Macroeconomic variable selection for creditor recovery rates

Abdolreza Nazemi [a,*], Frank J. Fabozzi [b]

[a] *School of Economics and Business Engineering, Karlsruhe Institute of Technology, Neuer Zirkel 3, Geb. 20.53, Karlsruhe, 76131, Germany*
[b] *EDHEC Business School, Nice, France*

ABSTRACT

We study the relationship between U.S. corporate bond recovery rates and macroeconomic variables used in the credit risk literature. The least absolute shrinkage and selection operator (LASSO) is used in selecting macroeconomic variables. The LASSO-selected macroeconomic variables are considered to be explanatory variables in ordinary least squares regressions, bootstrap aggregating (bagging), regression trees, boosting, LASSO, ridge regression and support vector regression techniques. We compare the out-of-sample predictive power of two types of models (LASSO-selected models with models that add principal components derived from 179 macroeconomic variables as explanatory variables). We find the recovery models with LASSO-selected macroeconomic variables outperform suggested models in the literature.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

The recent global financial crisis highlighted the importance of credit risk and regulatory requirements for addressing that risk. Financial institutions are allowed to use their internal risk parameters for calculation of capital requirements, conforming with Basel Accords II and III. Three key risk parameters in Basel II and III are used to calculate regulatory capital requirements: recovery rate or loss-given-default (LGD), probability of default (PD), and exposure at default (EAD). The recovery rate distribution for defaulted corporate bonds and corporate loans have been observed to be bimodal or multimodal. This could be one reason that prior studies of recovery rates using standard parametric models have not reported high predictive power for recovery rates. Recovery rates depend on the macroeconomic conditions, bond features and borrower characteristics. Consequently, macroeconomic conditions at the time of default are potential sources for recovery determination.

In recent years, the quantity and quality of available financial and macroeconomic data have increased due to improvements in computational and storage power. The new challenge, both theoretically and computationally, is to deal with the continually-increasing large datasets available for empirical applications. This paper investigates the importance of 179 macroeconomic variables in recovery rate modeling by applying LASSO to determine those that are most important in explaining corporate bond recovery rates.

We utilized LASSO to select macroeconomic variables because this econometric tools is more robust than those used in other standard econometric methods (such as forward or backward stepwise regression), for prediction and variable selection. Many studies assume that recovery rates depend linearly on available explanatory variables. Support vector regression techniques imply nonlinear dependency in the recovery rate for modeling purposes. To the best of our knowledge, our paper is the first to apply LASSO, bagging, boosting and ridge regression for recovery rate prediction.

In addition to being the first study to apply the econometric techniques just described to variable selection and recovery rate prediction, we make three further contributions. First, we provide the results of adding the 179 macroeconomic variables to recovery rate models. Second, we construct parametric and non-parametric models using only selected macroeconomic variables derived from shrinkage methods and compare their out-of-sample predictive power with models that include principal components of macroeconomic variables and models with macroeconomic (see, for example, Jankowitsch et al., 2014). Finally, to improve the predictive accuracy of the models, we test them with macroeconomic

* Corresponding author.
   E-mail addresses: abdolreza.nazemi@kit.edu (A. Nazemi),
frank.fabozzi@edhec.edu (F.J. Fabozzi).

variables that have been transformed and by doing so we ensure stationary variables.

The remainder of the paper is organized as follows. In the next section, we briefly review the related literature, emphasizing the macroeconomic aspects of recovery rates prediction. In Section 3 we describe the estimation and selection methods. Section 4 presents our dataset and summary statistics. Macroeconomic variable selection for recovery rate modeling using LASSO is described in Section 5. In Section 6 we investigate the ability of the parametric and non-parametric models to fit recovery rates of corporate bonds, including selected and principal components from macroeconomic variables. Section 7 provides our conclusions.

2. Related literature

Chen (2010) reports that recovery rates during the recessions of 1982, 1990, 2001 and 2008 were less than the average value of recovery rates in other economic conditions. Bruche and Gonzalez-Aguado (2010) propose the systematic time-variation model in recovery rate distributions and default rates, reporting that both the LGD of defaulted bonds and the number of defaulting firms increased during recessions. Cantor and Varma (2004) argue that macroeconomic variables play a significant role in estimating recovery rates. Comparing different parametric and non-parametric models with three macroeconomic variables for estimating recovery rates, Qi and Zhao (2011) find that non-parametric methods outperform parametric methods. Moreover, they report that recovery rates are lower when the 3-month U.S. Treasury bill rate and aggregate default rates are higher. Recovery rates are greater when market returns and industry distance-to-default values are higher.

In order to analyze corporate bond recovery rates, Jankowitsch et al. (2014) include as explanatory variables bond characteristics (e.g., liquidity measures, firm fundamentals and bond covenants), four macroeconomic conditions, industry dummy variables, default event types, and seniority classes. They find that macroeconomic variables – especially market-wide, industry-specific default rates and interest rates – are related to recovery rates. Altman et al. (2005) report that recovery rates and aggregate default rates are negatively correlated. Acharya et al. (2007) conclude that industries that are in distress have lower recovery rates.

Tobback et al. (2014) investigate the effects of 11 macroeconomic variables on corporate loan recovery rates. The predictive power of their models were improved significantly by adding macroeconomic variables. They mention that the macroeconomic variables had unexpected effects and varied influences on each model and dataset. Using five macroeconomic variables to model the influence of economic conditions on bond recovery rates, Chava et al. (2011) report that the logarithm of the amount of all defaulted debt and the 3-month Treasury bill rate were statistically significant in most recovery rate models. The "term spread" and the "credit spread" were found not to be statistically significant. Zhang (2009) stated that loans with stricter covenants have a higher recovery rate. He mentioned that by decreasing macroeconomic variables by one standard deviation, the expected LGD decreased by about 5% of its base value. Similarly, this study looks to find a relationship between recovery rates and the four macroeconomic variables affecting loans.

Mora (2015) incorporated several macroeconomic variables in her model: GDP growth, stock market return, housing price growth, housing price growth (state-level), 3-month Treasury bill rate, commercial paper spread (3-month commercial paper rate for high grade nonfinancial borrowers minus the 3-month Treasury bill rate) and corporate bond Baa-Aaa yield spread in recovery rate estimation for U.S. corporate bonds. She reports that macroeconomic variables are statistically significant in modeling recovery

rates but they do not have the same effect on every industry. Industries more reliant on external finance, and with sales growth closely correlated with GDP growth, tend to have lower recovery rates during a financial crisis.

Using four U.S. macroeconomic variables – GDP, unemployment rate, S&P 500 return, and 3-month Treasury Bill rate – to model the economic cycle's effect over the period 1985–2012, Yao et al. (2015) find that support vector regression techniques outperform other methods for predicting recovery rates of U.S. corporate bonds.

To model recovery rates, Nazemi et al. (2016) add principal components derived from 104 macroeconomic variables (from a broad range of categories, such as stock market conditions, credit market conditions, international competitiveness, business cycle conditions and micro-level conditions). They report that fuzzy decision fusion techniques significantly increased the predictive power of recovery rate modeling. Although they find that the predictive accuracy of fuzzy decision fusion techniques worsened with Box-Cox transformations of macroeconomic variables, the transformations improved the predictive accuracy of the linear regression model they use. Moreover, they show that a improvement in performance measures occurred by adding the principal components calculated from the 104 macroeconomic variables to all techniques. Because of the large number of macroeconomic variables they include in their model, it was not possible to interpret their results.

This study is close to Qi and Zhao (2011). The authors reported that regression trees and neural networks outperform parametric methods for predicting recovery rates of corporate bonds. Comparing different parametric and non-parametric models for estimating the LGD of defaulted leasing contracts, Hartmann-Wendels et al. (2014) find that model trees outperform all other methods when there is a large sample. Our paper has four main contributions compared to Qi and Zhao (2011) and Hartmann-Wendels et al. (2014) published in this journal. First, our paper applies LASSO, bagging, boosting, ridge regression and support vector regression algorithms for recovery rate prediction. We find that these models outperformed regression trees. Second, our study investigates the effects of adding the 179 macroeconomic variables to recovery rate models. Third, we improved out-of-sample predictive accuracy of recovery rate models by adding LASSO-selected macroeconomic variables from 179 macroeconomic variables. Fourth, we study the predictive accuracy of the models with the macroeconomic variables have been transformed in order to be confident that the resulting macroeconomic variables are stationary.

Table 1 summarizes the studies that incorporate macroeconomic variables in recovery rate models for U.S. corporate debt instruments, as well as the econometric methodology employed.

3. Estimation and selection methods

In this section, we describe the six estimation and selection methods used in this study: shrinkage methods, regression tree, bagging, boosting, principal components regression, and support vector regression.

3.1. Shrinkage methods

For a ridge regression, as introduced by Hoerl and Kennard (1970), regression coefficients are estimated by minimizing the following formula:

$$\sum_{i=1}^{N} \left(y_i - \alpha - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 + \lambda_2 \sum_{j=1}^{k} \beta_j^2 \tag{1}$$