Expert Systems with Applications 40 (2013) 4897–4905

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A statistical modeling methodology for the analysis of term structure of credit risk and its dependency

Jiashen You [a], Tomohiro Ando [b,*]

[a] 8125 Math Sciences, Box 951554, Los Angeles, CA 90034, USA
[b] 4-1-1 Hiyoshi, Kohoku-ku, Yokohama-shi, Kanagawa 223-8526, Japan

ARTICLE INFO

Keywords:
Credit risk dependency
Corporate bond pricing
Default probability
Loss given default

ABSTRACT

This paper presents a statistical modeling methodology for simultaneous estimation of the term structure for the risk-free interest rate, hazard rate, loss given default as well as credit risk dependency structure between bond-issuing industries. A model like this provides a realistic view for the market anticipation of credit risk for corporate bonds and the flexibility in capturing credit risk dependency between industries. Our statistical modeling procedure is carried out without specifying the model likelihood explicitly, and thus robust to the model mis-specification. An empirical analysis is conducted using the financial information on the Japanese bond market data. Numerical results confirm the practicality of the proposed methodology.

1. Introduction

Systematic credit default risk for corporate bonds has conventionally been studied using term structure of hazard rates. Much like the term structure of interest rates, the name suggests the dependency on time to maturity of and hints at an implicit pricing model for the underlying bonds. Many mathematical finance models for estimating term structure have placed emphases on market equilibrium or the no-arbitrage assumption as well as been in pursuit of Markovian properties through stochastic processes. While economic theories are crucial to the interpretability of some model parameters, there has been an increasing demand for practical and adaptive modeling procedures as the 2008 financial crisis has shown that market does not always behave as theoretical models predicts.

In this paper, we basically employ the statistical modeling framework of Ando (forthcoming) for the analysis of the term structures of risk-free interest rate and hazard, the loss given default (LGD). As a new contribution, we try to model the credit risk dependency structure between the industries. The LGD of corporate bonds, in the form of a fraction of unity maturity value, is modeled by a discrete recovery rate upon a default event. We recognize the importance of the validity of key economic theories while seeking applicability and ease of expansion for our approach. When estimating credit risk from corporate bonds, following Ando (forthcoming), we incorporate diverse information such as credit ratings, term structure for interest rate as well as the financial strength of the companies. Credit risk dependency structure between industries is estimated from the aggregated residuals while other parameters are divided into different credit rating classes allowing us to capture the mean structure of hazard functions for each rating class.

Modeling the term structure of interest rates from bonds with various time to maturity have long been of interest to researchers in financial econometrics and related disciplines. On the mathematical finance front, Vasicek (1977) and Cox et al. (1985) pioneered in using a Markov process with averaging continuous interest rate paths. As in many similar applications, the yield curve cannot be determined exogenously. While these models are capable of ensuring desirable properties such as mean-reversion, the theoretical nature also exposes them to less explainable yet possible scenarios such as negative interest for the Vasicek model.

McCulloch (1971, 1975) was the first to propose a statistical estimation technique, using spline functions to model the discount and yield curves from observations on prices of bonds with varying maturities and coupon rates. The use of least squares method for greatly speeds up the estimation process of spline coefficients, provided that the pricing formula for non-defaultable bonds is additive. Some notable extensions and successive improvements of such approximation models can be found in Vasicek and Fong (1982), Shea (1985), Chambers et al. (1984) and Krivobokova et al. (2006).

There is an extensive amount of finance literature on modeling risk structure on defaultable bonds. The Black and Scholes model (1973) for option pricing and Merton's immediately-followed

* Corresponding author. Tel.: +81 45 564 2039.
E-mail addresses: jiashen@stat.ucla.edu (J. You), andoh@kbs.keio.ac.jp (T. Ando).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.02.017

extensions Merton (1973, 1974) have been treated by many as sacrosanct in this field. Most of the similar stochastic models with equilibrium-theory assumption have reaped the benefit of Itô's Lemma (1951). While mathematical models may be solved analytically and in some cases, even closed-form solutions are obtained (Amin and Jarrow, 1991; Heston, 1993), oversimplifying the dynamics could hinder expressing complex market reality.

When a default event occurs, bond holders typically receive a fraction of the par value and write the remaining amount at a loss. By introducing hazard rate into the term structure for interest rate, Duffie and Singleton (1999) and Jarrow and Turnbull (1995) laid the foundation for the reduced-form model in a continuous-time Markovian setting. However, the LGD in this model is usually a given parameter.

From the situation-dependent industry perspective, making predictions for distance to defaults or LGD can be as important as estimating the numbers and probabilities of defaults. Both statistical (Altman et al., 2005; Covitz and Han, 2004) and mathematical (Hamilton et al., 2001; Jarrow et al., 2006) models have been employed. It is also worth noting that in recent artificial intelligence literature, various attempts have been made to model interest rate and default risk using neural network models (Jeong et al., 2012; Khashman, 2010), fuzzy-knowledge based models (Ju et al., 1997; Streit and Borenstein, 2012) or support vector machine (Li et al., 2012; Perko et al., 2011; Ribeiro et al., 2012; Yu et al., 2010, 2011). These new approaches offer insights from unconventional angles, but their estimates may suffer loss on interpretability.

This paper employs a recently invented model (Ando, forthcoming; Kariya, 2012) that can significantly reduce the model misspecification risk of underlying stochastic processes for corporate bond pricing. B-splines are used to estimate the mean hazard rate structure separated by each rating class for ease of computation as well as its smoothing effect. Covariates such as company financial strength are used to boost the practicality of our model. LGD can be approximated by using a logistic function of current model estimates and its value is well-defined.

In practice, bonds within the same corporate; companies within the same sector; industries within the same economy are all correlated. Therefore, it is imperative to model such dependency. Since these implied dependency is model-specific and is not observed directly, many studies have attempted to either analytically (Zhou, 1997) or empirically quantify (Cowan and Cowan, 2004; You and Ando, 2012; Zhou, 2001) default correlation. For the model discussed in this paper, industry effect is controlled as a covariate. Dimension-reduction technique is introduced to efficiently model the correlation structure in the error terms. Empirical results reveal that within-industry dependency is more influential than between-industry dependency.

The remaining of this paper is organized as follows. Section 2 overviews the basic model (Ando, forthcoming; Kariya, 2012) for the term structure of the forward instantaneous interest rate and hazard function. Model parameterization is also specified. Section 3 discusses several key assumptions needed for the parameter estimation procedure. Results from empirical analyses are presented in Section 4. Section 5 concludes this paper with some final remarks.

2. Model specification

In this section, we overview the results of Ando (forthcoming) and Kariya (2012) (see also references therein) for the pricing of government/corporate bonds. The main objective of this study is not on improving a well-studied mathematical pricing formula, but rather to integrate scattered information together with economically sound statistical modeling platform. We strive for the ease of computation, generalization and interpretability of our model parameters.

2.1. Term structure for risk-free instantaneous forward interest rate

We review the theoretical pricing formula of a non-defaultable bond. We assume for a risk-free government bond with fixed coupon amount C and a par value of M redeemable at maturity. A total of L successive payments for the bond are made at time $\boldsymbol{t} = (t_1,\ldots,t_L)'$, measured in years from the current time $t_0 = 0$. Let $r(t)$ be the term structure of the risk-free instantaneous forward rate, the present value (PV) of a government bond with a maturity date t_L years from now can be computed from the sum of its discounted cash flow:

$$\mathrm{PV}(r(\cdot),\boldsymbol{t}) = M\cdot\exp\left[-\int_{t_0}^{t_L} r(u)du\right] + \sum_{\gamma=1}^{L} C\cdot\exp\left[-\int_{t_0}^{t_\gamma} r(u)du\right]. \tag{1}$$

Denoting by $D_g(t)$ the attribute-dependent stochastic discount function, i.e.

$$D_g(t) = \exp\left\{-\int_{t_0}^{t} r(u)du\right\},$$

the Eq. (1) becomes

$$\mathrm{PV}(D_g(\cdot),\boldsymbol{t}) = M\cdot D_g(t_L) + \sum_{\gamma=1}^{L} C\cdot D_g(t_\gamma). \tag{2}$$

Ando (forthcoming) and Kariya (2012) assumed that $D_g(t)$ can be decomposed into the group mean function $\bar{D}_g(t)$ and the stochastic deviation term $\Delta_g(t)$ simply by:

$$D_g(t) = \bar{D}_g(t) + \Delta_g(t). \tag{3}$$

It is common knowledge that theoretical pricing models such as the Eq. (2) may not give accurate predictions for the prices of bonds. Therefore, error terms must be included in the corresponding statistical model to account for those variations.

$$\mathrm{PV}(D_g(\cdot),\boldsymbol{t}) = M\cdot \bar{D}_g(t_L) + \sum_{\gamma=1}^{L} C\cdot \bar{D}_g(t_\gamma) + \varepsilon, \tag{4}$$

where $\varepsilon = M\cdot\Delta_g(t_L) + \sum_{\gamma=1}^{L} C\cdot\Delta_g(t_\gamma)$ is regarded as a noise term.

2.2. Term structure for hazard

Let T be a random variable that measures the time to default. The survival function is defined as $D_h(t) = \Pr(T \geqslant t)$. A hazard function $h(t)$ that specifies the instantaneous rate of default at t conditioned on the survival up to time t. This $h(t)$, hereafter called as the hazard term structure (an instantaneous default intensity), can be considered as the standardization of the conditional probability of a bond which has not defaulted up till time t defaulting within the next time period $(t, t + \Delta t)$, where Δt is a small dimensionless measure of time. The survival probability with the above hazard term structure of a bond up to time t is therefore

$$D_h(t) = \exp\left\{-\int_{t_0}^{t} h(u)du\right\}. \tag{5}$$

Assuming that the hazard term structure as in Eq. (5), Ando (forthcoming) evaluated the present value of a corporate bond as