Measuring the Sensitivity of Credit Ratings to Macroeconomic Indicators across Business Sectors

A Study on the US Market

Hazem Ashour
Oliwer Silfverberg

Abstract This paper examines the sensitivity of credit ratings to macroeconomic indicators, with a focus on monetary policy tools, across eleven business sectors in the United States. We examine the data using ordered probit and random forest models, taking into account firm fundamentals that measure the business and the financial risk, in addition to the macroeconomic indicators. Employing quarterly credit ratings by Standard & Poor's for 299 American companies from 1985 till 2016, we find that firm-specific risk factors commonly have more explanatory power in determining credit rating classes. In addition, the findings suggest that business sectors respond differently to changes in the macroeconomic indicators, with some variables displaying high significance while others being indubitably insignificant.

Lund University School of Economics and Management
Master's Thesis (15 credits ECTS)
May 2019
Supervisor: Jens Forssbaeck

Acknowledgements

We would like to express our sincere gratitude to our supervisor Prof. Jens Forssbaeck and co-supervisor Dinh-Vinh Vo for their continuous support during our master's thesis, and immense knowledge. We would also like to thank our parents and families for supporting us throughout writing this thesis and our lives in general.

List of Abbreviations

Abbreviation	Meaning
ADF	Augmented Dickey-Fuller
AIC	Akaike Information Criterion
CD	Consumer Discretionary
CFO	Cash Flow From Operations
CS	Consumer Staples
E	Energy
F	Financials
FAVAR	Factor Augmented Vector Auto Regressive
Fed	The Federal Reserve System
FFO	Funds From Operations
FOMC	Federal Open Market Committee
FRED	Federal Reserve Economic Data
GDP	Gross Domestic Product
GICS	Global Industry Classification Standard
HC	Health Care
I	Industrials
IT	Information Technology
M	Materials
MDI	Mean Decrease Impurity
OOB	Out-Of-Bag
PCE	Personal Consumption Expenditures
PDP	Partial Dependence Plot
RE	Real Estate
ROC	Receiver Operating Characterisitc
SBC	Schwartz Bayesian Criterion
TS	Telecommunication Services
U	Utilities
WRDS	Wharton Research Database Services

Contents

List of Figures

List of Tables

1

Introduction

Credit risk, a key component in investment decisions and asset pricing models, has been on the research agenda of finance scholars for more than three decades. Duffie and Singleton (2003) define credit risk as the risk of a change in the market value of a security derived by a decline in its credit quality. In the process of raising capital, for example, banks and financial institutions require detailed information about the obligor's ability to fulfill its financial obligations in the future. One way to measure the creditworthiness of an obligor is by credit rating or credit score. Credit ratings, in general, represent the rating agencies' opinions of the creditworthiness of a borrower based on quantitative and qualitative factors that incorporate the obligor's business and financial risk (Trueck and Racev, 2009). In the process of assigning a rating, historical and current information regarding loans, debt repayments, missing payments, and defaults, if any, are analyzed. Besides, the obligor's future potential to honor its financial obligations is taken into consideration. Credit ratings transition adjusting to changes in the obligor's debt issues, overall firm performance, industry and country risk (see SPratings (2018) for a detailed description of credit ratings, rating methodologies, and categories). Moreover, rating agencies take into account "economic momentum indicators" (S&P Global, 2019*b*) such as disposable personal income per capita, industrial production, term spread, labor market indicators, among others, to adjust for forecasted changes in the business cycle and the state of the economy.

In this paper, we examine the impact of macroeconomic indicators on credit ratings across eleven business sectors in the United States, classified according to the Global Industry Classification System (GICS). We attempt to replicate S&P's process of assigning ratings, by analyzing the firm fundamentals that are assumed to gauge the firm's business and financial risk level, and we evaluate whether, in excess of these firm fundamentals, credit ratings are influenced by changes in the macroeconomic indicators across the eleven business sectors. We eventually evaluate which of these sectors is more sensitive to changes in the macroeconomic indicators and to which degree. The purpose of the paper is not to investigate whether ratings are procyclical, which has been studied by many researchers in the past few decades, but rather measure the sensitivity of the macroeconomic indicators across the eleven business sectors, in an attempt to classify the credit ratings across these sectors. We utilize a panel data consisting of credit ratings, firm fundamentals, and time series data of macroeconomic indicators. We conduct ordered probit and random forest models to forecast the firms' credit rating contingent on the firm-specific factors and the macroeconomic indicators. We compare the results between the two models as a robustness check and find that, although the results exhibit some variation between the two models, the models seem to agree with each other regarding the intuitive reliable results.

We find that, in general, firm-specific risk factors, represented by beta, idiosyncratic risk, profitability, and leverage ratios, have more explanatory power in determining credit rating classes across the eleven sectors. We, however, find that business sectors respond differently to changes in macroeconomic indicators, with some macroeconomic indicators exhibiting high significance in determining the rating classes across some sectors while others being indubitably insignificant. Although the effective Federal funds rate, for example, is found to be the most important macroeconomic indicator across ten out of eleven sectors, its importance in absolute terms displays huge variation. On the other hand, the trimmed mean PCE inflation rate appears to be the most important macroeconomic indicator in one business sector while being almost insignificant in the other ten sectors. The findings suggest, in general, that the important macroeconomic indicators can be easily used to classify the credit ratings, especially in the cyclical business sectors, while on the other hand, ratings in defensive sectors are better explained by firm fundamentals.

The structure of the paper is as follows. Section two provides a brief literature review of credit ratings, ratings' transitions, and the cyclicality of credit ratings. Section three describes the data, defines the dependent and the independent variables, and explains how the data is classified. Section four describes the ordered probit and random forest models, assumptions, sampling techniques, and how the data is used in the models. Section five reports the empirical findings, compares them across the business sectors, and interprets the results per each business sector in real life applications. Section six concludes.

2

Literature Review

Studying the behavior of ratings' transition has become an area of interest for many researchers for the past few decades. Many scholars attempt to research whether ratings' transitions are pro-cyclical. Others investigate the variation of rating fundamentals between investment grade and speculative grade bonds. One of the first papers to explore this area of research is Horrigan (1966), which utilizes financial ratios to classify the credit ratings for corporate bonds. Horrigan employs profitability, solvency, and asset efficiency ratios such as long-term solvency, short-term capital-turnover, long-term capital-turnover, and profit-margin, among others, and succeeds in forecasting one-half of the sample's ratings. Fama (1986) which, through comparing the expected return on bills and private-issuer money market securities, provides evidence of time-varying term premiums and default premiums in the expected returns on private-issuer securities for all maturities.

A pioneering paper in understanding the behavior of ratings' transition is Amato and Furfine (2004), which documents that credit ratings are pro-cyclical but suggest, based on empirical findings, that this finding is driven by cyclical changes in the firms' business and financial risk factors, and not to cycle-related changes in the rating standards. The paper also documents that lower-rated firms are subject to more intensive monitoring at critical points in the business cycle, particularly recessions. Altman and Kao (1992) suggest that there is serial autocorrelation in ratings when the initial rating change is a downgrade, but the autocorrelation is not evident when the initial rating change is an upgrade. Nickell et al. (2000) examine the ceteris paribus dependence of ratings migration probability on industries, countries, and stages of the business cycle using an ordered probit approach, and report variation of transition matrices between financial institutions and industrial sector, and the US v.s non-US obligors. They also report that the business cycle dimension is the most important in explaining the transition probabilities. Cheung (1996) conducts an ordered probit analysis on the 1969-70 to 1994-95 provincial data in Canada and finds that the debt-to-GDP ratio, the employment ratio, and provincial GDP as a share of total Canadian GDP, among other indicators, do affect the conditional distribution for provincial credit ratings, with the first variable affecting the credit rating non-linearly at different degrees for the nine provinces in the sample.

Guo and Bruneau (2014) investigate the impact of monetary policy on corporate default risk to identify the most significant macroeconomic indicators using a FAVAR model examining both US and European markets. In an attempt to answer the question of whether the subprime crisis was to some extent triggered by the rise in the FED rate during 2005, they document that common macroeconomic indicators explain only "19%-21% of the default rates of US corporate bonds, and 8% of default rates of corporate bonds (21% in volume) in the euro area", concluding that the obligor's specific risk factors explain the more significant portion of default probabilities. In both the US and European markets, they find that production, employment rates, and stock market indexes are the three macroeconomic indicators that contribute the most to default rates. They also document higher sensitivity to macroeconomic volatility among speculative-grade corporate bonds, and slightly more sensitive in the US than Europe. Furthermore, they find a higher sensitivity among the Financial and Real Estate sectors to changes in interest rates, but report that monetary policy tools overall are not the most reliable macroeconomic indicators to explain credit default probabilities.

Other researchers argue that countercyclical amendments in supervisory requirements and the stimulation of improved risk measurement could and should be used effectively to alleviate the impact of cyclicality and macroeconomic volatility on the financial stability of firms and households (Furfine and Lowe, 2001). The paper mentions banking regulations such as capital requirements and provisions as regulatory tools that could be used to minimize the effect of cyclicality on the financial sector adequately. They also discuss the importance of establishing supervisory rules and using monetary policy tools to make the system more immune to misevaluations of risk, and reduce the impact of financial volatility.

Dudian and Popa (2012) investigate the relationship between sovereign credit ratings and the gross domestic product in Europe, due to the limited international research regarding the relevance of sovereign risk for big international investors to the financial market and economic growth. They propose a more refined rating scale with three more categories: positive, stable, and negative within each rating class. They perform a fixed effect panel data regression with the economic growth as the dependent variable and the refined rating scale as the independent variable with observations from 10 countries in Central and Eastern Europe during the period 1996-2010. They find a negative correlation within all the ten countries between the economic growth and credit ratings, which validates the European Commission's conclusion of the procyclicality of the sovereign rating.

As evident from the review, previous literature has found that credit ratings are in fact procyclical and that autocorrelation exists in credit ratings in times of economic recessions. However, ratings are more influenced by financial ratios and firm fundamentals. Besides, speculative-grade bonds and newly rated bonds are more sensitive to changes in the economic conditions and business cycles than investment grade bonds and bonds that have stayed in the same rating class for over a year. We develop this by analyzing long-term issuer credit ratings for 299 companies across business sectors in the US market, using both firm-specific risk factors and macroeconomic indicators, with a focus on monetary policy, as our explanatory variables. We evaluate the sensitivity for each of these business sectors to changes in the macroeconomic conditions and monetary policy indicators, and we classify the ratings based on these two explanatory variables. In addition, we enrich previous literature by using both ordered probit and random forest models to sidestep the issue of contradicting results in the previous literature, compare the results as a robustness check, and draw a well-informed conclusion regarding the sensitivity of the business sectors and ratings' classification.

3

Data

The paper examines the sensitivity of credit rating classes to financial risk, business risk, and macroeconomic indicators across eleven business sectors in the United States. In order to conduct the empirical tests, we construct a database consisting of three sets of data at a quarterly frequency; credit ratings, firm fundamentals, i.e., business and financial risk factors, and macroeconomic indicators.

3.1 Credit Ratings

S&P applies credit ratings to both firms (long-term and short-term issuer credit ratings) and individual debt obligations (long term and short term issue credit ratings). The issue-specific credit ratings are based on the creditworthiness of a financial obligation, a class of financial obligations or a specific financial program of an obligor, while issuer's credit ratings, on the other hand, represent a forward-looking opinion regarding the obligor's fundamental creditworthiness and measure the firm's overall capacity to fulfill its financial obligations. Besides, the issuer's credit ratings reflect the likelihood of default, taking into account all financial obligations (see Standard & Poor's (2018), for detailed comparisons). We choose to examine the long-term issuer credit ratings for the reasons mentioned above.

The sample consists of quarterly domestic long-term issuer credit ratings for 299 North American firms for more than 30 years, starting from January 1985 to December 2016. We choose to start our sample from 1985 onwards due to major adjustments in S&P's rating methodology that took place at the time. The sample contains all rating categories across the rating scale except for the "C" class. In order to conduct both the ordered probit and the random forest models, we assign numerical values to the rating classes starting from AAA=1, AA=2, ..., and ending with D=9. The source of data on S&P long-term issuer credit ratings is Compustat – North America, extracted from Wharton Research Database Services (WRDS). Compustat North America, a part of Capital IQ from Standard Poor's, is a database of U.S. and Canadian fundamental and market information on active and inactive publicly held companies. A summary of the ratings in each of the eleven business sectors is provided

Table 3.1: The total number of ratings in each industry and rating class, 1985-2016. This table presents the total number of credit ratings in the eleven business sectors for each rating class. The ratings are S&P long-term issuer credit ratings collected from Compustat – North America, extracted from Wharton Research Database Services (WRDS) and then transformed to a numerical scale with nine categories. The sectors in each column are abbreviated in the following way: Consumer Discretionary is CD, Consumer Staples is CS, Energy is E, Financials is F, Health Care is HC, Industrials is I, Information Technology is IT, Materials is M, Real Estate is RE, Telecommunication Services is TS and Utilities is U (see abbreviations list for more details)

	CD	CS	E	F	HC	I	IT	M	RE	TS	U	Total
AAA	0	66	66	77	457	170	71	0	0	74	0	981
AA	139	529	147	409	277	803	229	235	0	48	419	3235
A	399	872	362	1682	594	2307	520	1010	298	554	1824	10422
BBB	545	522	363	855	250	2310	297	1322	751	527	2562	10304
BB	215	231	22	141	83	669	252	544	130	175	210	2672
B	225	80	0	41	53	174	173	229	0	205	28	1208
CCC	3	1	0	4	1	29	7	27	0	4	1	77
CC	3	0	0	0	0	17	0	0	0	0	4	24
D	0	4	0	1	0	10	0	3	0	1	9	28
Total	1529	2305	960	3210	1715	6489	1549	3370	1179	1588	5057	28951

in Table 3.1. The total number of ratings in the whole dataset is 28 951. The Industrials business sector is the most extensive sector with 6489 credit ratings. The two most common rating classes are A and BBB with over 10 000 observations each. The credit rating dataset only consists of 24 observations from class CC and 28 observations from class D, indicating that the overall dataset is imbalanced. For example, the Real Estate sector has 715 observed credit ratings in class BBB, but none in the classes B-D.

3.2 Firm Fundamentals

In assigning corporate credit ratings, S&P analyzes firms' business risk and financial risk (see S&P Global (2013) for a detailed description of the corporate rating criteria). The exact technique regarding how S&P combines all the risk factors is still unclear. In this paper, we attempt to replicate S&P risk corporate criteria methodology as closely as possible. As defined by S&P, business risk is built on analyzing country risk where the firm operates, market risk, organizational and managerial capabilities, competitive advantages, and disadvantages of the firm within its operating market. Brand identity, cost structure, distributional network, trademarks, and patents are all characteristics that are taken into account when assessing the competitive position of the company. Country risk, market (industry) risk, and competitive position are then combined for the assessment of the business risk. Leaning towards being a qualitative measure, rating agencies find it challenging to measure business risk, with no standard formula and plenty of possible customization.

We obtain the two measures of market risk from estimating the market model. Following Amato and Furfine (2004) and Blume et al. (1998), we separate equity risk into beta (systematic) and idiosyncratic (unsystematic) risk elements. By definition, systematic risk is the overall risk affecting all assets and securities on a large scale in an economy. Being a macroeconomic variable by nature, systematic risk cannot be reduced using diversification. Beta is a measure of stock volatility relative to the overall market risk. In the sample, beta with values between zero and one, all else equal, means the firm has low volatility towards market risk. Likewise, beta with a value of one means the firm moves in the exact direction of the market, with the same magnitude. Beta with values higher than one indicates high sensitivity of the firm to systematic risk, and moves in the same direction as the market risk with higher magnitude.

On the contrary, idiosyncratic risk is defined as the firm-specific risk and has little to no correlation with the market risk. In the sample, the market model is estimated using monthly observations. We use monthly observations due to the limited availability of daily observations. Data of beta and idiosyncratic risk is obtained from the beta suite by WRDS. Furthermore, S&P uses two financial ratios; return on capital defined as net income (loss) relative to total capital and EBITDA margin defined as earnings before interest, taxes, depreciation, and amortization relative to net sales. We utilize these two fundamentals for assessing profitability. EBITDA margins are also analyzed against historical performance to evaluate the volatility of profits.

With respect to the financial risk, S&P in accordance with their latest adjustments criteria in 2013, evaluates leverage based on two categories; core ratios and supplementary ratios. Core ratios consist of funds from operations (FFO) to debt and debt to EBITDA. We limit the core ratios category to debt relative to EBITDA only, defined as the total of the firm's financial obligations relative to earnings before interest taxes, depreciation, and amortization. This constraint is due to limited data on gains (losses) on sales of assets, which is part of the funds from operations to debt formula. In regard to the supplementary ratios, we utilize the cash flow from operations (CFO) relative to debt as a coverage ratio for assessing the financial strength of the firm and the ability to meet financial obligations. We obtain the data regarding the beta and the idiosyncratic risk from the beta suite by the Wharton Research Database Services and the financial ratios from Compustat North America - Daily.

3.3 Macroeconomic Indicators

The purpose of this paper is to evaluate whether, in excess of the aforementioned firm fundamentals that are assumed to describe the obligor's risk specific factors, credit ratings are influenced by changes in macroeconomic indicators across business sectors, and eventually evaluate which of these business sectors is more sensitive to changes in the macroeconomic indicators and to which degree. To capture the impact of the macroeconomic indicators, we choose eight independent variables. The time-series for all eight macroeconomic variables are pictured in figure 3.1 for the whole sample period 1985 - 2016.

The first is an indicator of the open market operations, the second and third are indicators of the change in the gross domestic product and unemployment rates respectively; the forth is an indicator of discount rates, the fifth indicator assesses the change in inflation and the last two gauge money supply in the economy. We also add an indicator related to the charge off rate on consumer loans, which is assumed to be a consequence of the effective Federal funds rate. We obtain the macroeconomic indicators from the Federal Reserve Economic Data (FRED), which is the database managed by the Federal Reserve Bank of St. Louis, as percentages or percentages' change.

FRED defines the Federal funds rate as the interest rate at which depository institutions, with excess reserve balance, trade federal funds (balances held at Federal Reserve Banks) with depository institutions in need of cash, on an overnight basis (FRED, 2019*a*). The weighted average for all the similar transactions is known as the effective Federal funds rate. Although technically determined by the market and the aforementioned interbank transactions, the effective Federal funds rate is regulated by the Federal Reserve through open market operations to reach the Federal funds target rate.

The Federal Reserve System (Fed) defines open market operations as "the purchase and sale of securities in the open market by a central bank" (Board of Governors of the Federal Reserve System, 2018) which is considered as the primary tool used by the Federal Open Market Committee (FOMC) for controlling monetary policy in the United States. The FOMC uses market operations to achieve a targeted Federal funds rate, which intrinsically affects most interest rates in the country, such as finance rates on personal, automobile, mortgage loans, among others.

In an expansionary monetary policy, the Fed purchases government securities such as treasury bills and treasury bonds from commercial banks, through private bond dealers, and deposits the transaction proceeds into the bank's account. The deposit, which goes into the bank's cash reserve increases the amount of money available to lend to customers, which lowers the interest rates, increasing spending level with affordable financing rates, supporting economic activity and job creation, and ultimately decreasing the effective Federal funds rate. On the other hand, in a contractionary monetary policy, the Fed sells government securities to banks and financial institutions, decreasing the banks' cash reserve and the amount of money available to lend, and ultimately raising the effective Federal funds rate.

The second macroeconomic indicator is the real gross domestic product (GDP). Real GDP is a seasonally adjusted annual rate expressed as percentage change from the preceding period (see The U.S. Bureau of

Economic Analysis (2019) for detailed description of the series.) The adjusted annual rate is equivalent to the growth rate over a year assuming a quarterly GDP constant growth rate for three more quarters or the average rate. Real GDP is defined as an inflation-adjusted indicator that factors in the final value of all goods and services produced by an economy within a well-defined geographical region in a given year expressed in the preceding year prices. GDP figures are produced by the Bureau of Economic Analysis in the United States and is reported on a quarterly basis. Real GDP percentage change has a local maximum in the sample of 7.5% in Q2 2000 and local minima of -8.4% in Q4 2008.

The third macroeconomic indicator is the unemployment rate, aged fifteen – sixty-four for all persons (not specific to a particular race or gender) in the United States. The unemployment rate is defined as the number of unemployed workers, yet actively seeking a job, expressed as a percentage of the total labor force, i.e., all persons eligible for a job in a country. The unemployment rate is a measure of the labor market performance, indicating the degree to which an economy is unable to utilize the unused workers to increase economic activity and generate profits. High unemployment rates are sign of a recession and may call for a decrease in the interest rates in the country. Likewise, low unemployment rates are a sign of economic expansion, which may be followed by inflation, and may require increasing interest rates. Despite having a direct unfavorable impact on economies, unemployment rates may have different effects on corporates in terms of both amplitude and magnitude. We discuss in detail the impact of unemployment rates on the sample's credit ratings in Section five.

Our fourth indicator is the spread between Moody's seasoned Baa corporate bond with maturities of twenty years and above, and Federal funds rate, and is expressed as a percentage at a quarterly frequency. By virtue of it representing the difference between the short-term interest rates set by the Fed and the interest rates on long term financial securities, the spread reflects the forecasts of economic conditions, as interpreted by analysts and bond market investors. Generally, the low yield spread is associated with a forecasted economic recession, indicating forecasts of less future growth of the current Fed effective short-term rates. On the contrary, a large spread is associated with economic expansion. These forecasts are explained by the ability of the spread to reflect the outlook of the monetary policy on an economy (Bonser-Neal and R Morley, 1997). For instance, tight monetary policy is associated with an increase in the short-term effective Federal fund's rate. Long-term interest rates generally rise as well, but with less magnitude resulting in a low spread, resulting ultimately in a lower economic activity and growth in the short-term.

The trimmed mean PCE (Personal Consumption Expenditures) inflation rate, the fifth indicator, is defined by FRED as an "alternative measure of core inflation in the price index for personal consumption expenditures." (FRED, 2019*c*). The trimmed mean PCE inflation provides more accurate information than the core inflation considering that it includes the prices of foods and energy. The use of a trimmed mean eliminates the effects of data outliers or fat tails that may biasedly alter the traditional mean. Computing the trimmed mean PCE inflation includes an analysis of the individual components of the personal consumption expenditures, sorting these components ascendingly from products with the most price decline to products with the most price incline, smoothing out the extreme values, and taking the weighted average of all the included components as the trimmed mean PCE inflation.

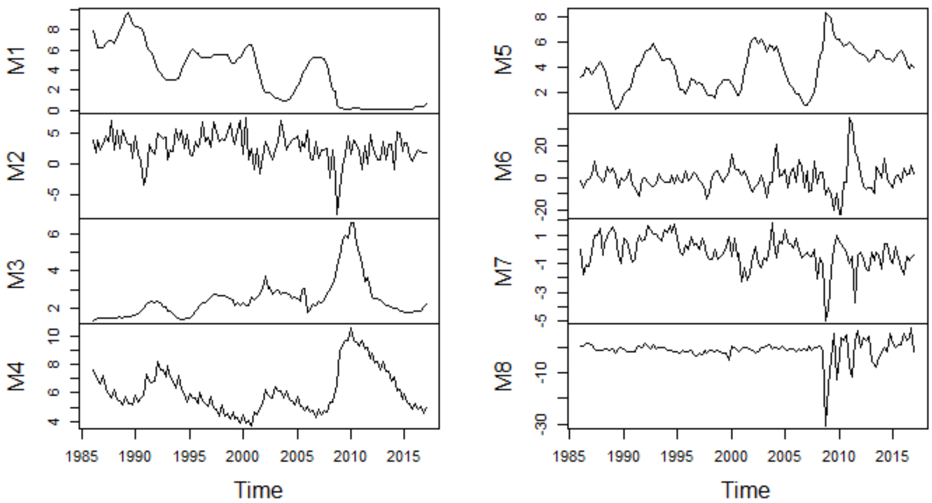


Figure 3.1: Macroeconomic indicators. Plot of the eight macroeconomic indicators, 1985-2016. The macroeconomic indicators are plotted for the whole sample of data consisting of quarterly observations collected from the Federal Reserve Bank of St. Louis database. The macroeconomic indicators are shortened to M1 up to M8 and expressed in percent, where M1 is the effective Federal funds rate, M2 is the real gross domestic product, M3 is the charge-off rate on consumer loans, M4 is the unemployment rate aged between 15 and 64, M5 is Moody's seasoned Baa corporate spread, M6 is the trimmed mean PCE inflation rate, M7 is the velocity of M2 money stock, and M8 is the M1 money multiplier.

The last two macroeconomic indicators: velocity of M2 money stock and M1 money multiplier assess money supply, providing insights regarding spending versus saving in an economy. FRED defines the velocity of money in general as the rate at which one unit of currency is exchanged within an economy to purchase domestically- produced goods and services within a given period (FRED, 2019*d*). While M1 represents the currency in circulation in an economy, such as coins, bills, travelers' checks, money held in checking accounts on demand in banks, i.e., checkable deposits, M2 on the other hand includes all the items included below M1 in addition to savings deposits, certificates of deposit, money market funds, and other time deposits. Being a less liquid, saving oriented money indicator, the velocity of M2 is a gauge of the rate at which

Table 3.2: The Augmented Dickey-Fuller test for all eight macroeconomic indicators. The Augmented Dickey-Fuller test is reported in Table 3.2 to test if the macroeconomic indicators have a unit root, against the alternative hypothesis of stationarity. In the table, the number of lags, test-statistic, and its corresponding p-value are reported.

Variable	Lags	Test-statistic	P-value
Federal Funds rate	5	-3.99	0.01
Gross Domestic Product	1	-5.08	0.01
Charge-off rate	3	-3.52	0.04
Unemployment rate	4	-4.97	0.01
Moody's Baa Spread	2	-4.13	0.01
Inflation rate	6	-5.31	0.01
Velocity of M2	1	-5.88	0.01
M1 Multiplier	3	-4.26	0.01

the economy is saving rather than spending. M1 money multiplier is the ratio of M1 relative to the St. Louis adjusted monetary base. FRED defines the adjusted monetary base as the sum of currency in circulation, plus deposits held by commercial banks at the Federal Reserve Bank, adjusted for the effects of changes in reserve requirements. Explained succinctly, M1 money multiplier represents the ratio of M1 to M1 plus M2 (FRED, 2019*b*). In the sample, we expect M1 to have a negative notch (i.e., -) attached to its coefficient, and a positive notch (i.e., +) to be attached to M2.

3.3.1 Test for Stationarity

A formal Augmented Dickey-Fuller (ADF) test is used to test whether all quarterly macroeconomic variables are stationary (see Appendix A for a more in detail description of stationarity). The ADF-test tests the null hypothesis of a unit root against the alternative hypothesis that the process is stationary. Akaike information criterion is used to determine the optimal lag length for each macroeconomic variable. The test-statistic and corresponding p-values from each ADF-test are presented in Table 3.2. The p-values are below 0.05 for all tests, indicating that we reject the null hypothesis of a unit root and conclude that all macroeconomic variables are stationary.

3.3.2 Test for Multicollinearity

By looking at the correlation matrix of the macroeconomic indicators, we can investigate whether there is any presence of a pair-wise correlation among the variables. If multicollinearity exists between more than two variables, the relationship will be more challenging to detect, and hence, the correlation matrix might miss it (see Appendix A for a more in detail description of multicollinearity). The correlation matrix for the eight macroeconomic indicators is present in Table 3.3. The highest correla-

Table 3.3: Correlation matrix for the eight macroeconomic indicators. The correlation matrix for all eight macroeconomic indicators is reported in Table 3.3. The macroeconomic indicators are shortened to M1 up to M8 and expressed in percent. M1 is the effective Federal funds rate, M2 is the real gross domestic product, M3 is the charge-off rate on consumer loans, M4 is the unemployment rate aged between 15 and 64, M5 is Moody's seasoned Baa corporate Spread, M6 is the trimmed mean inflation rate, M7 is the velocity of M2 money stock, and M8 is the M1 money Multiplier.

	M1	M2	M3	M4	M5	M6	M7	M8
M1	1							
M2	0.20	1						
M3	-0.51	-0.24	1					
M4	-0.47	-0.19	0.55	1				
M5	-0.77	-0.26	0.51	0.66	1			
M6	0.02	0.03	-0.24	-0.16	-0.12	1		
M7	0.30	0.71	-0.26	-0.07	-0.23	-0.02	1	
M8	0.02	0.31	-0.21	-0.10	-0.19	-0.01	0.25	1

tion in absolute value is between the Federal funds rate and Moody's Baa spread with a correlation of -0.77. Therefore, suspected multicollinearity, if any, could be a result of these two variables. The absolute correlation between the other pairs of variables, in general, is low. The variable M1 money multiplier, for example, has correlations between 0.01 and 0.31 in absolute value with all other macroeconomic indicators. Generally, the variables have both positive and negative signs in the pair-wise correlations, hence, we seem to have the right mix of variations without any highly correlated variables. Since the sample does not contain any highly or perfectly correlated macroeconomic indicators, we do not suspect problems with high multicollinearity.

3.4 Global Industry Classification Standard

This paper analyzes the sensitivity of credit ratings to macroeconomic indicators, across business sectors in the United States. We classify the firms in the sample according to the Global Industry Classification Standard into eleven sectors; Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Telecommunication Services, Real Estate, and Utilities. GICS is an industry taxonomy developed by S&P Dow Jones Indices and Morgan Stanley Capital International. We choose to categorize the sample according to GICS rather than the traditional industry-based classification systems for two main reasons. The first is that GICS provides four layers of classification categories for each firm; the hierarchy begins with eleven sectors, narrowed down to twenty-four industry groups, sixty-nine industries; and finally one hundred and fifty-eight sub-industries providing a clear and distinguished analysis resolving the overlapping issue that exists in several classification systems. The second and most significant reason is that GICS is a market-oriented rather than a production-oriented classification system. For instance, the traditional classification of consumer economy into "Consumer Products" and "Consumer Services" is replaced by "Consumer Discretionary" and "Consumer Staples" providing an indicator of the business cycle and state of the economy (see S&P Global (2019*a*) for a detailed description of the business sectors, the industries, and the sub-industries).

4

Methodology

In this section, we present the methodological framework used to investigate the relationship between the credit ratings and the fundamental- and macroeconomic indicators. The section is divided into two major sub-sections, one for each of the two different techniques that we use to process the data: the ordered probit and the random forest models.

4.1 The Ordered Probit Model

Credit ratings are specified as a discrete qualitative variable with a natural ordering, and thus, an ordered probit model is more suitable than a linear regression to handle the data (Amato and Furfine, 2004). The probit model is a regression-based model where the dependent variable can be either nominal or ordinal with more than two different categories and the independent variables can be both qualitative and quantitative. The model is designed to calculate the probability that an observation falls into a particular class of the dependent variable, based on the linear combination between the independent variables. The probit model is estimated using the maximum likelihood method, and all inferences are, therefore, based on the log-likelihood and chi-square test statistic. The general form of the probit model is given by:

$$Y = f(\beta_0 + \sum_{i=1}^{n} \beta_i X_i) \tag{4.1}$$

where $\beta_i' s$ are the estimated coefficients, $X_i' s$ are the independent variables, and Y is the dependent variable. In this paper, Y represents the credit ratings, and $X_i' s$ are the observable fundamental and macroeconomic indicators that influence the determination of the firms' credit rating.

The coefficients $\beta_i' s$ in a probit model, rarely have any direct interpretation (like the marginal effect in linear regression) and instead represent the change in the cumulative normal probability by a unit change in the corresponding independent variable X_i, that the dependent variable Y falls into a specific category (Trueck and Racev, 2009). To model different credit ratings, we use an ordered probit model to extend the framework to fit more than two categories for ordered data. This is easily done by defining $N-1$ thresholds t_N for the rating classes $n = 1, \ldots, N$:

$$f(x) = \begin{cases} 1 \text{ if } \mathrm{z}_i \leq t_1 \\ 2 \text{ if } \mathrm{t}_0 < z_i \leq t_2 \\ 3 \text{ if } \mathrm{t}_1 < z_i \leq t_3 \\ \ldots \\ N \text{ if } \mathrm{z}_i > t_{N-1} \end{cases}$$

where z_i is a latent linking function that generates the observed values of y_i. The function z_i is a linear combination of the independent variables:

$$z_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_n x_{ni} + \epsilon_i \tag{4.2}$$

where ϵ_i is a standard Gaussian distributed term (Blume et al., 1998). Using the latent variable z_i, we can simplify the interpretation of β_i a bit: one unit increase in x_i leads to a $\beta_i x_i$ increase in the latent variable z_i and thus leads to a lower credit rating since we map the highest ratings to the lowest numbers. On the other hand, a negative sign leads to a higher credit rating.

All thresholds t_N are unknown parameters that together define the range where the latent variable z_i may fall into and need to be estimated along with the β_i coefficients. The probabilities that y_i are assigned to category $n = 1, \ldots, N$ can then be estimated by:

$$P(\cdot) \begin{cases} P(y_i = 1) = \Phi(t_1 + \beta X_i) - \Phi(\beta X_i) \\ P(y_i = 2) = \Phi(t_2 + \beta X_i) - \Phi(t_1 + \beta X_i) \\ \ldots \\ P(y_i = N) = 1 - \Phi((t_{N-1}) - \beta X_i)) \end{cases}$$

where $\beta = (\beta_1, \ldots, \beta_n)'$, $X_i = (x_{1i}, \ldots, x_{ni})$ and $\Phi(\cdot)$ represents the cumulative normal distribution function. The marginal effect for a regressor x_i on the probability for the category j is defined as:

$$\frac{\partial p_{ij}}{\partial X_{ri}} = [\Phi'(t_{j-1} - X_i'\beta) - \Phi'(t_j - X_i'\beta)]\beta_r \tag{4.3}$$

The marginal effects for each variable on the different categories sums up to zero. The marginal effects are measured in percentage and depend on the values of all other explanatory variables and the estimated coefficients.

The ordered probit model does not have any coefficient of determination R^2, which measures the proportion of the total variance that is explained by the model. Instead, some similar pseudo-R-square statistics have been proposed. McFadden (1974) suggests R-square based on the log-likelihood values from a model with only an intercept and the full model:

$$R^2_{McFadden} = 1 - \frac{log\hat{L}(M_{Full})}{log\hat{L}(M_{Intercept})} \tag{4.4}$$

where the $log\hat{L}(M_{Intercept})$ is the total sum of squares, and the $log\hat{L}(M_{Full})$ is the sum of squared errors from the estimated model. McFadden's R-square is best to be used when comparing nested models. For a classification model, observations that are correctly classified will contribute less to the likelihood value compared to observations that are not correctly classified. Thus, if a model correctly classifies all observations, $log\hat{L}(M_{Full})$ will be close to zero and McFadden's R-square will be close to one, indicating a perfect prediction ability. However, values of $R^2_{McFadden}$ between 0.2 and 0.4 represent an excellent fit (McFadden, 1977).

4.2 The Random Forest Model

The random forest is a predictive approach used in statistics and machine learning for classification purposes. Random forest is used to make predictions of a response variable based on a set of predictor variables in a way robust to overfitting (Breiman (2001);Liaw, A and Wiener, M (2012)). The foundation is an ensemble of many trees to classify individual observations. Random forest obtains a class vote from each decision tree and then makes the overall classification by the majority vote. The trees are created by sampling individuals and variables from a prespecified training dataset by minimizing its expected prediction error, which is defined as:

$$Err(\varphi_\ell) = E_{X,Y}[L(Y, \varphi_\ell(X))] \tag{4.5}$$

where ℓ is the training set that φ_L is trained in, and L is a loss function that measures the deviation between its two arguments. Each observation is classified in every tree, and the final prediction is the most common outcome over all classification trees. To understand the underlying process that makes the predictions, it is important to identify which predictor variables are crucial to make these predictions.

4.2.1 Classification Tree: Basic Concepts

A classification tree can be used to predict qualitative response variables and is the building block in the random forest model. A simple prediction is that each observation belongs to the most commonly occurring class in the training dataset. The classification tree can be created from the following definitions:

Definition 1. *A tree is a graph $G = (V, E)$ where any two vertices or nodes are connected by exactly one path.*

Definition 2. *A rooted tree is a tree in which one of the nodes has been designated as the root.*

Definition 3. *If an edge exists from t_1 to t_2, then node t_1 is referred to as the parent of node t_2, while t_2 is said to be the child of node t_1.*

Definition 4. *A node is referred to as internal if it has at least one child, and a terminal if it has no children.*

Definition 5. *A binary tree is called a rooted tree where all internal nodes have exactly two children.*

The classification rooted tree is defined as a model $\varphi : X \to Y$, where every node t is a subspace $X_s \subseteq X$ of the input space and the root node t_0 represents X itself. A split follows internal nodes s_t, which divides the space X_t into disjoint subspaces by recursive binary splitting, corresponding to each of the node's children. The paths of the trees that connect the nodes are usually referred to as branches, and the terminal nodes that give the predicted class are called leaves.

In classification, the response variable Y is a finite set of values, $Y = (c_1, c_2, ..., c_j)$, which is a partition over the universe Ω. In the same way, a classifier φ is a partition of the universe Ω since it makes the prediction $\hat{Y}$ of Y and is defined on the predictor space X as

$$X = X^{\varphi}_{c_1} \cup X^{\varphi}_{c_2} \cup ... \cup X^{\varphi}_{c_j} \tag{4.6}$$

where $X^{\varphi}_{c_K}$ is a set of vectors $\mathbf{x} \in X$ such that $\varphi(\mathbf{x}) = c_k$. Learning a classifier can therefore be seen as learning a partition of X by matching as close as possible to the best possible separation.

4.2.2 Impurity

The impurity $i(t)$ is a measure of the goodness of any node at t, where a smaller impurity gives a purer and better prediction $\hat{y}_t(\mathbf{x})$ for all $\mathbf{x}$ that all learning samples falling into t. Formally, the impurity decrease can be defined in the following way:

Definition 6. *The impurity decrease of a binary split $s \in \Omega$ dividing node t into a left node t_L and a right node t_R is*

$$\Delta i(s,t) = i(t) - \frac{N_{t_L}}{N_t} i(i_L) - \frac{N_{t_R}}{N_t} i(i_R) \tag{4.7}$$

where $\frac{N_{t_L}}{N_t}$ and $\frac{N_{t_R}}{N_t}$ are the proportions of the learning sample from ℓ_t going to node t_L and t_R, respectively.

A split can be seen as a partition and is called binary if it divides t into two subsets. An optimal decision tree can grow by iteratively splitting each node t by using the split s^* that locally maximizes the decrease of impurity of the resulting child nodes.

Definition 7. *A split s at node t is a partition of X_t, which is a set of non-empty subsets of X_t such that every element of X_t is in precisely one of these subsets.*

One criterion to make each binary split is to minimize the classification error rate, which is the fraction of the training individuals in a region that do not belong to the most common class. In other words, this classification rule is the same as assigning an individual to the most likely class of the training data in that individual's given region. The impurity function $i(t)$ can then be defined as:

$$i_{CE}(t) = 1 - \max_{c \in Y} p(c \mid t) \tag{4.8}$$

where $p(c \mid t)$ is the proportion of the training observations belonging to c at each node t. The classification error rate suffers from two serious drawbacks. If the majority class is the same for both child nodes, it will lead to an impurity decrease of zero. Also, the classification error does not account for changes in subsequent class distributions. A better criterion to evaluate the quality of a split must account for further possible improvements deeper down the tree and makes $i(t)$ successively smaller when t gets more homogeneous and larger when t gets more heterogeneous. Gini (1912) suggests an impurity function based on an index that satisfies these requirements:

$$i_G(t) = \sum_{K=1}^{J} p(c_k \mid t)(1 - p(c_k \mid t)) \tag{4.9}$$

The Gini index is reliable and robust as a decision rule, which accounts for the total variance across the K classes and is a measure of node purity. A small value of $i_G(t)$ indicates that observations at that node dominantly belong to a single category.

4.2.3 Importance

Importance is a measure in the random forest that explains which features are essential in determining the classification rule. The importance measure helps us get a deeper understanding of the model by focusing only on the crucial variables. The importance is the improvement in the splitting-criterion attributed to the splitting variable and is calculated separately for all trees in the forest for each variable. Following Breiman et al. (1984), we define the importance measure of variable X_j as:

$$Imp(X_j) = \sum_{t \in \varphi} \Delta I(\bar{s}_t^j, t) \tag{4.10}$$

where $\bar{s}_t^j$ is the closest split of s_t defined on the variable X_j that can mimic the actual split s_t at the node t and is usually referred to as surrogate splits. Additionally, the importance measure for a random forest model is calculated using the sum of the weighted impurity decreases over all nodes t where X_j is used over the average for all trees φ_m:

$$Imp(X_j) = \frac{1}{M} \sum_{m=1}^{M} \sum_{t \in \varphi_m} 1(j_t - j)[p(t)\Delta i(s_t, t)] \tag{4.11}$$

where j_t is an identifier for the variable used in the split of node t. The importance in random forest models is typically referred to as the mean decrease impurity importance (MDI). The mean decrease in Gini impurity is a scale irrelevant measure, meaning that it is easy to compare across different variables. A disadvantage with the Gini-based importance measure is that splits are biased towards numerical variables with many split points.

The importance measure can also be calculated based on the accuracy in the out-of-bag sample. To measure the importance of one variable, first, the accuracy in the out-of-bag sample is measured. The values of the specific variable are then randomly shuffled in the out-of-bag sample while keeping all other variables constant. At last, a decrease in the prediction accuracy on the shuffled data is calculated as the difference in accuracy by removing the variable. The mean decrease in accuracy is then measured across all trees.

4.2.4 Sensitivity and Specificity

Sensitivity is generally the probability of predicting an event among true events, while specificity is the probability of predicting no event among no true events. Usually, sensitivity measures the performance of how good a model is to identify observations with the correct rating and specificity measures how good a model is to identify observations that do not belong to the specific tested rating. A higher value of both sensitivity and specificity are preferred. For binary classification, the sensitivity and specificity are calculated as:

$$Sensitivity = \frac{A}{A+C} \tag{4.12}$$

$$Specificity = \frac{D}{B+D} \tag{4.13}$$

where A, B, C, and D come from the $2x2$ matrix:

Predicted/Reference	Event	No Event
Event	A	B
No Event	C	D

In the multiclass setting, the sensitivity and specificity are derived from a "one-versus-all" configuration. Each rating class represents an event against all other rating classes that represent no event in separate tests. The multiclass sensitivity and specificity are then calculated as the average of the performance measures over all separate trials.

4.2.5 Detection Rate

The detection rate is the rate of true events that are predicted events:

$$DetectionRate = \frac{A}{A + B + C + D} \tag{4.14}$$

with A, B, C, and D refer to the values in the matrix above. Detection rate is the proportion of predicted events of the whole training set.

4.2.6 Area under the Receiver Operating Characteristic Curve

In the binary setting, the area under the Receiver Operating Characteristic (ROC) is widely used to measure the performance of any classification rule and is generally referred to as AUC. The AUC compares the overall distribution of the estimated probability $\hat{p}(x)$ between the classes. The AUC ranges between 0.5 for a purely random classifier and 1 for a perfect classifier.

Assume $f(\hat{p}) = f(\hat{p}(x) \mid 0)$ is the probability function of the estimated probability to be assigned to class 0 for class 0 points and $g(\hat{p}) = g(\hat{p}(x) \mid 1)$ is the probability function of the estimated probability to be assigned to class 0 for class 1 points. Also, assume that $F(\hat{p}) = F(\hat{p}(x) \mid 0)$ and $G(\hat{p}) = G(\hat{p}(x) \mid 1)$ are the associated cumulative distribution functions to f and g. The ROC curve is then obtained by plotting $G(\hat{p})$ on the vertical axis against $F(\hat{p})$ on the horizontal axis. The ROC curve lies in a unit square and can be seen in figure 4.1 (Venkateswaran, B and Ciaburro, C, 2019).

F is referred to as 1-specificity, and G is the sensitivity. A good classifier is corresponding to points where $G(\hat{p}) > F(\hat{p})$ and is visualized by a ROC curve in the upper left corner and is illustrated in figure 4.1. A perfect classification rule is characterized by sensitivity and specificity of 100 % while no better (worse) classifier than chance lies on (below) the straight line. The line representing random guessing is pictured from the lower left to the upper right corner in figure 4.1 (Hand and Till, 2001).

The area under the ROC curve (AUC) is simply an overall aggregated measure of the model's performance, across all possible classification outcomes and is defined by:

$$AUC = \int G(u)f(u)du \tag{4.15}$$

Therefore, the AUC represents the probability that a randomly selected observation of class 1 will have a smaller estimated probability to be assigned to class 0 than a randomly selected observation of class 0. AUC focuses on how well the classification rule differentiates between the distributions of the two classes and ignores the cost and influences of other external factors that could affect the classification (Hand and Till, 2001). AUC is a desirable performance measure, considering it is scale-invariant,

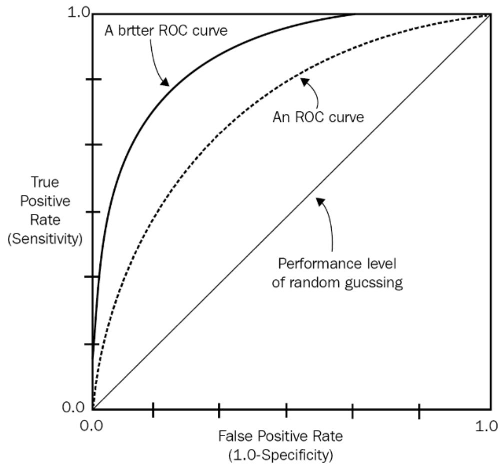


Figure 4.1: The Receiver Operating Characteristic Curve (Venkateswaran, B and Ciaburro, C, 2019): a hypothetical example. The ROC curve is obtained by plotting the Sensitivity on the vertical axis against 1-Specificity on the horizontal axis. A good classifier corresponds to a ROC curve closer to the upper left corner in the unit square plot. The area under the ROC curve (AUC) is then an overall aggregated measure of the model's performance.

easily compared between models, and measures the model's prediction power irrespective of the chosen classification threshold.

To generalize the AUC measure for a multiclass setting, one AUC value is calculated for each class when comparing that class to all others at the same time. The generalized AUC is calculated as the mean AUC over all pairwise class comparisons.

One way to interpret the AUC values is in line with the academic point system:

– 0.9-1 = Excellent (A)

– 0.8-0.9 = Good (B)

– 0.7-0.8 = Fair (C)

– 0.6-0.7 = Poor (D)

– 0.5-0.6 = Fail (F)

4.2.7 Out-of-Bag Classification Error

The out-of-bag (OOB) classification error is a valid and proven-to-be an unbiased test error as the number of trees added to the forest is increasing. Breiman and Cutler (2003) argue that there is no need for either cross-validation or using the validation set approach since it is estimated during the estimation of the random forest. If the number of training sets is sufficiently large, it has been proven that the OOB error is equivalent to the leave-one-out cross-validation error without a computational burden (James et al., 2013). An advantage of using OOB is that the random forest can be fitted in one sequence, since the cross-validation is being performed while the model is being trained.

In the construction of each decision tree, a different sample from the original data will be used to construct k trees. The OOB error is estimated from the observations that are left out (out-of-bag) in each construction by running them through the k trees to get a classification. As a result, observations not used to construct the random forest internally will be used to test the model's performance. Finally, assume that j is the class with the majority of votes every time n was out-of-bag, the OOB classification error is obtained by calculating the proportion that the class j is not equal to its actual class of n averaged over all observations.

4.2.8 Partial Dependence Plot

The partial dependence plot (PDP) is first used by Friedman (2001) to interpret machine learning algorithms. The PDP shows the marginal effect of features on the class probabilities in the fitted classification model. Typically, the graphical visualization of the marginal effects is restricted to one or two variables due to the difficulty to produce more than three dimensions. Mathematically, the partial dependence of the S predictors on the predictive function $f(X)$ is defined as:

$$f_S(X_S) = E_{X_C}[f(X_S, X_C)] \tag{4.16}$$

and can be estimated by

$$\bar{f}_S(X_S) = \frac{1}{N}\sum_{i=1}^{N}[f(X_S, X_{Ci})] \tag{4.17}$$

where S is denoted as a subset of the p predictor variables and C the complement to S, such that $S \subset (X_1, X_2, ..., X_p)$, $S \cup C = (X_1, X_2, ..., X_p)$ and the random forest predictor function f depends on all p predictor variables, $f(X) = f(X_S, X_C)$. Thus, to calculate the partial dependence, the entire training set must be used for every set of values in X_S, which can be computationally intensive in large datasets (Hastie et al., 2008).

The partial dependence curve at a particular value of the feature indicates the average predicted probability when all data points are forced to take on one feature value. It also shows how the predicted probability on average changes when X_i changes, given that the features are independently distributed. If the features are correlated, the computation of the PDP at a certain level of the feature involves taking the average over the marginal distribution of C, which might include unrealistic intervals for the complement variables. Thus, by looking at areas of the distribution where the probability mass is low, it can lead to problems when interpreting the predicted probabilities.

For classification problems, the partial dependence function is measured in logit of probabilities:

$$f(x) = logp_k(x) - \frac{1}{K}\sum_{j=1}^{K} logp_j(x) \qquad (4.18)$$

where K is the number of classes, and p_j is the proportion of votes for class j. Therefore, a negative value of the partial dependence function is associated with a lower probability than 0,5 and a positive value corresponds to a higher probability than 0.5. An increase in partial dependence function for a specific rating class leads to a higher likelihood of being assigned to that rating class.

4.2.9 Accuracy and the Kappa Coefficient

The overall accuracy is a performance measure of the proportion that is correctly classified. Accuracy is typically measured in percentage, where 100 % accuracy indicates that the classification is perfect and all the observations are correctly classified. Accuracy is a good measure when the data is balanced and has to be used with caution when working with imbalanced data, considering that a bad classifier that only predicts one outcome can get high accuracy. For example, in a binary setting where the ratio between class 0 and 1 is 99:1, a classifier that only predicts 0 no matter what will get an accuracy of 99 %.

On the other hand, a performance measure that is suitable for imbalanced data is Cohen's Kappa. The Kappa coefficient measures how well the classification rule performs compared to just randomly classified observations (Sim and Wright, 2005). The Kappa compares the model's accuracy to an expected accuracy in the following way:

$$\chi = \frac{Observed\ Accuracy - Expected\ Accuracy}{1 - Expected\ Accuracy} \qquad (4.19)$$

where the expected accuracy is the accuracy that can be expected from any random classifier and is based on the confusion matrix. The Kappa coefficient can take on any values between -1 to 1 where unity indicates a perfect accuracy of 100 % and a value of zero represents that the classifiers' accuracy is no better than those predicted by chance.

4.3 Model Selection Critera

Akaike Information Criterion (AIC) and the Schwartz Bayesian Criterion (SBC) are two of the most commonly used model selection criteria, which makes a trade-off reduction in the sum squared residuals to get a more parsimonious model (Enders, 2014). In small samples, AIC can perform better than SBC, while SBC has large sample superior properties and is asymptotically consistent. Generally, AIC is biased towards selecting overparameterized models.

$$AIC = Tln(sum\ of\ squared\ residuals) + 2n \qquad (4.20)$$

$$SBC = T ln(sum\ of\ squared\ residuals) + n ln(T) \tag{4.21}$$

where n is the number of parameters estimated, and T is the number of usable observations. When the fit of the model improves, both AIC and SBC will approach $-\infty$. By evaluating the models over the same sample and keeping T fixed, we can select the most appropriate model by choosing the model with the smallest value of any model selection criteria.

4.4 Model Formulations

In this section, the model formulations for both the random forest model and the ordered probit model are presented in detail.

4.4.1 Creating the Forest

One of the main objectives of this thesis is to investigate the importance of the macroeconomic indicators across different sectors. One way to do this is to implement a random forest model with the eight macroeconomic indicators presented in section 3.3. The indicators are: the effective Federal funds rate, gross domestic product, charge-off rate on consumer loans, unemployment rate, Moody's Baa corporate spread, trimmed mean PCE inflation rate, velocity of M2 Money stock and the M1 money multiplier as explanatory variables. To improve the model's accuracy and performance, we propose a model with these eight macroeconomic indicators along with all the firm fundamentals presented in section 3.2 as explanatory variable against the credit rating as the response variable. This variable setup will be trained in each of the eleven business sectors to investigate differences in the macroeconomic importance between sectors.

The variables are used as inputs in Breiman's random forest algorithm to each decision tree and the following random forest for each of the eleven business sectors (Hastie et al., 2008). By using the fundamental and macroeconomic indicators as explanatory variables in the random forest algorithm, the model produces a predicted credit rating for each firm in each period, for the given set of explanatory variables. The random forest implementations will be done using the *randomForest* package in the R statistical software environment, created by Andy Liaw.

The random sampling of m predictors in Step 1 of the algorithm works as a de-correlation of the trees in the random forest and is, in general, relatively insensitive to the selected value. The recommendation is to start with a value of m between 2 and the total number of predictors p while the recommended number of trees to start with is 1000 and then increase it if the performance of the model improves (Kuhn and Johnson, 2013).

To further investigate the performance of the trained random forest models, a baseline model is created to serve as a benchmark. The benchmark-

Random Forest Algorithm for Classification

1. For the number of trees $n = 1$ up to N:
    a) Randomly draw a bootstrapped sample Z^* with the size of B from the training set with replacement.
    b) A decision tree T_n is growing to the bootstrapped data, by repeating the following steps recursively at each terminal node of the tree, until the smallest node size b_{min} is reached.
        I. Select the number of m variables at random from the entire set of explanatory variables used as inputs in the model.
        II. Choose the best variable/split among the m randomly considered variables at each split.
        III. Split the node into two child nodes.
2. The process in 1 gives the ensemble of trees $\{T_n\}_1^N$. To make a credit rating prediction for a new point x: Let $\hat{C}_n(x)$ be the class prediction of the nth decision tree, then the Random Forest class prediction $\hat{C}_{rf}^N(x)$ is the majority vote $\{\hat{C}_n(x)\}_1^N$.

ing models are trained in random forest with a different set of variables. In the benchmarking model, we change all the firm fundamentals used by S&P to a new set of fundamentals that are also common indicators of the firm's credit-strength and default probability. The fundamentals used in the benchmarking model are similar, to a great extent, to the ratios used to calculate Altman's Z-score: long-term debt over asset, working capital over total assets, retained earnings over total assets, EBIT over total assets, return on assets, and market value of equity over total liability (Altman, 1968).

4.4.2 Training and Testing the Dataset

A meaningful way to evaluate a machine learning model is to divide the data into two different sets: a training set and a test set. Commonly, most of the observations are used in the training set, and a smaller portion of the data is used in the test set. The training set is a subset of the total data set used to fit the model, while the test set is a subset of the total data set used to test the trained model. There is no general rule on how to choose the proportions for the two different sets since it depends on the complexity of the models being fit to the data and the training sample size (Hastie et al., 2008). It is essential that the test set is large enough to yield statistically meaningful results and is a representative of the data set as a whole.

By dividing the whole dataset into a training and a test set, it is easier to check if the model is overfitting. Overfitting occurs in machine learning when a model is optimized and fits the training data too well by learning all the details and noises to the extent where it negatively affects the model's prediction performance in a new dataset. In this case, the model will be too complex containing more parameters than justified by the data and will therefore find it hard to generalize when the model is tested on a new dataset. Generally, all data has some degree of randomness within it, and by learning it in detail, the model will classify with substantial errors.

The trained model's performance in the test set should generally be lower than the model's performance in the training set, since the data is new. In addition, the model does not learn any additional patterns in the test set that are not included in the training set. However, if the accuracy is substantially higher in the training set compared to the test set, it is a sign of overfitting. To limit the extent of overfitting, using a resampling technique to model the accuracy, a validation set to test the trained model, and cross-validation to make sure the model can be useful and helps improve the model.

4.4.3 Tuning the Hyperparameters

Random forest handles three different categories of data to train the models. It handles the input data in the form of credit ratings, firm fundamentals, and the macroeconomic indicators to adjust the parameters. Typically, the parameters are the values that represent a model and can be used to distinguish the model from other models of the same type. At last, random forest use hyperparameters that govern the training process and represent configuration variables. Hyperparameters are not directly related to the inputted data and can be tuned to optimize the model's performance. To get the optimal setting of the hyperparameters, we run complete executions of the random forest model for different values of the hyperparameters. The performance of each model is subsequently evaluated by comparing each model's accuracy. We eventually adjust the hyperparameter until the model generates the best value of accuracy.

The hyperparameter that can be tuned in the random forest is the number of decision trees n in the forest and the number of features sampled by each tree when splitting a node m. Commonly, when the number of trees is increasing, the model performs better, and the OOB error predictions are more reliable. However, when n gets too large, there is a trade-off between the prediction performance and the computational costs for learning these trees (Oshiro et al., 2012). In this thesis, we select the numbers of trees to be 1000 for each random forest model to get reliable results.

When random forest is used for classification, the default value of features randomly considered at each split m is set to p, where p is the number of explanatory variables. To choose the optimal number of features at each split, we perform a search process that determines the optimal value of m for each random forest model. The optimal value of m in each industry is presented in Table 4.1, together with the chosen value of n, which is fixed to 1000 for all sectors.

4.4.4 Formulating the Ordered Probit Model

Regarding the ordered probit models, a model for each sector is estimated based on the whole sample period 1985 – 2016. The categorical dependent

Table 4.1: The optimal number of features m considered at each split and the number of decision trees. Table 4.1 reports the optimal tuned number of features m considered at each split. Also, the fixed number of decision trees 1000 *for all sectors, is reported. The industries in each column are abbreviated in the following way: Consumer Discretionary is CD, Consumer Staples is CS, Energy is E, Financials is F, Health Care is HC, Industrials is I, Information Technology is IT, Materials is M, Real Estate is RE, Telecommunication Services is TS and Utilities is U.*

	CD	CS	E	F	HC	I	IT	M	RE	TS	U
m	8	8	14	8	8	8	8	8	9	8	8
n	1000										

variable is the credit rating, while the independent variables are the firm fundamentals and macroeconomic indicators presented in section 3.2 and 3.3 respectively. This model is referred to as the basic model. In the second stage, all insignificant explanatory variables are removed from each of the eleven models obtained in the first stage, and the models are re-estimated only with the significant explanatory variables; this second stage is referred to as the refined model. The ordered probit estimations are done in the computer software program Stata.

By using the longitudinal data of the firm fundamentals together with the time-varying macroeconomic indicators, the relationship between the credit ratings and the independent variables is a time-fixed effect model with eight time-varying intercepts for each period. The reason for the time-fixed effect model is that the macroeconomic indicators only vary over time but are constant across all firms and business sectors at any given time-period. We are not using time-fixed effects because we are interested in classifying the credit ratings across the sectors rather than finding the causal relationship. As a result, the model does not capture the heterogeneity in the time dimension in the macroeconomic indicators

We are aware that by omitting the heterogeneity, the coefficients β in the probit model will not be consistently estimated (Wooldridge, 2010). However, the probit of the dependent variable on the independent variables will consistently estimate $\frac{\beta}{\sigma}$ under the assumption that the unobserved heterogeneity is normally distributed. A formal proof of this is given in Appendix B. Assuming normality, we can obtain the correct directions of the partial effects of any explanatory variable by investigating the signs of β. However, we will not be able to obtain the absolute magnitude and partial effects of the explanatory variables in the non-linear model.

Comparing the two models, we observe that one of the main advantages of the random forest model over the ordered probit model is that the former does not require any particular assumption of the data. In addition, the random forest is robust to overfitting and outliers, automatically selects important variables, and can handle missing data. The random forest is also not limited to linear relationships between the variables, which is the case in the ordered probit model. Generally, the random forest model is flexible to use, requires minimal data pre-processing, and gives highly accurate results with only a few hyperparameters to tune. On the other hand, it is more complex, harder to construct, less intuitive, and requires more training and tuning time in comparison to the probit model.

5

Empirical Findings

In this section, both the ordered probit and the random forest models are evaluated based on the performance and goodness-of-fit. The findings from both models are extracted and interpreted. This section ends with combining the statistical results and relating them to real-world economic applications for each of the eleven business sectors.

5.1 The Results from the Random Forest

A random forest model is trained in each sector to determine the importance of the macroeconomic indicators in different sectors and classify the credit ratings across the sectors. To get a more realistic model, all firm fundamentals and macroeconomic indicators are included in training the model in each of the eleven sectors. All models are evaluated with different performance measures, both in training and test sets.

5.1.1 Performance

The performance measures, previously mentioned in section 4.2, are presented in Table 5.1. The overall goodness of fit of the models should be interpreted carefully because the measures only evaluate the performances in the training set. Therefore, the measures only give us a deeper understanding of how well the random forest performs when the models are trained. Each column in Table 5.1 represents the performance measures for a random forest model in each sector.

The overall accuracy is ranging between 0.78 – 0.91, indicating that around four of five firms are classified correctly in sectors CD, I, M, TS, and U, while approximately nine of ten firms are correctly classified in sectors E, F, IT and RE.

The sensitivity and specificity range between 0.57 – 0.87 and 0.93 – 0.98, respectively, across all sectors. As mentioned before, a higher value of these measures is preferred compared to a lower one. The specificities are substantially high in all sectors, indicating that all trained models are good to correctly not classify a firm to a wrong rating on average. On the other hand, the models' performance to correctly classify firms to

Table 5.1: The performance measures: Accuracy, Kappa, Sensitivity, Specificity, and Detection Rate for the eleven business sectors. To evaluate the in-sample performance of the random forest classification rules, five different performance measures are reported in Table 5.1. Generally, a higher value of the performance measures indicates a better in-sample performance fit. The sectors in each column are abbreviated in the following way: Consumer Discretionary is CD, Consumer Staples is CS, Energy is E, Financials is F, Health Care is HC, Industrials is I, Information Technology is IT, Materials is M, Real Estate is RE, Telecommunication Services is TS and Utilities is U.

	CD	CS	E	F	HC	I	IT	M	RE	TS	U
Accuracy	0.83	0.86	0.90	0.89	0.88	0.78	0.89	0.82	0.91	0.80	0.80
Kappa	0.77	0.81	0.85	0.82	0.84	0.68	0.86	0.75	0.83	0.72	0.66
Sensitivity	-	-	0.78	-	-	-	0.81	0.57	0.87	-	-
Specificity	0.97	0.98	0.97	0.98	0.98	0.97	0.98	0.96	0.93	0.96	0.96
Detection rate	0.12	0.11	0.18	0.11	0.13	0.08	0.13	0.12	0.30	0.10	0.10

its actual rating in the training set is relatively worse, due to the lower values of sensitivity. The detection rates, representing the proportion of correctly classified ratings in the whole training set, range between 0.08 – 0.30 for all business sectors. This measure is dependent on the data and can be very misleading in imbalanced data sets. This is the case here since the proportion of firms that can be classified correctly is much smaller than the proportion of firms that do not belong to the tested rating class.

For imbalanced data, Cohen's Kappa is a suitable measure to evaluate the training of a model. The Kappa coefficients are shown in the second row of Table 5.1 and varying between 0.66 and 0.86 across the sectors, indicating that the classification rule in all the sectors is better than chance, since the values are far greater than zero. According to the Kappa coefficient, the random forest finds it harder, for example, to train any classification rule in the Utility sector (0.66) compared to the Information Technology sector (0.86).

In conclusion, the random forest model seems to fit the training set quite well, with far better prediction power than a random classifier. There is still room for improvement in the classification rules since the values of accuracy, kappa, and sensitivity can still be increased. The model can be refined, and the performance can be enhanced by adding more relevant variables that contain information which affects the rating decision. In general, the overall classification performance seems to be slightly better in the Real Estate and Information Technology sectors compared to the other sectors.

5.1.2 Out-of-Bag Error and Prediction Evaluation

The out-of-bag error rate is calculated during training the model. The out-of-bag error rate for each trained model can be seen in the first row in

Table 5.2. The smallest OOB error rate is observed for the Real Estate sector, while the largest OOB error is found for the Industrials sector with values of 0.06 and 0.21 respectively. This indicates that the worst trained models correctly classify around four out of five firms when the classification rule is tested in the out-of-bag sample while the best-trained model correctly classifies almost nineteen of twenty firms.

In order to evaluate the models' prediction power and investigate whether each trained model is useful to predict the credit ratings, we cross-validate the models by testing them in a test sample that is independent of the training sample. This is done by running the classification rules for each sector. The predicted accuracy denoted P-Accuracy is the accuracy for each trained model when the model is applied to the corresponding sectors test set. P-Accuracy is calculated as the proportion of correctly classified firms relative to the total number of firms in each sector. The values of P-Accuracy in Table 5.2 are associated with the prediction matrices presented in Appendix C, where the diagonal elements represent the number of correctly classified ratings for each rating class. Given that most firms are correctly classified, we can conclude that the overall prediction power of the random forest models is quite high. The predicted accuracy is ranging between 0.74 and 0.93 in different sectors. The model trained in the Real Estate sector performs best on its test set with an accuracy of 0.93, while the trained model in the Industrial sector performs worst with a predicted accuracy of 0.77, which means that the worst trained model still correctly classifies around three out of four firms in a new dataset.

We can also use the prediction matrices in Appendix C to see the rating classes where the models have the best prediction power. For example, in the Energy sector, all observations with AAA (21) and AA (37) are correctly classified, while for lower ratings, there is some misclassification. For rating A, 91 of 109 are correctly classified while five and four misclassifications occur for the ratings BBB and BB, respectively, resulting in a predicted accuracy of 0,91. From the prediction matrix, we can, therefore, conclude that the trained model in the Energy sector easily recognizes and correctly classifies the highest ratings, while it has some errors for lower ratings.

The multiclass AUC is presented in the third row in Table 5.2. All sectors have an AUC value above 0.9, with excellent prediction power (A), except the Materials sector, which has an AUC value of 0.88 and with good classification power (B). In general, all AUC values are high, ranging between 0.88 – 0.97 for all the sectors, indicating that all trained classification rules are exceedingly better performing than purely random classifications and are almost as good as a perfect classifier.

A summary of the benchmarking models' performances, corresponding to Table 4.1, 5.1, and 5.2 can be found in Appendix D. The benchmarking models are estimated in the same way as the main models. The classification rules are also estimated for each industry with a fixed number of the hyperparameter n of 1000 and the optimal tuned value of the

Table 5.2: The performance measures out-of-bag error, prediction accuracy and AUC. Table 5.2 reports three performance measure that evaluates the random forest model in the test sample for each of the eleven trained models. Better performance is obtained by lower values of OOB and higher values of P-Accuracy and AUC. The sectors in each column are abbreviated in the following way: Consumer Discretionary is CD, Consumer Staples is CS, Energy is E, Financials is F, Health Care is HC, Industrials is I, Information Technology is IT, Materials is M, Real Estate is RE, Telecommunication Services is TS and Utilities is U.

	CD	CS	E	F	HC	I	IT	M	RE	TS	U
OOB	0.16	0.14	0.10	0.11	0.12	0.21	0.11	0.17	0.06	0.20	0.20
P-Accuracy	0.74	0.85	0.77	0.78	0.82	0.78	0.91	0.83	0.93	0.81	0.80
AUC	0.91	0.97	0.96	0.93	0.97	0.91	0.96	0.88	0.94	0.92	0.91

hyperparameter m. The benchmarking model's accuracy and predicted accuracy vary between 0.73 – 0.89 and 0.74 – 0.91, respectively. The main models have slightly higher accuracy for all sectors except in the Telecommunication service sector and higher prediction accuracy for all sectors except the Industrial and Telecommunication service. The Kappa coefficient is the performance measure that is most improved in our main models compared to the benchmarking models in most of the sectors. For example, the Kappa coefficient increases from 0.68 in the benchmarking model to 0.85 in our model for the Energy sector. In addition, the main models have ten excellent classification rules in the test set compared to nine excellent classification rules in the benchmarking models. Also, the out-of-bag error is lower in the main model, compared to the benchmarking model, in all the sectors except the Telecommunication Services sector.

In conclusion, the main set of variables performs better than the benchmarking set of variables in all sectors except one. The overall performances for both the main model and the benchmarking model, however, give good performances in the test samples. So, we can verify that the main models are based on reliable classification rules that give accurate performances in the test samples. In the following sections, only the main model is be further investigated and analyzed.

5.1.3 Importance of the Features across the Sectors

There are two different measures to evaluate the importance of features in the model. The mean decrease in accuracy measures how much the accuracy decreases when the variable is excluded from the model. A higher value of the mean decrease in accuracy indicates that the feature is more important compared to a feature with a lower value. On the contrary, a negative value indicates a low prediction power and weaker classification than randomness. Further, the other measure of importance is the mean decrease in Gini, which is based on the mean decrease of the Gini impurity when a variable is used to split a node. A variable that

Table 5.3: The mean decrease in accuracy for all features in the eleven sectors. Table 5.3 reports the accuracy importance measure calculated as the mean decrease in accuracy for all the features in all sectors. A higher value of importance indicates that the feature is more important in the classification compared to an indicator with a lower value. The raised numbers 1 to 3 are there to highlight the most important, second most important, and third most important macroeconomic indicators in each of the eleven sectors. The sectors in each column are abbreviated in the following way: Consumer Discretionary is CD, Consumer Staples is CS, Energy is E, Financials is F, Health Care is HC, Industrials is I, Information Technology is IT, Materials is M, Real Estate is RE, Telecommunication Services is TS and Utilities is U.

	CD	CS	E	F	HC	I	IT	M	RE	TS	U
Beta	69	104	76	144	56	118	82	101	92	96	142
Idiosyncratic Risk	182	144	183	170	183	259	131	219	77	94	148
Debt/EBITDA	22	51	21	58	34	69	28	32	12	28	41
CFOD/Debt	30	31	24	69	43	57	42	50	73	40	23
RoC	53	74	26	68	62	88	48	83	46	53	59
EBITDA Margin	52	41	30	52	64	66	44	50	22	35	59
Federal Funds rate	55^1	41^1	40^1	49^1	35^1	55^1	55^1	43^1	26^1	35^1	52^1
Real Gross Domestic Product	6	0	3	9	2	4	12	0	2	7	7
Charge-Off rate	27	23^3	24^3	42^2	21^2	41^2	36^2	28^2	18^2	26^3	32^3
Unemployment rate	28^3	27^2	32^2	26	15	40^3	20^3	26^3	10	29^2	35^2
Moodys Baa spread	35^2	16	21	27^3	16^3	31	17	23	9	15	30
Trimmed Mean Inflation rate	4	-1	0	3	-3	2	1	5	2	2	3
Velocity of M2	13	12	5	15	8	17	12	14	5	13	13
M1 Money Multiplier	10	13	8	19	6	14	14	17	11^3	11	12

ends up in a node with higher purity will have a higher decrease in the Gini coefficient.

The mean decrease in accuracy and mean decrease in Gini are reported in Table 5.3 and 5.4 , respectively, for all variables in all business sectors. The corresponding importance plots for both importance measures are presented in appendix E. The importance plots are simply the ten most important variables from each measure in each sectors, ranked from top to bottom and used to get a quick overview of the importance ranking within each trained model. The first, second, and third most crucial macroeconomic indicators in each sector are raised with the corresponding number in both Tables 5.3 and 5.4 for convenient identification and comparison.

Generally, the idiosyncratic risk and the beta values are the two most important features when training all the models for the two measures of importance, except in the Health Care sector where the idiosyncratic risk is the most important, followed by EBITDA margin and RoC in the mean decrease in accuracy and mean decrease in Gini respectively. Regarding the two importance measures in most of the sectors, five or more firm fundamentals are more important for the classification rule than the most important macroeconomic indicator. This indicates that the firm fundamentals generally have more explanatory power to determine the credit rating, which is in line with Guo and Bruneau (2014) findings, investigating macroeconomic variables impact on default rate of US corporate bonds. They conclude that macroeconomic variables only explain

Table 5.4: The mean decrease in Gini for all features in the eleven sectors. Table 5.4 reports the Gini importance measure calculated as the mean decrease in Gini for all the features in all sectors. A higher value of importance indicates that the feature is more important in the classification compared to an indicator with a lower value. The Gini importance measure is scale-invariant, so it is convnient to compare the importance of different features. The raised numbers 1 to 3 highlight the most important, second most important, and third most important macroeconomic indicators in each of the eleven sectors. The sectors in each column are abbreviated in the following way: Consumer Discretionary is CD, Consumer Staples is CS, Energy is E, Financials is F, Health Care is HC, Industrials is I, Information Technology is IT, Materials is M, Real Estate is RE, Telecommunication Services is TS and Utilities is U.

	CD	CS	E	F	HC	I	IT	M	RE	TS	U
Beta	125	182	88	310	110	494	148	245	119	149	419
Idiosyncratic Risk	238	397	171	282	314	784	277	520	102	187	472
Debt/EBITDA	37	122	22	141	60	281	37	86	14	48	163
CFOD/Debt	46	48	19	106	59	265	58	131	73	63	121
RoC	81	166	20	159	125	363	76	190	43	94	229
EBITDA Margin	71	83	40	101	97	261	55	125	21	54	192
Federal Funds rate	73^1	53^3	38^1	84^1	39^1	167^1	75^1	88^1	23^1	46^1	137^1
Real Gross Domestic Product	13	15	6	28	12	69	13	32	3	17	46
Charge-Off rate	29^3	63^2	14^3	54^2	26^2	133^2	35^2	51^2	10^2	32^3	82^2
Unemployment rate	25	33	20^2	37	16	111^3	19^3	51	6	36^2	80^3
Moodys Baa spread	33^2	33	12	40^3	18^3	95	16	49^3	6	25	67
Trimmed Mean Inflation rate	10	127^1	4	23	10	71	11	37	5	17	45
Velocity of M2	16	42	4	30	12	82	13	42	4	23	50
M1 Money Multiplier	16	24	5	31	13	81	15	44	6^3	19	48

a small fraction of the default rate while the obligor's specific risk is a more powerful explanatory factor for the default rate.

According to the low values of the mean decrease in accuracy for gross domestic product, trimmed mean PCE inflation rate, velocity of M2 and M1 money multiplier, these variables are not substantially relevant in determining the credit ratings in any of the eleven sectors compared to the Federal funds rate, which is the most important macroeconomic indicator for all the sectors. The second and third most important macroeconomic indicators are shared among the charge-off rate, unemployment rate, and Moody's Baa spread, across the different sectors according to the mean decrease in accuracy.

By using the mean decrease in Gini as an importance measure, we can compare the features' relative importance. According to Table 5.4, the most important macroeconomic indicator is the Federal funds rate in all sectors except in Consumer Staples, where the trimmed mean PCE inflation rate has the highest value (127). This result is unexpected because this variable is not important according to the mean decrease in accuracy. As concluded before, the firm fundamentals are generally more important than the macroeconomic indicators. For example, if we look at the Industrials sector, the idiosyncratic risk (784) is over four times more important than the most important macroeconomic indicator (167). Even the EBITDA margin (261), which is the least important fundamental is more than one and a half times more important than the Federal funds rate in the Industrials sector. This pattern where the least important fundamental is more important than the most important macroeconomic indicator can be seen in the Financial, Health Care, and Telecommunication Service sectors as well. On the contrary, a different pattern is observed in some sectors, where the most important macroeconomic indicator is more important than some of the fundamentals. For example, in the Consumer Discretionary sector, the Federal funds rate (73) is more than one and a half times more important than both the debt over EBITDA and the operating cash flow over debt.

In general, the Industrials sector exhibits the highest values in firm fundamentals in terms of the mean decrease in Gini. For example, the beta in the mean decrease in Gini is 494 in the Industrials sector and only 88 in the Energy sector. In addition, the Industrials sector also has the highest mean decrease in Gini for all macroeconomic indicators except for the trimmed mean PCE inflation rate, where the Consumer Staples has the highest value (127 compared to 71).

In summary, the gross domestic product and velocity of M2 are not among the top three important macroeconomic indicators in any of the two importance measures, while the inflation rate and the M1 money multiplier are only among the top three important indicators in one sector each: Consumer Staples and Real Estate, respectively. Both importance measures conclude that the Federal funds rate is generally the most important macroeconomic indicator across the eleven sectors, while the charge off rate, unemployment rate, and Moody's Baa CB spread are most commonly ranked second and third among the macroeconomic indicators.

5.2 The Results from the Ordered Probit Model

In this section, we represent the results from the ordered probit model, compare the model performance across the eleven business sectors, and analyze the sensitivity of the variables.

5.2.1 Goodness-of-Fit

In the first step, the ordered probit model is estimated with all six firm fundamentals and the eight macroeconomic indicators as the explanatory variables against the credit ratings as the dependent variable. As the first goodness of fit measure, a likelihood test is calculated that tests if at least one of the explanatory variables have a significant effect on the credit ratings. The likelihood test-statistic and its associated p-value are shown in the lower part of Table 5.5. The p-values for all sectors are equal to zero, indicating that we can reject the null hypothesis that at least one of the coefficients of the explanatory variables is equal to zero.

To further investigate the model's performances, we can look at McFadden's R-square, which also can be found in the lower part in Table 5.5. McFadden's R-square is widely varying between 0.08 and 0.36 for the different sectors. According to this performance measure, the fundamentals and macroeconomic indicators can best predict the credit ratings in the Health Care sector with a value of 0.36. It also gives excellent fit in the Consumer Discretionary, Real Estate, Energy, Consumer Staples, Information Technology, Materials and Telecommunication Services sectors with R-square values ranging between 0.2 and 0.3, while the worst fit is obtained in the Utility sector with R-square value of 0.08.

5.2.2 Variable Sensitivity in the Basic Model

The models' estimated coefficients and their significance are presented in the upper part of Table 5.5. In Consumer Staples and Health Care, the credit ratings are only affected by five of the fourteen factors. On the contrary, credit ratings in Consumer Discretionary, Telecommunication Service and Utility sectors are affected by eleven different factors. Consumer Discretionary, Real Estate, Telecommunication Services, and Utilities are affected significantly by most of the macroeconomic variables (6), while Consumer Staples and Health Care are affected by the fewest (1).

By the nature of them being industries that provide products and services with almost zero price elasticity of demand, it is a predicted and an intuitive result that the Consumer Staples and Health Care sectors are not as dependent on the macroeconomic indicators as much as the other sectors. On the contrary, we find that the Consumer Discretionary, producing non-essential and luxurious products and services, displays higher sensitivity to changes in the macroeconomic indicators and the state of the business cycle.

The beta value is a significant risk factor in all sectors except Health Care, Materials, and Real Estate. The idiosyncratic risk and return on capital are the only explanatory variables that are significant to the credit rating in all sectors, while the trimmed mean PCE inflation rate does not affect the credit rating at all in any of them. In terms of the magnitude, we observe that Federal funds rate and Moody's Baa spread are the two macroeconomic variables that are significant in most of the sectors (9), closely followed by the charge off rate and unemployment rate (8). The significance of the unemployment rate is also in line with the findings of Cheung (1996) that the employment rate is one of the most important macroeconomic variables in explaining the changes in the provincial credit ratings in Canada.

*Table 5.5: Regression outputs for the eleven basic ordered probit models. Table 5.5 reports the estimated coefficients for the eleven ordered probit models of credit ratings as the dependent variable and all fundamentals and macroeconomic indicators as independent variables. The models are estimated over the whole sample period 1985-2016. In the analysis, higher ratings are assigned lower numbers. Thus, a negative coefficient indicates that an increase in the variable leads to a higher assigned credit rating and vice versa. * significant at 5%, **significant at 1% and ***significant at 0.1%. The lower part of Table 5.5 presents the number of observations in each test, McFadden's R-square, Likelihood-ratio test, and its corresponding p-value. The sectors in each column are abbreviated in the following way: Consumer Discretionary is CD, Consumer Staples is CS, Energy is E, Financials is F, Health Care is HC, Industrials is I,*

	CD	CS	E	F	HC	I	IT	M	RE	TS	U
Beta	0.23*	-1.71***	-2.11***	-1.51***	-0.11	0.62***	1.16***	-0.02	-0.14	-0.41***	0.25*
Idiosyncratic Risk	0.32***	0.58***	0.68***	0.21***	0.68***	0.27***	0.23***	0.36***	0.45***	0.35***	0.12***
Debt/EBITDA	0.00***	0.00	0.00	0.00	0.00	0.00	0.00	0.00**	0.00	0.00	0.00
CFOD/Debt	-0.46***	-0.27**	-0.50***	-0.28***	-0.85***	-0.06***	0.00	-0.24***	-1.35***	-0.62***	-0.84***
RoC	-9.74***	12.50***	-3.38**	3.45*	-1.69***	-7.44***	-5.26***	-4.93***	-5.28***	-4.10***	-0.15*
EBITDA Margin	-0.25***	0.01	0.17***	-0.05*	-0.65***	0.00	-0.24***	-0.07*	-0.11*	-0.14**	0.04*
Federal Funds rate	-0.24***	-0.03	-0.24***	-0.18***	0.02	-0.15***	0.07***	-0.06***	-0.25***	-0.11***	-0.23***
Real Gross Domestic Product	-0.10***	-0.02	-0.07**	-0.01	-0.01	-0.02	-0.03	-0.02	-0.06*	-0.07***	-0.03*
Charge-Off rate	0.18***	0.02	-0.05	-0.01	-0.09*	0.15***	0.09*	0.11***	0.17***	-0.12***	0.22***
Unemployment rate	-0.06*	-0.02	0.14***	0.10***	0.02	-0.10***	-0.07**	-0.02	-0.11**	0.13***	-0.06***
Moodys Baa spread	-0.24***	-0.08**	-0.35***	-0.22***	-0.02	-0.15***	0.03	-0.09***	-0.22***	-0.12***	-0.25***
Trimmed Mean Inflation rate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Velocity of M2	0.16***	0.02	0.01	0.02	0.02	0.01	0.12***	0.00	0.09	0.10**	0.00
M1 Money Multiplier	0.00	-0.01	-0.02**	-0.01	0.00	-0.01**	0.00	-0.01	-0.02*	0.00	-0.01*
McFadden's R-square	0.296	0.246	0.256	0.144	0.364	0.173	0.255	0.253	0.272	0.197	0.079
LR-test	1374.61	1706.11	635.18	1149.88	1938.73	3306.14	1318.8	2451.92	562.38	946.23	874.74
p-value	0	0	0	0	0	0	0	0	0	0	0
N	1529	2305	960	3210	1715	6489	1549	3370	1179	1588	5057

Table 5.6: The model selection criteria: Akaike information criterion (AIC) and Bayesian information criterion (SBC) for the basic and the refined model. This table consists of the Akaike information criterion and Bayesian information criterion for the basic and refined model. Both the information criteria are calculated for all the eleven sectors. The sectors in each column are abbreviated in the following way: Consumer Discretionary is CD, Consumer Staples is CS, Energy is E, Financials is F, Health Care is HC, Industrials is I, Information Technology is IT, Materials is M, Real Estate is RE, Telecommunication Services is TS and Utilities is U.

	Basic Model										
	CD	CS	E	F	HC	I	IT	M	RE	TS	U
AIC	3309.1	5274.8	1884.2	6896.8	3428.7	15834.4	3900.0	7266.5	1540.0	3888.7	10274.0
BIC	3415.8	5395.4	1971.8	7024.3	3537.6	15990.3	4006.9	7388.9	1621.2	4001.5	10411.1
	Refined Model										
AIC	3305.3	5267.6	1877.4	6888.6	3416.0	15830.4	3894.2	7263.3	1539.5	3883.3	10271.0
BIC	3396.0	5336.6	1945.5	6979.8	3475.9	15952.4	3969.0	7349.0	1595.3	3979.9	10388.6

To further improve the ordered probit models and get more precise results, we remove all the insignificant fundamentals and macroeconomic variables in each sector and re-estimate the probit models. For model selection purposes, when comparing different specifications, we use the AIC and SBC criteria to determine which model specification has the best prediction power. The AIC and SBC values for both the basic and refined models are presented in Table 5.6 where the basic model corresponds to the models in Table 5.5 with all variables included, and the refined model corresponds to the models with the insignificant variables removed.

Both the AIC and SBC values are lower for the refined model compared to the basic model for all sectors, indicating that removing all insignificant variables slightly improves the model specifications. In the following discussion, the refined model is analyzed and compared to the basic model.

5.2.3 Variable Sensitivity in the Refined Model

The refined model is presented in Table 5.7. The Likelihood test and its associated p-value can be seen in the lower part of Table 5.7. The p-values are below 0.05 in all sectors, indicating that there is at least one significant explanatory variable in all sectors. McFadden's R-square is also presented in the lower part of Table 5.7, compared to the basic model in Table 5.5, the R-square values exhibit only slight variation across the sectors. For example, the R-square in the Materials sector is 0.252 compared to 0.253 before. According to this measure, the fit for all the models is classified as excellent, except the models of the Financial, Industrials, Telecommunication Service, and Utility sectors.

Similar to the results in the basic model, Consumer Staples and Health Care are the sectors in the refined model that are least affected by macroeconomic indicators (1), while the Consumer Discretionary, Telecommunication Service and Utility sectors are the ones that are most affected by the macroeconomic indicators (6). Most sectors are affected by five to six macroeconomic indicators, where all macroeconomic variables are significant in four or more sectors each, except for trimmed mean PCE inflation rate, which is insignificant in all sectors.

The idiosyncratic risk is significant in all industries with a positive sign, indicating that an increase in the idiosyncratic risk increases the probability of a downgrade in all sectors. The return on capital is also significant in all sectors, with a negative sign in all industries except in the Consumer Staples and Financials sectors. This positive sign is unexpected since an increase in the return on capital, all else equal, normally leads to a higher probability of a higher rating. Also, EBITDA margin is significant with a negative sign in nine sectors but has a contradicting positive sign in the Energy and Utility sectors.

*Table 5.7: Regression outputs for the eleven different ordered probit models. Table 5.7 reports the estimated coefficients for the eleven ordered probit models of credit ratings as the dependent variable and the refined significant fundamentals and macroeconomic variables previously presented in Table 5.5. The models are estimated over the whole sample period 1985-2016. In the analysis, higher ratings are assigned lower numbers. Thus, a negative coefficient indicates that an increase in the variable leads to a higher assigned credit rating and vice versa. * significant at 5%, **significant at 1% and ***significant at 0.1%. The lower part of Table 5.7 presents the number of observations in each test, McFadden's R-square, Likelihood-ratio test and its corresponding p-value. The sectors in each column are abbreviated in the following way: Consumer Discretionary is CD, Consumer Staples is CS, Energy is E, Financials is F, Health Care is HC, Industrials is I, Information Technology is IT, Materials is M, Real Estate is RE, Telecommunication Services is TS and Utilities is U.*

	CD	CS	E	F	HC	I	IT	M	RE	TS	U
Beta	0.23*	-1.82***	-2.06***	-1.52***		0.61***	1.15***			-0.40***	0.24*
Idiosyncratic Risk	0.32***	0.58***	0.68***	0.21***	0.67***	0.27***	0.23***	0.36***	0.42***	0.35***	0.12***
Debt/EBITDA								0.00**			
CFOD/Debt	-0.46***	-0.25**	-0.50***	-0.28***	-0.84***	-0.06***		-0.24***	-1.33***	-0.62***	-0.83***
RoC	-9.71***	12.38***	-3.56***	3.41*	-1.65***	-7.42***	-5.30***	-4.91***	-5.08***	-4.09***	-0.15*
EBITDA Margin	-0.25***		0.18***	-0.05*	-0.63***		-0.22***	-0.08*	-0.10*	-0.14**	0.04*
Federal Funds rate	-0.24***		-0.24***	-0.17***		-0.15***	0.05***	-0.05***	-0.26***	-0.11***	-0.23***
Real Gross Domestic Product	-0.10***		-0.06***							-0.07***	-0.03***
Charge-Off rate	0.19***				-0.08**	0.15***	0.08**	0.10***	0.17***	-0.12***	0.21***
Unemployment rate	-0.06*		0.12***	0.09***		-0.10***	-0.05**		-0.11**	0.13***	-0.06***
Moodys Baa spread	-0.24***	-0.04**	-0.35***	-0.21***		-0.15***		-0.08***	-0.22***	-0.12***	-0.26***
Trimmed Mean Inflation rate											
Velocity of M2	-0.24***						0.08***			0.11**	
M1 Money Multiplier			-0.02**			-0.01***			-0.02**		-0.01**
McFadden's R-square	0.296	0.244	0.255	0.143	0.363	0.173	0.254	0.252	0.267	0.197	0.079
LR-test	1372.44	1695.21	634.05	1145.99	1933.42	3300.14	1312.64	2443.11	552.89	945.71	871.75
p-value	0	0	0	0	0	0	0	0	0	0	0
N	1529	2305	960	3210	1715	6489	1549	3370	1179	1588	5057

The beta is significant in eight out of the eleven sectors, with a positive and negative sign in four sectors each, which is unexpected. Cash flow from operating activities relative to debt is significant in all sectors except in the Information Technology sector with a negative sign, meaning that an increase in the value of operating cash flow relative to debt, all else equal, in these ten sectors will result in a higher probability of upgrading to a higher credit rating class, which is an intuitive result given that the ratio represents the cash flow available for the firm to meet its financial obligations. Debt relative to EBITDA is only significant in the Material sector with a small positive value.

The Federal funds rate and Moody's Baa spread are significant in the majority of the sectors (9), closely followed by the charge off rate and unemployment rate (8). This is consistent with the importance result from the random forest as well. Recall that the Federal funds rate is the most important macroeconomic indicator according to the mean decrease in accuracy, while the charge off and unemployment rate are the top three important in most sectors. On the other hand, the trimmed mean PCE inflation rate is insignificant in all sectors, which is in line with the low values of the mean decrease in accuracy in Table 5.3 for all sectors.

The Gross Domestic Product is only significant in four (CD, E, TS, and U) sectors with a negative sign according to Table 5.7, indicating that, all else equal, a higher gross domestic product leads to a higher credit rating in all the significant sectors and will not affect the credit rating in any of the insignificant sectors. Moody's Baa spread is also observed to have a negative sign to all of its significant business sectors, indicating that, all else equal, a higher spread leads to a higher probability of credit rating upgrade.

The velocity of M2 and M1 money multiplier are significant in three and four sectors, respectively. The M1 money multiplier is significant for the Energy, Industrials, Real Estate, and the Utility sectors, with a negative sign for all of them, indicating that an increase in the M1 money multiplier, all else equal, leads to a higher probability of a rating upgrade in these sectors while velocity of M2 is only significant for the Consumer Discretionary, Information Technology and the Telecommunication Service sectors with different signs.

5.3 Economic Applications

In this section, we analyze the abstract results presented in the previous section, and we attempt to connect these results to economic applications and business cycle factors in the real world.

5.3.1 Consumer Discretionary

As mentioned earlier in section 4, Consumer Discretionary represents the non-essential consumer goods and services. According to S&P, the sector comprises four main industry groups: Automobiles and Components, Consumer Durables and Apparel, Consumer Services, and Retailing. The sector is broken down into sub-industries such as Automobile Manufacturers, Home Furnishing, Leisure Products, Hotels, Resorts, and Cruise Lines, among others (S&P Global, 2019*a*) . Some of the largest consumer discretionary companies are Starbucks, Nike, and Mercedes Benz. The consumer discretionary sector tends to perform poorly during recessions, and the opposite is true during economic expansions. Consumer discretionary stocks are usually a rational investment option when analysts speculate future economic upturns. In our sample, we find that the most important macroeconomic indicators in the sector are the Federal funds rate, the spread between Moody's seasoned Baa corporate bond, and the charge off rate on consumer loans respectively. The importance of the fund's rate and the spread are attributable to the increase (decrease) in spending level and economic activity in times of low (high) interest rates and cheap (expensive) financing options. The charge off rate on consumer loans takes on a negative notch conveying the following: the lower the proportion of consumers defaulting on consumer loans, the higher the volume of sales of discretionary products expected to take place in an economy and hence, all else equal, a higher probability of the firm to upgrade to a higher rating category, and vice versa.

We also observe the insignificance of M1 money multiplier indicator and the trimmed mean PCE inflation rate. The insignificance of M1 in the consumer discretionary sector is explained by the fact that the luxurious products and services produced by the sector cost on average more than the capacity of M1 components. On the contrary, the velocity of M2 representing the rate of the savings in an economy (See definitions in section three) is another significant factor with a negative notch implying the positive correlation between the volume of discretionary income invested and the ability to purchase discretionary products using these investments.

5.3.2 Consumer Staples

Consumer Staples is the business sector representing essential products and services consumed by the general mass on a daily basis. Industries included below the Consumer Staples business sector are Food & Staples Retailing, Food, Beverage & Tobacco, and Household & Personal Products. The three industries are grouped into sub-industries such as Drug Retail, Food Retail, Brewers, Soft Drinks, Packaged Foods, and Meats, among others (S&P Global, 2019*a*). Some of the largest Consumer Staples companies are Procter Gamble, PepsiCo, Unilever, and Philip Morris, among others.

By the virtue of it being non-cyclical business sector representing the essential products and services, the Consumer Staples sector has little to zero correlation with changes in the macroeconomic conditions and the monetary policy indicators. Consumer Staples' products are in stable demand during bullish and bearish markets, which is explained by the theory of price elasticity of demand. Fueled by the consistent demand of its products, the sector is characterized by stable performance and steady returns during the different stages of the business cycles, and in general, outperforms other business sectors during economic recessions.

Our results in Table 5.4 suggest that inflation rate appears to be the most important macroeconomic indicator in explaining the credit rating classification in the sector, followed by the charge off rate on consumer loans and the effective Federal funds rate. The result is in line with the fact that, in Consumer Staples sector, extra costs resulting from inflation are passed onto consumers, resulting in a shift to cheaper and lower end products in times of economic recessions and higher interest rates.

5.3.3 Energy

The Energy business sector consists of all the companies involved in the process of producing energy products, starting with extraction, manufacturing, refining, and distribution of fuel, and ending with the production and sales of the final energy products (S&P Global, 2019*a*). Energy products in general are inputs to almost every part of the economic activity, affecting macroeconomic indicators such as production, labor market, and real wages, among others. The sector produces products that are considered highly inelastic in today's world. Industries, for example such as Agriculture, Plastics, Chemicals, and Health Care, among others are massive consumers of oil products. The performance of Energy companies, however, is determined in general by fluctuations in the oil prices, which are determined by decisions from oil exporting countries and geopolitical conflicts. Our results in Table 5.4 document that idiosyncratic risk is more than four times as important as the most important macroeconomic indicator, followed by beta, verifying that credit ratings in the Energy business sector are explained by firm-specific risk, industry and market risk rather than macroeconomic indicators.

5.3.4 Financials

The Financial sector is defined by GICS as all the firms that provide financial services to corporate and household customers. In S&P's 2018 latest classification adjustments, the sector is grouped into three industries: Banks, Diversified Financials, and Insurance. The five industries are further broken down into sub-industries such as Diversified Banks, Consumer Finance, and Capital Markets, amongst other sub-industries (S&P Global, 2019*a*). Some of the top financial services brands currently are Morgan Stanley, China Construction Bank, JPMorgan Chase Company, MasterCard Inc., and Ernst & Young LLP.

In the results presented in Table 5.4, we document that the most three crucial macroeconomic indicators in explaining the credit ratings in the Financial sector are the Federal funds rate, the charge off rate on consumer loans, and the spread, respectively. The results come in no surprise given the fact that business activity in the banking sector booms in times of low-interest rates and cheaper financing. Contrarily, the increase in the charge off rate on consumer rates means a higher portion of written off debt, i.e., charge offs, resulting in weak balance sheets and financial statements, which leads to higher leverage ratios and lower financial strength. Financial institutions could end up moving to a lower credit class, and in extreme cases, default happens. Although most of these loans are covered by collaterals, of which ownership is entitled to the bank in case of default, the process is hectic and extraordinarily lengthy, and the bank in the best-case scenario could end up with an asset in a low liquidity horizon.

5.3.5 Health Care

Similar to the Consumer Staples business sector in terms of the necessity of the products and services it provides, the Health Care sector is, to a great extent, insensitive to changes in the macroeconomic indicators. Health Care is considered a defensive-oriented sector, meaning its performance in absolute terms does not exhibit substantial variation across the different phases of the business cycle but will outperform the cycle-oriented business sectors during a recession. A quick look at Table 5.4 would, in fact, indicates that idiosyncratic risk, representing an obligor-specific risk factor, is three times as important as beta, a market-oriented risk factor, in explaining the credit rating and as eight times as important as the effective Federal fund's rate.

5.3.6 Industrials

The Industrials sector is comprised of companies that provide manufacturing and industrial products and services. The sector is grouped into three primary industries: Capital Goods, Commercial and Professional Services, and Transportation. The three industries are further broken down into sub-industries such as Building Products, Electrical Equipment, Aerospace Defense, Airlines, and Marine, among many others (S&P Global, 2019*a*). Unlike the Materials sector, the Industrials sector does not manufacture raw materials. Among the best-performing companies in the Industrials sector in 2019 are General Electric, Masco, TransDignm, Dover, among others (Imbert, 2019).

Being a capital-intensive sector that requires expensive capital expenditures, it comes as no surprise that, among all the explanatory variables, the idiosyncratic risk appears to be the most important variable in explaining the credit ratings. Among the macroeconomic indicators, effective Federal funds rate, charge off rate and unemployment rate are the three most important indicators. The sector is considered to be highly cyclical and is affected not only by macroeconomic indicators but also factors such as trade negotiations, political environment, and overall "global growth environment" (Imbert, 2019). In times of economic turmoil characterized by expensive financing, companies could not maintain their capital investment plans, explaining the high sensitivity to the

Federal funds rate. The importance of the unemployment rate could be interpreted by the fact that Industrials is a labor-intensive sector.

5.3.7 Information Technology

The sector is comprised of companies working on the research and development of technologically based goods and services. The sector is categorized into three industries: Software and Services, Technology Hardware Equipment, Semiconductors Semiconductor Equipment. The three industries are further broken down into sub-industries such as Electronic Equipment Instruments, Application Software, System Software, among others (S&P Global, 2019*a*). In the United States, the sector includes tech giants such as Microsoft, Oracle, IBM, and SAP. Major changes in 2018 have moved companies such as Google and Facebook from the Information Technology sector to the Communications Sector. The sector is characterized, in general, by strong financial statements and low leverage ratios compared to other sectors. The major investment concerns when it comes to the Information Technology sector are factors such as slow growth, increased global competition, and trade disputes. Although being a relatively cyclical sector, Federal funds rate, unemployment rate, and inflation rate appear not to be the most powerful explanatory tools to provide insights regarding credit rating classes in the sector.

5.3.8 Materials

The Materials sector consists of companies in the field of discovering, developing, and processing of raw materials. The industries operating in the business sector are Chemicals, Construction Materials, Containers & Packaging, Metals & Mining, and Paper & Forest Products (S&P Global, 2019*a*). The sector is heavily reliant on the prices of the raw material, whose prices exhibit huge variation during different phases of the business cycle, according to inflation pressures and interest rate changes. As a result, the sector is known to underperform during recessions. Looking at Table 5.4, we can observe that effective Federal fund's rate is the most important macroeconomic indicator, which is in line with the results from the ordered probit model, reported in Table 5.7.

5.3.9 Real Estate

Real Estate is believed to be one of the most complicated business sectors, if not the most, and the relationship between the sector and the business cycle is still ambiguous. While cyclicality factors mostly determine the performances of other business sectors, Real Estate, on the other hand, is assumed to have its own cycles, which determine the phases of the business cycle sometimes. In general, Real Estate is assumed to be one of the outperformers in times of economic expansion and lower financing options, which are mostly accompanied by a higher demand for housing. Contrarily, the sector is believed to underperform during economic contraction and higher financing options, which is typically loaded with low demand. However, real estate prices are assumed to vary according to the expected future prices, known as real estate speculation. This phenomenon is believed, by some researchers to be the reason for the volatility in real estate prices, followed in some cases by real estate bubbles (Malpezzi, S and Wachter S, 2002).

Intrinsically, we expect the performance of the real estate sector, and hence, the behavior of its credit ratings to be reliant on indicators such as the population growth, growth in gross domestic product, and interest rates. Examining the results from our sample, we can find that among macroeconomic indicators, effective Federal funds rate, charge off rate on consumer loans are the two crucial indicators followed by unemployment rate and the spread between Moody's Baa and Federal funds rate, which is almost in line with the probit model results reported Table 5.7.

5.3.10 (Tele) Communication Services

According to GICS, the Communication Services business sector includes companies working in telecommunication Services and Media Entertainment industries. The two industries are broken down into sub-industries such as Wireless Telecommunication Services, Diversified Communication Services, Advertising, Publishing, and Movies & Entertainment, among others (S&P Global, 2019*a*). The sector witnessed a few adjustments in 2018 by the Global Industry Classification System to become a communication services-oriented sector rather than telecommunication sector, coping with the increase in the market weight of communication companies in S&P 500. Being the youngest sector among all the eleven business sectors with less than twenty years of historical data for many companies, analysts find it challenging to forecast movements and performances in the Communication sector.

Communication Services is the sector where we find the most significant variation of results between the probit and the random forest models lies. According to the probit results, all the macroeconomic variables examined are significant, except for the trimmed mean PCE inflation and M1 money multiplier. Contrarily, the random forest results suggest that the effective Federal fund rate, unemployment rate, and charge off rate are the most important indicators, respectively. Since our sample ends in 2016 before the changes in the sector, prior to the fundamental changes in the sector, we expect the performance of this sector to be relatively similar to the performance of the Information Technology.

5.3.11 Utilities

The Utilities sector consists of the list of the companies providing essential services such as water, electricity, sewage, natural gas, among others (S&P Global, 2019*a*). The sector, in general is heavily regulated with high entry barriers. In the majority of the cases, companies operating in the sector are public companies that are characterized by stable performance and consistent dividends (Murphy, 2019). Similar to the

Consumer Staples and the Health Care sectors, the Utilities business sector is a defensive oriented sector. The sector usually overperforms relative to other sectors during the times of economic downturn and the products and services provided are essential products and services whose demand has little to zero correlation with changes in the phases of the business cycle. However, Utilities in general, is a capital-intensive sector, with big-budget capital expenditures and infrastructure that require debt financing. As a result, the sector is characterized by high leverage ratios making it sensitive to changes in interest rates specifically. By analyzing the results, we find that the effective Federal funds rate is as twice important as any other macroeconomic indicator, with the charge off rate coming in the second place, which lies in line with the interest rate risk theory mentioned above.

5.4 Investigating the Contradicting Signs

In this subsection, we use the partial dependence plot to investigate the contradicting signs in some of the explanatory variables in the ordered probit model. The Federal funds rate, for instance, is the most important macroeconomic variable for the classification rules in all the sectors according to the mean decrease in accuracy, and almost equally important in the mean decrease in Gini measure. In addition, the Federal funds rate is highly significant in all the sectors except the Consumer Staples and Health Care sectors, according to the refined probit model. The expected sign of the Federal funds rate is negative, taking into account the lagged-effect of the indicator.

During times of economic expansion, the Federal funds rate is raised by the Federal Reserve System, as a quantitative tightening tool, in an attempt to reduce the economic activity and consumption levels in order to control inflation. Eventually, this leads to lower profitability in the business sector and at last, lower credit ratings. It is worth mentioning that this cycle takes a few years, on average, and hence, it is hard to interpret the direct impact of the indicator's coefficient on the credit ratings instantly. However, in the Information Technology sector, the Federal funds rate has a positive coefficient, according to the refined ordered probit model presented in Table 5.7. To investigate this more in-depth, we can look at the partial dependence plots for each rating class against the Federal funds rate in the Information Technology sector, which are presented in Appendix F for each rating class.

Generally, for higher values of the Federal funds rate above 5, the average predicted probability to be assigned to the two highest rating classes (AAA and AA) increases, indicating that more firms are assigned to the highest rating classes during the times when the Federal funds rate are high. On the other hand, the same values of the Federal funds rate above 5 indicate, on average, that a lower probability is assigned to the ratings A, BBB, and BB since the PDP is decreasing in this region. Since the mass of the partial dependence function is more right-skewed (more of the mass is distributed in the left tail) for the BBB, BB and CCC rating, a decrease in the Federal funds rate to small values means that the average probability to be assigned to these three rating classes increases. The pattern is the opposite for the rating class B, where an increase in the Federal funds rate on all levels leads to a higher probability to be assigned to this rating class. The finding is as well consistent with the positive sign from the probit analysis. Again, we clarify that the model investigates the direct effect of the explanatory variables and does not take into account the lagged-effect of the monetary policy tools in this case.

5.5 Comparing the Two Models

Comparing the results from the ordered probit and the random forest models, many of the results appear to be in agreement for both methods. For example, both models state that the idiosyncratic risk is a significant and essential factor for determining the credit rating across all the sectors and that the Federal funds rate is, typically, an important macroeconomic indicator. On the other hand, some of the other results are in contradiction. For example, the random forest predicts the trimmed mean PCE inflation rate to be the most important macroeconomic indicator in the Consumer Staples sector, according to Table 5.4, while the ordered probit model predicts that the same variable is insignificant in the same industry.

In addition, some results from the ordered probit model are unpredictable. According to Table 5.7, the EBITDA margin is significant with a positive sign in both the Energy and Utility sectors. The result cannot be explained in economic terms since an increase in the EBIDTA margin, all else equal, typically leads to a higher probability of being assigned a higher rating. The same problem is briefly discussed in section 5.4 with the unexpected sign of the Federal funds rate in the Information Technology sector.

As mentioned in section 4.4.4, the ordered probit model suffers from inconsistent results if the unobserved heterogeneity is not normally distributed or independent of the explanatory variables in the regression model. Thus, even if we have an infinite amount of data, we will not be able to estimate the parameters and signs with certainty, which could be one potential reason for the unexpected signs in some of the probit models. Therefore, we believe that the results from the ordered probit model should be interpreted with caution. The findings speak in favor of the random forest model. Of course, in the cases where both models agree, for example regarding the significance and importance of the idiosyncratic risk, the result is more reliable compared to a result where the models draw different conclusions. Since we are more interested in investigating the macroeconomic variables' predicting power to assign firms to a credit rating rather than explaining the causal relationship between the macroeconomic indicators and the credit ratings, the unexplained signs are irrelevant given the high prediction power of the random forest model. In conclusion, the ordered probit, used in many of previous research in analyzing the procyclicality of credit ratings, generates contradicting results in some cases. We interpret the results from the ordered probit with caution and we rely more on the random forest in analyzing the results, especially in economic terms.

6

Conclusion

In this paper, we analyze long-term issuer credit ratings for 299 companies across eleven business sectors in the United States, using both firm-specific risk factors and macroeconomic indicators, with a focus on monetary policy. Our findings suggest that, although credit ratings are sensitive to changes in the business cycle and the macroeconomic conditions, firm-specific risk factors, represented by beta, idiosyncratic risk, profitability and leverage ratios have more explanatory power in determining credit rating classes in the majority of the sectors. Beta and idiosyncratic risk, for example, are more important than any of the most important macroeconomic indicators in determining the credit rating classes, across all the eleven business sector. Furthermore, we find that business sectors respond differently to changes in macroeconomic indicators, with some macroeconomic indicators exhibiting high significance in determining the rating classes across some sectors while others having no explanatory power. We find, for instance, that cycle-oriented sectors respond to changes mainly in interest rates, and charge off rate on consumer loans.

On the contrary, credit ratings in defensive sectors are well-explained by the firm-specific risk factors. In addition, we find that the magnitude of the macroeconomic indicators' importance level exhibits huge variation. The effective Federal funds rate, for example, appears to be the most important macroeconomic indicator across ten out of eleven business sectors. Its importance in absolute terms, however, displays huge variation. Comparing the results between the ordered probit and the random forest models, we find that the two models seem to generate contradicting results regarding a few of the explanatory variables, but, in general, seem to agree regarding the intuitive and reliable results.

Regarding the limitations, the two models could be improved by conducting the analysis using data from more than one country. In this case, the data set will be panel data of credit ratings, firm fundamentals, and macroeconomic indicators. Including data from several markets will result in cross-sectional variation in the macroeconomic indicators, making the relationship easier to investigate. In addition, including more relevant macroeconomic indicators and industry-specific variables for each business sector will enhance the model performance. In this paper, we examine the impact of macroeconomic indicators on the classification of credit rating across different business sectors in the United States. It would be interesting if future research could build on these findings and extend the rating classification framework to also forecast average credit rating upgrades and downgrades for each sector. The model could possibly be conducted by modeling the rating transitions as a function of the macroeconomic indicators in a survival type of study.

Bibliography

Altman, E. (1968), 'Financial ratios, discriminant analysis and the prediction of corporate bankruptcy', *The Journal of Finance* **23**(4), 589–609.

Altman, E. and Kao, D. (1992), 'The implications of corporate bond ratings drift', *Financial Analysts Journal* **48**(3).

Amato, J. and Furfine, C. (2004), 'Are credit ratings procyclical?', *Journal of Banking Finance* **28**(11), 2641–2677.

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), 1389–1413.

Board of Governors of the Federal Reserve System (2018), 'Federal Reserve Board - Open market operations', `https://www.federalreserve.gov/monetarypolicy/bst_openmarketops.htm`. Retrieved on 16/05/2019.

Bonser-Neal, C. and R Morley, T. (1997), 'Does the yield spread predict real economic activity? a multicountry analysis', *Economic Review - Federal Reserve Bank of Kansas City* **Issue Q**(III).

Breiman, L. (2001), 'Random forests', *Stat.berkeley.edu* .

Breiman, L. and Cutler, A. (2003), 'Random forests - classification description', *Stat.berkeley.edu* .

Breiman, L., Friedman, J, H., Olshen, R. A. and Stone, C. J. (1984), *Classification and Regression Trees*, Chapman Hall/CRC.

Brooks, C. (2014), *Introductory Econometrics for Finance*, 3rd edn, Cambridge: Cambridge University Press.

Cheung, S. (1996), 'Provincial credit ratings in canada: An ordered probit analysis', *SSRN Electronic Journal* .

Dudian, M. and Popa, R. (2012), 'Analysis on the relationship between rating and economic growth for the european union emergent economies', *International Journal of Economics and Management Engineering* **6**(4), 431–434.

Duffie, D. and Singleton, K. (2003), *Credit Risk: Pricing, Measurement, and Management*, Princeton University Press.

Enders, W. (2014), *Applied economic time series*, 4th edn, John Wiley Sons Inc.

Fama, E. (1986), 'Term premiums and default premiums in money markets', *Journal of Financial Economics* **17**(1), 175–196.

FRED (2019*a*), 'Effective Federal Funds Rate', `https://fred.stlouisfed.org/series/FEDFUNDS`. Retrieved on 16/05/2019.

FRED (2019*b*), 'St. Louis Adjusted Monetary Base', `https://fred.stlouisfed.org/series/BASE`. Retrieved on 16/05/2019.

FRED (2019*c*), 'Trimmed Mean PCE Inflation Rate', `https://fred.stlouisfed.org/series/PCETRIM1M158SFRBDAL`. Retrieved on 16/05/2019.

FRED (2019*d*), 'Velocity of M2 Money Stock', `https://fred.stlouisfed.org/series/M2V`. Retrieved on 16/05/2019.

Friedman, J. (2001), 'Greedy function approximation: A gradient boosting machine', *The Annals of Statistics* **29**(5), 1189–1232.

Furfine, C. and Lowe, P. (2001), 'Procyclicality of the financial system and financial stability: issues and policy options', *Bank for International Settlements* **1**.

Gini, C. (1912), 'Variabilità e mutabilità', *Memorie di metodologica statistica* .

Guo, L. and Bruneau, C. (2014), 'Macroeconomic variables and default risk: an application of the favar model', *Revue d'économie politique* **124**(5).

Hand, D. and Till, R. (2001), 'A simple generalisation of the area under the roc curve for multiple class classification problems', *Machine Learning* **45**, 171–186.

Hastie, T., Friedman, J. and Tibshirani, R. (2008), *The elements of statistical learning: Data Mining, Inference, and Prediction*, 2nd edn, New York: Springer.

Horrigan, J. (1966), 'The determination of long-term credit standing with financial ratios', *Journal of Accounting Research* **4**, 44–62.

Imbert, F. (2019), 'This sector should post earnings growth double the market, but the trade comes with some risks', `https://www.cnbc.com/2019/03/19/industrials-could-outperform-in-2019-as-their-earnings-grow.html`. Retrieved on 24/05/2019.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013), *An introduction to statistical learning: with Applications in R*, 1st edn, New York: Springer-Verlag New York: Heidelberg Dordrecht London.

Kuhn, M. and Johnson, K. (2013), *Applied predictive modeling*, 1st edn, New York: Springer-Verlag New York Inc.

Liaw, A and Wiener, M (2012), 'Classification and Regression by randomForest', `https://datajobs.com/data-science-repo/Random-Forest-[Liaw-and-Weiner].pdf`. Retrieved on 02/05/2019.

Malpezzi, S and Wachter S (2002), 'The Role of Speculation in Real Estate Cycles ', `http://realestate.wharton.upenn.edu/wp-content/uploads/2017/03/401.pdf`. Retrieved on 22/05/2019.

McFadden, D. (1974), *Conditional logit analysis of qualitative choice behavior*, pp. 105–142.

McFadden, D. (1977), *Cowles Foundation for reserach in Economics*, Cowles.yale.edu.

Murphy, C. B. (2019), 'How the Utilities Sector is used by Investors for Dividends and Safety', `https://www.investopedia.com/terms/u/utilities_sector.asp`. Retrieved on 24/05/2019.

Nickell, P., Perraudin, W. and Varotto, S. (2000), 'Stability of rating transitions', *Journal of Banking Finance* **24**(1-2).

Oshiro, T., Perez, P. and Baranauskas, J. (2012), 'How many trees in a random forest?', *Machine Learning and Data Mining in Pattern Recognition* **7376**(8), 154–168.

Sim, J. and Wright, C. (2005), 'The kappa statistic in reliability studies: Use, interpretation, and sample size requirements', *Physical Therapy* **85**(3), 257–268.

S&P Global (2013), 'Corporate Methodology', `https://www.spratings.com/scenario-builder-portlet/pdfs/CorporateMethodology.pdf`. Retrieved on 16/05/2019.

S&P Global (2019*a*), 'Global Industry Classification Standard', `https://www.spglobal.com/marketintelligence/en/documents/112727-gics-mapbook_2018_v3_letter_digitalspreads.pdf`. Retrieved on 16/05/2019.

S&P Global (2019*b*), 'U.S. Business Cycle Barometer', `https://www.spglobal.com/_media/documents/us-business-cycle-barometer_feb2019.pdf`. Retrieved on 23/05/2019.

SPratings (2018), 'S&P Guide to Credit Rating Essentials What are credit ratings and how do they work?', `https://www.spratings.com/documents/20184/774196Guide_to_Credit_Rating_Essentials_Digital.pdf`. Retrieved on 23/05/2019.

Standard & Poor's (2018), 'S&P Global Ratings Definitions', https://www.standardandpoors.com/en_US/web/guest/article/-/view/sourceId/504352. Retrieved on 16/05/2019.

The U.S. Bureau of Economic Analysis (2019), 'Gross Domestic Product', https://www.bea.gov/data/gdp/gross-domestic-product. Retrieved on 18/05/2019.

Trueck, S. and Racev, S. (2009), *Rating based modeling of credit risk*, 1st edn, Amsterdam [etc.]: Elsevier/Academic Press.

Venkateswaran, B and Ciaburro, C (2019), 'Receiver Operating Characteristic curve', https://www.oreilly.com/library/view/neural-networks-with/9781788397872/bf8d6e44-8ea1-4697-8268-995bac7867bd.xhtml. Retrieved on 24/05/2019.

Wooldridge, J. (2010), *Econometric analysis of cross section and panel data*, 2nd edn, Cambridge, Mass.: MIT.

Appendix A

Stationarity and Multicollinearity

In appendix A, the statistic properties stationarity and multicollinearity will be briefly explained. This section will help to understand the tests we are implementing on the raw observed macroeconomic data.

A.1 Stationarity

A stochastic process is said to be covariance stationary if it has a constant mean and variance for all t and $t-s$, meaning that the mean is finite and all autocovariances depends only on the relative position between the two observations and not on the time (Enders, 2014). Mathematically, a stochastic process X_t is covariance stationary if and only if:

$$E[X_t] = \mu, \forall t > 0 \tag{A.1}$$

$$Cov[X_t, X_{t-s}] = \gamma_s, \forall s \geq 0, t > s \tag{A.2}$$

for all finite μ and γ_s. While a process is said to be strongly stationary if all random variables of the stochastic process have the same distribution not depending on time. Formally, a process is strongly stationary if:

$$F_{t_1+k, t_2+k, ..., t_s+k}(x_1, x_2, ..., x_s) = F_{t_1, t_2, ..., t_s}(x_1, x_2, ..., x_s) \tag{A.3}$$

for all t_i. Meaning that a process X_t is strongly stationary if the joint distribution function of $(x_{t_1+k}, x_{t_2+k}, ..., x_{t_s+k})$ is the same as the one of $(x_{t_1}, x_{t_2}, ..., x_{t_s})$ for any indices $(t_1, t_2, ..., t_s)$ where t_i and k is any integer and s is a positive integer.

There are essential differences between stationary and non-stationary processes. A stationary process has the same statistical properties over time, which is easier to analyze and investigate compared to non-stationary process, where the statistical properties may change over time. For example, shocks to a stationary process are temporarily and the impact will decrease over time, so the process will revert to its long-run mean.

In regression analysis with time series observations, it is essential that the variables are stationary to draw correct conclusions. Using non-stationary variables can lead to the problem of spurious regression, resulting in a false good fit with a high R-square for a regression model when regressing two non-related variables that are both increasing with time (Brooks, 2014). So, the inference drawn from such a model is valueless and the conclusions can be totally wrong. Also, using non-stationary data violates the asymptotic assumptions that will no longer be valid. Both the t-ratios will no longer follow a t-distribution and the standard F-statistic will no longer follow an F-distribution, leading to wrong hypothesis testing and confidence intervals.

The augmented Dickey-Fuller (ADF) test will be used to test whether all macroeconomic variables are stationary. The ADF-test examines if the process has a unit root by regressing the first difference against the first lag and p lagged differences of the dependent variable, with the hypothesis:

$$H_0 : \psi = 0 \tag{A.4}$$

$$H_1 : \psi < 0 \tag{A.5}$$

in

$$\Delta x_t = \psi x_{t-1} + \mu + \sum_{i=1}^{p} \alpha_i \Delta x_{t-i} + u_t \tag{A.6}$$

The test-statistic is the ratio between the estimated ψ and its standard deviation and does not follow the t-distribution under the null hypothesis, because the process under the null hypothesis has a unit root (Enders, 2014). A rejection of the null hypothesis indicates that the process does not have a unit root and is, in fact, stationary. The test is called augmented since the inclusion of the p lagged differences make sure that the error terms u_t not are autocorrelated.

To make sure that the random forest classification and ordered probit model perform well, and in order to make correct statistical inference from the observed data, the macroeconomic time series that the two models are based on must be stationary.

A.2 Multicollinearity

In regression analysis, one wants to estimate the relationship between each of the explanatory variable and the dependent variable. If the explanatory variables are uncorrelated with each other, they are said to be orthogonal to one another, which mean that the coefficient value of the individual effect will not change when adding or removing a variable from the regression equation (Brooks, 2014). On the other hand, if two or more of the explanatory variables are moderate to highly correlated, a problem referred to as multicollinearity happens. The problem of multicollinearity could arise because of structural reasons when creating new predictors from already existing explanatory variables or from data-based observations.

Multicollinearity indicates that a change in one variable is associated with changes in other correlated variables as well, where a stronger correlation makes it more challenging to change one variable without changing the other. In this case, it will be hard for the model to distinguish between the correlated explanatory variables and estimate the marginal contribution of any of the variables. Therefore, the marginal contribution of the explanatory variable in reducing the error sum of squares will depend on which other variables are included in the model. This will result in high standard errors, which in turn leads to wider confidence intervals for individual parameters and inappropriate significance tests that can draw wrong conclusions about the explanatory variables.

The data-based multicollinearity could be reduced by removing one or more of the correlated variables, which could lead to the omitted variable bias if the removed variable was influential in the data generating process of the dependent variable (Brooks, 2014). Another approach to reducing the multicollinearity is to transform the highly correlated variables into a ratio and then use the ratio instead of the individual variables in the model.

Appendix B

The Ordered Probit Model: Consistency Proof

In Appendix B, a formal proof that the ratio $\frac{\beta}{\sigma}$ is consistent in a response model will be presented. The proof follows (Wooldridge, 2010) derivations closely. The following holds for the case of more than two outcomes as well, but for simplicity, only the binary case will be presented.

Suppose that the structural Binary Response model is defined as:

$$P(y = 1 \mid x, c) = \Phi(x\beta + \gamma c_t) \tag{B.1}$$

where x is a $1 \times K$ vector with $x_1 = 1$ and c_t is the unobserved heterogeneity.

Equation (B.1) can then be rewritten in a latent variable form as:

$$y^* = x\beta + \gamma c_t + e \tag{B.2}$$

where $y = 1[y^* > 0]$, $e \mid x, c \sim Normal(0, 1)$ and c is independent of x, $c_t \sim Normal(0, \tau^2)$. If these assumption holds, it follows that $\gamma c_t + e$ is independent of x and is normally distributed and:

$$P(y = 1 \mid x) = P(\gamma c_t + e > -x\beta \mid x) = \Phi(\frac{x\beta}{\sigma}) \tag{B.3}$$

where $\sigma^2 = \gamma^2\tau^2 + 1$. Therefore, the probit of y on x consistently estimates $\frac{\beta}{\sigma}$.

Meaning that

$$plim(\hat{\beta}_j) = \frac{\beta_j}{\sigma} \tag{B.4}$$

because $\sigma = (\gamma^2\tau^2 + 1)^{\frac{1}{2}} > 0$ if $\gamma \neq 0$ and $\tau^2 \neq 0$.

Appendix C

Prediction Matrices

In Appendix C, the prediction matrices are presented for each sector. The matrices represent the predicted rating classification on the vertical axis against the actual rating on the horizon axis, when running the trained random forest models on their respective test set. The prediction accuracy (P-Accuracy) is obtained by summing up the diagonal elements and divide by the total number of observations in each industry. The prediction matrices are associated with the predicted accuracies presented in table 5.2.

Consumer Discretionary									
	AAA	AA	A	BBB	BB	B	CCC	CC	D
AAA	0	0	0	0	0	0	0	0	0
AA	0	35	6	3	0	0	0	0	0
A	0	4	107	7	2	0	0	0	0
BBB	0	0	13	159	8	6	0	0	0
BB	0	0	2	8	36	9	0	0	0
B	0	0	0	0	2	48	1	2	0
CCC	0	0	0	0	0	0	0	0	0
CC	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0

Consumer Staples									
	AAA	AA	A	BBB	BB	B	CCC	CC	D
AAA	15	2	0	0	0	0	0	0	0
AA	4	119	17	4	1	0	0	0	0
A	0	16	235	15	7	0	0	0	0
BBB	0	2	8	150	11	0	0	0	0
BB	0	0	3	9	52	5	0	0	0
B	0	0	0	0	0	16	0	0	0
CCC	0	0	0	0	0	0	0	0	0
CC	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	1

Energy									
	AAA	AA	A	BBB	BB	B	CCC	CC	D
AAA	21	0	0	0	0	0	0	0	0
AA	0	37	7	0	0	0	0	0	0
A	0	0	91	5	1	0	0	0	0
BBB	0	0	11	106	3	0	0	0	0
BB	0	0	0	0	5	0	0	0	0
B	0	0	0	0	0	0	0	0	0
CCC	0	0	0	0	0	0	0	0	0
CC	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0

Financials									
	AAA	AA	A	BBB	BB	B	CCC	CC	D
AAA	22	0	0	0	0	0	0	0	0
AA	0	47	2	0	0	0	0	0	0
A	1	6	230	11	5	1	0	0	0
BBB	0	0	10	132	6	1	0	0	0
BB	0	0	0	5	18	1	1	0	0
B	0	0	0	0	0	8	0	0	0
CCC	0	0	0	0	0	0	0	0	0
CC	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0

Health Care									
	AAA	AA	A	BBB	BB	B	CCC	CC	D
AAA	120	10	5	0	0	0	0	0	0
AA	6	64	3	0	0	0	0	0	0
A	2	7	172	14	1	0	0	0	0
BBB	0	0	2	60	1	0	0	0	0
BB	0	0	2	1	30	0	0	0	0
B	0	0	0	0	0	14	0	0	0
CCC	0	0	0	0	0	0	0	0	0
CC	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0

Industrials									
	AAA	AA	A	BBB	BB	B	CCC	CC	D
AAA	48	2	1	0	0	0	0	0	0
AA	1	180	12	4	0	0	0	0	0
A	0	48	563	91	25	6	0	0	0
BBB	0	21	106	559	45	10	0	0	3
BB	0	1	3	24	132	17	0	2	0
B	0	0	0	2	5	22	1	0	1
CCC	0	0	0	0	0	1	5	1	1
CC	0	0	0	0	0	0	0	4	0
D	0	0	0	0	0	0	0	0	0

Information Technology									
	AAA	AA	A	BBB	BB	B	CCC	CC	D
AAA	28	1	0	0	0	0	0	0	0
AA	0	57	2	0	0	0	0	0	0
A	1	3	150	1	0	0	0	0	0
BBB	0	0	4	74	10	0	0	0	0
BB	0	0	1	4	63	9	1	0	0
B	0	0	0	2	4	50	0	0	0
CCC	0	0	0	0	0	0	0	0	0
CC	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0

Materials									
	AAA	AA	A	BBB	BB	B	CCC	CC	D
AAA	0	0	0	0	0	0	0	0	0
AA	0	55	2	0	0	0	0	0	0
A	0	17	257	23	4	2	0	0	0
BBB	0	7	35	360	40	5	2	0	0
BB	0	0	0	6	113	10	3	0	0
B	0	0	0	3	8	50	3	0	1
CCC	0	0	0	0	0	0	5	0	0
CC	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0

Real Estate									
	AAA	AA	A	BBB	BB	B	CCC	CC	D
AAA	0	0	0	0	0	0	0	0	0
AA	0	0	0	0	0	0	0	0	0
A	0	0	62	3	1	0	0	0	0
BBB	0	0	17	232	2	0	0	0	0
BB	0	0	0	1	36	0	0	0	0
B	0	0	0	0	0	0	0	0	0
CCC	0	0	0	0	0	0	0	0	0
CC	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0

Telecommunication Service									
	AAA	AA	A	BBB	BB	B	CCC	CC	D
AAA	24	1	3	1	0	0	0	0	0
AA	0	3	0	0	0	0	0	0	0
A	6	10	133	25	3	2	0	0	0
BBB	0	0	20	123	5	5	0	0	0
BB	0	0	1	4	45	1	0	0	0
B	0	0	0	0	3	56	2	0	0
CCC	0	0	0	0	0	0	0	0	0
CC	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0

Utility									
	AAA	AA	A	BBB	BB	B	CCC	CC	D
AAA	0	0	0	0	0	0	0	0	0
AA	0	81	14	3	0	1	0	0	0
A	0	23	397	73	3	0	0	0	0
BBB	0	6	131	705	34	2	0	0	0
BB	0	1	0	4	28	0	0	0	0
B	0	0	0	0	0	7	0	0	0
CCC	0	0	0	0	0	0	0	0	0
CC	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	2

Appendix D

The Benchmark Model: Altman Ratios

In Appendix D, the performance measures from both the training- and test set are presented for the benchmarking set of variables. The benchmarking models are estimated for each of the eleven sectors with the optimal value of m and a fixed constant value of n for each of the sectors. In the benchmarking models, the set of macroeconomic indicators are the same, while the firm fundamentals have changed to Long-term Debt over asset, Working capital over Total assets, Retained earnings over Total assets, EBIT over Total assets, Return on assets and Market value of equity over Total liability. The tables in this appendix are constructed similarly and correspond to table 4.1, 5.1, and 5.2.

	CD	CS	E	F	HC	I	IT	M	RE	TS	U
m	14	14	14	14	14	8	8	8	14	8	8
n	1000										

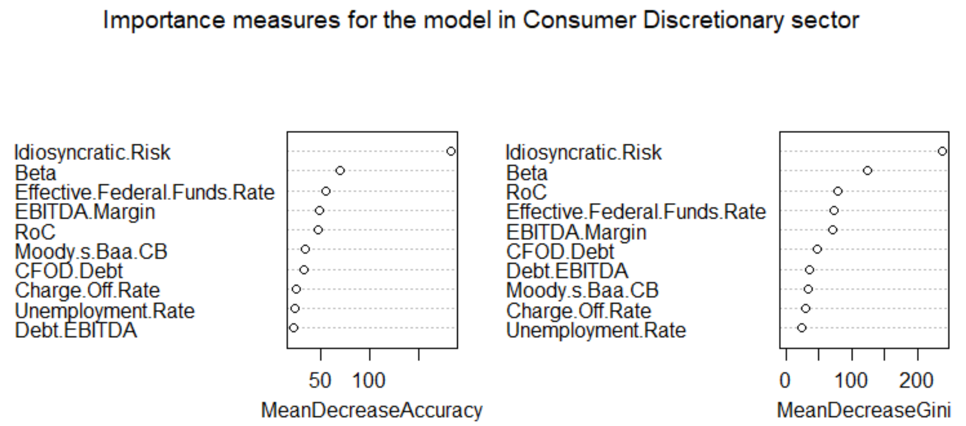
	CD	CS	E	F	HC	I	IT	M	RE	TS	U
Accuracy	0.75	0.79	0.76	0.76	0.79	0.76	0.79	0.75	0.89	0.83	0.73
Kappa	0.68	0.72	0.68	0.62	0.73	0.68	0.74	0.66	0.78	0.78	0.57
Sensitivity	0.54	0.58	0.67	-	0.64	0.69	-	0.56	0.80	-	0.61
Specificity	0.97	0.96	0.96	0.96	0.96	0.97	0.97	0.96	0.94	0.97	0.96
Detection Rate	0.08	0.10	0.08	0.08	0.10	0.08	0.09	0.08	0.22	0.09	0.07

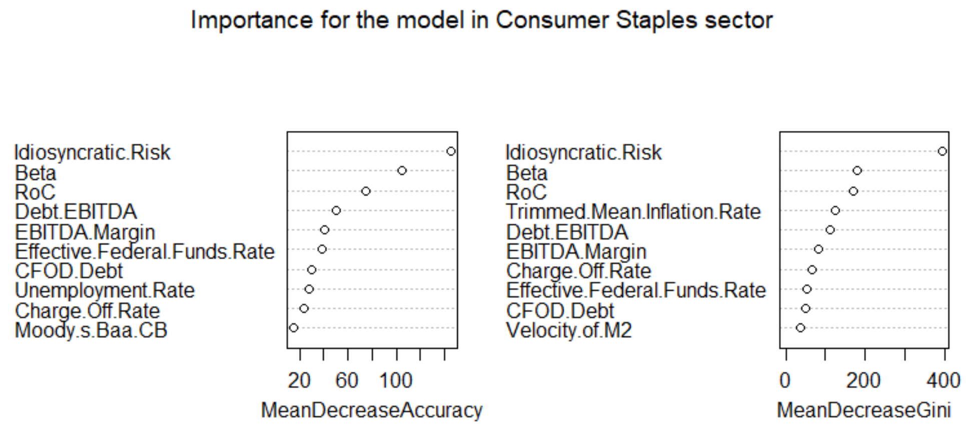
	CD	CS	E	F	HC	I	IT	M	RE	TS	U
OOB	0.25	0.20	0.24	0.23	0.19	0.23	0.19	0.24	0.11	0.16	0.27
P-Accuracy	0.74	0.82	0.77	0.78	0.82	0.79	0.82	0.77	0.91	0.83	0.74
AUC	0.93	0.92	0.95	0.87	0.97	0.90	0.91	0.90	0.90	0.95	0.87

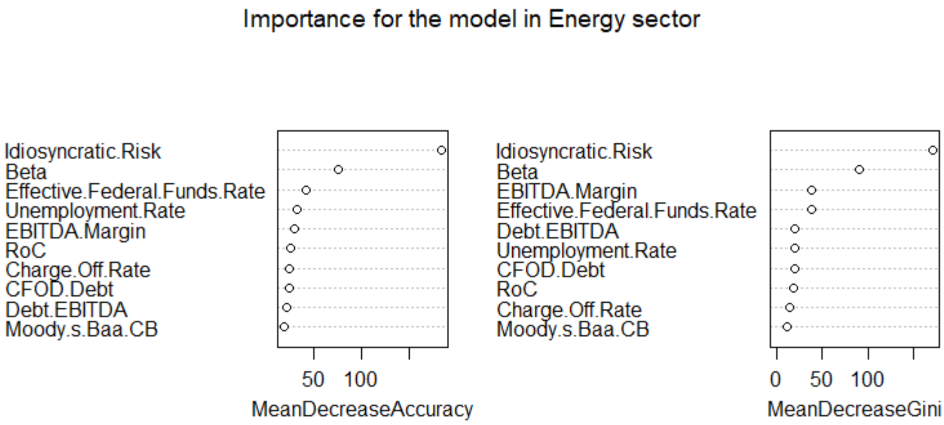
Appendix E

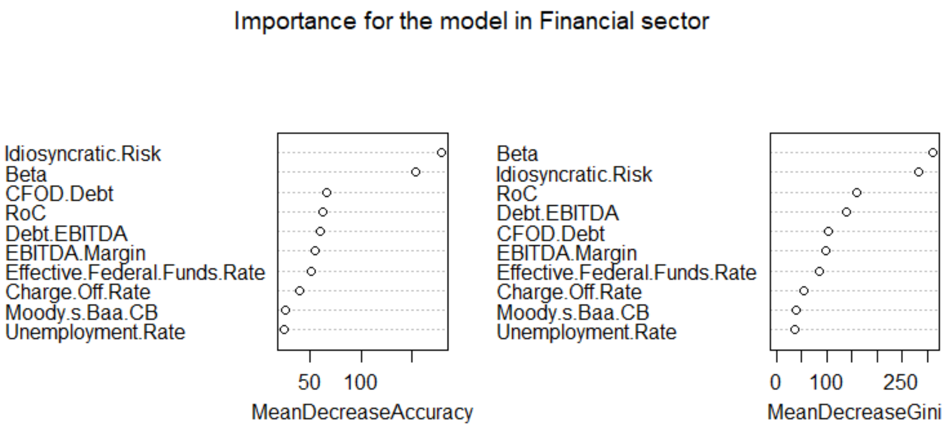
Importance Plots

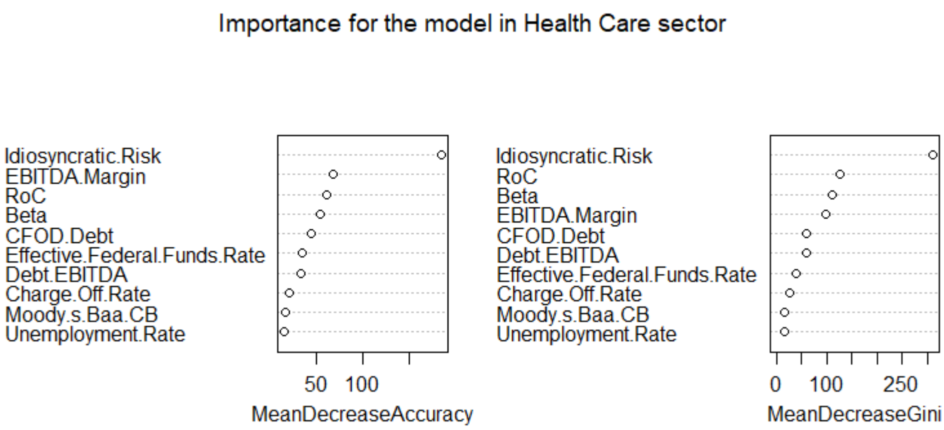
The plots in Appendix E corresponds to the Mean decrease in Accuracy in table 5.3 and Mean decrease in Gini in table 5.4. The Importance plots are simply the ten most important variables from each measure in each industry, ranked from top to bottom.



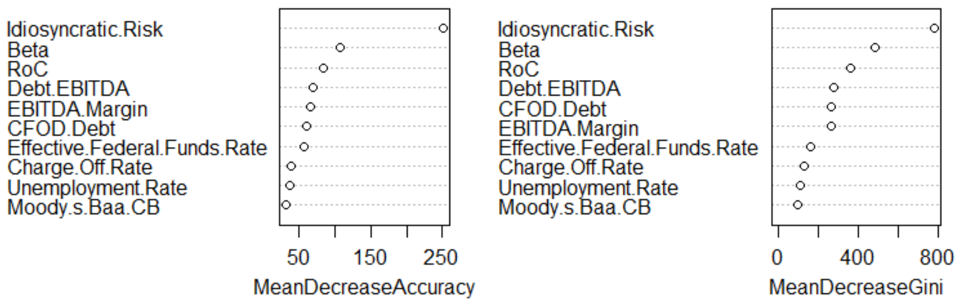



Importance for the model in Energy sector
Idiosyncratic.Risk
Beta
Effective.Federal.Funds.Rate
Unemployment.Rate
EBITDA.Margin
RoC
Charge.Off.Rate
CFOD.Debt
Debt.EBITDA
Moody.s.Baa.CB
50
100
MeanDecreaseAccuracy
Idiosyncratic.Risk
Beta
EBITDA.Margin
Effective.Federal.Funds.Rate
Debt.EBITDA
Unemployment.Rate
CFOD.Debt
RoC
Charge.Off.Rate
Moody.s.Baa.CB
0
50
100
MeanDecreaseGini

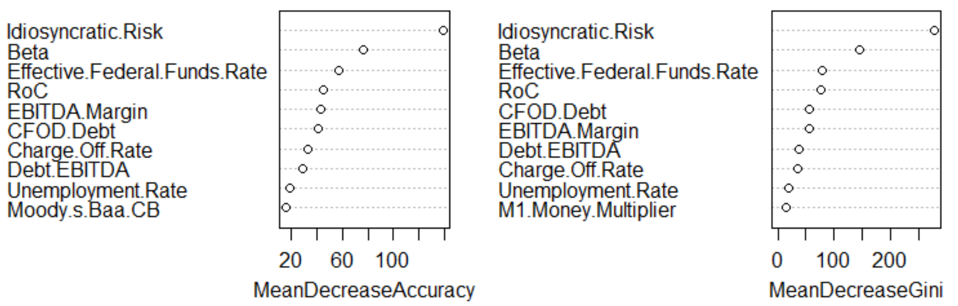

Importance for the model in Financial sector
Idiosyncratic.Risk
Beta
CFOD.Debt
RoC
Debt.EBITDA
EBITDA.Margin
Effective.Federal.Funds.Rate
Charge.Off.Rate
Moody.s.Baa.CB
Unemployment.Rate
50
100
MeanDecreaseAccuracy
Beta
Idiosyncratic.Risk
RoC
Debt.EBITDA
CFOD.Debt
EBITDA.Margin
Effective.Federal.Funds.Rate
Charge.Off.Rate
Moody.s.Baa.CB
Unemployment.Rate
0
100
250
MeanDecreaseGini


Importance for the model in Health Care sector
Idiosyncratic.Risk
EBITDA.Margin
RoC
Beta
CFOD.Debt
Effective.Federal.Funds.Rate
Debt.EBITDA
Charge.Off.Rate
Moody.s.Baa.CB
Unemployment.Rate
50
100
MeanDecreaseAccuracy
Idiosyncratic.Risk
RoC
Beta
EBITDA.Margin
CFOD.Debt
Debt.EBITDA
Effective.Federal.Funds.Rate
Charge.Off.Rate
Moody.s.Baa.CB
Unemployment.Rate
0
100
250
MeanDecreaseGini

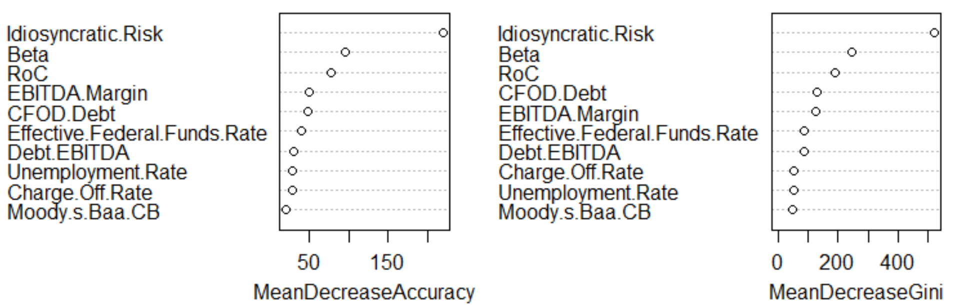
Importance for the model in Industrial sector



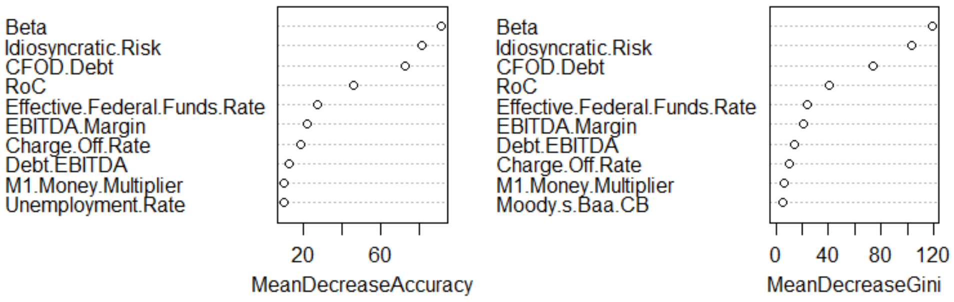
Importance for the model in Information Technology sector



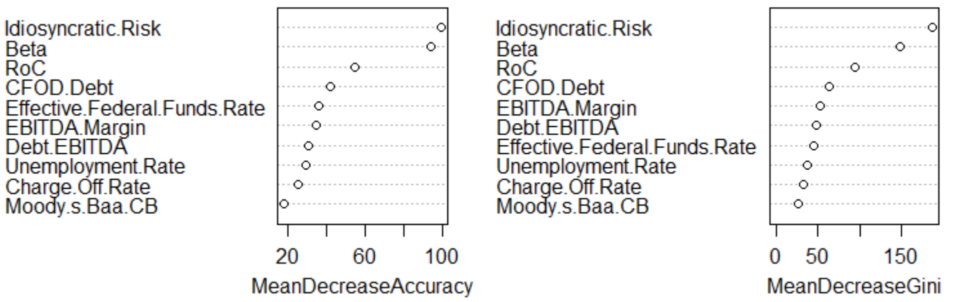
Importance for the model in Materials sector



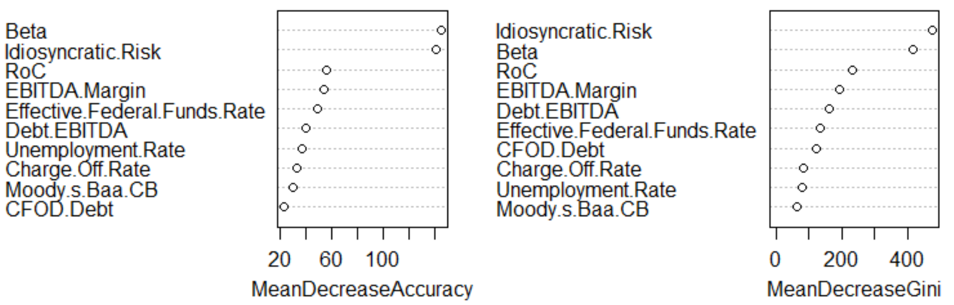
Importance for the model in Real Estate sector



Importance for the model in Telecommunication Service sector



Importance for the model in Utility sector



Appendix F

Partial Dependence Plots

In Appendix F, the partial dependence plots for each of the rating classes in the Information Technology sector is presented for changes in the macroeconomic indicator Federal funds rate. These plots correspond to section 5.4 to further investigate the unexpected positive sign that was obtained in the Information Technology sector from the ordered probit model in table 5.7.