i} \exp(X'_j \beta)}
\]

Each firm \( i \)'s contribution to the partial likelihood is the ratio of its true risk score at its moment of bankruptcy \( \exp(X'_i \beta) \) to the sum of the contemporaneous true risk scores among every firm still at risk. The corresponding log partial likelihood is

\[
l(\beta) = \log L(\beta) = \sum_{t_i \text{ uncensored}} \left\{ X'_i \beta - \log \sum_{t_j \geq t_i} \exp(X'_j \beta) \right\}
\]

Clearly, the partial likelihood depends only on the order in which events occur, not the actual time at which the event occur. So it can be viewed as the joint density function for subjects’ ranks in terms of event order. If we use the partial likelihood to estimate the model parameters, we are losing some information as we are not using the actual times of event even though they are known. Though some journal papers showed that the information lost by ignoring actual event time is smaller than one might expect.

Breslow’s estimator for baseline cumulative hazard function is given by

\[
\hat{H}_0(t) = \sum_{t_i \leq t} h_0(t) = \sum_{t_i \leq t} \frac{1}{\sum_{t_j \leq t} \exp(X'_j \beta)}
\]
Appendix B. Schoenfeld residuals

Schoenfeld residuals, also known as partial residuals, compare the observed and expected predictor values. For each predictor $X$, we define the Schoenfeld residual for firm $i$ as

$$
\hat{SR}_i(X) = X_i - \text{expected}_i(X)
$$

$$
= X_i - \frac{\sum_{j \geq t_i} X_j \exp(X_j \hat{\beta})}{\sum_{j \geq t_i} \exp(X_j \hat{\beta})}
$$

$$
= X_i - \bar{X}(t_i, \hat{\beta})
$$

where $\bar{X}(t_i, \hat{\beta})$ is the weighted average of the predictor $X$ for firms still in the risk set when firm $i$ experiences the event (default) at time $t_i$. Schoenfeld residuals are defined only for firms who actually default, as the residuals compare predictor values at observed default times only. Therefore, censored firms do not have Schoenfeld residuals. Note that each firm’s value of predictor $X$ is weighted by its risk score $\exp(X \beta)$. 

105
References:


Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. Econometrica 70 (1), 191-221.


G. Rodriguez (2007) Lecture notes, Chapter 7 survival models


Chapter 3: The Effect of Default Risk on Cross-section Stock Returns

Abstract

This chapter explores the effect of default risk on stock returns using default probability estimated from our one-year ahead hazard model in chapter 2. With this better measure for distress risk, our study captures a pattern different from that of current empirical literature. We note that, by and large, the return-risk in cross-sectional stock displays an inverted U-shaped relation. To uncover this credit risk puzzle (i.e., the negative risk-return relation for the distressed portfolios), we further investigate the portfolio returns after risk-adjusting by several empirical asset pricing models. We find that this anomaly could be explained by a momentum effect using Carhart’s (1997) four-factor asset pricing model. We also conduct tests for competing asset pricing models. We estimate stochastic discount factor (SDF) representation of linear factor pricing models by using generalized method of moments (GMM) approach, and find evidence that support default risk as a factor that should be considered in explaining stock returns. A four-factor asset pricing model augmented with a default risk factor is a superior model in explaining stock return in terms of Hansen and Jagannathan (1997) (HJ) distance.
1. Introduction

Most investors are comfortable with the notion that higher-risk assets should be expected to produce higher returns. This risk-return tradeoff underlies the conceptual framework of asset pricing models and investment decisions. We define a firm’s default risk the same as in chapter 2, i.e., a firm’s failure to meet its financial obligation, which includes bankruptcies, delisting for financial reasons, or D(default) ratings issued by a leading credit rating agency S&P. Then, intuitively, investors should be compensated for holding stocks of firms with higher default risk. And we expect to observe positive relation between default risk and stock return.

However, empirical studies that examine the relation between default risk and stock returns yield quite different results. And the results vary with the measure of financial health status that is used. Vassalou and Xing (2004), who calculated distance to default (a default risk measure based on the Merton (1974) model), find distressed stocks with a low distance to default have higher returns. Yet, Campbell et al. (2008), who use failure probability estimated from a dynamic logit model, and Avramov (et al.2009), who use credit ratings to measure firm’s financial health document a puzzling credit risk effect, i.e. low credit risk firms realize higher returns than riskier firms. This is puzzling because investors seem to pay a premium for bearing credit risk.

One purpose of this paper is to revisit this open issue (i.e. the relation between default risk and stock returns), using default probability generated by our best model from chapter two as an empirical measure of firm’s distress risk. With this better measure

---

1 The stock price formulation in Jarrow (2001) showed how the credit risk of a firm could affect its stock price, hence stock return, when stock is interpreted as a debt with the last seniority that regularly pays dividends as coupons. This relationship is also evident when we consider Merton (1974)’s structural model or its extension.
for distress risk, our study captures a pattern different from that of current empirical literature. We note that in our sample, the portfolios exhibit mean excess returns that do not increase, nor decrease monotonically with default risk dimension. By and large, the return-risk displays an inverted U-shaped relation. To uncover this credit risk puzzle (i.e., the negative risk-return relation for the distressed portfolios), we further investigate the portfolio returns after risk-adjusting by several empirical asset pricing models. We find that this anomaly could be explained by a momentum effect using Carhart’s (1997) four-factor asset pricing model. Our model’s result is more economically intuitive, since investors should be compensated for holding risky assets.

The default probability measure used in our study is more related with that of Campbell et al. (2008). But they do not model the impact of macroeconomic state on firms of different characteristics as we do. Our PD measure is derived from a hybrid model in chapter 2, incorporating as predictors firm specific characteristic, macroeconomic factors as well as interaction between these terms. So it may capture information missing in Campbell et al. (2008).

We do not using credit ratings data to examine the portfolio risk-return relation for several reasons. For one thing, the rating data may not capture changes in a firm’s credit healthiness in a timely manner. Typically, a firm experiences a substantial change in its default risk prior to its rating change. For another, the current rating of a firm does not incorporate much of the influence of the business cycle on default rates (Nickell, Perraudin and Varotto, 2000; Kavvathas, 2001; Wilson 1997a, b). Also, firms under the same rating category exhibit significant heterogeneity in terms of short-term default probabilities (Kealhofer, 2003)
Based on the historical pattern of cross-sectional stock risk-return relation, we conduct tests for competing asset pricing models, and provide answers for the following research questions:

(1) Is default risk a factor that should be considered in explaining stock return?

(2) Does the default risk carry a premium?

(3) Can we find a best asset pricing model?

To address the first question, we run generalized method of moments (GMM) estimation in stochastic discount factor (SDF) representation of linear factor pricing models. We find evidence that supports default risk as a factor that should be considered in explaining stock return. Hansen and Jagannathan (1997) distance measure is used to compare the performance of competing asset pricing models. We also apply Kan and Robotti’s (2009) test to check whether our best asset pricing model exhibits significantly different HJ-distance from other alternatives. We run Fama-MacBeth (1973) regression on the cross-sectional regression model of portfolio return, and find that default risk does carry a premium.

Our work is similar in spirit to that of Vassalou and Xing (2004). They use Merton’s (1974) option pricing model to compute default measures for firms and assess the effect of default risk on equity returns. Our main contribution to the current literature is that we use a better measure for default risk, and include it as a factor in asset pricing model tests. Our default risk measure is derived from a hybrid model, incorporating richer information as predictors, such as firm specific characteristics, macro economic factors as well as interaction between these terms. We also consider a momentum effect and
Carhart’s (1997) four-factor asset pricing model in analysis and model testing, which is missing in Vassalou and Xing (2004).

The remainder of this chapter is organized as follows. In section 2, we examine the historic pattern of portfolio return-risk relation. We report portfolio return before and after risk-adjusting by various risk-based asset pricing models. Section 3 presents the econometric specification and tests for asset pricing models; provides data description, and show the results of empirical analysis. Concluding remarks are given in section 4.

2. Default Risk and Cross Section Equity Return

In this section, we do some preliminary analysis on the historic pattern of stock portfolio return-default risk relation. We first report the raw return for default decile portfolios, and then we investigate returns on portfolios after risk-adjusting by several empirical asset pricing models.

2.1.1 Raw Return Pattern

Following Dichev (1998), Griffin and Lemmon (2002) and Vassalou and Xing (2004), we sort all stocks into 10 portfolios based on our predicted probability of default from chapter 2 at the beginning of each month. That is, the portfolios are rebalanced monthly in response to updated default probability.\(^2\) The sample period is January 1999-December 2008.

\(^2\) We also perform an additional analysis using returns on portfolios rebalanced in June of every year. The central message of the paper is not affected by the formation of portfolios.
Panel A of Table 1 shows the characteristics of default-risk sorted stock portfolios. We report mean excess return and standard deviation, as well as average size (in logarithm), book-to-market ratio, probability of default for each stock portfolio.

We note that in our sample, the portfolios exhibit mean excess returns that do not increase, nor decrease monotonically with default risk dimension. By and large, the return-risk displays an inverted U-shaped relation. The mean excess returns are lowest at the safest decile portfolios with 2.16% per year, and increases along the risk dimension till 7th decile portfolios at highest return of 8.4% per year, and then decreases with the risk dimension.

The return standard deviation displays a U-shaped relation. The standard deviations of the safest and the most distressed decile portfolios are the highest. For the median risk decile, average excess return is 6.6% per year, the standard deviation is 22%, so its sharpie ratio, i.e., the ratio of mean excess return to standard deviation, is comparable to that of the aggregate market. 3.

The patterns for the average size and book-to-market of different default decile are very clear. Size is monotonically decreasing along the risk dimension, while book-to-market is increasing instead. This pattern is consistent with the notion that small size firm or high book-to-market firms’ stocks have higher default risk exposure, in favor of Fama and French’s (1993) small-cap and value risk factors. In our sample, Stocks in the lowest default risk decile have an average size of 21.87 billion dollars, a book-to-market ratio of 0.24, while stocks in the highest default risk decide have an average size of 137.53 million dollars, a book-to-market ratio of 2.48.

3 According to Cochrane (2005), the 50 years postwar U.S. data, aggregate stock market has an average return of 8% per year in excess of treasury bills rate, while the standard deviation of the stock return has been about 16%, thus the historical annual market Shape ratio has been about 0.5.
Table 1: stock portfolios characteristics, alphas, and betas by default risk decile

From January 1999 to December 2008, at the beginning of each month, stocks are sorted into 10 portfolios based on the predicted probability of default from our hybrid model. Equally weighted average portfolio returns are reported in percentage terms. In panel A, we report annualized standard deviation and skewness of portfolio returns (in %), as well as average size (in logarithm), book-to-market ratio, mean probability of default for each portfolio. In panel B and panel C, we show results from time-series regressions of the risk decile excess return over the market on a constant, market return (MKT), as well as three (MKT, SMB, HML) and four (MKT, SMB, HML, UMD) Fama-French factors. Reported alphas and returns are annualized in % terms. The robust Newy and West (1987) t-statistics are in parentheses. * denotes significant at 10%, ** denotes significant at 5% level, *** denotes significant at 1% level.

<table>
<thead>
<tr>
<th></th>
<th>Lowest Risk</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Highest Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: 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>log(Size($MM))</td>
<td>9.99</td>
<td>9.05</td>
<td>8.36</td>
<td>7.90</td>
<td>7.50</td>
<td>7.05</td>
<td>6.60</td>
<td>6.11</td>
<td>5.71</td>
<td>4.92</td>
</tr>
<tr>
<td>Book-to-market ratio</td>
<td>0.24</td>
<td>0.34</td>
<td>0.42</td>
<td>0.49</td>
<td>0.56</td>
<td>0.65</td>
<td>0.74</td>
<td>0.89</td>
<td>1.17</td>
<td>2.48</td>
</tr>
<tr>
<td>Ave.Default prob. (%)</td>
<td>0.004</td>
<td>0.008</td>
<td>0.011</td>
<td>0.015</td>
<td>0.020</td>
<td>0.027</td>
<td>0.039</td>
<td>0.067</td>
<td>0.176</td>
<td>1.663</td>
</tr>
<tr>
<td>Portfolio SD</td>
<td>0.32</td>
<td>0.25</td>
<td>0.21</td>
<td>0.22</td>
<td>0.22</td>
<td>0.23</td>
<td>0.24</td>
<td>0.26</td>
<td>0.29</td>
<td>0.41</td>
</tr>
<tr>
<td>Portfolio Skewness</td>
<td>2.56</td>
<td>-0.42</td>
<td>-2.42</td>
<td>-2.65</td>
<td>-2.29</td>
<td>-2.65</td>
<td>-2.98</td>
<td>-2.00</td>
<td>0.81</td>
<td>5.41</td>
</tr>
<tr>
<td>Annualized Mean excess return</td>
<td>2.16</td>
<td>4.2</td>
<td>5.52</td>
<td>7.08</td>
<td>6.6</td>
<td>7.32</td>
<td>8.4</td>
<td>5.88</td>
<td>5.04</td>
<td>4.8</td>
</tr>
<tr>
<td>(in percent units)</td>
<td>(0.22)</td>
<td>(0.54)</td>
<td>(0.83)</td>
<td>(1.03)</td>
<td>(0.96)</td>
<td>(1.03)</td>
<td>(1.13)</td>
<td>(0.71)</td>
<td>(0.55)</td>
<td>(0.37)</td>
</tr>
<tr>
<td><strong>Panel B: Factor pricing models regression coefficients (alpha)</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>CAPM alpha</td>
<td>4.2</td>
<td>6.72</td>
<td>7.68</td>
<td>9.24</td>
<td>8.52</td>
<td>9.72</td>
<td>10.56</td>
<td>8.4</td>
<td>8.76</td>
<td>10.8</td>
</tr>
<tr>
<td></td>
<td>(0.72)</td>
<td>(2.07)**</td>
<td>(2.87)**</td>
<td>(2.95)***</td>
<td>(2.58)**</td>
<td>(2.65)**</td>
<td>(2.8)***</td>
<td>(1.87)*</td>
<td>(1.62)</td>
<td>(1.21)</td>
</tr>
<tr>
<td>Fama-French 3-factor alpha</td>
<td>1.92</td>
<td>2.88</td>
<td>2.76</td>
<td>3.12</td>
<td>1.68</td>
<td>2.04</td>
<td>2.52</td>
<td>-0.36</td>
<td>1.32</td>
<td>2.52</td>
</tr>
<tr>
<td></td>
<td>(0.6)</td>
<td>(1.65)</td>
<td>(1.44)</td>
<td>(1.37)</td>
<td>(0.73)</td>
<td>(0.84)</td>
<td>(1.02)</td>
<td>(-0.11)</td>
<td>(0.26)</td>
<td>(0.29)</td>
</tr>
<tr>
<td>Carhart (1997) 4-factor alpha</td>
<td>-0.36</td>
<td>2.28</td>
<td>3.6</td>
<td>4.32</td>
<td>3.12</td>
<td>4.08</td>
<td>4.68</td>
<td>2.64</td>
<td>5.76</td>
<td>10.32</td>
</tr>
<tr>
<td></td>
<td>(-0.14)</td>
<td>(1.34)</td>
<td>(1.98)*</td>
<td>(2.09)**</td>
<td>(1.58)</td>
<td>(2.02)**</td>
<td>(2.56)**</td>
<td>(1.14)</td>
<td>*</td>
<td>(1.64)</td>
</tr>
</tbody>
</table>

(To be cont’d)
Table 1 (cont'd): stock portfolios characteristics, alphas, and betas by default risk decile

<table>
<thead>
<tr>
<th>Portfolios</th>
<th>Lowest</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Highest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Risk</td>
</tr>
</tbody>
</table>

**Panel C: Factor pricing models regression coefficients (Loadings on the factor)**

**CAPM**

<table>
<thead>
<tr>
<th>MKT beta</th>
<th>1.3683</th>
<th>1.2448</th>
<th>1.1186</th>
<th>1.0453</th>
<th>1.0125</th>
<th>1.0142</th>
<th>1.0191</th>
<th>1.0779</th>
<th>1.2895</th>
<th>1.7135</th>
</tr>
</thead>
</table>

**Fama-French 3-factor model:**

<table>
<thead>
<tr>
<th>MKT beta</th>
<th>0.9938</th>
<th>1.0865</th>
<th>1.1015</th>
<th>1.0736</th>
<th>1.0958</th>
<th>1.0948</th>
<th>1.1052</th>
<th>1.1617</th>
<th>1.3234</th>
<th>1.6952</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(16.39)**</td>
<td>(32.4)**</td>
<td>(30.23)**</td>
<td>(24.74)**</td>
<td>(25.87)**</td>
<td>(23.04)**</td>
<td>(23.37)**</td>
<td>*</td>
<td>(14.33)**</td>
<td>(10.3)**</td>
</tr>
<tr>
<td>SMB beta</td>
<td>0.7898</td>
<td>0.5914</td>
<td>0.4724</td>
<td>0.5176</td>
<td>0.5047</td>
<td>0.5683</td>
<td>0.6016</td>
<td>0.6769</td>
<td>0.6358</td>
<td>0.7846</td>
</tr>
<tr>
<td></td>
<td>(11.16)**</td>
<td>(15.1)**</td>
<td>(11.1)**</td>
<td>(10.22)**</td>
<td>(10.21)**</td>
<td>(10.25)**</td>
<td>(10.9)**</td>
<td>(9.21)**</td>
<td>(5.9)**</td>
<td>(4.08)**</td>
</tr>
</tbody>
</table>

| HML beta  | -0.5847 | -0.0540 | 0.3008 | 0.4776 | 0.6407 | 0.6803 | 0.7223 | 0.7716 | 0.5839 | 0.5314 |
|           | (-7.16)**| (-1.2)**| (6.13)**| (8.17)**| (11.24)**| (10.64)**| (11.34)**| (9.1)**| (4.7)**| (2.4)**|

**Carhart (1997) 4-factor model**

<table>
<thead>
<tr>
<th>MKT beta</th>
<th>1.1453</th>
<th>1.1267</th>
<th>1.0451</th>
<th>0.9907</th>
<th>0.9991</th>
<th>0.9648</th>
<th>0.9607</th>
<th>0.9595</th>
<th>1.0267</th>
<th>1.1792</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(20.88)**</td>
<td>(32.02)**</td>
<td>(28.06)**</td>
<td>(23.16)**</td>
<td>(25.16)**</td>
<td>(23.52)**</td>
<td>(25.44)**</td>
<td>*</td>
<td>(14.61)**</td>
<td>(9.18)**</td>
</tr>
<tr>
<td>SMB beta</td>
<td>0.6940</td>
<td>0.5660</td>
<td>0.5081</td>
<td>0.5700</td>
<td>0.5658</td>
<td>0.6505</td>
<td>0.6930</td>
<td>0.8048</td>
<td>0.8235</td>
<td>1.1109</td>
</tr>
<tr>
<td>HML beta</td>
<td>-0.5191</td>
<td>-0.0366</td>
<td>0.2764</td>
<td>0.4418</td>
<td>0.5989</td>
<td>0.6240</td>
<td>0.6598</td>
<td>0.6842</td>
<td>0.4555</td>
<td>0.3082</td>
</tr>
<tr>
<td></td>
<td>(-7.55)**</td>
<td>(-0.83)**</td>
<td>(5.92)**</td>
<td>(8.24)**</td>
<td>(12.04)**</td>
<td>(12.14)**</td>
<td>(13.94)**</td>
<td>*</td>
<td>(5.17)**</td>
<td>(1.92)*</td>
</tr>
<tr>
<td>UMD beta</td>
<td>0.2849</td>
<td>0.0757</td>
<td>-0.1061</td>
<td>-0.1558</td>
<td>-0.1818</td>
<td>-0.2444</td>
<td>-0.2716</td>
<td>-0.3801</td>
<td>-0.5579</td>
<td>-0.9703</td>
</tr>
<tr>
<td></td>
<td>(7.15)**</td>
<td>(2.96)**</td>
<td>(-3.92)**</td>
<td>(-5.01)**</td>
<td>(-6.3)**</td>
<td>(-8.2)**</td>
<td>(-9.9)**</td>
<td>(-10.91)**</td>
<td>(-10.93)**</td>
<td>(-10.4)**</td>
</tr>
</tbody>
</table>

118
Overall, the most striking pattern is the variation in mean excess returns across the portfolios in Panel A of Table 1. To sum up, we did not observe a default risk effect (positive average return differential between high and low default risk firms) in highly distressed firms (from 8th to 10th decile) as is observed in firms below 8th risk decile. Instead, we document a negative risk-return relation for the distressed portfolios. This is puzzling because investors seem to pay a premium for bearing credit risk. We are motivated to do some further investigations next.

2.1.2 Risk-adjusted Return Pattern

Is it possible that the negative risk-return relation for the distressed portfolios is related to a firm’s systematic risk? To check if this puzzling credit risk effect is captured by risk-based pricing model, we also report returns on the 10 portfolios after risk-adjusting by several empirical asset pricing models. Here, we do not assume which asset pricing model is better or best. We leave the discussion for the test of these pricing models to section 3.

Panel B and panel C of Table 1 show results from time-series regressions of the risk decile portfolio return in excess of the risk-free rate on factors from the CAPM model, as well as Fama and French three-factor (MKT, SMB, HML) model and a four-factor (MKT, SMB, HML, UMD) model proposed by Carhart (1997)\(^5\) Reported alphas are

\(^4\) The 6th risk decile firms in our data have an average S&P rating of BBB-. Firms in the 7th decile and 8th decile have an average rating of BB+ and BB-, respectively. For firms for which we have S&P Long-term Domestic Issuer Credit Ratings data, we convert their S&P rating into conventional numerical scores according to the rule as follows: AAA=1, AA+=2, AA=3, AA-=4, A+=5, A=6, A-=7, BBB+=8, BBB=9, BBB-=10, BB+=11, BB=12, BB-=13, B+=14, B=15, B-=16, CCC+=17, CCC=18, CCC-=19, CC=20, C=21, D=22.

\(^5\) MKT is excess return on the market portfolio. SMB (Small Minus Big) is a measure of size risk, constructed as the return difference between portfolios of small and large stocks. HML (High Minus Big) captures the risk related to firm book-to-equity, defined as the return difference between portfolios of high
annualized in percentage terms. The robust Newy and West (1987) t-statistics are in parentheses. Figure 1 ~ Figure 4 graphically summarize the behavior of factor loadings and alphas.

**Factor loadings (betas) and alpha from asset-pricing models**

In risk-based asset pricing models, factor loadings (betas) measure the degree to which an asset’s returns vary relative to those (risk) factors. If the size effect (SMB) and value effect (HML) can be viewed as default effects, then high risk firms should have high factor loadings (exposure) to these risk factors. It is a pattern found in major empirical finance studies such as works of Fama and French (1993), Campbell et. al. (2008).

In our paper, we note by and large, market risk beta and the portfolio loadings on small cap risk factor SMB are increasing along the dimension of default risk, either in three-factor or four-factor models. It implies that MKT or SMB factors will not help explain the low average return in distressed stocks.

The portfolio loadings on value risk factor HML is decreasing for the financially distressed stocks (decile 8 through decile 10), indicating HML is not likely to be a distress factor. Yet financially distressed stocks still have much higher loadings on value risk factors HML than stocks with low failure risk (decile 1 through decile 3). For example, in Fama-French 3-factor model (Panel C of table 1), factor loadings on HML range from 0.53 to 0.77 for decile 8 through decile 10, from –0.58 to 0.30 for decile 1 through decile 3, respectively. Therefore, we notice that the pattern of HML loadings and low book-to-market ratios (i.e., the value placed on the firm by accountants relative to the value the public markets placed on the firm). UMD is the momentum factor introduced and validated by Carhart (1997). It measures the difference in return between the better and worse performing stocks, or “up minus down” (UMD). The time series data for MKT, SMB, HML and UMD are all obtained from Ken French’s website.
Mean excess return and Alpha of default-risk sorted portfolios

Figure 1 Mean excess return and Alphas of default-risk sorted portfolios

The figure plots the mean monthly excess return over the market, CAPM alpha, Fama-French three-factor alpha and Carhart’s four-factor alpha for the 10 default-risk sorted portfolios from 1999 to 2008. The portfolio return is equally weighted. Portfolios are formed at the beginning of each month using the model-predicted probability of default from chapter 2. Default is defined as failure to meet its financial obligation, including bankruptcies, delisting for financial reasons, or D (default) ratings issued by a leading credit rating agency S&P.

Factor loadings of default-risk sorted portfolios (CAPM)

Figure 2 Factor loadings of default-risk sorted portfolios (CAPM model)

The figure plots loadings on the excess market return (MKT) for the 10 default-risk sorted portfolios from 1999 to 2008. The portfolio return is equally weighted. Portfolios are formed at the beginning of each month using the model-predicted probability of default from Chapter 2.
Factor loadings of default-risk sorted portfolios (Fama-French 3-factor model)

The figure plots loadings on the excess market return (MKT), the value factor (HML), and the size factor (SMB) for the 10 default-risk sorted portfolios from 1999 to 2008. The portfolio return is equally weighted. Portfolios are formed at the beginning of each month using the model-predicted probability of default from chapter 2.

Figure 3 Factor loadings of default-risk sorted portfolios (Fama-French 3 factor model)

Factor loadings of default risk-sorted portfolios (Carhart’s 4 factor model)

The figure plots loadings on the excess market return (MKT), the value factor (HML), the size factor (SMB), and momentum effect (UMD) for the 10 default-risk sorted portfolios from 1999 to 2008. The portfolio return is equally weighted. Portfolios are formed at the beginning of each month using the model-predicted probability of default from chapter 2.

Figure 4 Factor loadings of default-risk sorted portfolios (Carhart’s 3 factor model)
helps to correct the negative risk-return relation in the distressed firm stocks, but cannot explain the overall underperformance of these stocks compared with stocks of low default risk. (See the alpha from Fama-French three-factor model in Panel C of Table 1 and Figure 3).

A more striking pattern is shown by the factor loadings on a momentum factor UMD in Carhart (1997) four-factor model. Beta for UMD is decreasing monotonically along the default risk dimension, from 0.285 for the safest decile to –0.97 for the most risky decile. Remember in our default prediction analysis of chapter 2, one significant predictor is recent past return. This suggests that the negative momentum might explain the low average return in distressed stocks. Indeed, this is what we find. After adjusting by four-factor model including a momentum effect, the average returns (alphas) are overall increasing along the default risk dimension. This result is more economically intuitive, since investors get compensated for holding risky assets. A long-short portfolio holding the lowest-default risk decile stocks and shorting the highest-default risk decile has an average return of 10.68 basis points per year, which is statistically significant at the 10% level.  

We note after adjusting by four-factor model, alpha is statistically significant for most default risk deciles. If we assume the market portfolio is the correct benchmark, then it suggests there might still be some mispricing for the asset return in the model. Possibly, an additional default risk factor is worth considering. We will examine it next in section 3.

---

6 When we sort stocks into five portfolios, the results are qualitatively the same, i.e., average returns appear to be higher the higher the default measure.
Sum Up

The credit risk puzzle we document in our paper, namely, negative risk-return relation for the distressed portfolios, can be explained by the Carhart (1997) four-factor asset pricing model including momentum effect. Given that mispricing still exists in the adjusted model, it is very meaningful to check if default risk is another factor worth considering in explaining cross-sectional stock return. We will do it next.

3. Pricing of Default Risk---Asset Pricing Model Test

One of the findings from chapter 2 suggests that default risk is affected by macroeconomic factors. Intuitively, the state of the economy will affect a firm’s performance. It may lead to unexpected changes in the market value of the firm’s assets, “generating market risk---this affects the probability of default---generating credit risk.” (Jarrow and Turnbull (2000), p.272). We note that firm’s bankruptcy and default do rise markedly when the economy is overall in downturn. So it invites us to suggest a systematic component in firm’s default risk. Another possible argument for the systematic default risk is the ripple effects or contagious effect the default firm may have on other distressed firms (see Jorion and Zhang (2007)).

If default risk is systematic and cannot be diversified away, it should therefore be priced in the cross-section of equity returns. On the other hand, defaults could be rare events in the sense that a small number of firms malfunction may be due to its own idiosyncratic risk. If the latter effect dominates, we may not see stockholders getting compensated for holding stocks with high default risk.
This section serves the goal of investigating through asset pricing models, whether default risk is a factor worth considering in pricing stock return, or equivalently, whether default risk is predominantly systematic, so it should be priced. Specifically, we test if the stochastic discount factor (SDF) parameter associated with the default risk factor is significantly different from zero in competing asset pricing models through the Generalized Method of Moments (GMM) methodology of Hansen (1982). We also check whether the default risk factor carries a premium through cross-sectional regression (CSR) method proposed by Fama and MacBeth (1973). In this paper, we focus on linear factor pricing models—the most popular models in the empirical asset pricing literature.

3.1 Econometric Specification and Tests of Linear Factor Pricing Models

There are two equivalent ways to represent linear factor pricing models including the classical beta representation and the stochastic discount factor (SDF) representation.\(^7\)

1) Linear beta representation using Fama-MacBeth regression method

The linear beta representation describes the expected return on an asset as a linear function of its factor betas.

\[
E(R_i) = \lambda_1 \beta_{i1}^1 + \lambda_2 \beta_{i2}^2 + \cdots + \lambda_K \beta_{iK}^K = \lambda' \beta
\]  

where \( \lambda \equiv (\lambda_1, \cdots, \lambda_K)' \) is the parameter vector, \( \beta \equiv (\beta_1, \cdots, \beta_K)' \) is the vector of factor betas, and \( R_{it} \) is the return on \( i\)-th test portfolios in excess of the risk-free rate. Estimation

\(^7\) For derivation of the two representations, interested readers may refer to Cochrane (2005).
and specification test can be performed using cross-sectional regression (CSR) method proposed by Fama and MacBeth (1973).

According to Fama and MacBeth (1973), parameters for asset pricing models are estimated in two steps:

In the first step, factor betas $\beta$ are estimated through time-series regression of each asset’ return against the proposed risk factors.

In the second step, the factor risk premiums $\lambda$ are estimated using cross-sectional regression of asset returns on the factor betas $\hat{\beta}$ computed from the first step for a fixed time period. If factor betas $\beta$ are invariant over time, we can pool all the observations in second step, and do one simple cross-sectional regression. However factor betas $\beta$ may be changing over time. For the latter case, we can repeat the second step for different time period, then compute time-series average of the cross-sectional estimates.

One obvious drawback of the above Fama-MacBeth procedure is the error-in-variable problem arising from the estimation error in second step. Because the regressors (factor betas $\hat{\beta}$) are estimates themselves. In this paper, we use portfolios instead of individual stocks as test assets. Our practice substantially decreases the estimation error of factor betas in the first step, as pointed out by Black, Jensen and Scholes (1972).

Following the approach suggested by Shanken (1992) and Jagannathan and Wang (1998), we also estimated the sampling errors of factor betas $\hat{\beta}$, and computed bias-corrected standard errors for the risk premium estimators $\hat{\lambda}$ in the Fama-MacBeth procedure. Then usual t- test for the significance of risk premium estimators $\hat{\lambda}$ can be applied as well.
2) Stochastic discount factor (SDF) representation using GMM estimation

The GMM approach has the advantage that it allows for estimation of model parameters in a single pass, thereby avoiding the error-in-variables problem arising from traditional second step estimation.

*SDF model and HJ distance*

Let the proposed SDF (or pricing kernel) \( m \) be a linear function of \( K \) systematic factors \( f \):

\[
m(b) = 1 + b_1 f_1 + b_2 f_2 + \cdots + b_K f_K = b'f
\]

where \( b \equiv (b_1, \cdots, b_K)' \) are the parameters and \( f \equiv (f_1, \cdots, f_K)' \) are the factors vector, and \( R \equiv (R_1, \cdots, R_N)' \) be a vector of returns on \( N \) test portfolios in excess of the risk-free rate. Then \( e(b) \equiv E(Rm(b)) \) is the \( N \)-vector of pricing errors of the model.

When the asset pricing model is correctly specified (i.e. SDF \( m \) correctly priced the \( N \) portfolios), the \( N \)-dimensional pricing errors \( E(Rm) \) should be zero, i.e.

\[
e(b) \equiv E(Rm(b)) = 0_N. \quad (3)
\]

This is the model for the moments. GMM approach can be utilized to test if some factor \( f_K \) is priced in the asset return by checking the parameters \( b_K \) for that factor.

When a model is misspecified, we can evaluate the relative performance of several model specifications by comparing the size of pricing errors. For this purpose, we use the Hansen and Jagannathan (1997) (HJ) distance measure to find the best model.

HJ distance is a scalar measure of the magnitude of the asset pricing model misspecification, defined as the square root of a quadratic form of the pricing errors.

---

\(^8\) See Cochrane (2005) for more explanation.
\[ Dist(b) = \sqrt{e(b)'G^{-1}e(b)} \]  

(4)

where weighting matrix \( G^{-1} \) is the inverse of second moment matrix of \( R \). Note that this weighting matrix remains the same across various model specifications, enabling the comparison of pricing error associated with different models. We do not use the most famous “optimal” weighting matrix suggested by Hansen and Singleton (1982) here, because it would yield misleading conclusion. See Hansen and Jagannathan (1997) for details.

Equation (4) shows that the HJ distance depends on the parameters \( b \). Empirically, the values for the vector \( b \) is chosen such that HJ distance is minimized. Under this choice of \( b \), the HJ-distance is defined as

\[ Dist(b) = \sqrt{\min_b e(b)'G^{-1}e(b)} \]  

(5)

We compute the \( p \)-value of the HJ distance through simulation.\(^{10}\)

**Testing Equality of HJ distance**

We apply Kan and Robotti’s (2009) test to investigate whether different asset pricing models exhibit significantly different HJ-distance measures. Kan and Robotti’s (2009) test proposes a misspecification robust version of standard errors for the estimates of the SDF parameters. This test can be applied to both correctly specified and misspecified models.

\(^9\) The optimal weighting matrix is useful for testing whether the pricing errors are zero though. In our context, it is of the form \( [\text{var}(Rm(b))]^{-1} \), i.e., the inverse of the variance matrix of the moment conditions. Note that if the model is noisier, i.e. \( [\text{var}(Rm(b))] \) is larger, then the value of the quadratic form will be smaller, implying a smaller model error. So it gives misleading conclusion.

\(^{10}\) Hansen and Jagannathan (1997) show the asymptotic distribution of \( \{T[Dist(b)]\}^2 \) is a weighted sum of \( \chi^2 \)-distributed random variables, each having 1 degree of freedom. So we simulate the weighted sum of \( N-K \times \chi^2(1) \) random variables 100,000 times to derive the \( p \)-value for HJ distance.
In our paper, we apply the test to nested models case. Assume model 2 nests model 1. Let \( y_1 = \begin{bmatrix} 1, f_1' \end{bmatrix} \), \( y_2 = \begin{bmatrix} 1, f_1', f_2' \end{bmatrix} \), where \( f_1 \) and \( f_2 \) are two sets of distinct factors, and \( f_i \) is of dimension \( K_i \times 1, i = 1, 2 \). We assume the SDF of model 1 is given by \( m_1 = 1 + \delta y_1 \), and the SDF of model 2 is given by \( m_2 = 1 + \gamma y_2 \). Let \( \gamma_2 \) be a vector the last \( K_2 \) elements of \( \lambda \). The test of equality of HJ-distance of the two models is simply a test of the null hypothesis \( H_0 : \gamma_2 = 0_{K_2} \). For proof, see Kan and Robotti’s (2009). Under the null hypothesis \( H_0 : \gamma_2 = 0_{K_2} \), \( T\hat{\gamma}_2 V(\hat{\gamma}_2)^{-1} \hat{x}_2 \sim x_{K_2}^2 \), where \( V(\hat{\gamma}_2) \) is a consistent estimator of the asymptotic variance of \( \sqrt{T}(\hat{\gamma}_2 - \gamma_2) \).

3.2. Empirical Analysis

3.2.1 Test Models

In our empirical analysis, we test the following asset pricing models.

(1) The first model is augmented CAPM, including as factors the excess return of the market portfolio \( MKT \), and the aggregate default risk measure \( APD \). The cross-sectional regression model takes the form of

\[
E(R_i) = \lambda_0 + \lambda_{MKT} \beta_{MKT} + \lambda_{APD} \beta_{APD} \tag{6}
\]

and the model for the moments

\[
E[R_i (1 + b_{MKT} MKT_i + b_{APD} APD_i)] = 0 \tag{7}
\]

The test of CAPM model is to test if the intercept is zero. If we assume the market portfolio is the correct benchmark, then the test of CAPM model can be interpreted as a test of whether asset return \( R_i \) is correctly priced.
where $R_u$ is the return of a stock or a portfolio in excess of the risk-free rate, $APD_t$ is defined as a simple average of default probabilities $PD_{it}$ of all firms at time $t$. $\beta_i^{MKT}$ is obtained through OLS regression of $R_u$ on a constant and $MKT_t$. The other betas are estimated in a similar way.

(2) The second model is Fama-French (1993) three factor (MKT, SMB, and HML) models augmented with the aggregate default risk measure $APD$. It assumes cross-sectional regression specification takes the form

$$E(R_u) = \lambda_0 + \lambda_i^{MKT} \beta_i^{MKT} + \lambda_i^{SMB} \beta_i^{SMB} + \lambda_i^{HML} \beta_i^{HML} + \lambda_i^{APD} \beta_i^{APD}$$

(8)

and the model for the moments

$$E[R_u (1 + b_{MKT} MKT_t + b_{SMB} SMB_t + b_{HML} HML_t + b_{APD} APD_t)] = 0$$

(9)

Fama and French (1996) argue that the SMB and HML factors proxy for financial distress. We can test this hypothesis by including $APD$.

(3) The third model is a four-factor (MKT, SMB, HML, UMD) model proposed by Carhart (1997), augmented with the aggregate default risk measure $APD$. It assumes the cross-sectional regression model takes the following form

$$E(R_u) = \lambda_0 + \lambda_i^{MKT} \beta_i^{MKT} + \lambda_i^{SMB} \beta_i^{SMB} + \lambda_i^{HML} \beta_i^{HML} + \lambda_i^{UMD} \beta_i^{UMD} + \lambda_i^{APD} \beta_i^{APD}$$

(10)

and the model for the moments

$$E[R_u (1 + b_{MKT} MKT_t + b_{SMB} SMB_t + b_{HML} HML_t + b_{UMD} UMD_t + b_{APD} APD_t)] = 0$$

(11)

---

12 The difference in market betas alone cannot explain the return differential between small and large firms, or the differential between stocks with high and low book-to-market values. To explain stock returns, Fama and French (1993) propose a three-factor model that includes two additional factors SMB, HML, related to firm size and book-to-market, respectively.

13 Logically, small firms are expected to be more sensitive to many risk factors because they have limited ability to buffer adverse financial troubles or economic shocks. On the other hand, the HML factor suggests value (high BM) stocks have higher risk exposure than growth (low BM) stocks, as the stock market often anticipates bankruptcy or other financial difficulties, driving down (up) the firm’s public value (BM ratio).
where UMD is the momentum factor, computed as the difference in return between the better and worse performing stocks, or “up minus down” (UMD).

For model 1 through 3, we also analyze its benchmark version without the aggregate default risk measure $APD$ included for comparison purposes. We also consider macroeconomic factors from the Chen, Roll and Ross’s (1986) model. The macro factors\textsuperscript{14} are not significant variables in our model after we including the aggregate default risk measure $APD$, so we do not report the result here.

3.2.2 Test assets

We choose portfolios as the test assets whose returns the asset pricing models will be used to explain. Eighteen portfolios are formed at the end of each month from the intersection of three independent sorting based on size, book-to-market and probability of Default. Specifically, all equities in our sample are sorted into two portfolios according to size. They are also sorted into three portfolios according to book-to-market ratio, and three portfolios based on probability of Default. Finally, the intersection of the three sorts yields altogether $2*3*3=18$ portfolios. We do the sorting to obtain maximum dispersion in the characteristic of portfolio, namely, size, book to market ratio and predicted hybrid PD. Summary statistics of the portfolios are reported in Table 2.

We perform the asset pricing tests on portfolios instead of individual stocks for the following reasons. First, the usage of portfolios as test assets in CRS regression minimized the errors-in-variable problem arising from estimated regressors $\hat{\beta}$. The measurement error in beta will be smaller if we expect that some of the individual noise

\textsuperscript{14} Namely, term premium, i.e., the return spread between long term government bond and Treasury bills; default premium, i.e., the return spread between long term corporate and long-term government bonds, the growth rate in monthly industrial production in the United States, and the change of inflation rate.
Table 2: Summary statistics on the 18 Size, BM and PD (probability of default) sorted portfolios
The eighteen portfolios are formed from the intersection of three independent sortings of all stocks into two size, three BM (book-to-market ratio) and three default risk portfolios. Default risk is measured by PD (the probability of default) derived from one-year-ahead model in chapter 2. Size is measured by taking natural logarithm of the market value of equity. Mean excess return is the equally-weighted average returns of portfolios in excess of risk-free rate, reported in percentage terms per month.

<table>
<thead>
<tr>
<th>group</th>
<th>Size</th>
<th>BM</th>
<th>PD</th>
<th>Mean Excess Return</th>
<th>Size</th>
<th>BM</th>
<th>PD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Small</td>
<td>High</td>
<td>High</td>
<td>0.5577%</td>
<td>4.8304</td>
<td>2.0995</td>
<td>0.600%</td>
</tr>
<tr>
<td>2</td>
<td>Small</td>
<td>High</td>
<td>Medium</td>
<td>1.0755%</td>
<td>5.4422</td>
<td>1.1278</td>
<td>0.007%</td>
</tr>
<tr>
<td>3</td>
<td>Small</td>
<td>High</td>
<td>Low</td>
<td>1.1194%</td>
<td>5.7090</td>
<td>1.0210</td>
<td>0.002%</td>
</tr>
<tr>
<td>4</td>
<td>Small</td>
<td>Medium</td>
<td>High</td>
<td>0.4201%</td>
<td>5.2362</td>
<td>0.5954</td>
<td>0.195%</td>
</tr>
<tr>
<td>5</td>
<td>Small</td>
<td>Medium</td>
<td>Medium</td>
<td>1.0124%</td>
<td>5.6783</td>
<td>0.5558</td>
<td>0.006%</td>
</tr>
<tr>
<td>6</td>
<td>Small</td>
<td>Medium</td>
<td>Low</td>
<td>1.5876%</td>
<td>5.8323</td>
<td>0.5120</td>
<td>0.002%</td>
</tr>
<tr>
<td>7</td>
<td>Small</td>
<td>Low</td>
<td>High</td>
<td>-0.0388%</td>
<td>5.2693</td>
<td>0.2235</td>
<td>0.359%</td>
</tr>
<tr>
<td>8</td>
<td>Small</td>
<td>Low</td>
<td>Medium</td>
<td>1.3866%</td>
<td>5.7201</td>
<td>0.2298</td>
<td>0.006%</td>
</tr>
<tr>
<td>9</td>
<td>Small</td>
<td>Low</td>
<td>Low</td>
<td>2.1786%</td>
<td>5.8755</td>
<td>0.2044</td>
<td>0.002%</td>
</tr>
<tr>
<td>10</td>
<td>Big</td>
<td>High</td>
<td>High</td>
<td>-0.0476%</td>
<td>7.4294</td>
<td>1.4171</td>
<td>0.150%</td>
</tr>
<tr>
<td>11</td>
<td>Big</td>
<td>High</td>
<td>Medium</td>
<td>0.5251%</td>
<td>7.7318</td>
<td>1.0311</td>
<td>0.006%</td>
</tr>
<tr>
<td>12</td>
<td>Big</td>
<td>High</td>
<td>Low</td>
<td>0.3887%</td>
<td>9.0372</td>
<td>0.9913</td>
<td>0.002%</td>
</tr>
<tr>
<td>13</td>
<td>Big</td>
<td>Medium</td>
<td>High</td>
<td>-0.5284%</td>
<td>7.3908</td>
<td>0.5894</td>
<td>0.063%</td>
</tr>
<tr>
<td>14</td>
<td>Big</td>
<td>Medium</td>
<td>Medium</td>
<td>0.6575%</td>
<td>7.4412</td>
<td>0.5348</td>
<td>0.005%</td>
</tr>
<tr>
<td>15</td>
<td>Big</td>
<td>Medium</td>
<td>Low</td>
<td>0.3643%</td>
<td>9.1283</td>
<td>0.4899</td>
<td>0.002%</td>
</tr>
<tr>
<td>16</td>
<td>Big</td>
<td>Low</td>
<td>High</td>
<td>-0.6658%</td>
<td>7.4016</td>
<td>0.1824</td>
<td>0.094%</td>
</tr>
<tr>
<td>17</td>
<td>Big</td>
<td>Low</td>
<td>Medium</td>
<td>-0.0797%</td>
<td>7.5224</td>
<td>0.2235</td>
<td>0.005%</td>
</tr>
<tr>
<td>18</td>
<td>Big</td>
<td>Low</td>
<td>Low</td>
<td>0.1522%</td>
<td>9.6752</td>
<td>0.2004</td>
<td>0.001%</td>
</tr>
</tbody>
</table>

to average out in the regression. Also, sample error variance for portfolio will be \( I/n \) as large as that for individual assets. Secondly, if we use GMM to circumvent the error-in-variable problem, then we have to limit our sample to a small number of portfolios. As Ferson and Foerster (1994) point out, GMM estimation has rather poor finite sample properties. The problem are especially pronounced when the sample is small and the number of assets, \( n \) is large.

In determining the number of assets used, we follow a useful rule of thumb that a saturation ratio (the number of observation per parameter) below 10 (or so) is likely to give poor performance of the test. Say for augmented four-factor model in equation (11),
we have $18\times120 = 2160$ observations in our sample, 5 parameters in factors, and $18\times(18+1)/2 = 171$ unique parameters in $S$ (the covariance matrix).\(^{15}\) So the saturation ratio is $2160/(5+171) = 12.27$, which is moderate.

3.2.3. Data

We use monthly returns on eighteen size, book-to-market and default probability ranked portfolios in excess of the one-month T-bill rate to compare various asset pricing models. The returns on the portfolios are constructed in section 3.2.2. The Fama-French factors HML, SMB, UMD and the market factor MKT (excess return on the market portfolio) are obtained from Kenneth French’s website.\(^{16}\) The one-month T-bill rate used in our asset pricing tests is also from the same website. The aggregate default risk measure $APD$ is defined as a simple average of default probabilities $PD_i$ of all firms. The default probabilities $PD_i$ are obtained from our duration model estimation in chapter 2. Our sample data are from January 1999 to December 2008 (120 monthly observations).\(^{17}\)

3.2.4 Results

Table 3 reports the estimates for competing asset pricing models presented in section 3.2.1., Eq. (6) ~Eq. (11). The tests are performed on the excess returns of 18 portfolios constructed in section 3.2.2. The $\lambda$ coefficients are risk premium implied for factors in cross-sectional regression (CSR) specification of Eq. (1). The $t$-statistics for $\lambda$ are

---

\(^{15}\) GMM relax the strict assumption of i.i.d. errors, hence estimation of the covariance matrix of asset returns is unavoidable

\(^{16}\) We acknowledge Ken French for making the data available. Details about the data can be obtained from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

\(^{17}\) We also perform an additional analysis using quarterly data. The central message of the paper is not affected by the choice of the return horizon.
Table 3: Evaluation of Competing Asset Pricing Models

This table gives the estimates for various asset pricing models. Each asset pricing model is estimated in two representations: cross-sectional regression (CSR) specification estimated by using the Fama-MacBeth procedure, and the stochastic discount factor (SDF) specification estimated by Generalized Method of Moments (GMM) with Hansen's (1982) optimal weighting matrix. The tests are performed on the excess returns of 18 portfolios of Table 2. The sample period is from January 1999 to December 2008. MKT is excess return on the market portfolio. SMB (Small Minus Big) is a measure of size risk. HML (High Minus Big) captures the risk related to firm book-to-equity. UMD is the momentum factor introduced and validated by Carhart (1997). The $\lambda$ coefficients are risk premium implied for factors in CRS, the $b$ coefficients are coefficients for factors in pricing kernel. $J$-statistics is used to test the null hypothesis: overidentifying restrictions of the model are satisfied. The Wald test is a joint significance test of the coefficients in pricing kernel. $HJ$-dist denotes the Hansen-Jagannathan (1997) distance, a scalar measure of the magnitude of the asset pricing model misspecification. $t$-value is in parenthesis. * denotes significant at 10%, ** denotes significant at 5% level, *** denotes significant at 1% level. APD is aggregate default probability.

Panel A: The CAPM

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$\lambda_{0}$</th>
<th>$\lambda_{MKT}$</th>
<th>$\lambda_{SMB}$</th>
<th>$\lambda_{BM}$</th>
<th>$\lambda_{UMD}$</th>
<th>$\lambda_{APD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>0.0104</td>
<td>-0.0038</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$-value</td>
<td>(1.6201)</td>
<td>(-0.6174)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$b_{MKT}$</th>
<th>$b_{SMB}$</th>
<th>$b_{HML}$</th>
<th>$b_{UMD}$</th>
<th>$b_{APD}$</th>
<th>J-statistics</th>
<th>Wald Test</th>
<th>$HJ$-dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>-1.2729</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16.7631</td>
<td>0.45</td>
<td>0.5632</td>
</tr>
<tr>
<td>$t$-value</td>
<td>(-0.67)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-0.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p$-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.470525</td>
<td>0.5001</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Panel B: The CAPM augmented by APD

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$\lambda_{0}$</th>
<th>$\lambda_{MKT}$</th>
<th>$\lambda_{SMB}$</th>
<th>$\lambda_{BM}$</th>
<th>$\lambda_{UMD}$</th>
<th>$\lambda_{APD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>0.0113*</td>
<td>0.0039</td>
<td>0.1035*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$-value</td>
<td>(1.7043)</td>
<td>(0.5654)</td>
<td>(1.9359)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$b_{MKT}$</th>
<th>$b_{SMB}$</th>
<th>$b_{HML}$</th>
<th>$b_{UMD}$</th>
<th>$b_{APD}$</th>
<th>J-statistics</th>
<th>Wald Test</th>
<th>$HJ$-dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>-3.715***</td>
<td>-12.193***</td>
<td>10.2102</td>
<td>130.91</td>
<td>0.1631</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$-value</td>
<td>(-5.22)</td>
<td>(-11.28)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.855423</td>
<td>&lt;0.0001</td>
<td>0.4843</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(to be cont'd)
Table 3(cont'd): Evaluation of Competing Asset Pricing Models

### Panel C: The Fama-French three-factor model

<table>
<thead>
<tr>
<th>Coefficient: $\lambda_0$</th>
<th>$\lambda_{MKT}$</th>
<th>$\lambda_{SMB}$</th>
<th>$\lambda_{BM}$</th>
<th>$\lambda_{UMD}$</th>
<th>$\lambda_{APD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>0.0036</td>
<td>-0.0083</td>
<td>0.0239***</td>
<td>0.0123</td>
<td></td>
</tr>
<tr>
<td>$t$-value</td>
<td>(0.6985)</td>
<td>(-1.0263)</td>
<td>(2.9129)</td>
<td>(1.6417)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>coefficient:</th>
<th>$b_{MKT}$</th>
<th>$b_{SMB}$</th>
<th>$b_{HML}$</th>
<th>$b_{UMD}$</th>
<th>$b_{APD}$</th>
<th>$J$-statistic</th>
<th>Wald Test</th>
<th>HJ-dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>2.9851</td>
<td>-18.601***</td>
<td>-13.171***</td>
<td></td>
<td></td>
<td>9.3801</td>
<td>83.15</td>
<td>0.4036</td>
</tr>
<tr>
<td>$t$-value</td>
<td>(1.35)</td>
<td>(-7.87)</td>
<td>(-4.37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8568</td>
<td>&lt;0.0001</td>
<td>0.1915</td>
</tr>
</tbody>
</table>

### Panel D: The Fama-French three-factor model augmented by APD

<table>
<thead>
<tr>
<th>Coefficient: $\lambda_0$</th>
<th>$\lambda_{MKT}$</th>
<th>$\lambda_{SMB}$</th>
<th>$\lambda_{BM}$</th>
<th>$\lambda_{UMD}$</th>
<th>$\lambda_{APD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>0.0049</td>
<td>0.0001</td>
<td>0.0247***</td>
<td>0.0123</td>
<td>0.1221**</td>
</tr>
<tr>
<td>$t$-value</td>
<td>(0.9309)</td>
<td>(0.0158)</td>
<td>(2.943)</td>
<td>(1.6348)</td>
<td>(2.2302)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>coefficient:</th>
<th>$b_{MKT}$</th>
<th>$b_{SMB}$</th>
<th>$b_{HML}$</th>
<th>$b_{UMD}$</th>
<th>$b_{APD}$</th>
<th>$J$-statistic</th>
<th>Wald Test</th>
<th>HJ-dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>-2.5159**</td>
<td>-4.4421**</td>
<td>-3.7696**</td>
<td></td>
<td></td>
<td>10.2297***</td>
<td>7.1442</td>
<td>144.95</td>
</tr>
<tr>
<td>$t$-value</td>
<td>(-2.34)</td>
<td>(-2.51)</td>
<td>(-1.99)</td>
<td></td>
<td></td>
<td>(-7.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9290</td>
<td>&lt;0.0001</td>
<td>0.9071</td>
</tr>
</tbody>
</table>

### Panel E: The Carhart (1997) four-factor model

<table>
<thead>
<tr>
<th>Coefficient: $\lambda_0$</th>
<th>$\lambda_{MKT}$</th>
<th>$\lambda_{SMB}$</th>
<th>$\lambda_{BM}$</th>
<th>$\lambda_{UMD}$</th>
<th>$\lambda_{APD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>-0.0119**</td>
<td>0.0227**</td>
<td>0.0182**</td>
<td>0.0286***</td>
<td>0.0240**</td>
</tr>
<tr>
<td>$t$-value</td>
<td>(-1.9959)</td>
<td>(2.2292)</td>
<td>(2.4061)</td>
<td>(3.1374)</td>
<td>(2.5799)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>coefficient:</th>
<th>$b_{MKT}$</th>
<th>$b_{SMB}$</th>
<th>$b_{HML}$</th>
<th>$b_{UMD}$</th>
<th>$b_{APD}$</th>
<th>$J$-statistic</th>
<th>Wald Test</th>
<th>HJ-dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>-0.8434</td>
<td>-14.2336***</td>
<td>-13.4801***</td>
<td>-2.5927</td>
<td></td>
<td>9.6212</td>
<td>51.32</td>
<td>0.3972</td>
</tr>
<tr>
<td>$t$-value</td>
<td>(-0.30)</td>
<td>(-4.32)</td>
<td>(-4.03)</td>
<td>(-1.31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.7893</td>
<td>&lt;0.0001</td>
<td>0.194</td>
</tr>
</tbody>
</table>

### Panel F: The Carhart (1997) four-factor model augmented by APD

<table>
<thead>
<tr>
<th>Coefficient: $\lambda_0$</th>
<th>$\lambda_{MKT}$</th>
<th>$\lambda_{SMB}$</th>
<th>$\lambda_{BM}$</th>
<th>$\lambda_{UMD}$</th>
<th>$\lambda_{APD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>-0.0073</td>
<td>0.0222**</td>
<td>0.0202**</td>
<td>0.0247***</td>
<td>0.0185*</td>
</tr>
<tr>
<td>$t$-value</td>
<td>(-1.2035)</td>
<td>(2.1948)</td>
<td>(2.4819)</td>
<td>(2.9949)</td>
<td>(1.9609)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>coefficient:</th>
<th>$b_{MKT}$</th>
<th>$b_{SMB}$</th>
<th>$b_{HML}$</th>
<th>$b_{UMD}$</th>
<th>$b_{APD}$</th>
<th>$J$-statistic</th>
<th>Wald Test</th>
<th>HJ-dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$-value</td>
<td>(-2.59)</td>
<td>(-2.13)</td>
<td>(-2.15)</td>
<td>(-1.65)</td>
<td>(-6.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9557</td>
<td>&lt;0.0001</td>
<td>0.9195</td>
</tr>
</tbody>
</table>
obtained using the Fama-Macbeth standard errors correcting for estimation error in $\beta$.

The $b$ coefficients are coefficients for factors in pricing kernel or stochastic discount factor (SDF) specification of Eq. (2).

As shown in panel B, D and F, the $t$-value for $b_{APD}$ (the parameter of $APD$ factor in pricing kernel) is large, and $b_{APD}$ is significant at 1% level in all versions of asset pricing models augmented by a default risk factor. We conclude that default risk is predominantly systematic, and should be considered as a factor in pricing stock return. Default risk premium $\lambda_{APD}$ is about 0.1, and statistically significant at 1%~10% level. It indicates that investors will get compensated for holding stock of risky firms.

Also, both J-statistics and HJ-distance measure favor an augmented model over its basic version, i.e., model in panel B (D, F) preferred over model in Panel A (C, E). For example, J-statistics takes the value of 9.6212 with a p-value of 0.7893 in Panel E (four-factor model), and of 5.7186 with a p-value of 0.9558 in Panel F (augmented four-factor model). $0.9558 > 0.7893$, implying that the null hypothesis of over-identifying restrictions in GMM is more likely satisfied in augmented four-factor model. Similarly, the value of HJ-distance is smaller in Panel F than in Panel E ($0.1125 < 0.1174$); p-value for null hypothesis of no pricing error is larger in Panel F than in Panel E ($0.9195 < 0.9071$), favoring augmented four-factor model over its basic version.

A Wald test of joint significance of coefficients cannot reject all but one model (CAPM). Indeed, the t-value for $b_{MKT}$ is only $-0.67$, very small. The market risk premium $\lambda_{MKT}$ in CAPM is not significantly different from zero either, even after correcting for the estimation error in $\beta$. These evidences suggest that conventional specification of the CAPM is inconsistent with the data.
Next, we compare the performance of different augmented asset pricing models in Panel B, D and F by using HJ-distance measures. We find that augmented Carhart (1997) four-factor model outperforms augmented CAPM or augmented Fama-French three-factor model in terms of HJ-distance. HJ-distance (its p-value) is smallest (largest) in augmented four-factor model, with a value of 0.1125 (0.9195).

Could the difference in HJ-distance be purely by chance? We also apply Kan and Robotti’s (2009) test to check whether our best asset pricing model (augmented four-factor model) exhibit significantly different HJ-distance from other alternatives (augmented CAPM or augmented Fama-French three-factor model). For nested models case in our paper, the test statistics is chi-squared distributed under the null hypothesis of equal HJ-distance in two models. For details, see description in section 3.1. The p-value for the tests of augmented four-factor model and augmented three-factor model (augmented four-factor model and augmented CAPM) is 0.0198 (<0.0001). Therefore, we find significant evidence that support the augmented four-factor model as a superior model in terms of HJ-distance.

Lastly, we note that \( \lambda_0 \), the mispricing term in the augmented four-factor model of Panel F is insignificant, both economically and statistically. Overall, it suggests our augmented four-factor model is a good fit for the stock return data.

4. Conclusions

This chapter explores the effect of default risk on stock returns using default probability generated by our best model in chapter two. With this better measure for distress risk, our study captures a pattern different from that of current empirical
literature. We note that, by and large, the return-risk in cross-sectional stock displays an inverted U-shaped relation. To uncover this credit risk puzzle (i.e., the negative risk-return relation for the distressed portfolios), we further investigate the portfolio returns after risk-adjusting by several empirical asset pricing models. We find that this anomaly could be explained by momentum effect using Carhart’s (1997) four-factor asset pricing model. We also conduct tests for competing asset pricing models. We estimate stochastic discount factor (SDF) representation of linear factor pricing models by using generalized method of moments (GMM) approach, and find evidence that support default risk as a factor worth considering in pricing stock return. A four-factor asset pricing model augmented with default risk factor is a superior model in pricing stock return in terms of Hansen and Jagannathan (1997) (HJ) distance.
References:


