Sermpinis, G. , Tsoukas, S. and Zhang, P. (2018) Modelling market implied ratings using LASSO variable selection techniques. *Journal of Empirical Finance*, 48, pp. 19-35. (doi:[10.1016/j.jempfin.2018.05.001](https://doi.org/10.1016/j.jempfin.2018.05.001))

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/164498/>

Deposited on 25 June 2018

Enlighten – Research publications by members of the University of
Glasgow

<http://eprints.gla.ac.uk>

Modelling market implied ratings using LASSO variable selection techniques¹

by

George Sermpinis, Serafeim Tsoukas, Ping Zhang
University of Glasgow

January 2018

Abstract

Making accurate predictions of corporate credit ratings is a crucial issue to both investors and rating agencies. In this paper, we investigate the determinants of market implied credit ratings in relation to financial factors, market-driven indicators and macroeconomic predictors. Applying a variable selection technique, the least absolute shrinkage and selection operator (LASSO), we document substantial predictive ability. In addition, when we compare our LASSO-selected models with the benchmark ordered probit model, we find that the former models have superior predictive power and outperform the latter model in all out-of-sample predictions.

Key words: Market implied ratings, LASSO, Financial ratios, Forecasting

JEL: G24, G33, C25, O16

¹ We are grateful for comments and inputs from an anonymous referee, Oldrich Vasicek and from participants at the 2017 European Financial Management Association Conference, the 2017 European Meetings of the Econometric Society, the 2016 International Conference on Forecasting Financial Markets and the Adam Smith Business School seminars. Corresponding Author: Serafeim Tsoukas, email: serafeim.tsoukas@glasgow.ac.uk. Adam Smith Business School, University of Glasgow, G12 8QQ, Tel: +44 141 3306325. The third author is grateful to China Scholarship Council for funding (Grant number: 201508060273). Any remaining errors are our own.

1. Introduction

Recent turmoil in the financial markets has focused attention on the rating agencies and the process by which they assign ratings to firms and their financial obligations. A careful management of credit risk is high the agenda for both market participants and regulators. It is well known that the credit ratings of the top three nationally recognized statistical rating organizations (NRSROs), Standard and Poor's, Moody's and Fitch play a key role in the pricing of credit risk and in the delineation of investment strategies since they measure the firm's long-term ability and willingness to meet debt servicing obligations. As such, the ratings indicate the probability that a given borrower will default. However, the accuracy and the timing of the ratings have been heavily criticized, especially during the most recent financial crisis. It has been argued that the standard agency ratings do not adjust quickly to price changes and therefore may be out of date. In response to these concerns, Fitch has recently developed a new model to derive Market Implied Ratings (MIRs) from bond and equity prices. The obvious advantage of these ratings compared to the conventional agency ratings is that they adjust instantaneously to price changes.

This study offers methodological extensions applying a variable selection approach, the least absolute shrinkage and selection operator (LASSO), and its most promising derivation, the Elastic net, into ordered probit and continuation ratio models, to the task of predicting Fitch's CDS and Equity implied ratings (CDSIRs and EQIRs respectively hereafter). The research aims to exploit the LASSO properties and unveil the underlying structure of CDSIRs and EQIRs.² There are several studies that use accounting ratios and other publicly available information in reduced-form models in order to predict credit ratings. These studies use various techniques (OLS, multinomial and ordered logit/probit models) to identify the most important characteristics for predicting bond ratings (see for instance the early studies by Poldue and Soldofsky, 1969, Pinches and Mingo, 1973, Kaplan and Urwitz, 1979 and Kao and Wu, 1990). The upshot is that financial healthiness is associated with ratings determination and prediction of default. Another line of research advocates the importance of estimating the

² With respect to the latter aim of this study, we do not report estimated coefficients of the prediction models to preserve space. These results are available upon request.

models in a dynamic setting and documents a noticeable improvement in the predictive ability of the models once state dependence is controlled for (see Mizen and Tsoukas, 2012). One drawback of the reduced-form models, discussed above, is that they tend to employ many rating predictors as inputs despite the fact that only a sub-set is relevant. This has two critical implications. First, this approach can omit potentially important determinants leading to a decrease in prediction accuracy. Second, given the large number of predictors included, it does not provide a sparse representation, implying that these models cannot be readily used by market participants and rating agencies.

Our approach is mostly related to the literature that examines the determinants of credit ratings, but we add to it in two important ways. First, we make a methodological contribution by deriving a simple and more intuitive, yet innovative model, which is based on the variable selection technique, pioneered by Tibshirani (1996)—the least absolute shrinkage and selection operator (LASSO). It is well accepted that this selection approach not only helps in identifying the most relevant predictors from an extensive set of candidate variables, but also improves the predictive power (see Fan and Li, 2001 and Tian et al, 2015). In addition, LASSO does not require strict assumptions such as a pre-selection of the variables considered and it is consistent statistically as the number of observations approach infinity (Van de Geer, 2008). Importantly, LASSO can potentially sidestep the problem of multicollinearity, which is fairly common in probit/logit models, and it is computationally efficient even when considering a large set of potential predictors. Our study is the first, as far as we know, to provide a systematic empirical analysis of LASSO in MIRs forecasts. In doing so, we explore the relative importance of several time-varying covariates from an extensive set of firm-, market-specific and macro-economic explicators used to predict market implied ratings. This is important as we provide a parsimonious set of predictors that can be readily implemented by investors, managers and credit risk agencies.

Second, we use a data-set made up by MIRs instead of the standard long-term ratings. The former type of ratings represents an innovation to the ratings industry to address the issue of staleness in their long-term counterparts. The market implied ratings rely on proprietary and data-intensive rating models that incorporate market information into a model-based credit

assessment (see for instance, Rösch, 2005 and Tsoukas and Spaliara, 2014). The most appealing characteristic of these ratings is that they can adjust instantly to market changes. Hence, we build on the foundations of the literature on implied (or point-in-time ratings) by investigating the forecasting power of models that capture volatile market changes.

To preview our findings, we show that several financial factors along with market-driven and macroeconomic variables contain information about market implied ratings. Importantly, when applying the LASSO techniques, we are able to significantly improve the predictive power of our models in out-of-sample predictions compared to the ordered probit model, which is commonly adopted in the literature. Moreover, we note that the LASSO models with BIC-type tuning parameter selector outperform their LASSO counterparts with AIC-type selector for the dataset and periods under study. Thus, LASSO-selected models display improved forecasting power.

The rest of the paper is laid out as follows. In section 2 we discuss the relevant literature. Section 3 presents the data and summary statistics. In section 4 we describe our methodology. In Section 5 we report the empirical results and robustness tests. Section 6 concludes the paper.

2. Related literature

The issue of how rating agencies use public information in setting quality ratings has attracted considerable attention in the literature. In fact, the literature goes as far back as Horrigan (1966). This study presents the first attempt to predict ratings based on the characteristics of the bonds and the issuing firms. The author concentrates on accounting data and financial ratios in order to find the most appropriate predictors. The set of preferred variables contains total assets, net worth to total debt, net operating profit to sales and working capital. Poldue and Soldofsky (1969) also assign ordinal numbers to ratings and investigate different accounting variables as potential determinants. They conclude that the most significant independent variables are long-term debt to total assets, the coefficient of variation of

earnings, and total assets.³ West (1970) challenge Horrigan's study by using another set of explanatory variables namely earnings volatility, capital structure, reliability and marketability. Based on values of the obtained R-squared, the author claims that the proposed model has a better explanatory power.

Pinches and Mingo (1973) adopt a two-stage approach to assign ratings to bond issues. This study attempts to test the predicting ability of a small number of explanatory variables using multiple regression and discriminant analysis. The proportion of correct predictions lie in the region of 70 percent. Moreover, Kaplan and Urwitz (1979) confirm the above studies by using an ordered probit analysis. They show that ratings may be reasonably well predicted using balance sheet information. Other studies that use a small number of explanatory variables (e.g leverage, profitability, interest coverage, firm's size and subordination status) to predict credit ratings include Ederigton (1985) and Gentry et al (1988). The former study uses long-term debt, subordination, total assets and interest coverage as explanatory variables, while the later focuses on subordination, size of issue, debt ratio, cumulative years that dividends were paid and net income to interest.

Blume et al. (1998) depart from the traditional examination of credit ratings determinants by considering whether there is any tendency for a company that maintains the same values of accounting ratios over time to receive a lower rating due to worsening of rating standards. Using an ordered probit analysis they find that rating agencies have changed the way in which they evaluate credit standing and they report a secular tightening of rating agency standards. They conclude that it became more difficult for firms to obtain improved ratings in the mid-1980s and early 1990s.

More recently, ordered probit methodologies are employed by Hwang et al., (2009) and Hwang et al., (2010) to forecast credit ratings. Both studies show that several predictors are important in forecasting credit ratings such as the size of the company, balance sheet position, stock market performance and industry effects. In addition, modelling long-term ratings in a dynamic setting has shown improvements in forecasting (see Hwang, 2011 and Hwang, 2013).

³ The study shows that both profitability and coverage ratio are insignificant and quantitatively unimportant.

In a similar vein, Mizen and Tsoukas (2012) find that allowing for persistence in ratings significantly improves the forecasting power of long-term ratings. In a subsequent study, Tsoukas and Spaliara (2014) use the market implied ratings and the ordered probit modelling strategy to investigate the role of financial constraints. They conclude that financial variables are more important in predicting credit ratings for firms likely to face financing constraints.

The literature on market implied ratings has focused on the comparison between long-term agency ratings and market implied ratings. Breger et al. (2003) use bond spreads to find suitable thresholds categorizing bonds. They find that implied ratings are a superior application to identify default probability in the rating system. Rösch (2005) documents that implied ratings can provide more accurate default probability forecasts than long-term ratings. Castellano and Giacometti (2012) note that MIRs can be regarded as early warning signals of credit rating changes.

Moving to the line of work on variable selection techniques, there is only a handful of papers investigating default probabilities in various settings. Härdle and Prastyo (2013) employ the LASSO approach to predict default probability in a sub-set of Asian economies. Amendola et al, (2012) evaluate the default risk in the limited liability sector in Italy. Finally, Tian et al., (2015) evaluate the probability of bankruptcy using a comprehensive sample of US firms. The authors conclude that the accuracy in the out-of-sample prediction can be superior to previous studies of estimating default by combining reduced-form models with the LASSO procedure.

The studies discussed above provide a useful background on the credit ratings procedure and the selection process of relevant predictors. In the sections that follow we turn to our data and estimation strategy.

3. Data and summary statistics

3.1 Data sources

The data on market implied ratings are taken from Fitch's database and refer to solicited ratings for all traded US corporations. This database provides information on the CDS and

Equity implied ratings assigned to each issuer as well as the date that the rating became available. Both CDS and Equity implied ratings are reported on a monthly frequency and span the period 2002 to 2008.⁴ In keeping with the normal practice in the literature, we categorize our firms into rating buckets without consideration of notches (i.e. + or -). Amato and Furfine (2004) and Mizen and Tsoukas (2012) note that this classification takes into account large cumulative changes of ratings rather than small movements notch by notch, and avoids generation of rating categories with very few observations. Therefore, we consider seven rating categories, ranging from AAA to Below CCC, which are assigned numerical values in Table 1, starting with 1 to AAA, 2 to AA..., 7 to Below CCC.⁵

Insert Table 1

Firm-specific accounting data are extracted from Fitch's Peer Analysis Tool. Corporate historical data for all firms rated by Fitch are available on a quarterly basis from this database. For these firms with credit ratings, we link their ratings to Fitch's balance sheet statements and profit and loss accounts. Hence, our dataset is constructed by merging monthly market implied ratings data and quarterly firm-level accounting data.⁶ In other words, we have an entry for each firm-month with CDSIRs and EQIRs data and financial and market data. Following commonly used selection criteria in the literature; we exclude companies that do not have complete records on our explanatory variables and firm-months with negative sales and assets. To control for the potential influence of outliers, we winsorize the regression variables at the 1st and 99th percentiles.

Data on market indicators and macroeconomic variables are sourced from Bloomberg. These data items are reported monthly. Our combined sample contains data for 211 firms that

⁴ The research aims to study the structure and the predictability of the implied ratings in the years preceding the recent global financial crisis. Our choice to focus on a time window ending in 2008 is motivated by two important considerations. First, the global financial crisis and the collapse of Lehman Brothers constitute a shock that may have acted as a confounding factor to the determination of credit ratings. In fact, it can be argued that the misinterpretation of the credit ratings by investors was one of the main contributors to the crisis. Second, the data were downloaded early in 2010 owing to a research project supported by Fitch: the coverage period is therefore 2002-2008.

⁵ In EQIRs we do not observe any ratings in the last category, hence six groups are generated for this type of implied ratings.

⁶ For each firm, the quarterly data is repeated every month during the same quarter if the monthly data is not observed.

operate in all sectors of the US economy except agriculture, forestry and fishing and public administration. The panel has an unbalanced structure with the number of observations on each firm varying between 1 and 63. Our sample presents two characteristics that make it especially appealing for our analysis. First, it includes both investment and speculative grades ratings in line with previous studies (see for instance Amato and Furfine, 2004). This is particularly beneficial since we can cover the entire spectrum of firms. Second, the distribution of agency (long-term) ratings in CDS data is very similar to the distribution of agency ratings in the general bond population (see Fitch 2007). Thus, both the CDSIRs and the EQIRs databases can provide a representative base for conducting our empirical analysis.

3.2 Choice of explanatory variables

Prior empirical research on the determinants of credit ratings has considered both business and financial risks. The former type of risk includes an assessment of industry characteristics, firm size, management capability and organizational factors. The latter concerns the quality of a firm's accounting procedures, profitability, cash flow situation and its overall financial policy. In the market implied ratings, models typically consider market-related information in addition to the above-mentioned factors. With that in our mind, we also turn to rating agencies and in particular to Fitch, to find out what matters when assigning a market implied rating. In other words, the selection of our explanatory variables is guided both from the existing empirical literature (see for example Kaplan and Urwitz, 1979; Ederington, 1985; Poon, 2003; Chava and Jarrow, 2004; Amendola et al, 2011; Mizen and Tsoukas, 2012; Hwang, 2013; Creal et al, 2014; Doumpos et al, 2015 and Tian et al, 2015), and the common practice of rating agencies (see Fitch 2007 and Liu et al. 2007).⁷

3.2.1 Firm-specific variables

We consider 16 firm-specific accounting variables as potential predictors of ratings. These variables are intended to measure different aspects of firms' financial health, these are size, leverage, coverage, cash flow, profitability and liquidity. Specifically, we employ the firm size

⁷ The expected relationship between these variables and MIRs is presented in Table A1 in on-line Appendix A. The Table also provides a detailed description of the variables used in this study.

(DETA) as measured by the natural logarithm of firms' real total assets. Size accounts for the scale of the firm and would be expected to improve the rating. Next, we proxy leverage using a number of ratios: The ratio of long-term debt to total assets (LDA), the ratio of short-term debt to total assets (SDA), the ratio of total debt over total assets (TDA), the ratio of total assets over equity (AE), and the ratio of total debt to earnings before interest, taxes, depreciation, amortization, and restructuring or rent costs (TDEBITDA). Higher values of the above ratios are likely to increase financial risk and hence should worsen the rating. The next two measures capture the creditworthiness of the firm as they show the firm's ability to generate income to meet interest rate obligations: the ratio of earnings before interest and tax over interest expenses (EBITINT) and the ratio of total debt to earnings before interest, taxes, depreciation, amortization, and restructuring or rent costs to interest expenses (EBITDAINT). Both ratios would improve the credit rating if they were to increase. Further, cash flow is measured by the following ratios: cash flow from operating activities over total assets (CFOA), and cash and equivalent over total assets (CASHEQA). We expect firms with higher cash flows to have improved ratings. The following five ratios measure firm profitability: The ratio of operating income to net sales (OM), the ratio of net income without dividends over total capital (ROC), the ratio of net income over shareholders' equity (ROE), the ratio of net income over total assets (ROA) and the ratio of the funds from operations to total debt (FFD). An increase in the above-mentioned profitability ratios should be associated with an improvement of ratings. Finally, liquidity is measured by the ratio of cash from operations to liabilities (LIQ), which indicates a firm's ability to satisfy its short-term obligations as they become due. Higher levels of liquidity should improve credit ratings.

3.2.2 Market-driven indicators

As noted above, market implied ratings are likely to be determined by market-related conditions. Therefore, we employ the following market indicators: Excess return (EXRET) as measured by the monthly stock return on the firm minus the S&P 500 index return. The relative size of a firm in market (RSIZE) measured by each firm's market equity value divided by the total market equity value. The above-mentioned variables should be positively correlated with ratings upgrades. Next, we use the volatility of stock return (STD) which is calculated as the standard deviation of each company's monthly stock returns. We also use

the systematic risk of each firm (BETA), measured by the Capital Asset Pricing Model for each firm. Finally, we extract the 1-year and 5-year default probabilities (PD1 and PD5) from the Fitch's Peer Analysis Tool. All three variables should worsen ratings if they were to increase.

3.2.3 Macroeconomic influences

We also consider an extensive list of macro-economic covariates as potential predictors of market implied ratings. Specifically, the stock market performance is evaluated by the S&P 500 return, which calculates returns on the S&P 500 index (RLSP). The short-term interest rate as measured by the three-month commercial paper rate (CPFFM), three-month Treasury bill rate minus federal funds rate (TB3) and the one-year constant maturity treasury rate (GS1). We also employ the general price level, as measured by the growth rate in the narrow money stock (MB) and inflation rate (INFL). The aggregate economic activity is captured by the rate of change in industrial production (DLIP), the index of the growth rate of real GDP (DLGDP), the average of monthly Chicago Fed National Activity Index over the year (CFNA), the average of monthly unemployment rate over the year (UNRATE) and the Chicago Board Options Exchange (CBOE) volatility index (VIX). All macro-variables, apart from VIX, are reported in percentages. The eleven macroeconomic variables measure different aspects of the aggregate economy's performance. Their relationship with the market implied ratings could be either positive or negative as ratings tend to improve during good times but agencies have been observed to tighten their standards during these periods. Hence, the relationship between ratings and macro-economic variables is an issue that will be determined empirically.

3.3 Summary statistics

Tables 2 and 3 illustrate the distribution of firms by rating category for CDSIRs and EQIRs, respectively. It can be observed that the distribution of firms across the rating categories is quite stable and that most companies are assigned A and BBB ratings.

Insert Table 2

Insert Table 3

At the next stage, we report summary statistics for our explanatory variables in Tables 4 and 5. We present statistics splitting the sample between investment grade and non-investment grade to gauge any differences across ratings categories. P-values for the tests of equality of means across the above-mentioned groups are reported in the last columns of the tables. We observe, as expected, that firms in the investment grade group display better financial characteristics, as measured by the balance sheet indicators. The tests point to significant differences between the two groups, which indicate that there is a correlation between better financial health and an improved rating. In other words, there is considerable cross-sectional variation in market implied ratings. Moving to the market indicators, we find that improved market conditions are associated with investment grade ratings, which also suggests a link between the market climate and the ratings.

Insert Table 4

Insert Table 5

4. Methodology

We predict the changes in market implied ratings with ordered probit (OP) and continuation ratio (CR) models combined with LASSO or the Elastic net. The proposed methodology aims at selecting the most important predictors and at providing accurate MIRs forecasts. LASSO, originally proposed by Tibshirani (1996), is an extended form of an OLS regression which performs both variable selection and regularization through a shrinkage factor. It is capable of enhancing the accuracy and the interpretability of classical regression methods (Tibshirani, 1996). To maintain the properties of LASSO and capture the ordinal ranking of MIRs, penalty functions from LASSO or its variant (Elastic net) are added into OP or CR models. This helps us unveil the relation between the potential predictors (at the firm and macro level) and identify their significance in predicting MIRs. As benchmark to our study, we rely to the standard OP model. A description of the empirical modelling strategy follows.

4.1 OP

MIRs as a branch of credit ratings are discrete-valued signs and have an ordinal ranking. To meet the ordinal property of MIRs, OP is applied naturally as a benchmark in the relevant literature (Kaplan and Urwitz, 1979; Gentry, 1988; Blume et al., 1998; Amato and Furfine,

2004 and Hwang et. al., 2009). OP takes into account both the existence of ordinal ranking and the difference between any two adjacent ratings⁸. We define the categorical variable $y_{it} = 1, 2, \dots, 7$ according to the rating assigned to each firm. We assume that there is an unobservable dependent variable y_{it}^* associated with y_{it} , which can be expressed as:

$$y_{it}^* = X_{it}\beta_1 + X_{it-1}\beta_2 + X_{it-2}\beta_3 + X_{it-3}\beta_4 + W_{it}\beta_5 + W_{it-1}\beta_6 + W_{it-2}\beta_7 + W_{it-3}\beta_8 + Z_{it}\beta_9 + Z_{it-1}\beta_{10} + Z_{it-2}\beta_{11} + Z_{it-3}\beta_{12} + y_{it-1}\beta_{13} + \varepsilon_{it}, \quad (1)$$

where $i = 1, 2, \dots, N$ represents firms, and $t = 1, 2, \dots, T$ represents different time periods. In this context, t is the month end for monthly data. β are vectors of unknown parameters to be estimated. X denotes a set containing 16 accounting variables, which can be divided into broad groups of size, leverage, coverage, cash flow, profitability and liquidity. W and Z contain 6 market-driven variables and 11 macroeconomic variables, respectively. Following the literature (Güttler and Wahrenburg, 2007), all predictors in X , W and Z are lagged three periods denoted by $t-1$, $t-2$ and $t-3$ to mitigate potential time tendency. To capture potential non-linear influences, we allow for non-linear transformation of the variables and therefore the square of each predictor is considered and included⁹. Thus, the total number of firm-specific accounting, market driven and macroeconomic variables is 264¹⁰. y_{it-1} is an indicator of the firm's rating in the previous time periods. We consider 4 lags to control for state dependence. The concern of persistency in ratings is an important dimension of time-series variation and the use of models with lagged rating categories is the standard way of addressing this issue¹¹. The error term ε_{it} in equation (1) is assumed to be a normally

⁸ For details on the exposition of the OP, see Maddala (2008, pp 47-48).

⁹ This approach is justified theoretically since some variables may have a positive effect up to a certain (turning) point and negative thereafter.

¹⁰ We consider 16 accounting variables, 6 market-driven variables, 11 macroeconomic variables and their first three lags (in total, 132 variables). We also consider the square of these variables, thus in equation (1) we have in total 268 variables made up by 264 firm-specific accounting, market driven and macroeconomic variables and 4 state dependence variables. We use the same set of predictors throughout the paper.

¹¹ State dependence captures previous rating state and indicates the realization of a rating in the previous time period. Following Contoyannis et. al. (2004) and Mizen and Tsoukas (2012), the state dependence is controlled for by applying dummy variables representing first lags of each category in the dependent variable. Given that we observe limited observations in the extreme high and low rating categories, these ratings in the state dependence variables are merged into 5 groups, such as above AA ratings, A rating, BBB rating, BB rating and below B ratings. In addition, to avoid the dummy variable trap we omit one baseline rating category. Therefore, the lagged rating category BBB is not included in the models because it is regarded as the baseline category of lagged MIRs.

distributed residual with a zero mean and unit variance. In our data y_{it}^* is not observed. Thus what is observed are the market implied ratings assigned to firms, which can take M values. The relation between the observed variable y_{it} and the latent variable y_{it}^* it is assumed to be given by:

$$y_{it} = m \text{ if } \alpha_{m-1} < y_{it}^* \leq \alpha_m \text{ for } m = 1, \dots, M \quad (2)$$

For a set of parameters α_0 to α_M , where $\alpha_0 < \alpha_1 < \dots < \alpha_M$, $\alpha_0 = -\infty$ and $\alpha_M = \infty$. Assuming a standard normal distribution for ε_{it} , the conditional probabilities can be derived as:

$$\begin{aligned} Pr(y_{it} = m) = & \Phi(\alpha_m - X_{it}\beta_1 - X_{it-1}\beta_2 - X_{it-2}\beta_3 - X_{it-3}\beta_4 - W_{it}\beta_5 - W_{it-1}\beta_6 - \\ & W_{it-2}\beta_7 - W_{it-3}\beta_8 - Z_{it}\beta_9 - Z_{it-1}\beta_{10} - Z_{it-2}\beta_{11} - Z_{it-3}\beta_{12} - y_{it-1}\beta_{13}) - \Phi(\alpha_{m-1} - \\ & X_{it}\beta_1 - X_{it-1}\beta_2 - X_{it-2}\beta_3 - X_{it-3}\beta_4 - W_{it}\beta_5 - W_{it-1}\beta_6 - W_{it-2}\beta_7 - W_{it-3}\beta_8 - Z_{it}\beta_9 - \\ & Z_{it-1}\beta_{10} - Z_{it-2}\beta_{11} - Z_{it-3}\beta_{12} - y_{it-1}\beta_{13}), \end{aligned} \quad (3)$$

Where $\Phi(\cdot)$ is the standard normal distribution function. We can evaluate the above probabilities for any combination of parameters in the vectors α and β .

4.2 OP with LASSO

According to Tibshirani (1996), LASSO is a method of regression that enables estimation and variable selection simultaneously in a non-orthogonal setting. Under a suitable choice of penalty power, LASSO selects variables by forcing some coefficients to zero and shrinking others. This reduces the variance of the estimated value and increases the accuracy of the regression prediction. The LASSO estimator resolves the l_1 -penalized OP problem of estimating β by maximizing a likelihood function. In particular, the maximization of the likelihood proceeds subject to the constraint $\sum_{q=1}^p |\beta_q| \leq s$, where s is a user-specified tuning parameter and $q = 1, 2, \dots, p$ indicates the number of surviving predictors with non-zero estimated coefficients. This penalty corresponds to the L1 norm, and therefore it is often referred to as L1 penalized model. The OP with LASSO (L1 penalized OP model) can then be expressed as:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \left(\ell(\beta|y_{it}, A) - \lambda \sum_{q=1}^p |\beta_q| \right), \quad (4)$$

where $\ell(\beta|y_{it}, A)$ is the likelihood function of OP and A contains the pool of the potential predictors $(X_{it}, X_{it-1}, X_{it-2}, X_{it-3}, W_{it}, W_{it-1}, W_{it-2}, W_{it-3}, Z_{it}, Z_{it-1}, Z_{it-2}, Z_{it-3}, y_{it-1})$. All explanatory variables are standardized before applying the LASSO estimator. In equation (4), λ stands for the tuning parameter. As λ increases the sum of absolute values of estimated coefficients is reduced and shrinkage of coefficients is achieved. If λ exceeds a threshold value in the corresponding models, some estimated coefficients are set to zero ultimately. This “L1 norm penalty” or the constraint formulation in LASSO generates a more interpretable and sparse model.

As already noted, compared with other independent variable selection methods, LASSO can provide more stable and restricted models (Tibshirani, 1996; Fan and Li, 2001; Zou, 2006 and Tian et.al., 2015). It is also a computationally simple and efficient method (Efron et al., 2004). Several approaches, such as cross validation and information criteria, have been proposed in selecting latent models with minimum prediction errors or maximum log-likelihood estimation. Zou et.al. (2007) provide an efficient approach to obtain the optimal LASSO fit with the Akaike information criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Schwartz, 1978). Sun and Zhang (2012) note that the computational cost of applying cross-validation in penalized models is considerable, while the theory of applying cross-validation is poorly understood. Therefore, AIC and BIC are used in the present study to select the tuning parameter and further detect the “best” model among a series of candidate models in OP with LASSO¹². The “best” model in variable selection procedure will be chosen as the one that achieves the minimum AIC or BIC value. The exact algorithm behind this process is presented in on-line Appendix C. These models are benchmarked with their LASSO 10-cross validation counterparts.

¹² It is well-known that AIC and BIC have different properties in model selection (for details see Yang, 2005; Shao, 1997 and Zhang et al, 2010).

4.3 OP with Elastic net

Elastic net is a LASSO variant introduced by Zou and Hastie (2004) that can further improve the accuracy of the estimation in the presence of highly correlated predictors. The Elastic net allows “grouping” variables in the model by adding a l_2 -penalty from ridge regression. The OP with Elastic net resolves the l_1 -penalized and l_2 -penalized OP problem of estimating β to maximize the OP likelihood function. The maximization of the likelihood proceeds subject to the following constraints (penalty functions) $\sum_{q=1}^p |\beta_q| \leq s_1$ and $\sum_{q=1}^p (\beta_q^2) \leq s_2$, where s_1 and s_2 are user-specified tuning parameters. These penalty functions correspond to the L1 and L2 norm. The OP with Elastic net model is presented below:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \left(\ell(\beta|y_{it}, A) - \lambda_1 \sum_{q=1}^p |\beta_q| - \lambda_2 \sum_{q=1}^p (\beta_q^2) \right), \quad (5)$$

where $\ell(\beta|y_{it}, A)$ is the likelihood function of the OP and A contains the pool of the potential predictors $(X_{it}, X_{it-1}, X_{it-2}, X_{it-3}, W_{it}, W_{it-1}, W_{it-2}, W_{it-3}, Z_{it}, Z_{it-1}, Z_{it-2}, Z_{it-3}, y_{it-1})$.

Equation (5) can be converted as follows:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \left(\ell(\beta|y_{it}, A) - \lambda \sum_{q=1}^p \left\{ \alpha |\beta_q| + \frac{1}{2} (1 - \alpha) (\beta_q^2) \right\} \right), \quad (6)$$

where $0 \leq \alpha \leq 1$.

Equation (5) is the vanilla version of Elastic net. The factor $\lambda_2 \sum_{q=1}^p (\beta_q^2)$ allows correlated variables in the corresponding models, which is drawn from ridge regression. If α is equal to 0, the Elastic net keeps the l_2 -penalty in the model in equation (6) (ridge regression). Similarly, if α is equal to 1, the l_1 -penalty will only be kept in the Elastic net and equation (6) reduces to a simple LASSO estimator. The l_2 -norm constraint ensures a unique global minimum in the strictly convex loss function. As in the OP with LASSO estimator, the AIC-type and BIC-type

tuning parameter selectors are employed to the task of selecting the model with the minimum value (see on-line Appendix C). As before, all predictors are standardized before applying the Elastic net estimator.

4.4 CR with LASSO

The CR, originally proposed by Fienberg (1980), was designed for ordinal outcomes in which the categories represent the progression of events or stages in some process. It estimates the probability of one particular category given the categories preceding or succeeding this one. More specifically, it is centred to the binary choice on each ordinal category, which provides the conditional probability of estimating categories. Fienberg (1980), Hardin and Hilbe (2007) and Long and Freese (2006) argue that the CR is superior compared to the binary logistic regression. It is applicable in multi-classification problems where an individual can jump to the discrete rating category without having to pass the intermediate rating categories¹³. Similar to the binary logistic regression, the CR creates binary choices on each ordinal category which allow the calculation of the relevant conditional probabilities. The conditional probability that an individual falls in a level, given that this individual has been beyond that level, is based on “conditional incremental thresholds”. The CR may be regarded as an advanced version of the proportional odds model (ordered logistical model), which preserves the parsimony of the cumulative odds model and considers the ordinal categories of MIRs. These assigned integer values of categories in the CR can be controlled by users, implying that the estimated coefficients in the CR are influenced by the direction chosen for modelling the response variable. In our work, the backward formulation of CR of Archer and Williams (2012) is applied. The progression through the levels of MIRs from investment grade quality (AAA-BBB) to sub-investment grade quality (BB-Below CCC) is expressed by increasing integer values. This helps estimate the odds of lower MIRs rating compared with higher MIRs rating. The above can be expressed as follows:

$$\text{logit}(Pr(y_{it} = m | y_{it} \leq m, X = A)) = \log \left(\frac{Pr(y_{it} = m | y_{it} \leq m, X = A)}{Pr(y_{it} < m | y_{it} \leq m, X = A)} \right)$$

¹³ Market implied ratings share this property.

$$= X_{it}\beta_1 + X_{it-1}\beta_2 + X_{it-2}\beta_3 + X_{it-3}\beta_4 + W_{it}\beta_5 + W_{it-1}\beta_6 + W_{it-2}\beta_7 + W_{it-3}\beta_8 + Z_{it}\beta_9 + Z_{it-1}\beta_{10} + Z_{it-2}\beta_{11} + Z_{it-3}\beta_{12} + y_{it-1}\beta_{13}, \quad (7)$$

where $A = X_{it}, X_{it-1}, X_{it-2}, X_{it-3}, W_{it}, W_{it-1}, W_{it-2}, W_{it-3}, Z_{it}, Z_{it-1}, Z_{it-2}, Z_{it-3}, y_{it-1}$. In equation (7), the dependent variable y_{it} belongs to one of the ordinal rating categories m described in equation (2). For each unit observation, rather than modeling the response y_{it} directly, each variable y_{it} is equal to 1 if the response falls in category m , and 0 otherwise. Thus, conditional likelihood is calculated as in multiple logistic regressions. The above equation can be transformed into the following version to derive the conditional probability:

$$Pr(y_{it} = m | y_{it} \leq m) = \frac{e^a}{1 + e^a}, \quad (8)$$

where $a = X_{it}\beta_1 + X_{it-1}\beta_2 + X_{it-2}\beta_3 + X_{it-3}\beta_4 + W_{it}\beta_5 + W_{it-1}\beta_6 + W_{it-2}\beta_7 + W_{it-3}\beta_8 + Z_{it}\beta_9 + Z_{it-1}\beta_{10} + Z_{it-2}\beta_{11} + Z_{it-3}\beta_{12} + y_{it-1}\beta_{13}$.

The parameters can be estimated with maximum likelihood. Similar to OP with LASSO, the maximization of the likelihood proceeds subject to the constraint $\sum_{q=1}^p |\beta_q| \leq s$, where s is a user-specified tuning parameter. This algorithm combined with LASSO can produce shrinkage coefficients that improve the model's predictive ability. The resultant model, a CR model, in which a l_1 -penalized constraint is added into the corresponding likelihood function, is the L1-penalized continuation ratio model. The estimation is presented in equation (9) below:

$$\hat{\beta} = \underset{\beta}{argmax} \left(\mathcal{L}(\beta | y_{it}, A) - \lambda \sum_{q=1}^p |\beta_q| \right), \quad (9)$$

where $\mathcal{L}(\beta | y_{it}, A)$ is the likelihood function of CR and A contains the pool of the potential predictors $(X_{it}, X_{it-1}, X_{it-2}, X_{it-3}, W_{it}, W_{it-1}, W_{it-2}, W_{it-3}, Z_{it}, Z_{it-1}, Z_{it-2}, Z_{it-3}, y_{it-1})$. All explanatory variables are standardized before applying the LASSO estimator. In line with the abovementioned LASSO counterparts, the AIC-type and the BIC-type tuning parameter

selectors assist with selecting the best model¹⁴. All predictors are standardized before applying the LASSO estimator.

4.5 CR with Elastic net

To achieve CR with Elastic net, both l_1 -penalty and l_2 -penalty are added into maximum likelihood of CR to obtain estimated coefficients, are explained by equations (10) and (11) below:

$$\hat{\beta} = \operatorname{argmax}_{\beta} \left(\mathcal{L}(\beta | y_{it}, A) - \lambda_1 \sum_{q=1}^p |\beta_q| - \lambda_2 \sum_{q=1}^p (\beta_q^2) \right). \quad (10)$$

Similar to the OP with Elastic net, equation (10) is converted as follows:

$$\hat{\beta} = \operatorname{argmax}_{\beta} \left(\mathcal{L}(\beta | y_{it}, A) - \lambda \sum_{q=1}^p \left\{ \alpha |\beta_q| + \frac{1}{2} (1 - \alpha) (\beta_q^2) \right\} \right), \quad (11)$$

where $0 \leq \alpha \leq 1$, $\mathcal{L}(\beta | y_{it}, A)$ is the likelihood function of CR and A contains the pool of the potential predictors ($X_{it}, X_{it-1}, X_{it-2}, X_{it-3}, W_{it}, W_{it-1}, W_{it-2}, W_{it-3}, Z_{it}, Z_{it-1}, Z_{it-2}, Z_{it-3}, y_{it-1}$). All explanatory variables are standardized before applying the Elastic net estimator. Similar to the aforementioned LASSO counterparts, the AIC-type and the BIC-type tuning parameter selectors assist with selecting the best model. Once again, before implementing the Elastic net estimator, all predictors are standardized.

5. Empirical results

5.1 Accuracy

In Table 6 we evaluate the forecasts of the models under study for firms' CDSIRs and EQIRs, using Accuracy Ratios (ARs). The AR related to the percentage of correct forecast is the sum

¹⁴ For more details, see on-line Appendix C.

of all diagonal terms divided by the total number of observations in each contingency table (see Appendix B) $AR = \frac{1}{T} \sum_{t=1}^T 1(\hat{q}_t = q_t)$ where $\hat{q}_t$ denotes the predicted rating and q_t represents the actual outcome. We report statistics for all candidate models for both in- and out-of-sample predictions. For the out-of-sample predictions of ratings we use the past and current information available up to time T . We use an expanding window method, which allows the successive observations to be included in the initial sample prior to forecast of the next one-step ahead prediction of the rating while keeping the start date of the sample fixed. By this method, we forecast future ratings $\hat{q}_{t+1}$, $\hat{q}_{t+2}$ etc. The initial estimation window is 2002 to 2005 and the first prediction date is year 2006. We then increase T by one each time until T reaches year 2008. In addition, we report at the foot of each panel the number of surviving variables¹⁵.

Insert Table 6

To begin with the in-sample exercise, we find no notable differences between the competing models since they present a similar in-sample performance for both types of market implied ratings. With respect to CDSIRs, approximately 90% predictions are correct, while for EQIRs we find that the models have approximately 95% correct predictions. Moving to the out-of-sample prediction, the results suggest that the LASSO models clearly outperform their OP benchmark. When considering the CDSIRs, our results indicate that the percentage of correct predictions increase from 22% in the OP model to 84% in the LASSO models. With reference to EQIRs, the percentage of correct predictions improves from 49% in the OP model to 91% in the LASSO models.

Next, we compare the within performance of the LASSO models, by considering alternative LASSO specifications. Starting with CDSIRs, the models with BIC-type tuning parameter selector provide more accurate out-of-sample forecasts compared to the LASSO models with AIC-type tuning parameter selector. For EQIRs, there is little difference in the accuracy ratios of the various LASSO candidate models. It is interesting to note that the LASSO models with BIC-type tuning parameter selector select consistently a smaller number of predictors than

¹⁵ The surviving variables are defined as predictors with non-zero estimated coefficients after the penalized procedure. In the benchmark model, we do not drop any variables since OP does not penalise regression coefficients.

their AIC-type counterparts. This does not seem to affect their forecasting performance for EQIRs but leads to more accurate predictions for CDSIRs.¹⁶

5.2 Statistical significance

To evaluate the relative performance of the models presented in the sub-section above, we employ the Stuart–Maxwell test. This approach will help us formally test for the statistical significance of the difference between forecasts and to further validate our main findings. The Stuart–Maxwell test (Stuart, 1955; Maxwell, 1970) is a generalized version of McNemar’ test (McNemar, 1947), which is associated with multiple (k) categories and tests whether the difference between two related samples from an ordinal field is statistically different from zero. The Stuart–Maxwell tests the null hypothesis of equal marginal proportion for each category between the forecasts of two models (model A vs model B). Under the null hypothesis, the statistic is distributed as chi-square with $k - 1$ degrees of freedom. A statistically significant Stuart–Maxwell test statistic indicates that the forecasts of the first model (A) are different from those of the second model (B).

For the CDSIRs, as can be seen from Table 6, there is evidence of a statistically significant difference in out-of-sample predictions between the OP model and other competing models. For the EQIRs, the results of the out-of-sample forecasts paint a very similar picture. However, this tendency cannot be clearly observed in the in-sample forecasts of all competing models for both CDSIRs and EQIRs. Combining these statistics with the accuracy ratios, it can be confirmed that adding LASSO or the Elastic net estimator in the OP model or the CR model, can produce different out-of-sample forecasts that are more accurate than those generated by the OP.

5.3 Robustness tests

The findings of the previous Section are further validated by carrying-out several robustness tests. In the first test, we consider an alternative choice of tuning parameters by using the

¹⁶ Tables B1 to B36 in on-line Appendix B illustrate the contingency tables of the predicted against the actual outcome both in- and out-of-sample results for the various models presented in Table 6.

cross validation. Next, we rely on an alternative benchmark for forecasting, namely the Principal Component Analysis. In a further test we employ a tuning-free version of the LASSO estimator. As a fourth test, we use alternative measures of predictive ability. Furthermore, we allow only investment grade ratings. Finally, we consider random effects in panel data.

5.3.1 Cross validation

It is well accepted that the choice of tuning parameter is crucial for the finite-sample performance of estimators such as LASSO or elastic net. Motivated by this consideration, we present forecasts when cross validation is employed, which is one of the most commonly used model selection criteria (see for instance, Stone, 1974 and Yang, 2007). Specifically, we use the ten-fold stratified cross validation (10-fold CV) as a general application of tuning parameter selector. This choice is in line with the relevant literature of model selection (see for example, Kohavi, 1995).

Insert Table 7

The Accuracy Ratios and Stuart–Maxwell test statistics, presented in Table 7, corroborate our main findings. In the in-sample forecast exercises, all models present similar predictive ability. In the out-of-sample prediction, however, for the CDSIRs, all competing LASSO models provide more accurate forecasts than the benchmark model. This tendency can be also observed in the EQIRs. In sum, it appears that there are significant gains in predictive ability once the LASSO is applied even when cross validation is employed. We conclude that our main results are robust to alternative tuning choices.

5.3.2 Principal Component Analysis with OP

One of the most popular statistical procedures for variable selection is the Principal Component Analysis (PCA). PCA converts a set of possibly correlated variables to a smaller set of uncorrelated variables called Principal Components (PC). The first PC accounts as much of the variability in the dataset as possible and each succeeding component turn attains the highest variance possible under the constraint that it is orthogonal to the preceding

components. PCA is probably the most popular dimension reduction procedure in Economics and Finance and has been applied successfully to a series of forecasting problems (see Stock and Watson, 2002, Ludvigson and Ng, 2009 and Bailey et al, 2016). In our application, we are dealing with a large set of possibly correlated variables and thus a natural candidate to benchmark our procedure is the PCA. We select the PCs that account of the 70%, 80% and 90% of the variability in our dataset and combine them with OP. The relevant accuracy ratios are presented in Table 8.

Insert Table 8

Comparing Tables 8 and 6 (where the accuracy ratios of OP and OP with LASSO are presented), we note that PCA naturally improves the out-of-sample accuracy of the single OP model. For our OP combined with LASSO models, the only case that PCA manages to beat our models in terms of out-of-sample accuracy is for CDSIRs. In that case, OP based on the PCs that explain the 70% and 90% of the variability, present a higher accuracy ratio compared to LASSO model with AIC-type tuning parameter selector (but not under the BIC-type selector). It is also interesting to note that PCA with OP displays lower in-sample accuracy ratios at all cases. These results allow to argue that our LASSO formulations are robust to an alternative benchmark.

5.3.3 A tuning-free version of the LASSO

While the results presented so far are robust to different tuning choices, including cross validation, it is important to note that the latter is computationally costly and theoretically less well developed, especially for the purpose of variable selection and the estimation of regression coefficients (see Sun and Zhang, 2012). Thus, to further alleviate potential concerns regarding the choice of the tuning parameter, we employ the scaled LASSO, developed by Sun and Zhang (2012), without depending on model selection criteria such as AIC, BIC or CV¹⁷¹⁸.

Insert Table 9

¹⁷ Another tuning-free version of the LASSO estimator is the square-root LASSO by Belloni et al. (2011).

¹⁸ For the scaled LASSO, the authors generated the gradient descent algorithm in a convex minimization of a penalized joint loss function.

Table 9 presents the relevant Accuracy Ratios and the realizations of the Stuart–Maxwell statistical tests. We note that on in-sample evidence all models continue to display similar accuracy. In the out-of-sample evaluation, for the CDSIRs and EQIRs the predictive performance of the scaled LASSO is superior compared with that of the OP. Once again, this finding is consistent with our main results, indicating that our findings are robust to using the scaled LASSO.

5.3.4 Alternative measures of predictive ability

5.3.4.1

Thus far, the relative performance of the estimated models is evaluated in terms of an informal goodness of fit indicator, by comparing predicted and observed ratings. It is possible, however, to give a more quantitative measure of the predictive ability of our models. We therefore check the robustness of our measure of predictive power by using a measure based on a technique proposed by Merton (1981) and used in Henriksson and Merton (1981), Pesaran and Timmermann (1994), Kim et al. (2008) and Mizen and Tsoukas (2012). Specifically, let CP_j be the proportion of the correct predictions made by $\hat{q}_t$ when the true state is given by $q_t = j$. From the definition of conditional probability, CP is computed as $CP_j = \frac{\frac{1}{T} \sum_{t=1}^T 1(\hat{q}_t=j)(q_t=j)}{\frac{1}{T} \sum_{t=1}^T 1(q_t=j)}$ and the Merton's correct measure expressed CP is given by $CP = \frac{1}{J-1} [\sum_{j=1}^J CP_j - 1]$ where J is the number of categories, and $-\frac{1}{J-1} < CP < 1$. In the contingency table (see Appendix B) CP is the unweighted average of CP_j s minus one (to correct for the phenomenon that certain categories are over-represented). The CP_j s are calculated as the proportion of correct predictions divided by the total of each row. This modifies the measure of predictive ability to discount the influence of the dominant outcome. A high CP score indicates that the predictor is accurate for all rating categories.

Insert Table 10

The Accuracy Ratios when we account for the influence of the dominant outcome by reporting the Merton correct predictions are shown in Table 10. We also report the corresponding Stuart–Maxwell statistical tests. The test produces CP ratios that confirm our

main findings. In the in-sample exercise, there is little difference between the OP and the LASSO models. In contrast, the predictive ability of the out-of-sample predictions is superior when penalty functions are applied.

5.3.4.2

An alternative measure for our forecasts can be constructed based on the misclassification rate of Hastie et al. (2009): $T^{-1} \sum_{t=1}^T 1\{\hat{q}_t \neq q_t\}$. In the spirit of Diebold and Mariano (1995), we estimate the misclassification rate for each forecast and then we test if the mean difference of these rates between two models is 0 or not. If this difference is statistically different than 0, then this mean that the two models generate different forecasts. Tables 11 and 12 presents the relevant p-values for our out-of-sample forecasts.

Insert Tables 11

Insert Tables 12

We note that in almost all cases, our forecasts are statistically different. These results supplement the previous section and further demonstrate the superiority of LASSO as a variable selection technique and the effectiveness of BIC criterion in tuning the LASSO parameters. To sum up, the results are robust to using a test that calculates correct predictions using the proportion of correct predictions for each of the various rating categories.

5.3.5 Investment grade ratings

Much of the previous related literature studies employ data with investment grade ratings. However, as noted by Amato and Furfine (2004), restricting attention to one category is likely to induce selection bias. On the other hand, pooling together both categories may result in misspecification of our model if changes in financial and business risk have a different impact on creditworthiness across the groups of firms. Therefore, we drop all speculative grade ratings and re-estimate our models.

Insert Table 13

The results in Table 13 corroborate our main findings. In the in-sample forecast exercises, all models present similar performance. For the CDSIRs, both versions of the CR with LASSO provide more accurate forecasts. On the other hand, for EQIRs all LASSO models display better predictive performance than their OP benchmark. To sum up, even when limiting our sample to investment grade ratings only, the out-of-sample predictions evidence that the LASSO models outperform the benchmark model.

5.3.6 Accounting for the panel data dimension

As a final robustness test we consider a random-effects version of the ordered probit model to take into account the panel data dimension of the data-set. The Accuracy Ratios and the corresponding Stuart–Maxwell statistical tests are reported in Table 14. As can be seen, the main findings remain unchanged: when we apply the LASSO or Elastic net estimator the models have superior predictive ability compared to the ordered probit model, even when random effects are included. We conclude that our findings are robust to estimating the models with random effects to deal with the panel data nature of the sample.

Insert Table 14

5.4 Discussion

In the previous Sections, a forecasting exercise on CDSIRs and EQIRs prediction is presented. For both types of market implied ratings all models present similar in-sample accuracy. In the out-of-sample evaluation, for CDSIRs, we observe the LASSO models controlled by BIC-type tuning parameter selector outperform their benchmarks. We also observe that these models, select a smaller number of surviving variables than their counterparts with the AIC-type selector. This lends support to the argument that the models with BIC-type tuning parameter selector make better use of the available information. For the EQIRs, we note a similar pattern. The models with the BIC-type tuning parameter selector outperform their counterparts with the AIC-type selector in terms of accuracy while at the same time they select a smaller number of variables.

To sum up, we note that the LASSO models are able to provide more accurate out-of-sample forecasts on the CDSIRs and EQIRs ratings that outperform the OP model. This is of particular interest given that the OP model dominates the related literature in predicting credit ratings. From the LASSO models under study, the optimized models with BIC-type tuning parameter selector seem able to provide better forecasts while at the same time they use less predictors. These results are robust to modifying the tuning parameters, to considering a tuning-free version of LASSO, to evaluating the predictive performance of the models using a different statistical measure and to restricting the dataset to investment grade ratings.

6 Conclusion

The ability to predict credit ratings within a reasonable margin of accuracy is of vital importance for both market participants and rating agencies. The focus on market implied ratings is even more justified as long-term ratings have been heavily criticized about their performance during the recent global financial crisis. We model the prediction of market implied ratings applying a variable selection technique, the least absolute shrinkage and selection operator (LASSO), and its most promising derivation, the Elastic net, into ordered probit and continuation ratio models. All LASSO models select the most relevant predictors from a set of 268 variables and forecast the MIRs for a period of six years (2002 to 2008). This marks a break with the existing literature which typically relies on discrete limited dependent variable models.

Our results using monthly data from the US offer several interesting results. First, we show that financial factors along with market-driven and macroeconomic variables contain information about market implied ratings. Second, the LASSO models perform better in out-of-sample prediction than do ordered probit models, mostly adopted in previous studies. Finally, the optimized LASSO models with BIC-type tuning parameter selector outperform their counterparts with AIC-type selector for the dataset and periods under study. Hence, LASSO-selected models attain an improved forecasting power.

These results should go further in convincing risk managers and academics to explore variable selection models when assessing credit risk. The structure of credit ratings is unknown and likely to vary through time. Limited dependent variable models require a-priori knowledge on the explanatory variables set, which can lead to misspecifications. On the other hand, variable selection models such as LASSO, are more flexible and can unveil the underlying structure of the problem leading to superior estimations and improved predictive ability.

References

- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6), pp.716-723.
- Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), p.589.
- Archer, K. and Williams, A. (2012). L1 Penalized Continuation Ratio Models for Ordinal Response Prediction Using High-Dimensional Datasets. *Statistics in Medicine*, 31(14), pp.1464-1474.
- Amato, J. and Furfine, C. (2004). Are Credit Ratings Procyclical?. *Journal of Banking & Finance*, 28(11), pp.2641-2677.
- Amendola, A., Restaino, M. and Sensini, L. (2011). Variable Selection in Default Risk Models. *The Journal of Risk Model Validation*, 5(1), pp.3-19.
- Bailey, N., Holly, S. and Pesaran, M. H. (2016). A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-Sectional Dependence. *Journal of Applied Econometrics*, 31(1), pp. 249-280.
- Breger, L., Goldberg, L. and Cheyette, O. (2003). Market Implied Ratings. *Risk Magazine*, pp.85-89.
- Belloni, A., Chernozhukov, V., and Wang, L. (2011). Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika*, 98(4):791–806.
- Blume, M., Lim, F. and Mackinlay, A. (1998). The Declining Credit Quality of U.S. Corporate Debt: Myth or Reality?. *The Journal of Finance*, 53(4), pp.1389-1413.
- Castellano, R. and Giacometti, R. (2012). Credit Default Swaps: Implied Ratings Versus Official Ones. *4OR*, 10(2), pp.163-180.
- Chava, S. and Jarrow, R. (2004). Bankruptcy Prediction with Industry Effects. *European Finance Review*, 8(4), pp.537-569.
- Contoyannis, P., Jones, A. and Rice, N. (2004). The Dynamics of Health in the British Household Panel Survey. *Journal of Applied Econometrics*, 19(4), pp.473-503.

- Creal, D., Gramacy, R. and Tsay, R. (2014). Market-Based Credit Ratings. *Journal of Business & Economic Statistics*, 32(3), pp.430-444.
- Diebold, F. and Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3), p.253.
- Doumpos, M., Niklis, D., Zopounidis, C. and Andriosopoulos, K. (2015). Combining Accounting Data and a Structural Model for Predicting Credit Ratings: Empirical Evidence from European Listed Firms. *Journal of Banking & Finance*, 50, pp.599-607.
- Duffie, D., Saita, L. and Wang, K. (2007). Multi-Period Corporate Default Prediction with Stochastic Covariates. *Journal of Financial Economics*, 83(3), pp.635-665.
- Ederington, L. (1985). Classification Models and Bond Ratings. *The Financial Review*, 20(4), pp.237-262.
- Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least Angle Regression. *The Annals of Statistics*, 32(2), pp.407-499.
- Fan, J. and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association*, 96(456), pp.1348-1360.
- Fienberg, S. E. (1980) *The Analysis of Cross-Classified Categorical Data*. Cambridge, MA: The MIT Press.
- Fitch (2007) *Fitch CDS Implied Ratings Model*. Available at: https://www.fitchratings.com/web_content/product/methodology/cdsir_methodology.pdf (Accessed: 20 October 2016).
- Gentry, J., Whitford, D. and Newbold, P. (1988). Predicting Industrial Bond Ratings with a Probit Model and Funds Flow Components. *The Financial Review*, 23(3), pp.269-286.
- Güttler, A. and Wahrenburg, M. (2007). The Adjustment of Credit Ratings in Advance of Defaults. *Journal of Banking & Finance*, 31(3), pp.751-767.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2nd ed. New York: Springer, pp.219-221.
- Härdle, W. and Prastyo, D. (2013). Default Risk Calculation Based on Predictor Selection for the Southeast Asian Industry. *SSRN Electronic Journal*.
- Hardin, J.W., Hilbe, J.M. and W, H. (2007) *Generalized Linear Models and Extensions*. 2nd edn. College Station, TX: Stata Press.
- Henriksson, R. and Merton, R. (1981). On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills. *The Journal of Business*, 54(4), p.513.
- Horrigan, J. (1966). The Determination of Long-Term Credit Standing with Financial Ratios. *Journal of Accounting Research*, 4, p.44.
- Hwang, R. (2011). Predicting Issuer Credit Ratings Using Generalized Estimating Equations. *Quantitative Finance*, 13(3), pp.383-398.

- Hwang, R. (2013). Forecasting Credit Ratings with the Varying-Coefficient Model. *Quantitative Finance*, 13(12), pp.1947-1965.
- Hwang, R., Cheng, K. and Lee, C. (2009). On Multiple-Class Prediction of Issuer Credit Ratings. *Applied Stochastic Models in Business and Industry*, 25(5), pp.535-550.
- Hwang, R., Chung, H. and Chu, C. (2010). Predicting Issuer Credit Ratings Using a Semiparametric Method. *Journal of Empirical Finance*, 17(1), pp.120-137.
- Kao, C. and Wu, C. (1990). Two-Step Estimation of Linear Models with Ordinal Unobserved Variables: The Case of Corporate Bonds. *Journal of Business & Economic Statistics*, 8(3), p.317.
- Kaplan, R. and Urwitz, G. (1979). Statistical Models of Bond Ratings: A Methodological Inquiry. *The Journal of Business*, 52(2), p.231.
- Kim, T., Mizen, P. and Chevapatrakul, T. (2008). Forecasting Changes in UK Interest Rates. *Journal of Forecasting*, 27(1), pp.53-74.
- Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the 14th international joint conference on Artificial intelligence*, 2, pp.1137-1143.
- Liu, B., Kocagil, A.E. and Gupton, G.M. (2007) *Fitch Equity Implied Rating and Probability of Default Model*. Available at: https://www.fitchratings.com/web_content/product/methodology/eir_methodology.pdf (Accessed: 20 October 2016).
- Long, S.J. and Freese, J. (2006) *Regression Models for Categorical Dependent Variables Using Stata*. 2nd edn. College Station, TX: StataCorp LP.
- Ludvigson, S. C. and Ng, S. (2009). Macro Factors in Bond Risk Premia. *The Review of Financial Studies*, 22(12), pp.5027–5067.
- Maddala, G. (2008). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge Univ. Press, pp.47-48.
- Maxwell, A. (1970). Comparing the Classification of Subjects by Two Independent Judges. *The British Journal of Psychiatry*, 116(535), pp.651-655.
- McNemar, Q. (1947). Note on the Sampling Error of The Difference Between Correlated Proportions or Percentages. *Psychometrika*, 12(2), pp.153-157.
- Merton, R. (1981). On Market Timing and Investment Performance. I. An Equilibrium Theory of Value for Market Forecasts. *The Journal of Business*, 54(3), p.363.
- Mizen, P. and Tsoukas, S. (2012). Forecasting US Bond Default Ratings Allowing for Previous and Initial State Dependence in an Ordered Probit Model. *International Journal of Forecasting*, 28(1), pp.273-287.
- Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), p.109.

- Pesaran, M. and Timmermann, A. (1994). A Generalization of the Non-Parametric Henriksson-Merton Test of Market Timing. *Economics Letters*, 44(1-2), pp.1-7.
- Pinches, G. and Mingo, K. (1973). A Multivariate Analysis of Industrial Bond Ratings. *The Journal of Finance*, 28(1), pp.1-18.
- Pogue, T. and Soldofsky, R. (1969). What's in a Bond Rating. *The Journal of Financial and Quantitative Analysis*, 4(2), p.201.
- Poon, W. (2003). Are Unsolicited Credit Ratings Biased Downward?. *Journal of Banking & Finance*, 27(4), pp.593-614.
- Rösch, D. (2005). An Empirical Comparison of Default Risk Forecasts from Alternative Credit Rating Philosophies. *International Journal of Forecasting*, 21(1), pp.37-51.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), pp.461-464.
- Shao, J. (1997). An Asymptotic Theory for Linear Model Selection. *Statistica Sinica*, 7(2), pp.221-242.
- Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, 74(1), pp.101-124.
- Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460), pp.1167-1179.
- Stone, M. (1977). An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), pp.44-47.
- Stuart, A. (1955). A Test for Homogeneity of the Marginal Distributions in a Two-Way Classification. *Biometrika*, 42(3/4), p.412.
- Sun, T. and Zhang, C. (2012). Scaled Sparse Linear Regression. *Biometrika*, 99(4), pp.879-898.
- Tian, S., Yu, Y. and Guo, H. (2015). Variable Selection and Corporate Bankruptcy Forecasts. *Journal of Banking & Finance*, 52, pp.89-100.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), pp.267-288.
- Tsoukas, S. and Spaliara, M. (2014). Market Implied Ratings and Financing Constraints: Evidence from US Firms. *Journal of Business Finance & Accounting*, 41(1-2), pp.242-269.
- van de Geer, S. (2008). High-Dimensional Generalized Linear Models and The Lasso. *The Annals of Statistics*, 36(2), pp.614-645.
- West, R. (1970). An Alternative Approach to Predicting Corporate Bond Ratings. *Journal of Accounting Research*, 8(1), p.118.

- Yang, Y. (2005). Can the Strengths of AIC And BIC Be Shared? A Conflict Between Model Identification and Regression Estimation. *Biometrika*, 92(4), pp.937-950.
- Yang, Y. (2007). Prediction/Estimation with Simple Linear Models: Is It Really That Simple? *Econometric Theory*, 23(01), pp.1-36.
- Zhang, Y., Li, R. and Tsai, C. (2010). Regularization Parameter Selections via Generalized Information Criterion. *Journal of the American Statistical Association*, 105(489), pp.312-323.
- Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476), pp.1418-1429.
- Zou, H. and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), pp.301-320.
- Zou, H., Hastie, T. and Tibshirani, R. (2007). On The “Degrees of Freedom” of the Lasso. *The Annals of Statistics*, 35(5), pp.2173-2192.

Table 1: Rating categories

Market Implied Ratings	Corresponding numerical values
AAA	1
AA	2
A	3
BBB	4
BB	5
B	6
Below CCC	7

Notes: The table presents the rating categories and the corresponding numerical values.

Table 2: CDSIRs of firms by year

	AAA	AA	A	BBB	BB	B	Below CCC	Number of Observations
2002	4	35	46	32	20	1	0	138
2003	10	44	68	65	32	7	0	226
2004	11	41	60	70	35	10	0	227
2005	9	34	57	71	29	10	1	211
2006	9	24	52	63	26	4	1	179
2007	14	45	54	63	28	13	6	223
2008	12	19	13	33	13	3	1	94
Number of Observations	69	242	350	397	183	48	9	1298

Notes: The table presents the distribution of firms by rating category for CDSIRs by year.

Table 3: EQIRs of firms by year

	AAA	AA	A	BBB	BB	B	Below CCC	Number of Observations
2002	1	13	55	103	76	16	0	264
2003	2	19	91	106	43	5	0	266
2004	1	20	100	99	42	6	0	268
2005	0	10	70	87	38	4	0	209
2006	0	11	80	74	29	4	0	198
2007	1	19	91	80	41	8	0	240
2008	0	5	29	37	23	1	0	95
Number of Observations	5	97	516	586	292	44	0	1540

Notes: The table presents the distribution of firms by rating category for EQIRs by year.

Table 4: Descriptive statistics-CDSIRs

Variable	Mean	Standard Deviation	Minimum	Maximum	p-value
	(1)	(2)	(3)	(4)	(5)
DETA					
Investment grade	9.6388	1.0132	7.2272	12.2084	
Non-investment grade	8.9536	0.9959	6.2539	12.2087	0.0000
AE					
Investment grade	3.2256	3.6336	1.3324	73.7340	
Non-investment grade	5.0224	8.8729	1.3283	123.5602	0.0000
LDA					
Investment grade	20.2416	11.2733	0.0000	79.3983	
Non-investment grade	30.1868	19.4641	0.0000	110.4453	0.0000
SDA					
Investment grade	3.5941	3.9838	0.0000	23.5410	
Non-investment grade	3.2961	4.1573	0.0000	23.4635	0.0000
TDA					
Investment grade	24.3546	11.9238	1.3017	87.0595	
Non-investment grade	37.4629	20.8746	0.6518	126.8760	0.0000
TDEBITDA					
Investment grade	2.3395	1.2107	0.3200	19.6400	
Non-investment grade	4.1208	2.8918	0.3200	23.0700	0.0000
EBITINT					
Investment grade	13.3805	19.0270	0.1541	209.3023	
Non-investment grade	7.2685	15.2730	0.1248	210.4054	0.0000
EBITDAINT					
Investment grade	16.0444	21.9950	0.6500	235.4000	
Non-investment grade	9.0872	14.5213	0.3100	196.7000	0.0000
CFOA					
Investment grade	6.9186	6.2625	-41.0623	38.4372	
Non-investment grade	7.1077	7.2097	-13.4106	65.9955	0.0000
CASHEQA					
Investment grade	8.6116	9.6531	0.0238	71.8277	
Non-investment grade	6.5258	8.1009	0.0030	64.0272	0.0000
OM					
Investment grade	13.9496	9.7304	-13.5155	53.0189	
Non-investment grade	10.3893	9.9617	-20.2276	52.6046	0.0000
ROC					
Investment grade	3.7919	5.3598	-34.1330	35.7771	
Non-investment grade	2.1389	7.0866	-36.3880	34.5209	0.0000
ROE					
Investment grade	12.5162	35.3349	-361.1511	452.2565	0.0000

Non-investment grade	9.5597	46.9441	-359.7868	516.7883	
ROA					
Investment grade	3.9430	4.4180	-26.4074	22.0518	
Non-investment grade	2.5165	5.5633	-23.8888	23.2336	0.0000
FFD					
Investment grade	40.7412	32.5483	-16.2800	267.1700	
Non-investment grade	24.4291	26.0776	-17.8000	225.3200	0.0000
LIQ					
Investment grades	12.02064	10.94679	-13.3151	59.8763	
Non-investment grade	11.55889	12.65565	-13.56745	62.18169	0.0000
EXRET					
Investment grade	0.0117	0.0646	-0.3525	0.4527	
Non-investment grade	0.0195	0.1108	-0.3526	0.4673	0.0149**
RSIZE					
Investment grade	0.2170	0.2641	0.0103	1.7570	
Non-investment grade	0.1107	0.1812	0.0041	1.7495	0.0000
STD					
Investment grade	0.0159	0.0079	0.0041	0.1134	
Non-investment grade	0.0256	0.0151	0.0041	0.1141	0.0000
BETA					
Investment grade	0.9366	0.6368	-0.8663	4.4144	
Non-investment grade	1.0795	0.9672	-0.8799	4.9196	0.0000
PD1					
Investment grade	24.0820	70.0007	2.0000	3000.0000	
Non-investment grade	162.7669	489.4410	2.0000	3000.0000	0.0000
PD5					
Investment grade	260.5300	310.3217	14.0000	4495.0000	
Non-investment grade	793.0404	907.6053	14.0000	5464.0000	0.0000

Notes: The Table reports the summary statistics of the explanatory variables used in the empirical models. Column 5 reports the p-value for the test of equality of means between the investment grade and non-investment grade categories. Investment grade refers to ratings from AAA to BBB. Non-investment grade refers to ratings BB and below. A detailed description of the variables used in this study is given in Table A1 in the on-line Appendix A.

Table 5: Descriptive statistics-EQIRs

Variable	Mean	Standard Deviation	Minimum	Maximum	p-value
	(1)	(2)	(3)	(4)	(5)
DETA					
Investment grade	9.5314	1.0515	7.0031	12.2084	
Non-investment grade	8.8701	0.9461	6.2539	12.2087	0.0000
AE					
Investment grade	3.4552	4.7315	1.3283	116.1204	
Non-investment grade	5.3173	9.3934	1.3283	123.5602	0.0000
LDA					
Investment grade	20.9162	12.5949	0.0000	110.1548	
Non-investment grade	32.5301	19.9743	0.0000	110.4453	0.0000
SDA					
Investment grade	3.7397	4.1512	0.0000	23.5410	
Non-investment grade	3.0151	3.9714	0.0000	23.4635	0.0000
TDA					
Investment grade	25.3924	13.5320	1.3017	126.8209	
Non-investment grade	40.3718	21.1606	0.6518	126.8760	0.0000
TDEBITDA					
Investment grade	2.5062	1.4809	0.3200	20.4000	
Non-investment grade	4.4812	3.0369	0.3200	23.0700	0.0000
EBITINT					
Investment grade	13.2588	20.3637	0.1601	210.4054	
Non-investment grade	5.3897	10.2101	0.1248	157.3792	0.0000
EBITDAINT					
Investment grade	15.7294	21.3681	0.4100	235.4000	
Non-investment grade	7.2437	11.9978	0.3100	171.8000	0.0000
CFOA					
Investment grade	7.8973	6.5051	-41.0623	40.5546	
Non-investment grade	5.9413	7.0742	-41.0623	65.9955	0.0000
CASHEQA					
Investment grade	8.6834	9.6845	0.0030	71.8277	
Non-investment grade	5.7724	7.3325	0.0033	58.4741	0.0000
OM					
Investment grade	13.0445	9.5560	-16.9133	53.0189	
Non-investment grade	10.3826	10.3917	-20.2276	52.5504	0.0000
ROC					
Investment grade	3.8570	5.6971	-35.0754	35.7771	
Non-investment grade	1.5077	7.1322	-36.3880	34.5140	0.0000
ROE					
Investment grade	13.3395	36.3731	-361.1511	473.0769	
Non-investment grade	7.5924	48.9666	-359.7868	516.7883	0.0000
ROA					
Investment grade	4.1096	4.6843	-26.4074	23.2190	
Non-investment grade	1.8536	5.4585	-26.4074	23.2336	0.0000
FFD					
Investment grade	38.8160	31.7571	-16.2800	267.1700	
Non-investment grade	21.6264	24.5682	-17.8000	224.0300	0.0000
LIQ					
Investment grade	13.5432	11.8937	-13.5675	62.1817	
Non-investment grade	9.495881	11.69661	-13.56745	57.22284	0.0000
EXRET					
Investment grade	0.0140	0.0723	-0.3525	0.4648	
Non-investment grade	0.0181	0.1089	-0.3526	0.4673	0.5831

RSIZE					
Investment grade	0.2047	0.2687	0.0088	1.7570	
Non-investment grade	0.0922	0.1307	0.0041	1.6261	0.0000
STD					
Investment grade	0.0172	0.0085	0.0041	0.1134	
Non-investment grade	0.0250	0.0155	0.0041	0.1141	0.0000
BETA					
Investment grade	0.9081	0.7224	-0.8757	4.8848	
Non-investment grade	1.1028	0.9284	-0.8799	4.9196	0.0000
PD1					
Investment grade	24.1369	54.1782	2.0000	3000.0000	
Non-investment grade	215.5352	567.0797	2.0000	3000.0000	0.0000
PD5					
Investment grade	250.8945	299.6934	14.0000	4495.0000	
Non-investment grade	1010.6400	970.2219	17.0000	5464.0000	0.0000

Notes: The Table reports the summary statistics of the explanatory variables used in the empirical models. Column 5 reports the p-value for the test of equality of means between the investment grade and non-investment grade categories. Investment grade refers to ratings from AAA to BBB. Non-investment grade refers to ratings BB and below. A detailed description of the variables used in this study is given in Table A1 in the on-line Appendix A.

Table 6: Accuracy Ratios and selected variables

		OP	OP_LASSO		OP_ELASTIC NET		CR_LASSO		CR_ELASTIC NET	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CDSIRs	In-sample prediction	90.23%	90.30%*	89.39%***	90.30%	89.39%***	89.80%***	89.48%***	89.63%***	89.48%***
	Out-of-sample prediction	22.02%	31.56%***	84.53%***	31.56%***	84.53%***	58.80%***	83.73%***	60.92%***	84.26%***
	Surviving variables	268	143	45	144	45	75	48	78	51
EQIRs	In-sample prediction	95.05%	95.00%	94.69%*	95.05%	94.52%**	95.19%	94.77%	95.16%	94.73%
	Out-of-sample prediction	48.98%	84.77%***	90.95%***	80.88%***	90.78%***	85.45%***	91.03%***	80.80%***	90.95%***
	Surviving variables	268	167	95	181	87	152	83	187	91

Notes: The Table reports the Accuracy Ratios and the number of surviving variables for each model under study. "OP" stands for the ordered probit model. "OP_LASSO" refers to the ordered probit model with LASSO estimator. "OP_ELASTIC NET" stands for the ordered probit model with Elastic net estimator. "CR_LASSO" indicates the continuation ratio model with LASSO estimator. "CR_ELASTIC NET" is the continuation ratio model with Elastic net estimator. "AIC" is the AIC-type tuning parameter selector. "BIC" is the BIC-type tuning parameter selector. *** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 1% significance level. ** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 5% significance level. * denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 10% significance level.

Table 7: Accuracy Ratios and selected variables (10-fold cross validation)

		OP	OP_LASSO	OP_ELASTIC NET	CR_LASSO	CR_ELASTIC NET
CDSIRs	In-sample prediction	90.23%	90.32%**	90.33%**	89.53%***	89.48%***
	Out-of-sample prediction	22.02%	35.10%***	35.28%***	84.53%***	84.26%***
	Surviving variables	268	130	134	48	51
EQIRs	In-sample prediction	95.05%	94.88%	94.90%	95.13%	95.08%
	Out-of-sample prediction	48.98%	90.86%***	90.86%***	89.68%***	90.36%***
	Surviving variables	268	135	136	137	145

Notes: The Table reports the Accuracy Ratios and the number of surviving variables for each model under study. "OP" stands for the ordered probit model. "OP_LASSO" refers to the ordered probit model with LASSO estimator. "OP_ELASTIC NET" stands for the ordered probit model with Elastic net estimator. "CR_LASSO" indicates the continuation ratio model with LASSO estimator. "CR_ELASTIC NET" is the continuation ratio model with Elastic net estimator. "AIC" is the AIC-type tuning parameter selector. "BIC" is the BIC-type tuning parameter selector. *** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 1% significance level. ** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 5% significance level. * denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 10% significance level.

Table 8: PCA combined with OP

		PCA with OP		
		(1)	(2)	(3)
CDSIRs	In-sample prediction	45.94%	51.36%	57.39%
	Out-of-sample prediction	45.71%	32.01%	46.33%
	Selected principal components	16	25	45
	PCs cumulative percentage of total variation	70.92%	80.00%	90.31%
EQIRs	In-sample prediction	63.61%	72.00%	76.93%
	Out-of-sample prediction	63.37%	54.74%	51.69%
	Selected principal components	15	23	41
	PCs cumulative percentage of total variation	71.23%	80.43%	90.20%

Notes: The Table reports the Accuracy Ratios and the number of PCs. "OP" stands for the ordered probit model, "PC" for principal components and "PCA" for principal component analysis.

Table 9: Accuracy Ratios and selected variables (scaled LASSO)

		OP	SCALED_LASSO
CDSIRs	In-sample prediction	90.23%	87.98***
	Out-of-sample prediction	22.02%	39.26***
	Surviving variables	268	124
EQIRs	In-sample prediction	95.05%	93.93***
	Out-of-sample prediction	48.98%	89.26***
	Surviving variables	268	121

Notes: The Table reports the Accuracy Ratios and the number of surviving variables for each model under study. "OP" stands for the ordered probit model. "SCALED_LASSO" refers to the scaled LASSO for tuning-free parameter. *** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 1% significance level. ** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 5% significance level. * denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 10% significance level.

Table 10: CP Ratios and selected variables

		OP	OP_LASSO		OP_ELASTIC NET		CR_LASSO		CR_ELASTIC NET	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CDSIRs	In-sample prediction	69.29%	66.89%*	61.04%***	66.95%	61.05%***	63.29%***	61.84%***	62.74%***	61.84%***
	Out-of-sample prediction	18.22%	25.49%***	61.01%***	25.49%***	61.01%***	45.70%***	64.16%***	47.94%***	63.42%***
	Surviving variables	268	143	45	144	45	75	48	78	51
EQIRs	In-sample prediction	78.30%	77.16%	74.46%*	77.19%	74.25%**	78.58%	74.51%	78.45%	74.39%
	Out-of-sample prediction	35.49%	66.50%***	70.10%***	63.42%***	69.93%***	67.29%***	70.23%***	62.39%***	70.10%***
	Surviving variables	268	167	95	181	87	152	83	187	91

Notes: The Table reports the CP ratios and the number of surviving variables for each model under study. "OP" stands for the ordered probit model. "OP_LASSO" refers to the ordered probit model with LASSO estimator. "OP_ELASTIC NET" stands for the ordered probit model with Elastic net estimator. "CR_LASSO" indicates the continuation ratio model with LASSO estimator. "CR_ELASTIC NET" is the continuation ratio model with Elastic net estimator. "AIC" is the AIC-type tuning parameter selector. "BIC" is the BIC-type tuning parameter selector. *** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 1% significance level. ** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 5% significance level. * denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 10% significance level.

Table 11: Equal performance tests of out-of-sample prediction of CDSIRs

		OP	OP_LASSO		OP_ELASTIC NET		CR_LASSO		CR_ELASTIC NET	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
OP		~	7.60***	38.38***	7.60***	38.38***	23.66***	38.57***	24.79***	38.75***
	AIC		~	32.58***	~	32.58***	19.31***	32.75***	20.40***	32.90***
OP_LASSO	BIC			~	-32.58***	~	-17.08***	-1.62	-16.01***	-0.69
	AIC				~	32.58***	19.31***	32.75***	20.40***	32.90***
OP_ELASTIC NET	BIC					~	-17.08***	-1.62	-16.01***	-0.69
	AIC						~	17.36***	4.95***	17.43***
CR_LASSO	BIC							~	-16.27***	1.73*
	AIC								~	16.35***
CR_ELASTIC NET	BIC									~

Note: The table reports the test of equality of the mean difference in losses for two models. *** denotes that the null hypothesis of equal performance of two models is rejected at the 1% significant level. ** denotes that the null hypothesis of equal performance of two models is rejected at the 5% significant level and * denotes that the null hypothesis of equal performance of two models is rejected at the 10% significant level. ~ indicates that the two models generate the same set of forecasts.

Table 12: Equal performance tests of out-of-sample prediction of EQIRs

		OP	OP_LASSO		OP_ELASTIC NET		CR_LASSO		CR_ELASTIC NET	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
OP		~	23.65***	26.97***	21.38***	26.80***	23.72***	27.02***	21.26***	26.97***
	AIC		~	6.03***	-5.83***	5.82***	1.79*	6.09***	-5.82***	6.03***
OP_LASSO	BIC			~	-8.79***	-1.41	-5.46***	1.00	-8.66***	~
	AIC				~	8.64***	6.29***	8.85***	-0.26	8.79***
OP_ELASTIC NET	BIC					~	-5.25***	1.73*	-8.51***	1.41
	AIC						~	5.57***	-6.37***	5.46***
CR_LASSO	BIC							~	-8.71***	-1.00
	AIC								~	8.66***
CR_ELASTIC NET	BIC									~

Note: The table reports the test of equality of the mean difference in losses for two models. *** denotes that the null hypothesis of equal performance of two models is rejected at the 1% significant level. ** denotes that the null hypothesis of equal performance of two models is rejected at the 5% significant level, * denotes that the null hypothesis of equal performance of two models is rejected at the 10% significant level. ~ indicates that the two models generate the same set of forecasts.

Table 13: Accuracy Ratios and selected variables (investment-grade ratings)

		OP	OP_LASSO		OP_ELASTIC NET		CR_LASSO		CR_ELASTIC NET	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CDSIRs	In-sample prediction	90.76%	90.45%**	89.38%***	90.14%***	89.42%***	90.56%**	89.44%***	90.74%	89.40%***
	Out-of-sample prediction	49.18%	77.04%***	84.33%***	84.22%***	84.87%***	82.81%***	85.53%***	80.41%***	84.66%***
	Surviving variables	266	143	19	94	43	98	39	122	28
EQIRs	In-sample prediction	95.80%	95.61%*	95.02%**	95.49%	94.95%**	95.59%	95.05%	95.72%	94.97%*
	Out-of-sample prediction	57.82%	87.72%***	92.28%***	90.99%***	92.28%***	90.69%***	92.28%***	86.44%***	92.28%***
	Surviving variables	266	137	74	143	77	135	61	153	68

Notes: The Table reports the Accuracy Ratios and the number of surviving variables for each model under study. "OP" stands for the ordered probit model. "OP_LASSO" refers to the ordered probit model with LASSO estimator. "OP_ELASTIC NET" stands for the ordered probit model with Elastic net estimator. "CR_LASSO" indicates the continuation ratio model with LASSO estimator. "CR_ELASTIC NET" is the continuation ratio model with Elastic net estimator. "AIC" is the AIC-type tuning parameter selector. "BIC" is the BIC-type tuning parameter selector. *** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 1% significance level. ** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 5% significance level. * denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 10% significance level.

Table 14: Accuracy Ratios and selected variables (random effects)

		OP_re	OP_re_LASSO		OP_re_ELASTIC NET	
			AIC	BIC	AIC	BIC
CDSIRs	In-sample prediction	83.62%	84.10%***	83.00%***	84.03%**	83.00%***
	Out-of-sample prediction	21.84%	25.29%***	65.08%***	25.46%***	65.08%***
	Surviving variables	268	143	45	144	45
EQIRs	In-sample prediction	93.17%	94.27%***	93.71%***	93.31%***	93.98%***
	Out-of-sample prediction	48.98%	82.99%***	89.59%***	12.44%***	89.26%***
	Surviving variables	268	167	95	181	87

Notes: The Table reports the Accuracy Ratios and the number of surviving variables for each model under study. "OP_re" stands for the ordered probit model with random effects. "OP_re_LASSO" is the ordered probit model with LASSO estimator and random effects. "OP_re_ELASTIC NET" stands for the ordered probit model with Elastic net estimator and random effects. "AIC" is the AIC-type tuning parameter selector. "BIC" is the BIC-type tuning parameter selector. *** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 1% significance level. ** denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 5% significance level. * denotes that the Stuart-Maxwell null hypothesis of no difference of each category between two predictions is rejected at the 10% significance level.

On-line Appendix A

Table A1: Expected signs and variables definition

Covariates	Predicted relationship	Definition
Firm-specific variables (16)		
Size (1)		
DETA	+	Logarithm of real total assets
Leverage (5)		
AE	-	Total assets/Equity
LDA	-	Long-term debt/Total assets
SDA	-	Short-term debt/Total assets
TDA	-	Total debt/Total assets
TDEBITDA	-	Total debt/Earnings before interest, taxes, depreciation, amortization, and restructuring or rent costs
Coverage (2)		
EBITINT	+	Earnings before interest and tax/Interest expenses
EBITDAINT	+	Total debt/earnings before interest, taxes, depreciation, amortization, and restructuring or rent costs/Interest expenses
Cash flow (2)		
CFOA	+	Cash flow from operating activities/Total assets
CASHEQA	+	Cash and equivalent/Total assets
Profitability (5)		
OM	+	Operating income/Net sales
ROC	+	Net income less dividends/Total capital
ROE	+	Net income/Shareholders' equity
ROA	+	Net income/Total assets
FFD	+	Funds from operations/Total debt
Liquidity (1)		
LIQ	+	Cash from operations/Liabilities
Market-driven Variables (6)		
EXRET	+	Monthly stock return-the S&P 500 index return
RSIZE	+	Firm equity value/Total market equity value
STA	-	The standard deviation of a company's monthly stock returns
BETA	-	Systematic risk in the Capital Asset Pricing Model
PD1	-	1-year default probability
PD5	-	5-year default probability
Macroeconomic Variables (11)		
RLSP	~	Return on S&P 500 index
CPFFM	~	3-month commercial paper rate
TB3	~	3-month Treasury bill rate minus federal funds rate
GS1	~	1-year constant maturity treasury rate
MB	~	Growth rate in the narrow money stock
INFL	~	Inflation rate
DLIP	~	Rate of change in industrial production
DLGDP	~	Real GDP growth
CFNA	~	Average Chicago Fed National Activity Index
UNRATE	~	Average unemployment rate
VIX	~	The Chicago Board Options Exchange volatility index

Notes: "+" indicates that the Market Implied Ratings would improve if the covariates rise. "-" indicates that the Market Implied Ratings would worsen if the covariates rise. "~" indicates uncertainty in the sign.

On-line Appendix B

We cross tabulate predicted against observed CDSIRs outcomes in contingency Table B1 to B12 for the in-sample prediction.

Table B1: In-sample Prediction in Ordered Probit Model in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	64	169	1	0	0	0	0	234
AA	34	879	71	11	0	0	0	995
A	0	45	1591	69	0	0	0	1705
BBB	0	1	67	1804	23	0	0	1895
BB	0	0	0	50	783	8	0	841
B	0	0	0	0	11	133	4	148
Below CCC	0	0	0	0	0	5	2	7
Total	98	1094	1730	1934	817	146	6	5825

AR= 90.23%

CP= 69.29%

Table B2: In-sample Prediction in Ordered Probit Model with LASSO by AIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	59	174	1	0	0	0	0	234
AA	26	887	70	12	0	0	0	995
A	0	45	1592	68	0	0	0	1705
BBB	0	1	67	1805	22	0	0	1895
BB	0	0	0	53	781	7	0	841
B	0	0	0	0	11	135	2	148
Below CCC	0	0	0	0	0	6	1	7
Total	85	1107	1730	1938	814	148	3	5825

AR= 90.30%

CP= 66.89%

Table B3: In-sample Prediction in Ordered Probit Model with LASSO by BIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	6	227	1	0	0	0	0	234
AA	6	905	56	28	0	0	0	995
A	0	41	1567	97	0	0	0	1705
BBB	0	1	60	1813	22	0	0	1895
BB	0	0	0	55	779	7	0	841
B	0	0	0	0	11	137	0	148
Below CCC	0	0	0	0	0	7	0	7
Total	12	1174	1684	1993	812	151	0	5825

AR= 89.39%

CP= 61.04%

Table B4: In-sample Prediction in Ordered Probit Model with Elastic Net by AIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	60	173	1	0	0	0	0	234
AA	27	886	70	12	0	0	0	995
A	0	45	1592	68	0	0	0	1705
BBB	0	1	67	1805	22	0	0	1895
BB	0	0	0	53	781	7	0	841
B	0	0	0	0	11	135	2	148
Below CCC	0	0	0	0	0	6	1	7
Total	87	1105	1730	1938	814	148	3	5825

AR= 90.30%

CP= 66.95%

Table B5: In-sample Prediction in Ordered Probit Model with Elastic Net by BIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	6	227	1	0	0	0	0	234
AA	6	905	56	28	0	0	0	995
A	0	41	1567	97	0	0	0	1705
BBB	0	0	60	1813	22	0	0	1895
BB	0	0	0	55	779	7	0	841
B	0	0	0	0	11	137	0	148
Below CCC	0	0	0	0	0	7	0	7
Total	12	1173	1684	1993	812	151	0	5825

AR= 89.39%

CP= 61.05%

Table B6: In-sample Prediction in Continuation Ratio Model with LASSO by AIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	40	193	1	0	0	0	0	234
AA	18	893	61	23	0	0	0	995
A	0	42	1,569	94	0	0	0	1705
BBB	0	0	60	1,813	22	0	0	1895
BB	0	0	0	55	779	7	0	841
B	0	0	0	0	11	137	0	148
Below CCC	0	0	0	0	0	7	0	7
Total	58	1128	1691	1985	812	151	0	5825

AR= 89.80%

CP= 63.29%

Table B7: In-sample Prediction in Continuation Ratio Model with LASSO by BIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	19	214	1	0	0	0	0	234
AA	14	897	55	29	0	0	0	995
A	0	41	1567	97	0	0	0	1705
BBB	0	0	60	1813	22	0	0	1895
BB	0	0	0	55	779	7	0	841
B	0	0	0	0	11	137	0	148
Below CCC	0	0	0	0	0	7	0	7
Total	33	1152	1683	1994	812	151	0	5825

AR= 89.48%

CP= 61.84%

Table B8: In-sample Prediction in Continuation Ratio Model with Elastic Net by AIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	33	200	1	0	0	0	0	234
AA	21	890	61	23	0	0	0	995
A	0	41	1569	95	0	0	0	1705
BBB	0	0	60	1813	22	0	0	1895
BB	0	0	0	55	779	7	0	841
B	0	0	0	0	11	137	0	148
Below CCC	0	0	0	0	0	7	0	7
Total	54	1131	1691	1986	812	151	0	5825

AR= 89.63%

CP= 62.74%

Table B9: In-sample Prediction in Continuation Ratio Model with Elastic Net by BIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	19	214	1	0	0	0	0	234
AA	14	897	56	28	0	0	0	995
A	0	41	1567	97	0	0	0	1705
BBB	0	0	60	1813	22	0	0	1895
BB	0	0	0	55	779	7	0	841
B	0	0	0	0	11	137	0	148
Below CCC	0	0	0	0	0	7	0	7
Total	33	1152	1684	1993	812	151	0	5825

AR= 89.48%

CP= 61.84%

Table B10: In-sample Prediction in Principal Component Analysis_16 PCs with Ordered Probit model in CDSIRs

Actual CDSIRs	Predicted CDSIRs							Total
	AAA	AA	A	BBB	BB	B	Below CCC	
AAA	37	156	40	1	0	0	0	234
AA	35	271	446	243	0	0	0	995
A	3	200	774	721	7	0	0	1705
BBB	5	33	536	1,255	64	2	0	1895
BB	0	1	55	433	304	48	0	841
B	0	0	5	39	66	35	3	148
Below CCC	0	0	0	1	3	3	0	7
Total	80	661	1856	2693	444	88	3	5825

AR= 45.94%

Commented [Office1]: Should I add CP ratio in following tables related PCA?

Table B11: In-sample Prediction in Principal Component Analysis_25 PCs with Ordered Probit model in CDSIRs

Actual CDSIRs	Predicted CDSIRs							Total
	AAA	AA	A	BBB	BB	B	Below CCC	
AAA	31	186	17	0	0	0	0	234
AA	42	440	420	93	0	0	0	995
A	1	184	873	626	21	0	0	1705
BBB	0	9	619	1188	74	5	0	1895
BB	0	0	9	356	417	57	2	841
B	0	0	0	19	81	43	5	148
Below CCC	0	0	0	0	4	3	0	7
Total	74	819	1938	2282	597	108	7	5,825

AR= 51.36%

Table B12: In-sample Prediction in Principal Component Analysis_45 PCs with Ordered Probit model in CDSIRs

Actual CDSIRs	Predicted CDSIRs							Total
	AAA	AA	A	BBB	BB	B	Below CCC	
AAA	47	184	3	0	0	0	0	234
AA	70	483	397	45	0	0	0	995
A	2	109	1027	550	17	0	0	1705
BBB	0	17	582	1240	54	2	0	1895
BB	0	0	10	245	514	68	4	841
B	0	0	0	37	77	32	2	148
Below CCC	0	0	0	0	4	3	0	7
Total	119	793	2019	2117	666	105	6	5825

AR= 57.39%

We cross tabulate predicted against observed CDSIRs outcomes in contingency Table B13 to B24 for the out-of-sample prediction.

Table B13: Out-of-sample Prediction in Ordered Probit Model in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	9	14	4	1	23	22	11	84
AA	24	38	51	8	43	48	22	234
A	1	35	50	65	8	42	45	246
BBB	0	0	36	82	58	40	139	355
BB	0	0	0	10	37	14	76	137
B	0	0	0	0	5	13	34	52
Below CCC	0	0	0	0	0	3	20	23
Total	34	87	141	166	174	182	347	1131

AR= 22.02%

CP= 18.22%

Table B14: Out-of-sample Prediction in Ordered Probit Model with LASSO by AIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	1	20	15	22	20	6	0	84
AA	0	62	64	49	39	20	0	234
A	0	1	94	71	45	33	2	246
BBB	0	0	1	107	109	75	63	355
BB	0	0	0	0	57	20	60	137
B	0	0	0	0	1	17	34	52
Below CCC	0	0	0	0	0	4	19	23
Total	1	83	174	249	271	175	178	1131

AR= 31.56%

CP= 25.49%

Table B15: Out-of-sample Prediction in Ordered Probit Model with LASSO by BIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	0	84	0	0	0	0	0	84
AA	0	212	22	0	0	0	0	234
A	0	12	224	10	0	0	0	246
BBB	0	0	12	340	3	0	0	355
BB	0	0	0	5	132	0	0	137
B	0	0	0	0	4	48	0	52
Below CCC	0	0	0	0	0	23	0	23
Total	0	308	258	355	139	71	0	1131

AR= 84.53%

CP= 61.01%

Table B16: Out-of-sample Prediction in Ordered Probit Model with Elastic Net by AIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	1	20	15	22	20	6	0	84
AA	0	62	64	49	39	20	0	234
A	0	1	94	71	45	33	2	246
BBB	0	0	1	107	109	75	63	355
BB	0	0	0	0	57	20	60	137
B	0	0	0	0	1	17	34	52
Below CCC	0	0	0	0	0	4	19	23
Total	1	83	174	249	271	175	178	1131

AR= 31.56%

CP= 25.49%

Table B17: Out-of-sample Prediction in Ordered Probit Model with Elastic Net by BIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	0	84	0	0	0	0	0	84
AA	0	212	22	0	0	0	0	234
A	0	12	224	10	0	0	0	246
BBB	0	0	12	340	3	0	0	355
BB	0	0	0	5	132	0	0	137
B	0	0	0	0	4	48	0	52
Below CCC	0	0	0	0	0	23	0	23
Total	0	308	258	355	139	71	0	1131

AR= 84.53%

CP= 61.01%

Table B18: Out-of-sample Prediction in Continuation Ratio Model with LASSO by AIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	3	42	39	0	0	0	0	84
AA	0	137	89	8	0	0	0	234
A	0	2	170	73	1	0	0	246
BBB	0	0	6	218	131	0	0	355
BB	0	0	0	2	93	42	0	137
B	0	0	0	0	1	32	19	52
Below CCC	0	0	0	0	0	11	12	23
Total	3	181	304	301	226	85	31	1131

AR=58.80%

CP=45.70%

Table B19: Out-of-sample Prediction in Continuation Ratio Model with LASSO by BIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	2	81	1	0	0	0	0	84
AA	0	206	28	0	0	0	0	234
A	0	12	215	19	0	0	0	246
BBB	0	0	10	339	6	0	0	355
BB	0	0	0	4	133	0	0	137
B	0	0	0	0	4	46	2	52
Below CCC	0	0	0	0	0	17	6	23
Total	2	299	254	362	143	63	8	1131

AR= 83.73%

CP= 64.16%

Table B20: Out-of-sample Prediction in Continuation Ratio Model with Elastic Net by AIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	3	46	35	0	0	0	0	84
AA	0	142	87	5	0	0	0	234
A	0	2	173	70	1	0	0	246
BBB	0	0	6	227	122	0	0	355
BB	0	0	0	2	98	37	0	137
B	0	0	0	0	1	34	17	52
Below CCC	0	0	0	0	0	11	12	23
Total	3	190	301	304	222	82	29	1131

AR= 60.92%

CP= 47.94%

Table B21: Out-of-sample Prediction in Continuation Ratio Model with Elastic Net by BIC-typing Tuning Parameter Selector in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	2	81	1	0	0	0	0	84
AA	0	209	25	0	0	0	0	234
A	0	12	218	16	0	0	0	246
BBB	0	0	10	341	4	0	0	355
BB	0	0	0	5	132	0	0	137
B	0	0	0	0	4	47	1	52
Below CCC	0	0	0	0	0	19	4	23
Total	2	302	254	362	140	66	5	1131

AR= 84.26%

CP= 63.42%

Table B22: Out-of-sample Prediction in Principal Component Analysis_16 PCs with Ordered Probit model in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	15	47	21	1	0	0	0	84
AA	2	85	122	25	0	0	0	234
A	4	24	145	72	1	0	0	246
BBB	0	10	124	211	10	0	0	355
BB	0	0	12	70	53	2	0	137
B	0	0	2	11	31	8	0	52
Below CCC	0	0	0	3	13	7	0	23
Total	21	166	426	393	108	17	0	1131

AR= 45.71%

Table B23: Out-of-sample Prediction in Principal Component Analysis_25 PCs with Ordered Probit model in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	67	17	0	0	0	0	0	84
AA	89	130	15	0	0	0	0	234
A	45	98	81	22	0	0	0	246
BBB	46	114	126	64	5	0	0	355
BB	4	29	21	64	18	1	0	137
B	0	5	5	17	22	2	1	52
Below CCC	0	0	0	5	10	8	0	23
Total	251	393	248	172	55	11	1	1131

AR= 32.01%

Table B24: Out-of-sample Prediction in Principal Component Analysis_45 PCs with Ordered Probit model in CDSIRs

Actual CDSIRs	Predicted CDSIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	6	42	29	7	0	0	0	84
AA	0	77	108	49	0	0	0	234
A	0	12	99	131	4	0	0	246
BBB	0	2	44	237	70	2	0	355
BB	0	0	0	29	87	18	3	137
B	0	0	0	6	31	11	4	52
Below CCC	0	0	0	0	8	8	7	23
Total	6	133	280	459	200	39	14	1131

AR= 46.33%

We cross tabulate predicted against observed EQIRs outcomes in contingency Table B25 to B36 for the in-sample prediction.

Table B25: In-sample Prediction in Ordered Probit Model in EQIRs

Actual EQIRs	Predicted EQIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	3	12	0	0	0	0	0	15
AA	0	409	32	0	0	0	0	441
A	0	19	2492	115	0	0	0	2626
BBB	0	0	70	2915	56	0	0	3041
BB	0	0	0	46	1273	14	0	1333
B	0	0	0	0	14	173	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	3	440	2594	3076	1343	187	0	7643

AR= 95.05%

CP= 78.30%

Table B26: In-sample Prediction in Ordered Probit Model with LASSO by AIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	2	13	0	0	0	0	0	15
AA	0	410	31	0	0	0	0	441
A	0	16	2495	115	0	0	0	2626
BBB	0	0	72	2910	59	0	0	3041
BB	0	0	0	50	1269	14	0	1333
B	0	0	0	0	12	175	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	2	439	2598	3075	1340	189	0	7643

AR=95.00%

CP=77.16%

Table B27: In-sample Prediction in Ordered Probit Model with LASSO by BIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	0	15	0	0	0	0	0	15
AA	0	412	29	0	0	0	0	441
A	0	18	2492	116	0	0	0	2626
BBB	0	0	84	2895	62	0	0	3041
BB	0	0	0	55	1262	16	0	1333
B	0	0	0	0	11	176	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	0	445	2605	3066	1335	192	0	7643

AR=94.69%

CP=74.46%

Table B28: In-sample Prediction in Ordered Probit Model with Elastic Net by AIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	2	13	0	0	0	0	0	15
AA	0	410	31	0	0	0	0	441
A	0	16	2498	112	0	0	0	2626
BBB	0	0	71	2911	59	0	0	3041
BB	0	0	0	49	1269	15	0	1333
B	0	0	0	0	12	175	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	2	439	2600	3072	1340	190	0	7643

AR=95.05%

CP=77.19%

Table B29: In-sample Prediction in Ordered Probit Model with Elastic Net by BIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	0	15	0	0	0	0	0	15
AA	0	412	29	0	0	0	0	441
A	0	20	2490	116	0	0	0	2626
BBB	0	0	92	2888	61	0	0	3041
BB	0	0	0	58	1259	16	0	1333
B	0	0	0	0	12	175	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	0	447	2611	3062	1332	191	0	7643

AR=94.52%

CP=74.25%

Table B30: In-sample Prediction in Continuation Ratio Model with LASSO by AIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	3	12	0	0	0	0	0	15
AA	0	409	32	0	0	0	0	441
A	0	15	2501	110	0	0	0	2626
BBB	0	0	72	2914	55	0	0	3041
BB	0	0	0	47	1273	13	0	1333
B	0	0	0	0	12	175	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	3	436	2605	3071	1340	188	0	7643

AR=95.19%

CP=78.58%

Table B31: In-sample Prediction in Continuation Ratio Model with LASSO by BIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	0	15	0	0	0	0	0	15
AA	0	412	29	0	0	0	0	441
A	0	15	2495	116	0	0	0	2626
BBB	0	0	82	2897	62	0	0	3041
BB	0	0	0	55	1263	15	0	1333
B	0	0	0	0	11	176	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	0	442	2606	3068	1336	191	0	7643

AR=94.77%

CP=74.51%

Table B32: In-sample Prediction in Continuation Ratio Model with Elastic Net by AIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	3	12	0	0	0	0	0	15
AA	0	409	32	0	0	0	0	441
A	0	16	2497	113	0	0	0	2626
BBB	0	0	70	2918	53	0	0	3041
BB	0	0	0	49	1272	12	0	1333
B	0	0	0	0	13	174	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	3	437	2599	3080	1338	186	0	7643

AR=95.16%

CP=78.45%

Table B33: In-sample Prediction in Continuation Ratio Model with Elastic Net by BIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	0	15	0	0	0	0	0	15
AA	0	412	29	0	0	0	0	441
A	0	16	2494	116	0	0	0	2626
BBB	0	0	83	2896	62	0	0	3041
BB	0	0	0	56	1263	14	0	1333
B	0	0	0	0	12	175	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	0	443	2606	3068	1337	189	0	7643

AR=94.73%

CP=74.39%

Table B34: In-sample Prediction in Principal Component Analysis_15 PCs with Ordered Probit model in EQIRs

Actual EQIRs	Predicted EQIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	0	14	1	0	0	0	0	15
AA	3	152	278	8	0	0	0	441
A	0	148	1651	817	10	0	0	2626
BBB	0	4	559	2318	160	0	0	3041
BB	0	0	0	612	674	47	0	1,333
B	0	0	0	7	113	67	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	3	318	2489	3762	957	114	0	7643

AR= 63.61%

Table B35: In-sample Prediction in Principal Component Analysis_23 PCs with Ordered Probit model in EQIRs

Actual EQIRs	Predicted EQIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	0	15	0	0	0	0	0	15
AA	0	174	258	9	0	0	0	441
A	1	102	1929	589	5	0	0	2626
BBB	0	1	433	2440	164	3	0	3041
BB	0	0	0	408	864	61	0	1333
B	0	0	0	5	86	96	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	1	292	2620	3451	1119	160	0	7643

AR= 72.00%

Table B36: In-sample Prediction in Principal Component Analysis_41 PCs with Ordered Probit model in EQIRs

Actual EQIRs	Predicted EQIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	0	15	0	0	0	0	0	15
AA	0	286	154	1	0	0	0	441
A	0	74	2052	500	0	0	0	2626
BBB	0	0	393	2465	183	0	0	3041
BB	0	0	0	272	990	71	0	1333
B	0	0	0	9	91	87	0	187
Below CCC	0	0	0	0	0	0	0	0
Total	0	375	2599	3247	1264	158	0	7643

AR= 76.93%

We cross tabulate predicted against observed EQIRs outcomes in contingency Table B37 to B48 for the out-of-sample prediction.

Table B37: Out-of-sample Prediction in Ordered Probit Model in EQIRs

Actual EQIRs	Predicted EQIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	0	1	0	0	0	0	0	1
AA	7	51	45	11	0	0	0	114
A	28	13	248	172	38	0	0	499
BBB	11	26	17	185	123	34	0	396
BB	0	9	14	11	78	40	0	152
B	0	0	0	1	2	17	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	46	100	324	380	241	91	0	1182

AR= 48.98%

CP= 35.49%

Table B38: Out-of-sample Prediction in Ordered Probit Model with LASSO by AIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	0	1	0	0	0	0	0	1
AA	0	93	21	0	0	0	0	114
A	0	7	434	58	0	0	0	499
BBB	0	0	21	334	41	0	0	396
BB	0	0	0	17	121	14	0	152
B	0	0	0	0	0	20	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	101	476	409	162	34	0	1182

AR=84.77%

CP=66.50%

Table B39: Out-of-sample Prediction in Ordered Probit Model with LASSO by BIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	0	1	0	0	0	0	0	1
AA	0	109	5	0	0	0	0	114
A	0	8	475	16	0	0	0	499
BBB	0	0	39	349	8	0	0	396
BB	0	0	0	26	124	2	0	152
B	0	0	0	0	2	18	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	118	519	391	134	20	0	1182

AR=90.95%

CP=70.10%

Table B40: Out-of-sample Prediction in Ordered Probit Model with Elastic Net by AIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	0	1	0	0	0	0	0	1
AA	0	85	29	0	0	0	0	114
A	0	5	411	82	1	0	0	499
BBB	0	0	18	319	59	0	0	396
BB	0	0	0	16	121	15	0	152
B	0	0	0	0	0	20	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	91	458	417	181	35	0	1182

AR=80.88%

CP=63.42%

Table B41: Out-of-sample Prediction in Ordered Probit Model with Elastic Net by BIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	0	1	0	0	0	0	0	1
AA	0	109	5	0	0	0	0	114
A	0	9	474	16	0	0	0	499
BBB	0	0	39	349	8	0	0	396
BB	0	0	0	27	123	2	0	152
B	0	0	0	0	2	18	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	119	518	392	133	20	0	1182

AR=90.78%

CP=69.93%

Table B42: Out-of-sample Prediction in Continuation Ratio Model with LASSO by AIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	0	1	0	0	0	0	0	1
AA	0	93	21	0	0	0	0	114
A	0	6	436	57	0	0	0	499
BBB	0	0	21	335	40	0	0	396
BB	0	0	0	15	126	11	0	152
B	0	0	0	0	0	20	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	100	478	407	166	31	0	1182

AR=85.45%

CP=67.29%

Table B43: Out-of-sample Prediction in Continuation Ratio Model with LASSO by BIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	0	1	0	0	0	0	0	1
AA	0	109	5	0	0	0	0	114
A	0	8	475	16	0	0	0	499
BBB	0	0	39	349	8	0	0	396
BB	0	0	0	25	125	2	0	152
B	0	0	0	0	2	18	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	118	519	390	135	20	0	1182

AR=91.03%

CP=70.23%

Table B44: Out-of-sample Prediction in Continuation Ratio Model with Elastic Net by AIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	0	1	0	0	0	0	0	1
AA	0	86	28	0	0	0	0	114
A	0	7	410	82	0	0	0	499
BBB	0	0	18	321	57	0	0	396
BB	0	0	0	18	119	15	0	152
B	0	0	0	0	1	19	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	94	456	421	177	34	0	1182

AR=80.80%

CP=62.39%

Table B45: Out-of-sample Prediction in Continuation Ratio Model with Elastic Net by BIC-typing Tuning Parameter Selector in EQIRs

Actual EQIRs	Predicted EQIRs						Below	Total
	AAA	AA	A	BBB	BB	B	CCC	
AAA	0	1	0	0	0	0	0	1
AA	0	109	5	0	0	0	0	114
A	0	8	475	16	0	0	0	499
BBB	0	0	39	349	8	0	0	396
BB	0	0	0	26	124	2	0	152
B	0	0	0	0	2	18	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	118	519	391	134	20	0	1182

AR=90.95%

CP=70.10%

Table B46: Out-of-sample Prediction in Principal Component Analysis_15 PCs with Ordered Probit model in EQIRs

Actual EQIRs	Predicted EQIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	0	1	0	0	0	0	0	1
AA	0	51	61	2	0	0	0	114
A	0	18	357	119	5	0	0	499
BBB	0	0	84	269	43	0	0	396
BB	0	0	7	80	65	0	0	152
B	0	0	0	1	12	7	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	0	70	509	471	125	7	0	1182

AR= 63.37%

Table B47: Out-of-sample Prediction in Principal Component Analysis_23 PCs with Ordered Probit model in EQIRs

Actual EQIRs	Predicted EQIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	0	1	0	0	0	0	0	1
AA	4	76	34	0	0	0	0	114
A	2	120	355	22	0	0	0	499
BBB	0	16	206	173	1	0	0	396
BB	0	0	32	83	36	1	0	152
B	0	0	0	1	12	7	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	6	213	627	279	49	8	0	1182

AR= 54.74%

Table B48: Out-of-sample Prediction in Principal Component Analysis_41 PCs with Ordered Probit model in EQIRs

Actual EQIRs	Predicted EQIRs						Below CCC	Total
	AAA	AA	A	BBB	BB	B		
AAA	0	1	0	0	0	0	0	1
AA	10	80	24	0	0	0	0	114
A	2	150	323	24	0	0	0	499
BBB	0	14	210	171	1	0	0	396
BB	0	0	25	92	35	0	0	152
B	0	0	0	4	14	2	0	20
Below CCC	0	0	0	0	0	0	0	0
Total	12	245	582	291	50	2	0	1182

AR= 51.69%

On-line Appendix C

In this section, we report the procedure for selecting tuning parameters.

In our study, we select the tuning parameter with the aid of the AIC (Akaike, 1974) and the BIC (Schwartz, 1978).

These are presented below:

$$AIC = -2 * \log(\text{maximum likelihood}) + 2 * (\text{the number of used parameters within the model})$$

and,

$$BIC = -2 * \log(\text{maximum likelihood}) + 2 * (\text{the number of used parameters within the model}) * \log(\text{the number of observations})$$

The number of used parameters within the model is the number of non-zero estimated coefficients of used parameters (Efron et al., 2004 and Zou et al., 2007). We follow the parametrization procedure of Wurm et al., 2017.

The exact algorithm for the LASSO tuning parameter based on the AIC and the BIC criteria is presented below:

1. Estimate the intercept-only model by maximum likelihood (see, equation (4) for the OP with LASSO or equation (9) for the CR with LASSO).
2. Select a sequence of tuning parameters λ_i ($i = 1$ to n).
For each tuning parameter λ_i , there exists a set of selected independent variables and corresponding estimated coefficients. In other words, different potential models including different selected predictors controlled by the sequence of tuning parameters $\lambda_1, \lambda_2, \dots, \lambda_n$ are constructed.
3. Calculate the AIC or BIC for each potential model.
4. Choose the models with minimum value of AIC or BIC from the aforementioned series. These two models (one based on AIC-type tuning parameter selector and one based on BIC-type tuning parameter selector) are applied in our study.

Similarly, for the Elastic net estimator we follow Wurm et al., 2017:

1. Estimate the intercept-only model by maximum likelihood (see, equation (6) for the OP with Elastic net or equation (11) for the CR with Elastic net).
2. Select a sequence of tuning parameters $\alpha_1, \alpha_2, \dots, \alpha_{10}$.
3. For each α_i ($i = 1$ to 10), a sequence of tuning parameters λ_i ($i = 1$ to n) is selected. For each tuning parameter λ_i , there is a set of selected independent variables and the corresponding estimated coefficients. For each one of these models, we generate an AIC and BIC value.
4. The min AIC and BIC values are assigned to the corresponding α_i .
5. The α_i ($i = 1$ to 10) with the minimum AIC and the one with the minimum BIC value are selected. These two models are applied in our study.