Norwegian School of Economics
Bergen, Fall 2024

Rating the Raters

Investigating Credit Rating Agency Behavior

Kasper Rau Andersen og Theodor Bjerkvik

Supervisor: Øivind Anti Nilsen

Master thesis

Master of Science in Economics and Business Administration

Major: Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

This thesis was written in the fall of 2024 as part of our MSc in Economics and Business Administration at the Norwegian School of Economics (NHH).

We wish to express our gratitude to all who contributed to the completion of our master's thesis in Financial Economics.

First and foremost, we express our gratitude to our supervisor, Øivind Anti Nilsen, for his feedback, guidance, and support throughout the research and writing process. We appreciate the time and energy he devoted to encouraging us and challenge us to make the thesis the best it could be.

We wish to express our gratitude to each other for fostering a great partnership, which made this semester both productive and engaging, and for the friendships we have built with fellow students during our time at NHH.

Finally, we would like to express our gratitude to our families, for their support and understanding during this demanding but rewarding experience.

Norwegian School of Economics
Bergen, 2024

Kasper Rau Andersen

Theodor Bjerkvik

Abstract

This thesis investigates the role of credit ratings in financial markets and their reliability and consistency in assessing corporate default risk. Credit ratings serve as vital tools for reducing information asymmetry between issuers and investors, influencing investment decisions, corporate financing, and regulatory frameworks. While most previous studies analyze aggregated credit ratings across agencies, this thesis adds to the literature by comparing the performance of Fitch Ratings (Fitch) and S&P Global (S&P) in assessing corporate credit risk. To achieve this, we build upon the work of Machek and Hnilica (2013) and Engelmann, Hayden, and Tasche (2003) to evaluate rating performance, using credit rating adjustments from 2014 to 2023.

The findings reveal that there are disparities between Fitch and S&P. Specifically, there are statistically significant differences in the default probabilities associated with similar ratings issued by the agencies. Furthermore, the results indicate that Fitch more frequently positions companies approaching default in the poorest rating categories, whereas S&P tends to distribute these companies more broadly across the speculative grade. Lastly, in the year leading up to default, our findings reveal that Fitch issues larger downgrades upfront, while S&P applies more gradually larger downgrades as default approaches.

Table of Contents

List of Figures

List of Tables

1.0 Introduction

Credit Rating Agencies (CRAs) play an essential role in financial markets by providing credit risk assessments that address the issue of asymmetric information. In an environment where investors and stakeholders often lack the tools or expertise to independently evaluate issuers, CRAs offer a systematic approach to assessing risk. These assessments are vital for enhancing transparency and comparability, supporting informed decision-making, and ensuring the effective allocation of capital in increasingly complex financial systems.

This thesis investigates whether Fitch and S&P, two of the leading CRAs, differ in their ability to evaluate credit risk and identify companies at risk of default. Specifically, it examines the placement of defaulting companies within rating categories and the patterns of downgrades leading up to default events. By exploring these differences, the study provides valuable insights into how rating practices influence perceptions of risk and shape the decisions of market participants.

To analyze these topics the thesis employs a robust quantitative framework, incorporating Markov transition matrices, Cumulative Accuracy Profiles (CAP), and Receiver Operating Characteristic (ROC) curves to evaluate rating performance. The analysis is based on a dataset of 13,983 credit rating updates for all U.S.-based companies rated by either S&P or Fitch from 2014 to 2023. By focusing on long-term issuer ratings, the study ensures a consistent and meaningful comparison across both agencies.

The thesis is structured as follows. Section 2 provides a historical overview of CRAs and the credit rating industry, highlighting their role in addressing asymmetric information in financial markets. Section 3 presents the theoretical framework, reviewing relevant literature and introducing the research hypotheses. Section 4 describes the dataset, key variables, and agency-specific comparisons, while Section 5 outlines the methodological framework, including the transition matrix model, CAP, and ROC analysis. Section 6 details the results and discussion, including rating patterns and discriminatory power assessments. Finally, Section 7 present the conclusion of our study. The thesis concludes with the bibliography and appendix, which include supporting data and materials used.

2.0 Background

The following chapter present a description of the history of credit ratings, and how the credit rating industry operates. Furthermore, we explain why we have explicitly chosen to only evaluate rating changes for US-companies.

2.1 History of Credit Rating Agencies

The origins of credit ratings trace back to the economic turbulence of the 19th century, particularly following the Panic of 1837. The financial crisis exposed significant flaws in informal systems of determining creditworthiness, which relied heavily on personal relationships and subjective judgments. These challenges stemmed from information asymmetry, where creditors lacked sufficient insight into borrowers' financial conditions, leading to inefficient and risky lending practices. CRAs emerged to address this issue by providing standardized and objective assessments of credit risk. (Listokin and Taibleson, 2010). In response to, Lewis Tappan, a silk merchant who faced financial ruin during the crisis, founded The Mercantile Agency in 1841 (Partnoy, 1999). Tappan's agency represented a shift by providing systematic and detailed credit reports collected from correspondents across the country. These reports evaluated the financial health of businesses, marking the birth of structured credit risk assessment. Historian James H. Madison highlighted the Mercantile Agency's innovative service to American businessmen, describing it as a fundamental shift in risk evaluation (Grinder et al., 2017).

In 1909, John Moody revolutionized the financial landscape by introducing the first publicly accessible bond ratings, focusing exclusively on railroad bonds. This marked the inception of modern credit ratings, where Moody's firm evaluated the creditworthiness of bonds, offering insights for investors during an era of limited transparency (White, 2010). These ratings provided a systematic approach to assessing risk, a significant advancement from informal credit evaluations prevalent before Moody's work. To support this innovation, Moody adopted an "investor-pays" model, where investors paid for access to bond ratings, setting the foundation for monetizing credit rating services (White, 2010).

However, significant changes in the industry occurred in the 1970s with the shift from an "investor-pays" to an "issuer-pays" model. White (2010) attributes this transition to several factors: fears of revenue loss from widespread photocopying, increased issuer demand for ratings after the 1970 Penn-Central bankruptcy, regulatory changes requiring ratings for institutional portfolios, and the nature of bond rating as a two-sided market. As a consequence of the reasons outlined by White (2010), it became standard practice for bond issuers to pay for their ratings.

2.2 Credit Rating Industry

The credit rating industry is dominated by three major players: S&P, Moody's Investors Service, and Fitch, which is collectively referred to as the "Big Three." Together, they account for over 95% of global credit rating revenues (Vipond, T., n.d.), reflecting their established role in the financial system. These agencies assess the creditworthiness of issuers, including corporations, municipalities, and sovereign nations, providing ratings that range from high-grade investments to speculative categories. (Vipond, T., n.d.)

Credit ratings play a vital role in financial markets by enhancing transparency and providing an independent assessment of credit risk, which helps mitigate information asymmetry between issuers and investors (Moody's Analytics, 2003). By ranking the likelihood of default, ratings influence investor confidence and borrowing terms. Higher ratings, such as AAA or AA, signal financial stability and repayment capacity, reducing borrowing costs and broadening access to capital, while lower ratings increase costs and restrict access to financing (Claessens, Law, Wang, 2018). Furthermore, credit ratings hold regulatory significance as they are integrated into frameworks such as capital requirements, aligning risk assessments with regulatory compliance (Basel Committee on Banking Supervision, 2010).

Credit ratings are categorized into two broad segments:

1. *Investment-Grade Ratings*

Indicating low default risk, these ratings (AAA to BBB-) indicate relatively low to moderate credit risk (Fitch Ratings, n.d. ; S&P Global Ratings, n.d.).

2. *Speculative or "Junk" Ratings*

Representing higher risk, these ratings (BB+ to D) indicate higher levers of credit risk (Fitch Ratings, n.d. ; S&P Global Ratings, n.d.).

Table 2.2: *Rating System*

S&P and Fitch	*Numerical Value*	*Meaning*	
AAA	1	Highest quality	
AA+	2	High quality	
AA	3		
AA-	4		
A+	5	Strong payment capacity	
A	6		
A-	7		
BBB+	8	Adequate payment capacity	
BBB	9		
BBB-	10		INVESTMENT GRADE
BB+	11	Likely to repay; ongoing	SPECULATIVE GRADE, NON-
BB	12	uncertainty	INVESTMENT GRADE
BB-	13		
B+	14	High risk obligations	
B	15		
B-	16		
CCC+	17	Vulnerable to default	
CCC	18		
CCC-	19		
CC	20	Default imminent with little prospect for recovery	
C	21		
D	22	In bankruptcy, or default	

Credit ratings are systematically divided into investment grade and speculative grade categories, ranging from the highest rating, AAA, to default, D. Investment-grade ratings include levels, AAA, AA, A, and BBB, indicating lower credit risk. Speculative-grade ratings, include BB, B, CCC, and lower, reflecting higher risk. Each rating level is further refined with modifiers like + and – to denote relative strength within the category. This structure allows for direct comparison of ratings from agencies like Fitch and S&P, as their methodologies, while independently defined, produce similar thresholds and implications (Fitch Ratings, n.d.; S&P Global Ratings, n.d.).

The credit rating industry has faced significant criticism due to the issuer-pays model, where agencies are paid by the companies of which they issue ratings. Evidence shows that issuers are willing to pay a premium to rating agencies to inflate their ratings, highlighting concerns about conflicts of interest and the potential compromise of objectivity in the ratings process (Partnoy, 1999). Furthermore, the 2008 financial crisis highlighted vulnerabilities in the reliance on CRAs, particularly concerning the inaccurate ratings of structured financial products. This failure led to widespread criticism of CRAs and underscored conflicts of interest in their methodologies (Bush, C., 2022). In response, regulatory frameworks such as the *Dodd-Frank Act* in the United States and new *European Union regulations* were introduced to improve transparency, address conflicts of interest, and enhance accountability in the practices of CRAs (Binici et al., 2018).

Credit ratings remain central to modern financial systems, despite ongoing questions about their informational value. Binici et al. (2018) find that while the market's reaction to credit rating announcements has diminished since the 2008 financial crisis, the ratings continue to play a significant role in financial systems. Similarly, Partnoy (1999) argues that their prominence stems more from regulatory reliance than their ability to assess risk accurately. The introduction of Basel II by the Basel Committee on Banking Supervision in 2004 further ingrained the use of credit ratings in banking practices. Under this framework, banks were allowed to use ratings assigned by recognized CRAs to determine credit risk weights for many of their institutional credit exposures, aligning minimum capital requirements more closely to the risk of economic loss (Sy, 2009). This regulatory reliance has established credit ratings as an essential component in financial systems.

2.3 Why US Based Companies

CRAs aim to apply standardized methodologies to ensure consistent credit assessments, aligning with the principles of Code of Conduct Fundamentals for CRAs to enhance the reliability of external ratings across sectors and regions (Basel Committee on Banking Supervision, 2010). However, a study from Moody's Analytics (2003) found that Continental European ratings provide a more accurate rank ordering of default risk compared to U.S. ratings. The study highlights that European ratings exhibit stronger correlations with market-based credit measures, suggesting regional differences in rating accuracy and effectiveness (Moody's Analytics, 2003).

To control for variations, this study focuses on U.S.-based companies. The standardized regulatory framework in the United States, overseen by the Securities and Exchange Commission (SEC) and the Financial Accounting Standards Board (FASB), ensures consistent financial reporting through the application of Generally Accepted Accounting Principles (GAAP). The SEC enforces compliance with GAAP, while the FASB is responsible for developing and updating these standards, promoting transparency and comparability in financial disclosures to support efficient markets (U.S. Securities and Exchange Commission, 2002).

3.0 Theoretical Framework and Literature Review

One of the tools to evaluate the predictive powers of CRAs is calculating the default probability for each rating category, while also making CAP, and ROC curves. These measures assess how effectively CRAs discriminate between defaulters and non-defaulters, offering valuable insights into their differences. This section reviews studies that have employed CAP and ROC curves to evaluate credit ratings and identifies gaps in the literature, including the lack of direct comparisons between agencies.

3.1 Through-the-Cycle vs. Point-in-Time Ratings

CRAs use two primary methodologies to assign ratings: Through-the-Cycle (TTC) and Point-in-Time (PIT) (Kiff et al., 2013). The TTC approach aims to balance the need for stable ratings with accurate default estimates by smoothing out short-term volatility, ensuring that ratings remain relatively stable over time (Kiff et al., 2013). In contrast, the PIT approach focuses on current information to estimate default risk, providing a more accurate and timely reflection of an issuer's immediate creditworthiness but resulting in more frequent rating changes (Kiff et al., 2013).

In an interview, Rui Pereira, Fitch's Head of Structured Finance for North America, explains that they use a TTC approach for their top three rating categories, focusing on long-term creditworthiness and aiming to provide stable ratings that withstand economic fluctuations. However, for ratings below single A, Fitch's ratings become more responsive to macroeconomic changes (Fitch Ratings, 2020), reflecting a PIT approach.

In 2003, S&P emphasized the importance of rating stability, stating that the value of its ratings lies in focusing on the long term and avoiding fluctuations due to short-term performance changes (Altman & Rijken, 2006). However, recent statements reflect a shift toward prioritizing timeliness, as S&P now emphasizes a forward-looking approach that promptly adjusts ratings when their credit quality outlook changes. Rather than waiting for economic cycles to play out, S&P now highlights the importance of ensuring ratings reflect up-to-date information, arguing that markets function best when participants have access to current credit opinions and evolving risk assessments (S&P Global Ratings, 2020). This approach aligns more closely with PIT, as it reflects immediate credit risks and current conditions.

No agency operates exclusively within a TTC or PIT framework. Even agencies valuing long term stability within its ratings, respond to significant economic shifts to account for altered repayment capacities (Kiff et al., 2013). Historically, a TTC approach have been less reactive to macroeconomic changes, often delaying rating adjustments. However, advancements in technology and improved data access enable CRAs to update more frequently and employ a more PIT-oriented approach when necessary (Kiff et al., 2013). This focus on CRAs valuing timeliness is important for the research framework employed in this thesis. The PIT approach is designed to deliver timely credit indicators. If an agency were to use a TTC approach it would make it difficult for us to use a timeframe, because new ratings may have been delayed and therefore issued later than they perhaps should have. This focus on timely credit indicators used by both agencies enables us to use a set timeframe when calculating the default probabilities of each rating, and when designing the CAP and ROC models, as rating adjustments are issued timely.

3.2 Markov Chains Transition Matrix

Andrey Andreyevich Markov, a Russian mathematician, developed and formalized the concept of Markov Chains, a foundational concept in probability theory. His work primarily focuses on extending the Law of Large Numbers to stochastically dependent sequences, which later laid the foundation for Markov Chains as a mathematical concept used in probability theory. (Seneta, 2006).

The Markov Chains Transition Matrix is based on the Markov Property, which assumes that the probability of moving to a future state depends only on the current state (Machek & Hnilica, 2013). Markov Chains Transition Matrices are in this thesis used to analyze credit rating changes. In other words, capture upgrades, downgrades, and defaults from each state. Belkin et al. (1998) and Kiefer and Larson (2004) showcase their effectiveness in modeling rating migrations and estimating risk transitions, helping to understand how different ratings transition into others. Machek and Hnilica (2013) expand on this by exploring default probabilities and corporate lifespans, though their reliance on specific assumptions may limit flexibility in capturing diverse credit scenarios. Our approach directly examines observed transitions between all ratings for both S&P and Fitch. This method focuses on estimating credit stability, default probabilities, and rating trends based on our dataset. By centering on actual transitions, our framework provides a practical and adaptable tool for understanding credit risk dynamics.

3.3 Default Probabilities and Default Ratings

Credit ratings serve as ordinal indicators of credit risk, where lower ratings correspond to higher probabilities of default (Cantor & Mann, 2003). S&P publishes default probabilities based on issuers' ratings as of the start of each calendar year, tracking defaults over the subsequent year (S&P Global Ratings, 2024b). While widely used, this approach has limitations, as it does not account for the precise timing of rating assignments within the year. This may distort the relationship between newly assigned ratings and their observed default probabilities. To address this, our study calculates default probabilities based on a fixed 365-day horizon following the assignment of a new rating. This ensures that the same fixed timeframe of the issued rating is incorporated, providing a more accurate measure of default risk.

Furthermore, while S&P's reports, focus solely on its own ratings, as far as we are aware, no existing study compares default probabilities across rating agencies. Our study conducts a comparative analysis of Fitch and S&P, offering new insights into potential inter-agency differences.

Based on this, the following hypothesis has been made:

H1: Default probabilities for comparable ratings do not differ significantly between Fitch and S&P.

3.4 CAP and ROC

Among the tools used to measure their predictive power are the CAP and ROC curves. These metrics provide insights into how well CRAs differentiate between defaulters and non-defaulters, highlighting the preciseness of their issued ratings.

The Accuracy Ratio (AR), derived from the CAP model measures the ability of a rating system to distinguish between defaulters and non-defaulters. An AR of 1 indicates perfect discrimination, while an AR of 0.5 reflects performance equivalent to random guessing

(Engelmann et al., 2003). In contrast, the ROC curve is constructed by plotting the hit rate (true positive rate) against the false alarm rate (false positive rate) across all possible thresholds, providing a graphical representation of a rating system's ability to discriminate between defaults and non-defaults. The performance of the model improves as the ROC curve steepens at the left end and approaches the point (0,1). The AUC summarizes this performance, with a larger AUC indicating stronger discriminatory power across all threshold values (Çorbacıoğlu et al,. 2023). A higher AUC value indicates greater predictive accuracy, making ROC curves an effective tool for evaluating credit risk models (Krämer & Güttler, 2008).

The CAP curve rises cumulatively, measuring how defaults are placed throughout all ratings. A perfect 90-degree slope would require all defaulters to be ranked strictly in the lowest category, however, defaults are spread across different ratings, making the perfect model less steep in its initial phase. In contrast, the ROC curve appears as a sharp 90-degree slope because it evaluates performance rating by rating. It rewards agencies that place defaulters in the worst categories, as this maximizes the True Positive Rate (TPR) at the lowest thresholds. This steep initial rise occurs because ROC focuses on how quickly defaults are captured, rather than their cumulative order. Using both models allows us to identify differences between agencies: CAP highlights an agency's ability to rank defaults consistently across categories, while ROC shows how effectively defaults are isolated in the poorest ratings. Additionally, Irwin, R. J., & Irwin, T. C. (2013), points out that the ROC is often considered more practical due to its established theoretical models for fitting curves and its robust framework for determining optimal decision thresholds, enhancing its application in evaluating credit ratings.

3.4.1 Previous Applications of CAP and ROC in Default Prediction

Engelmann, Hayden, and Tasche (2003) introduced the CAP as a widely used tool to evaluate the discriminative power of credit rating systems, summarizing their performance through the Accuracy Ratio (Engelmann et al., 2003). They demonstrated that the CAP effectively captures a rating system's ability to differentiate between defaulters and non-defaulters by comparing its performance to that of random models and perfect models (Engelmann et al, 2003). However, the study relied on aggregated data, which does not differentiate between CRAs (Engelmann et al., 2003).

Krämer and Güttler (2008) examined the predictive power of credit ratings, focusing on Moody's and S&P. Using ROC curves and accuracy ratios, they confirmed that ratings from both agencies were valuable predictors of default. However, the study assumed equivalence in different rating scales and definitions, which may have obscured structural differences between the agencies. While split ratings were noted, the study did not explore whether differences in predictive accuracy arose from actual performance disparities or methodological variations (Krämer & Güttler, 2008). Instead, the primary focus remained on assessing the overall reliability of credit ratings rather than providing an in-depth comparison of inter-agency performance.

3.4.2 Identified Gaps in the Literature

While existing research demonstrates the usefulness of Markov models, default probabilities, CAP and ROC curves, gaps remain. There is a lack of systematic inter-agency comparisons, particularly regarding rating transitions, default probabilities, and predictive accuracy. Studies often focus on specific approaches or rely on aggregated data, limiting their practical applicability. Our study addresses these gaps by offering new insights into differences between S&P and Fitch, including how effectively each agency identifies and places defaulting companies within their rating systems. Based on the literature the following hypothesis has been made:

H2: Both agencies demonstrate predictive accuracy that is better than random guessing.

4.0 Data

This chapter provides an overview of the dataset and variables used to evaluate the predictive performance of S&P and Fitch credit ratings. The dataset covers 13,983 credit rating changes for 4,341 U.S.-based companies from 2014 to 2023. It includes downgrades, upgrades, and watch instances, where ratings are reassessed without changes. Watch instances are used by CRAs to signal heightened uncertainty or potential adjustments. For example, S&P employs "CreditWatch" to indicate a review of a rating for a possible upgrade, downgrade, or change in outlook, often resolving these within one to three months (Bhatia, 2002). The data focuses on ratings issued by S&P and Fitch. It also includes descriptive statistics highlighting the distribution of rating changes, defaults, and upgrades, an agency comparison to assess differences in rating practices, and a discussion of data limitations.

4.1 Data Sources

Data was extracted from the Bloomberg Terminal (Bloomberg L.P., 2024), to analyze all credit rating adjustments within our timeframe. The dataset includes S&P's Long-Term Foreign Issuer Credit Rating and Fitch's Long-Term Issuer Default Rating.

4.1.1 Types of Ratings

- *S&P's Long-Term Foreign Issuer Credit Rating:*

The rating evaluates a company's ability to meet its long-term financial obligations denominated in foreign currencies. It considers factors such as international credit risk, exposure to foreign exchange fluctuations, and broader economic conditions in global markets. The rating serves as an indicator of the issuer's overall capacity and willingness to repay foreign currency debt (S&P Global Ratings, n.d.).

- *Fitch's Long-Term Issuer Default Rating:*

The rating assesses an entity's likelihood of defaulting on its financial commitments, irrespective of currency denomination. It provides a broad view of the issuer's overall credit risk by integrating macroeconomic conditions, industry trends, and firm-specific financial factors (Fitch Ratings, n.d.).

While there are nuances in the focus and methodology of each rating type, both are reflecting a company's financial stability and susceptibility to credit risk. They are both widely used by investors, creditors, and other stakeholders to understand the default risk associated with long-term debt obligations.

4.1.2 Data Collection and Criteria

Time frame

The period 2014–2023 was chosen to ensure relevance by focusing on recent economic events while avoiding biases from earlier data. Reforms introduced after the 2008 financial crisis, such as those adopted by the SEC in 2009, significantly changed CRA practices. These reforms aimed to improve transparency, strengthen disclosure requirements, and address conflicts of interest (Hathaway et al., 2010). Including data from before these changes could skew results, as earlier methodologies may not align with the enhanced standards now governing credit rating processes. This timeframe ensures the analysis reflects modern and reformed CRA practices.

Rating Changes

All upgrades, downgrades and watches were included from the period.

Geographical Focus

Only U.S.-based companies were included to maintain consistency in economic and regulatory environments.

Only companies with at least one credit rating change during the study period were included, and incomplete or non-relevant data was excluded to maintain dataset integrity.

4.1.3 Variables

This section details the variables created to analyze credit rating adjustments, including initial and updated ratings, days relative to the default event, size of credit rating adjustments, and a binary variable for whether a company defaulted or not. These variables form the foundation for our analysis, when assessing whether CRAs issues appropriate ratings.

Days relative to default event: Captured as *"daysuntildefault"*.

The variable *"daysuntildefault"* provides a measure of timing relative to a company's default event. For companies that default, it records the number of days between each previous rating change and the default date. For instance, if a downgrade occurs 100 days before default, *"daysuntildefault"* is set to (*100)* for that recorded rating change, while it is set to (*0)* on the actual default date. This allows for a clear timeline of how close each rating change is to the default event. For non-defaulting companies, "d*aysuntildefault"* is uniformly set to (*0)*, as these companies do not default within our timeframe. This variable is essential for analyzing the timing and effectiveness of issuing rating changes leading up to default events. It enables proximity analysis, showing whether downgrades occur closer to default events. Additionally, it can be used to evaluate whether CRAs issue timely and predictive downgrades, offering insight into the performance of credit rating adjustments in the lead-up to defaults.

Initial Credit Rating: Captured as *"ratingfrom"* and *"ratingfromnumeric"*

The variable "r*atingfrom"* captures the initial credit rating assigned to a company before any rating adjustment occurs. It reflects the starting point for a rating change, providing insight into the company's creditworthiness as assessed by the CRA at that time. This categorical variable is recorded in the standard format used by the agency, such as BBB- or A+.

To facilitate quantitative analysis, *"ratingfromnumeric"* converts these categorical ratings into numerical values. The different values are detailed in Table 2.2. This numerical representation allows for easier integration into statistical models and calculations, enabling comparisons and analyses based on the size or progression of rating changes.

Updated Credit Rating: Captured as *"ratingto"* and *"ratingtonumeric"*

"ratingto" denotes the credit rating assigned to a company following a rating adjustment, representing the revised evaluation of its creditworthiness. To support quantitative analysis, *"ratingtonumeric"* translates this updated rating into a numerical format, enabling comparison with the initial rating captured by *"ratingfromnumeric."* Together, these

variables provide a clear view of the change in credit rating, allowing for an analysis of the size and direction of rating adjustments. The different values for *"ratingtonumeric"* and *"ratingfromnumeric"* are detailed in Table 2.2.

Size of Rating Adjustments: Represented as *"ratingchangenumeric"*

"ratingchangenumeric" measures the size and direction of a credit rating adjustment by calculating the difference between *"ratingtonumeric"* (the updated rating) and *"ratingfromnumeric"* (the initial credit rating). Positive values indicate a downgrade, where the creditworthiness of a company has decreased, reflecting heightened default risk. Negative values represent an upgrade, signifying improved creditworthiness and reduced perceived risk. This variable captures not only the direction but also the severity of rating adjustments.

Default Indicator: Represented as *"default"*

"default" is a binary variable that identifies whether a company defaulted during a credit rating adjustment. A value of *(1)* indicates that the company defaulted in a specific rating change, while *(0)* signifies that the company did not.

Eventual Default Indicator: Represented as "eventualdefault"

"eventualdefault" is a binary variable created which signifies whether a company defaulted within 365 days, i.e when *"daysuntildefault"* ≤ 365. A value of *(1)* indicates that default occurred within this timeframe, while *(0)* signifies no default within the next 365 days.

4.1.4 Ensuring Data Quality

To ensure data quality, we excluded non-U.S. companies, filtered out irrelevant rating types, and removed those that were no longer rated by the agencies, which were marked as "NR" (Not Rated). By structuring the dataset this way, the analysis can focus specifically on whether downgrades, upgrades, watch instances, reflect differences between the two agencies.

4.2 Descriptive Statistics

4.2.1 General Overview

The dataset contains a total of 13,983 credit rating changes recorded between 2014 and 2023, consisting of 7,109 downgrades, 4,981 upgrades, and 1,893 watch instances as seen in Table 4.2.1.

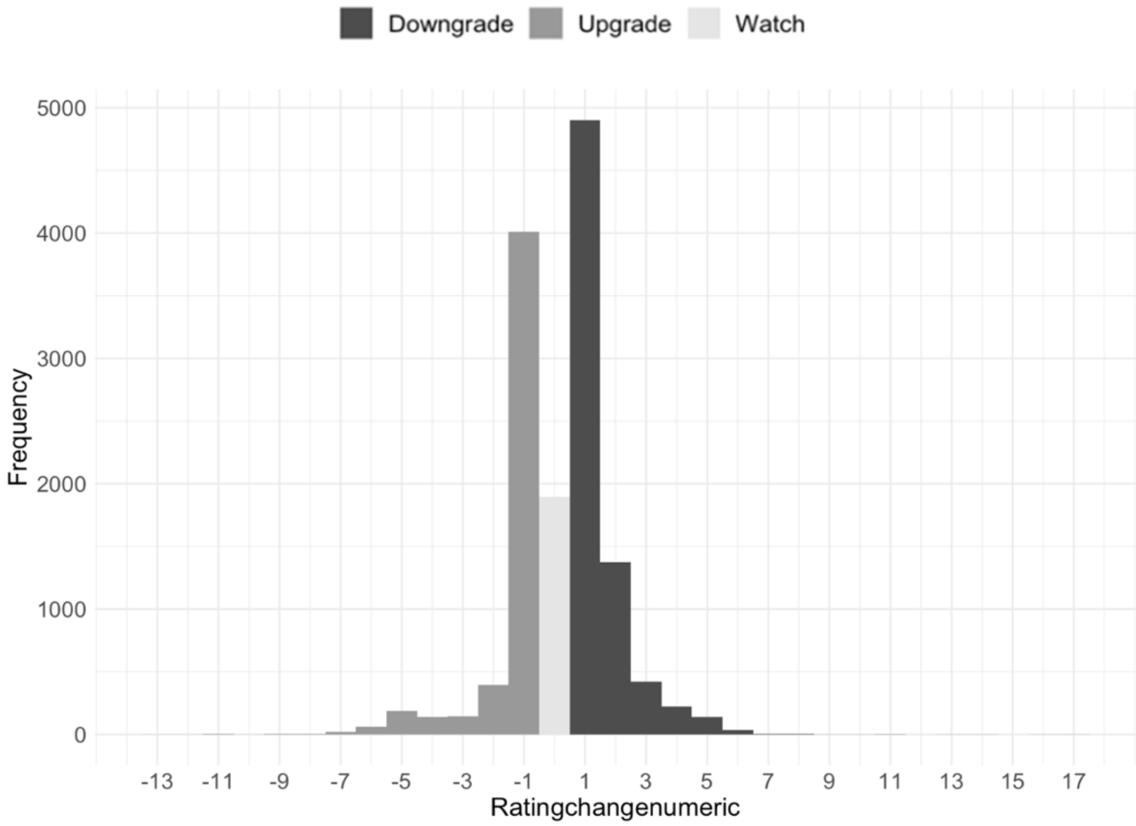
Table 4.2.1: *Credit Rating Changes by Year*

Year	*Total*	*Upgrades*	*Downgrades*	*Watch Instances*
2014	1 115	481	412	222
2015	1 328	458	631	239
2016	1 585	509	851	225
2017	1 347	504	648	195
2018	1 284	472	629	183
2019	1 318	401	764	153
2020	2 273	351	1 583	339
2021	1 265	735	393	137
2022	1 143	557	485	101
2023	1 325	513	713	99
Total	**13 983**	4 981	7 109	1 893

Note: The table includes all upgrades, downgrades and watch instances for the specified rating types, covering every US company that experienced a credit rating change between 2014 and 2023.

Figure 4.2.1 shows the distribution of credit rating adjustments, where most changes are small. Downgrades are represented by positive values, while upgrades are represented by negative values. The average rating adjustment is 0.25, indicating a slight tendency toward downgrades, backed by the numbers shown in Table 4.2.1. Despite the higher frequency of downgrades, the median adjustment is 1, confirming that most changes are small. The standard deviation of 1.74 highlights the variability in rating changes.

Figure 4.2.1: *Distribution of Credit Rating Adjustments*



4.2.2 Defaults

Table 4.2.2 presents the annual defaults between 2014 and 2023, totaling 795 defaults across the ten-year period. The number of defaults fluctuates significantly, with peaks in 2016, 2020, and 2023. These spikes correspond to the 2015 oil price collapse (Curtis, 2015), the COVID-19 crisis (ESMA, 2021) and the rising interest rates in 2023 (S&P Global Ratings, 2024a).

Table 4.2.2: *Default distribution by Year*

Year	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	*Total*
Defaults	34	58	115	65	66	82	177	44	38	116	795

Note: The table includes all defaults for the specified rating types, covering every US company that experienced a credit rating change between 2014 and 2023.

4.2.3 Upgrades, Downgrades and Watch by Year

Figure 4.2.3: *Frequency of Upgrades, Downgrades, Watch and Defaults by Year*

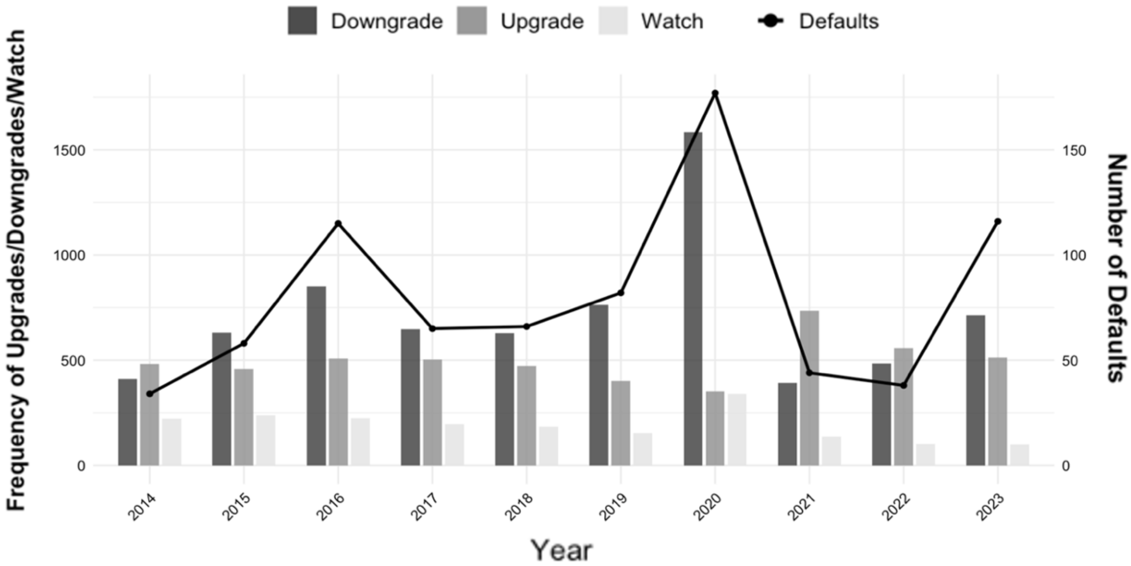


Figure 4.2.3 presents the frequency of downgrades, upgrades, watches, as well as the number of default events from 2014 to 2023. From the figure, we observe that downgrades generally had the highest frequency, in 2016 and 2020, indicating periods of increased financial distress. The spike in downgrades observed in 2016 can be attributed to the oil price crash, which placed significant financial strain on U.S. shale oil companies. As prices fell, many operators faced difficulties generating cash flow, leading to increased debt burdens, reduced investment in operations, and rising bankruptcies (Curtis, 2015). The spike in 2020 reflects a challenging year tied to global economic disruptions caused by the COVID-19 pandemic, which severely weakened the debt sustainability of businesses and governments (ESMA, 2021). The number of upgrades remained consistent over time, whereas downgrades occurred in sharp, concentrated surges, particularly during periods of economic decline caused by the pandemic. (ESMA, 2021). Additionally, the number of defaults rose significantly in 2020, closely aligning with the rapid rise in downgrades during the first and second waves of the pandemic (ESMA, 2021). After 2020, both downgrades and defaults decreased, showing a recovery trend in financial stability by 2021. However, defaults and downgrades rose again in 2023, driven by higher interest rates, which increased borrowing costs and financial pressures on issuers. Many firms, particularly those rated 'CCC+' or below, experienced weak liquidity, elevated leverage, and negative cash flows, leading to a rise in defaults (S&P Global Ratings, 2024a).

4.3 Descriptive Statistics Agencies Comparison

The interquartile range (IQR) contains the second and third quartile of values within "*Ratingtonumeric*". The horizontal line within the IQR indicates the median. The "whiskers" extend to 1.5 times the IQR from the lower and upper quartiles. Any points outside the whiskers are outliers, shown as individual dots, represents unusually high or low values.

Figure 4.3.1: *Distribution of Credit Ratings by Year and Agency*

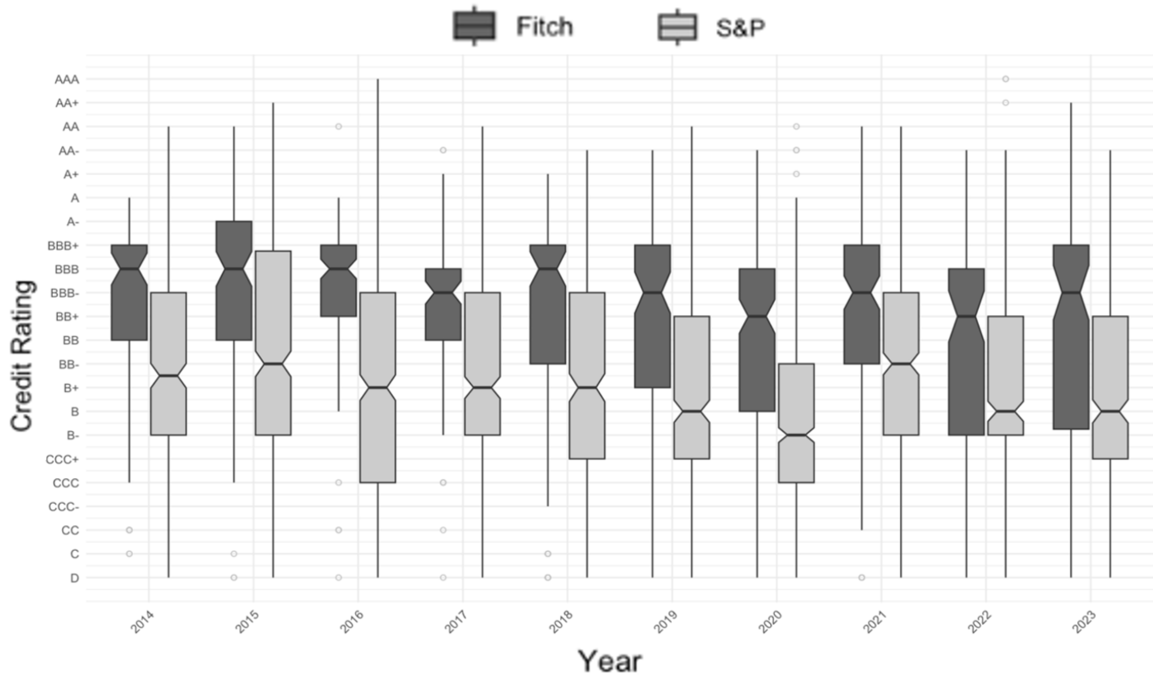


Figure 4.3.1 illustrates the distribution of credit ratings issued by S&P and Fitch from 2014 to 2023, highlighting the differences between the two agencies in their rating distribution. Fitch issues higher median ratings in every year compared to S&P. Additionally, the figure also shows that Fitch's ratings tend to cluster around higher credit ratings, whereas S&P more often covers a wider range. There are differences in the observed outliers, where Fitch has outliers in their lower rating categories, while S&P shows outliers in higher rating categories.

Table 4.3.1: *Credit Rating Distribution by Agency*

Agency	*Total*	*Upgrades*	*Downgrades*	*Watch Instances*
S&P	11 074	3 848	5 795	1 431
Fitch	2 909	1 133	1 314	462
Total	**13 983**	4 981	7 109	1 893

Note: The table includes all upgrades, downgrades and watch instances for the specified rating types, covering every US company in our dataset that experienced a credit rating update between 2014 and 2023.

Table 4.3.1 provides an overview of the credit rating distribution by agency, highlighting differences in the proportions of downgrades, upgrades, and Watch instances issued by S&P and Fitch. While S&P has a larger market share (Vipond, T., n.d.), covering more companies and issuing 11,074 total rating updates compared to Fitch's 2,910, the relative proportions of upgrades and downgrades reveal key distinctions. For S&P, downgrades reflect 52.33% of their total changes, compared to 45.17% for Fitch. Conversely, Fitch shows a higher proportion of upgrades at 38.95%, compared to 34.75% for S&P.

Figure 4.3.2: *Annual Downgrades, Upgrades and Watch*

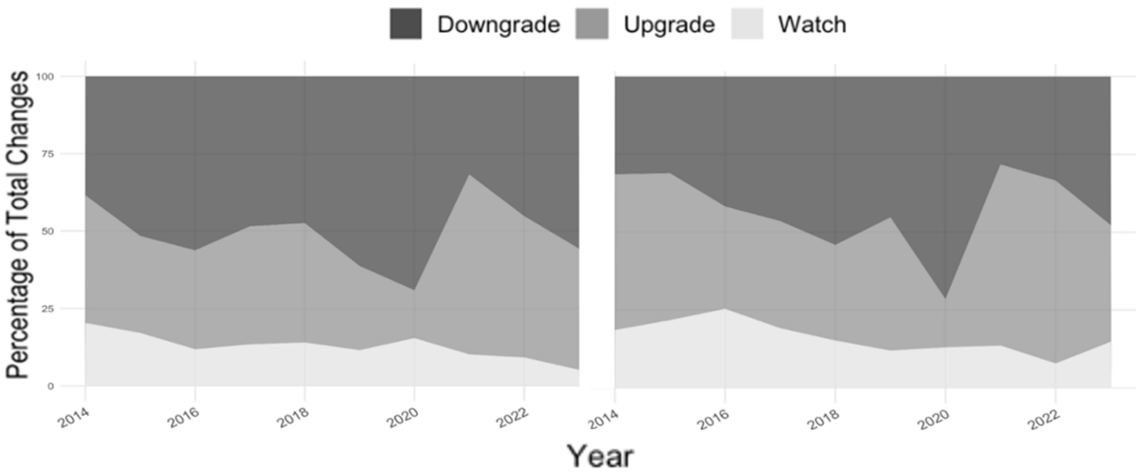


Figure 4.3.2 present the annual distribution of upgrades, downgrades, and watch instances for both agencies as percentages of the total changes each year. The visualization shows the differences in rating adjustment patterns between the two agencies.

Table 4.3.2: *Average rating size, median and standard deviation*

Agency	*Average Rating Change*	*Median Rating Change*	*Standard Deviation*
S&P	0.2626	1	1.7858
Fitch	0.1812	0	1.5205

Table 4.3.2 highlights the average size, median and standard deviation of rating adjustments for S&P and Fitch. S&P exhibits a larger average rating change of 0.26 notches compared to Fitch's 0.18. Furthermore, S&P's higher standard deviation of 1.79, compared to Fitch's 1.52, reflects greater variation in the size of its rating changes, indicating that S&P's adjustments are larger than those of Fitch.

4.4 Data Limitations

This section outlines the key limitations of the dataset and methodology used in the analysis, focusing on constraints that may affect the generalizability and interpretation of the findings.

4.4.1 Focus on S&P and Fitch

The analysis focuses exclusively on credit ratings issued by S&P and Fitch, two of the three largest CRAs (Vipond, T., n.d.). While these agencies are highly influential, the exclusion of other CRAs, including the second largest agency, Moody's, means that the findings may not fully capture the broader landscape of credit rating practices. By focusing exclusively on S&P and Fitch, the analysis offers valuable insights into their practices but may not fully represent the broader credit rating industry.

4.4.2 Data Specificity

Our study focuses on rating changes and does not incorporate company-specific financial data, i.e., balance sheet metrics, income statements, or market valuations. This limits the ability to apply methods such as financial ratio analysis, regression models with firm-specific predictors, or stress-testing credit ratings against financial indicators. It also restricts the exploration of causality between financial health and rating changes or comparative analyses across industries or time periods based on financial resilience.

5.0 Methodology

This chapter presents the methodology used to evaluate the predictive accuracy of S&P and Fitch credit ratings. Drawing on Machek & Hnilica (2013), we employ a Transition Matrix Model using Markov chains to analyze rating dynamics. Further, we examine significant differences in default probabilities across credit ratings, by employing the two-sample proportion z-test (Bobbitt, 2020). The CAP framework, introduced by Engelmann et al. (2003), measures the discriminatory power of credit ratings by assessing how well defaulters are ranked relative to non-defaulters. In contrast, the ROC curve, based on Swets (1988) and Engelmann et al. (2003), evaluates the ability of ratings to distinguish between defaulters and non-defaulters across all thresholds, providing a broader assessment of accuracy.

5.1 Transition Matrix Model Using Markov Chains

The methodology outlined is based on the study described in Machek & Hnilica, J. (2013).

5.1.1 Constructing the Transition Matrix

Step 1: Define States

The states in our analysis correspond to the various credit ratings assigned to entities, such as AAA, AA+, AA, and AA-. Additionally, the "*default*" (D) state is included as a terminal state.

Step 2: Count Transitions

We analyzed the observed transitions between credit rating states within our dataset. Specifically, we tracked how frequently entities moved from one rating to another, such as transitions from AAA to AA, or from AA to BBB, capturing the dynamics of credit rating changes during the period.

Step 3: Compute Transition Probabilities

The likelihood of moving from one credit rating state i to another j is represented by the transition probability:

$$P_{ij} = \frac{N_{ij}}{N_i} \tag{1}$$

Where:

- P_{ij} is the probability of transitioning from state *i* to state *j*,
- N_{ij} is the observed number of transitions from state *i* to state *j*,
- N_i is the total number of transitions from state *i*.

Step 4: Building the Transition Matrix

We organized the P_{ij} into a transition matrix *P*, where each row represents an initial credit rating, and each column shows the probability of transitioning to a specific rating directly from the initial rating.

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix} \tag{2}$$

Step 5: Validate the Matrix

We ensured that all transition probabilities P_{ij} were non-negative $P_{ij} \geq 0$ and that the sum of probabilities in each row equaled 1:

$$\sum_{j=1}^{n} P_{ij} = 1 \quad for\ all\ i \tag{3}$$

5.1.2 Transition Matrix Heatmap Rating Categories

The transition matrix heatmap groups credit ratings into broader categories by consolidating sub-levels into their respective groups. This approach aligns with the methodology of Machek

& Hnilica (2013) and simplifies the analysis while preserving key distinctions. Grouping ensures sufficient data points for meaningful trends, distinguishes between investment-grade and speculative-grade ratings, and consolidates all default scenarios under the "*Default*" category. The grouped categories, as outlined in Table 2.2, simplify the transitions by consolidating them into broader rating groups, making the matrices easier to interpret and visualize.

Table 5.1.2: *Grouped Ratings*

Grouped Ratings	*Original Ratings*		
AAA		AAA	
AA	AA+	AA	AA-
A	A+	A	A-
BBB	BBB+	BBB	BBB-
BB	BB+	BB	BB-
B	B+	B	B-
CCC	CCC+	CCC	CCC-
CC	CC		C
Default	D	RD	SD

5.2 Default Probability

To calculate the default probabilities, we used the total number of *"ratingto"* for all ratings and identified how many percent of the changes also has "*eventualdefault*" = 1. The default probability is then calculated as the proportion of defaults relative to the total ratings in that category. We then compute p-values to assess the statistical significance of differences in default probabilities between Fitch and S&P. This test examines the null hypothesis from Chapter 3.3, checking whether default probabilities of the two agencies are equal. We use the z-statistic formula provided by Bobbitt (2020):

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} \quad (4)$$

Where:

- $\hat{p}_1$ and $\hat{p}_2$ are the observed default probabilities for Fitch and S&P, respectively,
- $\hat{p}$ is the pooled proportion, calculated as $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$,

- x_1 and x_2 are the counts of defaults for Fitch and S&P,
- n_1 and n_2 are the total observation for Fitch and S&P.

The resulting z-statistic was then used to derive the p-value, based on the standard normal distribution.

5.3 CAP

According to Bhalla, D. (2019), the CAP is a quantitative measure of the discriminatory power of a rating or scoring system in distinguishing between different outcomes, where higher AR values indicate a greater ability to separate high-risk and low-risk groups. In our case it measures how well the system ranks companies, with higher ratings indicating lowers risk and lower ratings indicating higher risk. A higher AR means that the rating system has good discriminatory power, being able to separate high risk from low risk.

5.3.1 Grouping and Cumulative Calculations

To calculate the CAP curve, the dataset is grouped by the initial credit ratings ("*ratingfromnumeric"*) to compute the cumulative distributions of issuers and defaults. The cumulative proportions provide insights into how defaults and issuers are distributed across risk categories, which is essential for assessing the discriminatory power of the rating systems.

1. *Arrange Ratings in Descending Order:* The ratings "*ratingfromnumeric"* are arranged in descending order, starting with the poorest rating. This ordering ensures that the cumulative calculations reflect the intuitive progression from higher-risk to lower-risk ratings.

2. *Calculate Cumulative Proportions:*
   - *n(i)*: Cumulative proportion of issuers up to each rating level *i*, representing the percentage of the issuers at or above each level of risk.
   - *d(i)*: Cumulative proportion of defaults up to each rating level *i*, representing the cumulative percentage of defaults at or above each risk level.

Mathematically:

$$n(i) = \frac{\sum_{j<i} N_j}{N} \tag{5}$$

$$d(i) = \frac{\sum_{j<i} D_j}{N} \tag{6}$$

Where:

- N_j is the number of issuers at rating level j,
- D_j is the number of defaults at rating level j,
- N and D represent the total number of issuers and defaults, respectively.

5.3.2 Calculating The AR

The AR quantifies how well the rating model discriminates between defaulters and non-defaulters. To calculate the AR for S&P and Fitch, we follow a systematic three-step approach derive from Bhalla, D. (2019):

Step 1: Calculate the AUC for the Full Model

The Area Under the Curve (AUC) for the full model represents the discriminatory power of the rating system. It is calculated using the Trapezoidal Rule, which approximates the area under the curve as a sum of trapezoidal sections. The formula for each trapezoid is:

$$Area\ of\ Trapezoid = (x_{i+1} - x_i) \times \left(\frac{y_i + y_{i+1}}{2}\right) \tag{7}$$

Where:

- x_i and x_{i+1} are the consecutive points on the x-axis $n(_i)$,
- y_i and y_{i+1} are the corresponding points on the y-axis $d(_i)$.

The total AUC for the full model is the sum of all trapezoidal areas under the CAP curve:

$$\text{AUC}_{\text{model}} = \sum_{i=1}^{k} (x_{i+1} - x_i) \times \left(\frac{y_i + y_{i+1}}{2}\right) \tag{8}$$

Step 2: Calculate the AUC for the Random Model

The Random Model assumes that defaults are distributed randomly across all credit ratings. The cumulative distribution for such a model is represented by a straight diagonal line from *(0,0)* to *(1,1)*. The area under the curve is always:

$$\mathrm{AUC}_{\mathrm{random}} = 0.5 \tag{9}$$

Step 3: Calculate the AR

The AR is obtained by comparing the AUC of the Full Model with the AUC of the Random Model, normalized to account for the maximum possible improvement (i.e., the area above the random model). The AR is calculated as follows:

$$\mathrm{AR} = 2 \ \times \ (\mathrm{AUC}_{\mathrm{model}} - AUC_{random}) \tag{10}$$

Where:

- $\mathrm{AUC}_{\mathrm{model}}$ is the area under the CAP curve of the actual rating model.
- $\mathrm{AUC}_{\mathrm{random}} = 0.5$ is the area under the diagonal line.

According to Engelmann et al. (2003) the rating method is the better the closer AR is to 1. This scaling ensures that the AR lies within the range $[-1,1]$, where:

- *AR = 1:* Perfect discrimination (defaults occur only in the highest-risk categories).
- *AR = 0:* No discrimination (defaults are randomly distributed across all categories).
- *AR = -1:* Inverse discrimination (defaults occur only in the safest categories).

5.4 ROC

The ROC curve is a statistical tool used to evaluate the discriminatory power of binary classification models, particularly in predicting defaults in credit ratings. In this thesis, the ROC curve was constructed to assess the ability of ratings assigned by S&P and Fitch to differentiate between defaulting and non-defaulting entities. The AUC, which summarizes the model's performance, and its confidence intervals were calculated to quantify and validate the models' discriminatory power (Engelmann et al., 2003).

The methodology outlined here is based on the study described in Engelmann et al. (2003).

5.4.1 Determining Cut-Off Values Using the Youden Index

The Youden Index, as introduced by Youden (1950), is a statistical tool for evaluating the effectiveness of diagnostic tests by maximizing the trade-off between correctly identifying positive and negative cases. According to Yin, J., & Tian, L. (2014), the Youden Index is utilized to identify optimal cut-off values for S&P and Fitch ratings by maximizing J, where:

$$J = Sensitivity + Specificity - 1 \tag{11}$$

Definitions and Calculations:

1. *Sensitivity (True Positive Rate, TPR):*
   Sensitivity measures the proportion of actual positives (defaulting entities) that are correctly identified:

$$Sensitivity\ (TPR) = \frac{True\ Posivitves\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)} \tag{12}$$

2. *Specificity (True Negative Rate):*
   Specificity measures the proportion of actual negatives (non-defaulting entities) that are correctly identified:

$$Specificity = \frac{True\ Negatives\ (TN)}{True\ Negatives\ (TN) + False\ Positives\ (FP)} \tag{13}$$

3. *False Positive Rate (FPR):*
   The complement of specificity, calculated as:

$$FPR = Specificity - 1 \tag{14}$$

4. *Youden Index:*

The formula for J integrates the sensitivity and specificity to determine the optimal threshold:

$$J = \max[Sensitivty + Specificity - 1] \tag{15}$$

Using these definitions, we calculated J at each possible rating threshold by summing the sensitivity and specificity and subtracting one. The threshold that maximized J was identified as the optimal cut-off.

5.4.2 ROC Curve Construction

To construct the ROC curve, we plot *FAR(threshold)* on the x-axis and *HR(threshold)* on the y-axis for each rating. The ROC curve provides a visual representation of the trade-off between correctly identifying defaults (high HR) and minimizing false alarms (low FAR) (Engelmann et al., 2003). An ideal model with strong discriminatory power will have an ROC curve that moves steeply towards the top-left corner, indicating a high HR with a low FAR. This signifies that the model effectively separates defaulters from non-defaulters (Swets, 1988). In contrast, a model with no discriminatory power will produce an ROC curve that aligns closely with the 45-degree diagonal, equivalent to random guessing (Swets, 1988).

The ROC curve was constructed by calculating two key metrics across all thresholds:

1. Hit Rate (HR):

$$HR(threshold) = P(S_D \leq threshold) \tag{16}$$

Where, S_D represents the score distribution of a defaulter. HR measures the proportion of actual defaulters correctly classified as defaulters.

2. False Alarm Rate (FAR):

$$FAR(threshold) = P(S_{ND} \leq threshold) \tag{17}$$

where S_{ND} represents the score distribution of a non-defaulter. FAR measures the proportion of non-defaulters incorrectly classified as defaulters.

For each *"ratingtonumeric"*, $P(S_D \leq threshold)$ and $P(S_{ND} \leq threshold)$ were calculated, and the HR and FAR values were plotted to create the ROC curve, with FAR on the x-axis and HR on the y-axis. The curve represents the trade-off between correctly identifying defaulters and minimizing false alarms.

5.4.3 AUC

The AUC summarizes the overall discriminatory power of the ROC curve. It is defined as:

$$AUC = P(S_D < S_{ND}) + \frac{1}{2} P(S_D = S_{ND}) \tag{18}$$

Where S_D is the score of a defaulter and S_{ND} is the score of a non-defaulter. This probability was estimated using the Mann-Whitney (1947) U-statistic:

$$U = \frac{1}{N_D N_{ND}} \sum_{(D,ND)} u_{D,ND} \tag{19}$$

Where:

$$u_{D,ND} = \begin{cases} 1, if\ S_D < S_{ND} \\ \frac{1}{2} if\ S_D = S_{ND} \\ 0, if\ S_D = S_{ND} \end{cases} \tag{20}$$

Here N_D, and N_{ND}; are the numbers of defaulters and non-defaulters, respectively.
For discrete distributions, the AUC can also be calculated as the sum of trapezoidal areas:

$$AUC = \sum_{i=1}^{k} \frac{1}{2} (CD_i^D + CD_{i-1}^D)(CD_i^{ND} + CD_{i-1}^{ND}) \tag{21}$$

Where:

- CD_i^D: The cumulative distribution function (CDF) of scores for defaulters up to score S_i, defined as:

$$CD_i^D = \sum_{j=1}^{i} p_j^D \quad (22)$$

- CD_i^{ND}: The CDF of scores for non-defaulters up to score S_i, defined as:
-

$$CD_i^{ND} = \sum_{j=1}^{i} p_j^{ND} \quad (23)$$

- k: The number of unique score categories.

5.4.4 Confidence Intervals for AUC

Confidence intervals for the AUC were calculated to assess the uncertainty of the estimates. Using the asymptotic normal approximation of the Mann-Whitney (1947) U-statistic, the variance of $\hat{U}$ was estimated as:

(24)

$$\sigma_{\hat{U}}^2 = \frac{1}{4(N_D - 1)(N_{ND} - 1)}\left[P_{D\neq ND} + (N_D - 1)P_{D,D,ND} + (N_{ND} - 1)P_{ND,ND,D} - 4(N_D + N_{ND} - 1)\left(\hat{U} - \frac{1}{2}\right)^2\right]$$

The confidence interval was then constructed as:

$$[\hat{U} - \hat{\sigma}_{\hat{U}}\Phi^{-1}\left(1 + \frac{\alpha}{2}\right), \hat{U} + \hat{\sigma}_{\hat{U}}\Phi^{-1}\left(1 + \frac{\alpha}{2}\right)] \quad (25)$$

Where:

- Φ^{-1} is the quantile function of the standard normal distribution,
- α = 0.05 for a 95% confidence level or 0.01 for a 99% confidence level.

5.4.5 Testing Discriminative Power

To assess whether a rating system has meaningful discriminatory power, it is necessary to determine if its AUC is significantly greater than a set threshold. An AUC of 0.5 indicates no ability to distinguish between defaulters and non-defaulters, equivalent to random guessing. The null hypothesis (H_0) assumes that the model has no discriminative power. This hypothesis was tested using the Mann-Whitney (1947) U-statistic across multiple thresholds surpassing the null hypothesis, including 0.7, 0.8, and 0.9, to evaluate the model's performance under increasingly stringent criteria.

$$H_0 : AUC = 0.5$$

Testing at these thresholds allows for a nuanced understanding of the rating system's ability to discriminate between defaulters and non-defaulters under varying levels of difficulty, providing insights into its robustness and practical applicability.

The test statistic was calculated as:

$$Z = \frac{U - threshold}{\sigma_U} \tag{26}$$

If $Z > Z_\alpha$ where Z_α is the critical value for a standard normal distribution, the null hypothesis was rejected, indicating significant discriminatory power.

5.4.6 Interpreting AUC

AUC values are interpreted as follows: (Çorbacıoğlu et al,. 2023).

- $0.9 \leq AUC < 1$: *excellent*
- $0.8 \leq AUC < 0.9$: *considerable*
- $0.7 \leq AUC < 0.8$: *fair*
- $0.6 \leq AUC < 0.7$: *poor*
- $0.5 \leq AUC < 0.6$: *fail*
- $AUC = 0.5$: *Random guessing*

6.0 Results

This section presents and discusses the results obtained through the application of the methodology to analyze the performance of Fitch and S&P. Using the transition matrices derived from Machek and Hnilica (2013), these matrices provide valuable insights into how each agency structures and adjusts ratings over time, particularly in speculative-grade categories. The CAP and ROC curves, as outlined by Engelmann, Hayden, and Tasche (2003), are employed to assess the discriminatory power of Fitch and S&P in separating defaulting from non-defaulting companies. These results, helps us calculate the AR and AUC values, which are further discussed to evaluate each agency's effectiveness in identifying credit risk.

This chapter combines the results with discussion, providing an integrated analysis of the findings and their implications.

6.1 Transition Matrix Heat Map

The Y-axis represents the initial rating, while the X-axis shows the adjusted rating. The diagonal cells, running from the top left to the bottom right, represent the probabilities of ratings remaining unchanged, reflecting stability within each category, while the off-diagonal cells capture the probabilities of downgrades (above the diagonal) and upgrades (below the diagonal). In the following, we compare and interpret the results from Fitch and S&P's transition matrices, highlighting key differences in rating movements, stability, and the distribution of transitions across categories.

Figure 6.1.1: *S&P Transition Matrix Heatmap*

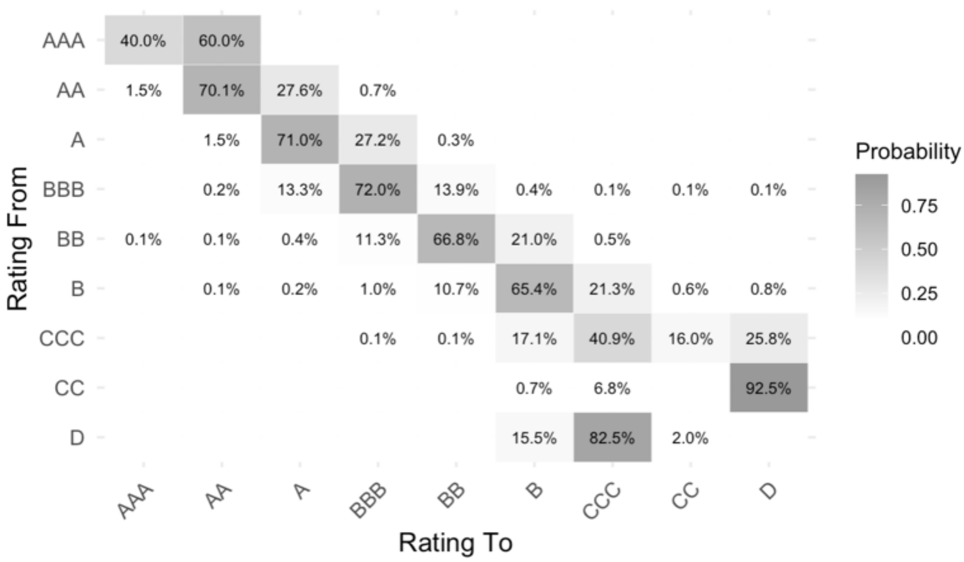
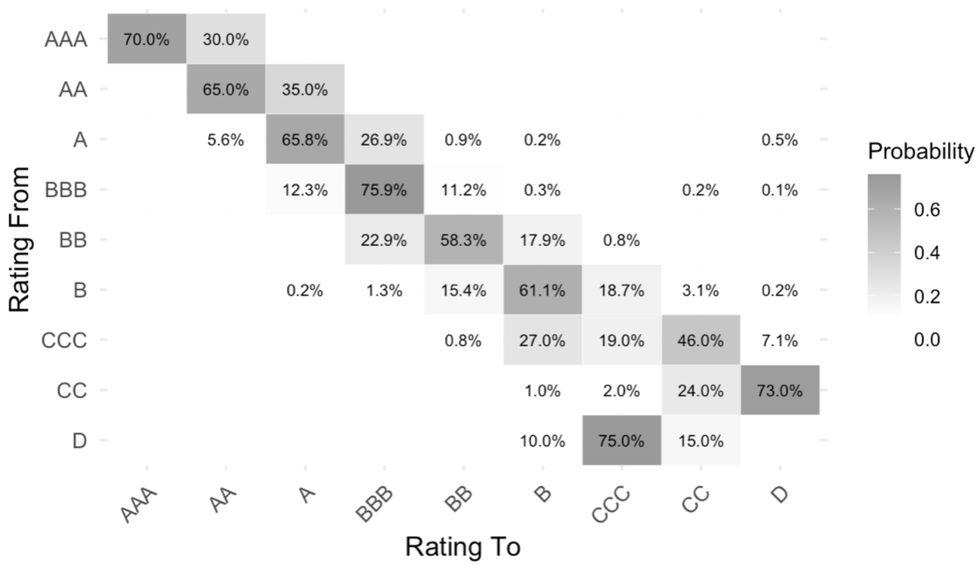


Figure 6.1.2: *Fitch Transition Matrix Heatmap*



Fitch and S&P's transition matrices highlight differences in rating movements across top-grade (AAA, AA, A), medium-grade (BBB, BB, B), and bottom-grade (CCC, CC, D) categories. In top-grade ratings, Fitch retains 70% of AAA ratings, transitioning the remaining 30% of the time to AA. S&P retains only 40% of AAA ratings within the same category, transitioning 60% to AA. Fitch also records a direct transition from A to *"Default,"* an anomaly not mirrored in any of S&P's top-grade ratings.

Fitch and S&P's transition matrices reveal differences in medium-grade ratings, particularly in stability. S&P exhibits more stability in this category, with 66.8% of BB ratings and 65.4% of B ratings remaining unchanged, compared to Fitch's 58.3% and 61.1%, respectively. However, S&P also demonstrates notable anomalies, such as a rare upward transition from BB to AAA, where no Fitch rating has gone from a medium-grade rating to higher than A. Fitch, by contrast, has less stability but smaller, more evenly distributed changes, with transitions like 22.9% of BB upgrading to BBB and 17.9% downgrading to B. In the bottom-grade we observe that Fitch's default probability at CC, is 73%, while S&P's is 92.5%. This can be largely put down to the fact that S&P's C rating is absent and therefore they have a larger proportion of defaulters in their second poorest rating.

The key takeaway from the transition matrices is that Fitch's rating transitions are more concentrated, with adjustments typically confined to a single rating category shift within aggregated levels. For example, Fitch's BB ratings transition with 22.9% upgrading to BBB and 17.9% downgrading to B, while multi-category jumps are rare. In contrast, S&P exhibits a broader spread of transitions, particularly in the BBB and BB ranges, where movements are more distributed across multiple adjacent and non-adjacent ratings. For instance, in the BB category, 11.3% of ratings remain stable, but 21% downgrade to B, and a small fraction transitions to even lower ratings like CCC and CC. This broader spread in S&P's ratings, compared to Fitch's more incremental changes, aligns with Table 4.3.2, which shows Fitch issuing smaller rating adjustments. The observed differences may also arise from our dataset containing more rating changes for S&P, leading to anomalies such as transitions from BB to AAA or B to AA, which are non-existing for Fitch.

6.2 Default Probability

Our analysis shows that there are differences in default probability between investment-grade and speculative-grade ratings, a pattern consistent with Cantor and Mann's (2003) findings, where default rates increase as ratings are poorer, particularly when comparing investment-grade from speculative-grade categories (Cantor & Mann, 2003). This section delves deeper into these differences, presenting results that underscore the near-zero default rates for investment-grade ratings and the increasing risks within speculative-grade categories. Furthermore, it explores whether there are statistically significant differences between Fitch and S&P in their default probabilities, particularly within speculative-grade ratings.

Table 6.2.1: *Default Probabilities Investment Grade*

Ratings	*Fitch*	*S&P*	*P-Value*
AAA	0.000	0.000	NA
AA+	0.000	0.000	NA
AA	0.000	0.000	NA
AA-	0.000	0.000	NA
A+	0.017	0.000	0.128
A	0.008	0.000	0.146
A-	0.000	0.002	0.436
BBB+	0.006	0.004	0.663
BBB	0.005	0.005	0.963
BBB-	0.006	0.004	0.671

Note: 'NA' indicates either (1) the data is not available or not reported for the respective rating grade, or (2) the default probabilities are 0.000 for both agencies, making it impossible to derive p-values.

A higher number of defaulting companies is observed in Fitch's investment-grade ratings (A+, A, BBB+, and BBB-). Although the default probabilities for these ratings are low and not statistically significant, Fitch’s default probabilities for all investment-grade categories are equal to or higher than those of S&P. However, the default probabilities observed align closely with the historical data reported by S&P Global Ratings (2023). Over the years, S&P has consistently observed default rates of approximately 0% for investment-grade categories, validating the reliability of investment-grade ratings in signaling creditworthiness. This external consistency aligns with our findings, as Fitch and S&P with few exceptions, assign non-defaulting companies to investment-grade ratings.

In contrast, our results shows that speculative-grade ratings have significantly higher default probabilities. This finding aligns with S&P Global Ratings (2023) and Cantor & Mann (2003). Not surprisingly, both agencies are able to place defaulting companies within the speculative grade. While S&P shows default probabilities greater than zero for all speculative grades, these probabilities are notably low at the higher end of the speculative spectrum.

Table 6.2.2: *Default Probabilities Speculative Grade*

Ratings	*Fitch*	*S&P*	*P-Value*
BB+	0.000	0.004	0.374
BB	0.006	0.003	0.567
BB-	0.011	0.008	0.612
B+	0.000	0.015	0.134
B	0.042	0.038	0.781
B-	0.120	0.068	0.02
CCC+	0.226	0.229	0.971
CCC	0.459	0.379	0.180
CCC-	0.375	0.707	0.005
CC	0.750	0.921	0.001
C	0.915	NA	NA

Note: 'NA' indicates either (1) the data is not available or not reported for the respective rating grade, or (2) the default probabilities are 0.000 for both agencies, making it impossible to derive p-values.

One of the primary objectives of our study was to assess whether there are statistically significant differences in default probabilities between Fitch and S&P across comparable rating categories. As far as we are aware, this has not been done, and therefore our hypothesis suggests that default probabilities are similar for both agencies (see Chapter 3.3). Our results indicate that Fitch assigns a notably high default probability of 0.898 (89.8%) to its C rating, while S&P omits the C category entirely (within our dataset), redistributing this risk across speculative-grade ratings. While S&P includes the C category within its rating scale (Table 2.2), it was not attributed to any U.S. company during the period analyzed. This absence introduces two key implications. First, it creates uncertainty regarding the operational meaning of S&P's "C" rating and its implications for companies receiving this rating in the future. Without historical data, the attributes and significance of this rating remain unclear. Secondly, the absence of a "C" rating within S&P's assigned data redistributes default probabilities across its other speculative-grade categories, resulting in higher default probabilities within these lower ratings in the speculative grade. The omission of rating C highlights a methodological difference between the agencies.

Our p-values derived from Tables A.5 and A.6 indicate that for investment-grade ratings, both agencies show consistent default probabilities, with no significant inter-agency differences observed. However, in the speculative-grade segment, there are significant differences. Specifically, the CCC-, CC, and B- rating categories have p-values below 0.05, highlighting statistically significant differences between the agencies' default probabilities. Consequently, we reject *H1* based on these findings.

6.3 Rating Changes before Eventual Default

This section examines the timing and size of rating changes issued by S&P and Fitch in the days leading up to default for companies which has "*eventualdefault*"=1 (from 1 to 365 days prior, excluding the default date itself). By analyzing patterns of these rating changes, this section checks whether one of the CRAs downgrades companies approaching default later than the other and perhaps with larger adjustments.

Table 6.3.1: *Fitch Rating Adjustments Prior to Default*

Coefficients			
	Estimate	*Pr (>| z |)*	
(Intercepts)	2.2697	<2e-16	***
daysuntildefault	-0.0032	0.0017	**

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1*

Table 6.3.1 and Figure 6.3.1 for Fitch highlight a significant relationship between the timing of downgrades and their size of the downgrades as defaults approach. The intercept is statistically significant given the p-value and represents the average downgrade size one day prior to default. The variable *"daysuntildefault",* is also statistically significant and represents the number of days a change has been made before its default date. Its negative coefficient indicates that as the default date approaches, Fitch increases the severity of its downgrades by a "*ratingchangenumeric*" of 0.0032.

Figure 6.3.1: *Average Rating Change vs. Days Until Default Fitch*

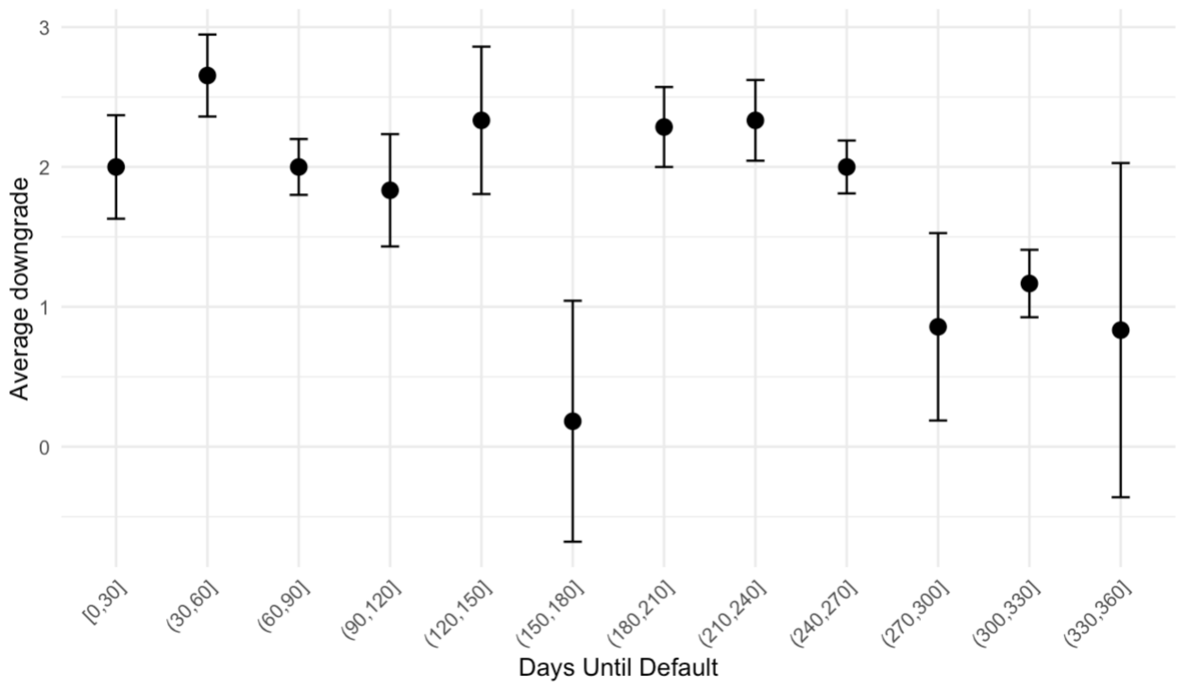


Figure 6.3.1 presents the average downgrade sizes across the defaulting companies, and with the error bars indicating the downgrade size's standard deviation. There is less consistency in Fitch's downgrades earlier in the timeline, while approaching default date, the deviations become narrower.

Table 6.3.2: *S&P Rating Adjustments Prior to Default*

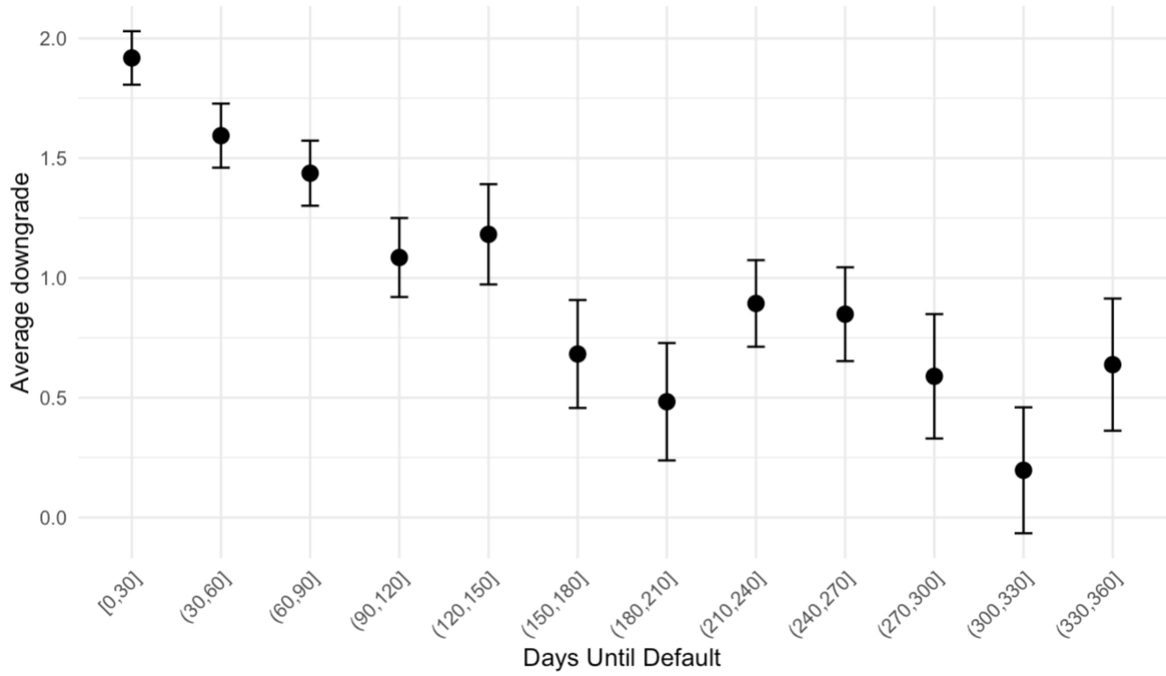
Coefficients			
	Estimate	*Pr (>| z |)*	
(Intercepts)	*1.7482*	*<2e-16*	*****
daysuntildefault	*-0.0042*	*2.93e-16*	*****

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1*

Table 6.3.2 and Figure 6.3.2 for S&P also reveal a statistically significant relationship between downgrade timing and size, with a larger coefficient for *"daysuntildefault"*, compared to Fitch. This indicates that S&P's downgrade sizes changes more than Fitch's, as defaults approach. The intercept suggests smaller downgrade sizes the day before default in comparison to Fitch.

Figure 6.3.2 demonstrates a consistent increased downgrade size as defaults approach. While earlier intervals show smaller average downgrade sizes, they also display greater variability compared to the months closer to default. This pattern was also observed for Fitch, where downgrade sizes became larger and variations narrower the closer the default.

Figure 6.3.2: *Average Rating Change vs. Days Until Default S&P*



Fitch's larger intercept and less steep *"daysuntildefault"* coefficient suggest that downgrades are issued earlier, and downgrade sizes are more consistent the year leading up to default. This may reflect its ability to identify potential defaulting companies earlier in the timeline. Conversely, S&P's steeper *"daysuntildefault"* coefficient, and lower standard deviation highlight a clearer and more consistent pattern of downgrades leading up to default. S&P's larger dataset likely contributes to its lower variability and more precise results, whereas Fitch's smaller dataset may result in greater variability.

6.4 CAP

As discussed in the theory and methodology, the CAP presents three models: the Random Model (diagonal line), the Perfect Model (dashed line) and the Actual Rating Model (solid line). The Actual Rating Model shows the performance of S&P and Fitch in distinguishing high-risk from low-risk issuers, with the proximity of the curve to the Perfect Model indicating the effectiveness of each system. When examining Figure 6.4.1 and Figure 6.4.2, we compare their CAP curves and calculate their AR to evaluate the discriminatory power of each rating system.

Figure 6.4.1: *CAP S&P*

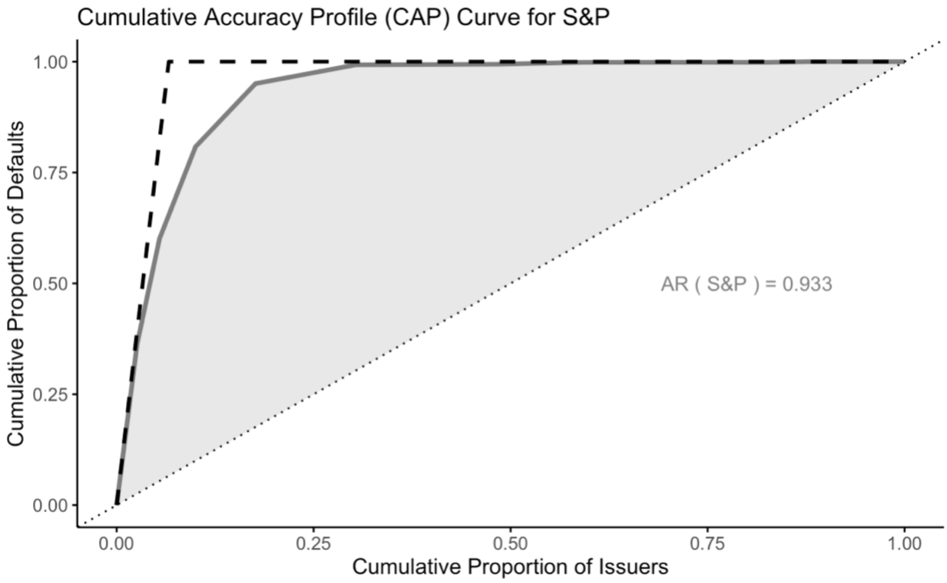
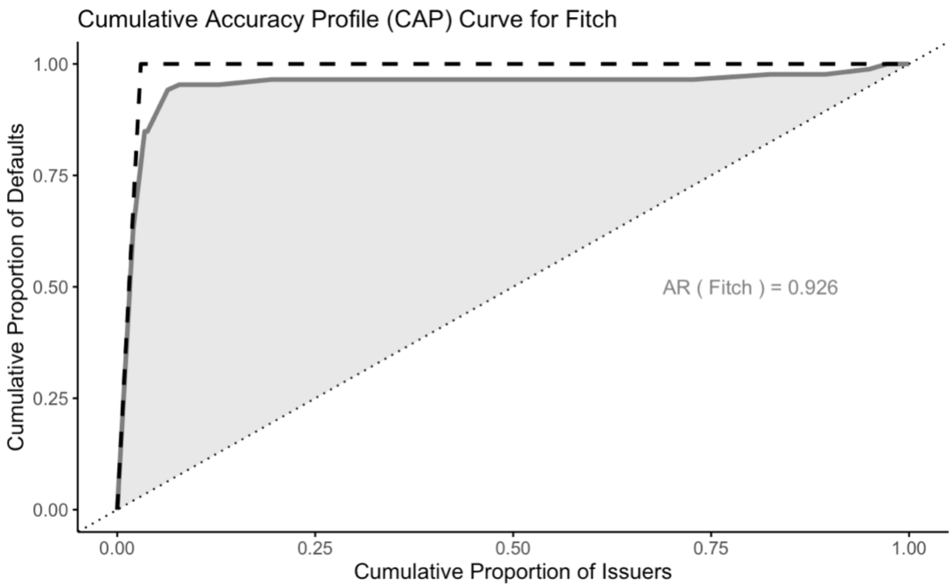


Figure 6.4.2: *CAP Fitch*



As outlined in Chapter 3.4, strong performance in identifying defaulting entities within the lowest-rated categories is reflected in the steep initial slopes of their CAP curves. Fitch's CAP curve closely follows the Perfect Model in the early phase, indicating that a high proportion of defaults are concentrated in its poorest-rated categories. S&P's curve, while also capturing defaults early, shows a broader distribution across the lower rating categories, suggesting that defaults are spread more evenly within speculative grades. In the second phase, Fitch's curve deviates from the Perfect Model, reflecting more instances of defaults among investment grade entities. In contrast, S&P's curve aligns more closely with the perfect model throughout the later phase, with fewer defaults occurring in higher-rated categories.

Fitch and S&P's predictive accuracy highlights their strong discriminatory power, with Fitch achieving an AR value of 0.926, whilst S&P's AR is 0.934, consistent with Engelmann et al. (2003), who assert that the effectiveness of a rating method improves as the AR approaches 1. Fitch demonstrates strong performance in concentrating a significant proportion of defaulting companies within the lowest-rated categories. However, S&P's AR are higher because there are less instances where defaulting companies are found within S&P's investment-grade ratings. These results are complemented by our findings in Table 6.2.1.

6.5 ROC and AUC

To delve deeper into the performance of S&P and Fitch, we analyze the results given by their ROC curves, as Engelmann et al. (2003) and Irwin and Irwin (2013) highlight that the properties of ROC curves are more intuitive and interpretable compared to those derived from the CAP.

As detailed in the methodology, the threshold for the analysis is determined using Youden's Index (Youden, 1950), which identifies the rating where the balance between TPR and FPR is optimized. The cutoff value is retrieved from (Tables A.3 and A.4), which provide TPR, FPR, and their respective values for all thresholds. By using Youden's Index (*J)* (Youden, 1950), the resulting threshold reflects the balance between maximizing the difference between TPR and FPR. As emphasized by Nahm (2022), such an approach is effective in refining threshold criteria to reduce overestimation of default risk while maintaining focus on the most vulnerable categories. Using the data from Tables A.3 and A.4, we calculated the *J* value for all thresholds, resulting in a cutoff value of B- for Fitch and CCC+ for S&P.

The ROC curve provides a clear visualization of how well the rating models by Fitch and S&P separate defaulting from non-defaulting companies. While the CAP provides a straightforward measure of a model's discriminatory power, the ROC curve offers a more comprehensive evaluation by illustrating the trade-off between the TPR and FPR across different thresholds. (Engelmann et al., 2003), with TPR on the Y-axis and FPR on the X-axis.

Engelmann et al., (2003) points out that for a Perfect Rating Model, Figures 6.5.1 and 6.5.2 would show a clear separation between the left and right distributions. Specifically, all defaults would lie to the left of the threshold, while all non-defaults would be positioned to the right (defaults are shown as dark gray, non-default are shown as white) Engelmann et al., (2003).

Figure 6.5.1: *Classification of S&P Defaulters and Non-Defaulters*

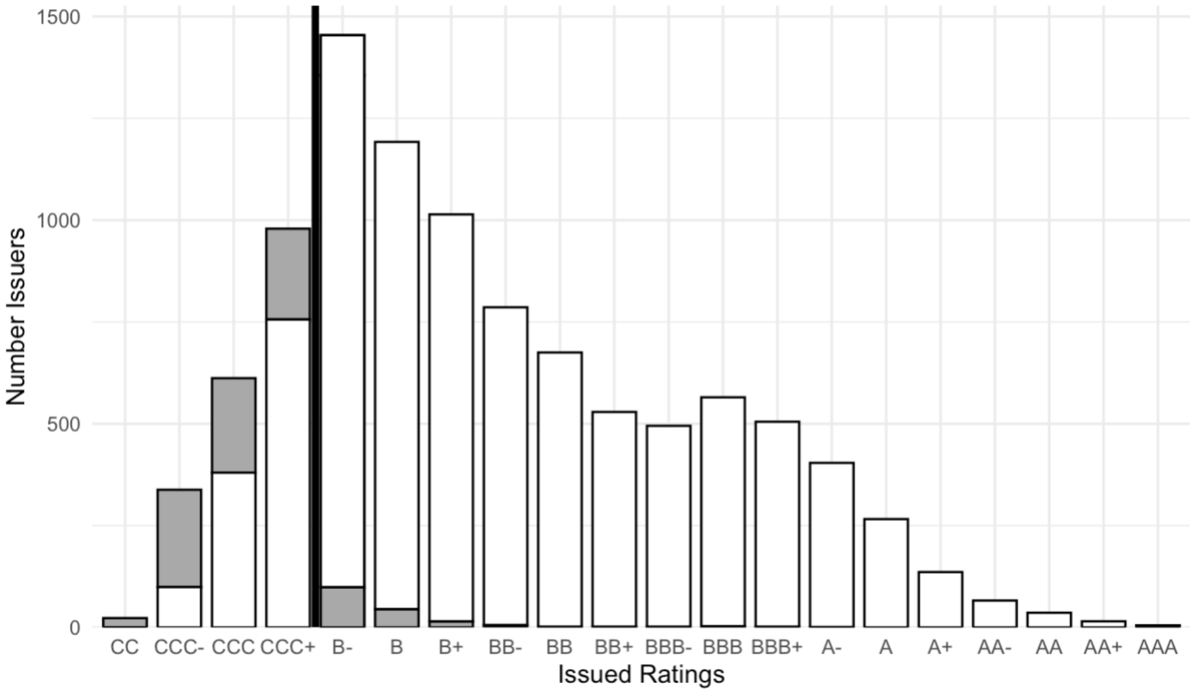
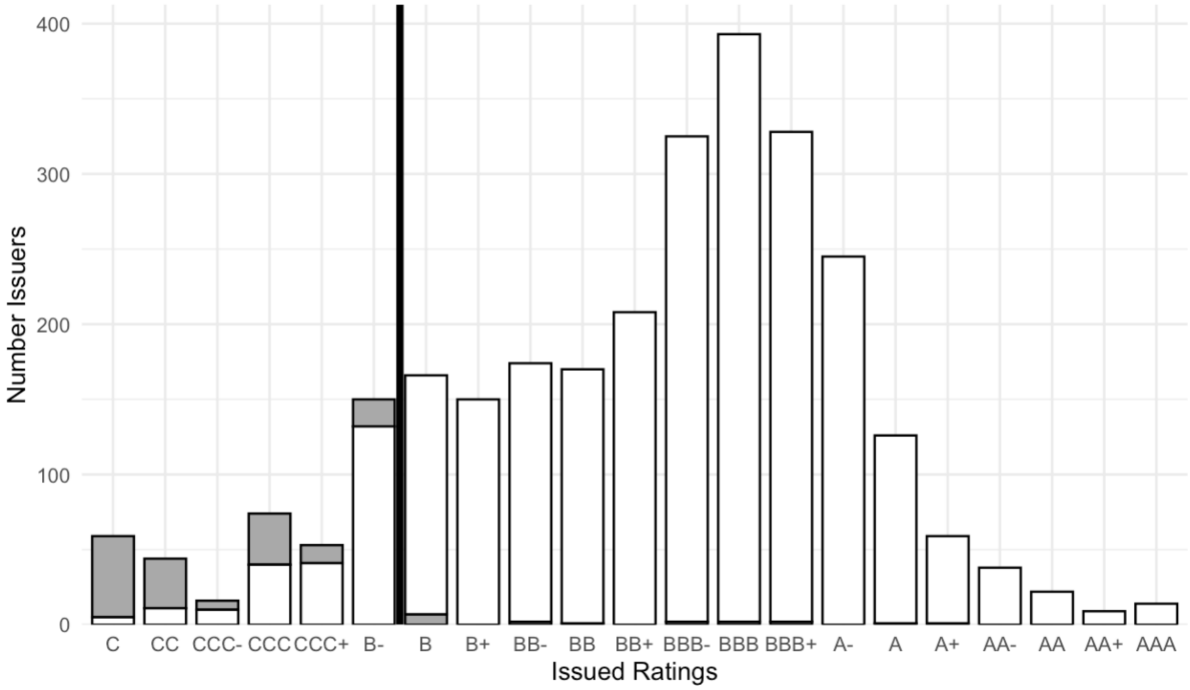


Figure 6.5.2: *Classification of Fitch Defaulters and Non-Defaulters*



Figures 6.5.1 and 6.5.2 illustrates how defaults are concentrated by S&P and Fitch. S&P's defaults are more evenly spread across its lower ratings below the threshold, except for the CC rating, which has relatively few defaults. In contrast, Fitch shows a higher concentration of defaults in its lowest ratings, particularly in C and CC. Above the threshold, both S&P and Fitch show a predominance of non-defaults. Around the threshold, Fitch has a smaller proportion of defaults compared to S&P, which in contrast shows a higher number of defaults in this overlapping area. The overlap near the threshold confirms Engelmann et al.'s (2003) observation that perfect separation between defaulting and non-defaulting companies is not achievable.

Table 6.5: *Results given cut off value retrieved by (Youden, 1950)*

Fitch

True Positive	*False Positive*	*True Negative*	*False Negative*
157	*239*	*2409*	*18*

S&P

True Positive	*False Positive*	*True Negative*	*False Negative*
964	*1258*	*7966*	*177*

At the threshold, Fitch identifies 157 true positives (defaults which is rated B- or worse) and 239 false positives (non-defaults but also given rating B- or worse) while S&P identifies 964 true positives and 1258 false positives. These figures, calculated cumulatively from the worst-rated entities to the best, directly inform the TPR and FPR values plotted on the ROC curve (see Tables A.3 and A.4 in the appendix).

At the CCC+ cutoff, the results for S&P reveal a TPR of 0.8449, indicating that 84.49% of defaults are placed within categories CCC+ ≥. The corresponding FPR is 0.1364, meaning 13.64% of non-defaults are misclassified as defaults. The maximized *J*, calculated to be 0.7085, shows that while many defaults have correctly been identified, also many have been classified left of the threshold. In comparison, Fitch's results at the B- threshold show a TPR of 0.8971, higher than that of S&P, as 89.71% of defaults are placed within category B- ≥. In contrast to S&P, Fitch's FPR is lower with a value of 0.0903, meaning that only 9.03% of non-defaults was placed left of the threshold. This yields a maximized *J* of 0.8069, better than S&P, and demonstrating a more favorable balance between true positives and false positives.

Figure 6.5.3: *ROC Curve S&P*

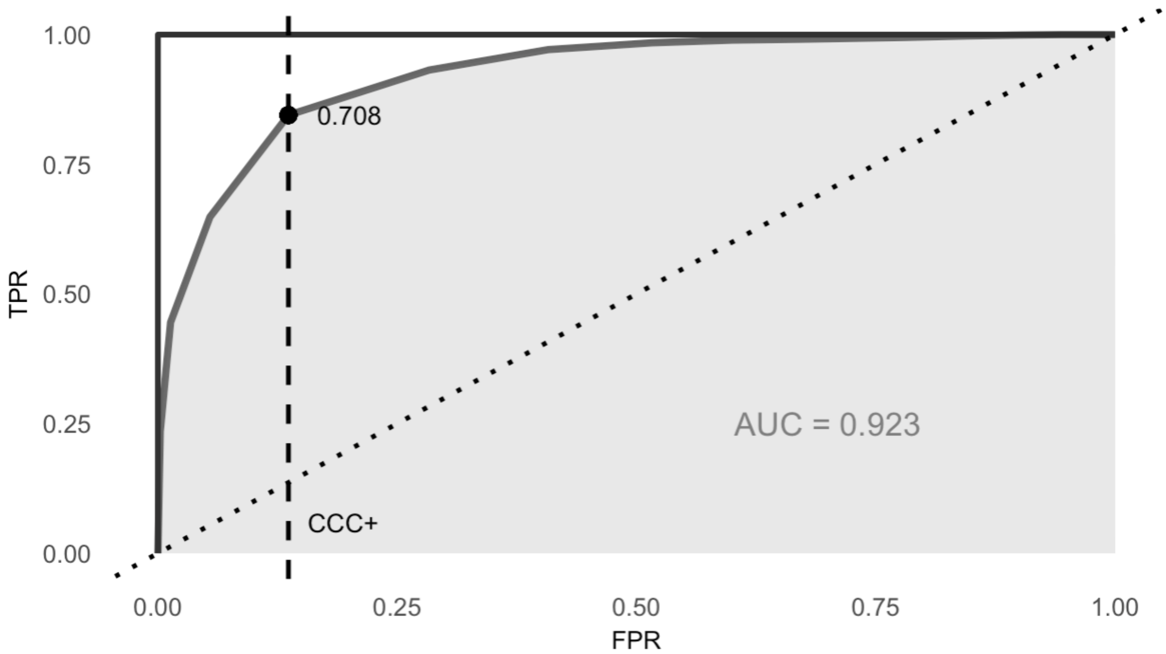
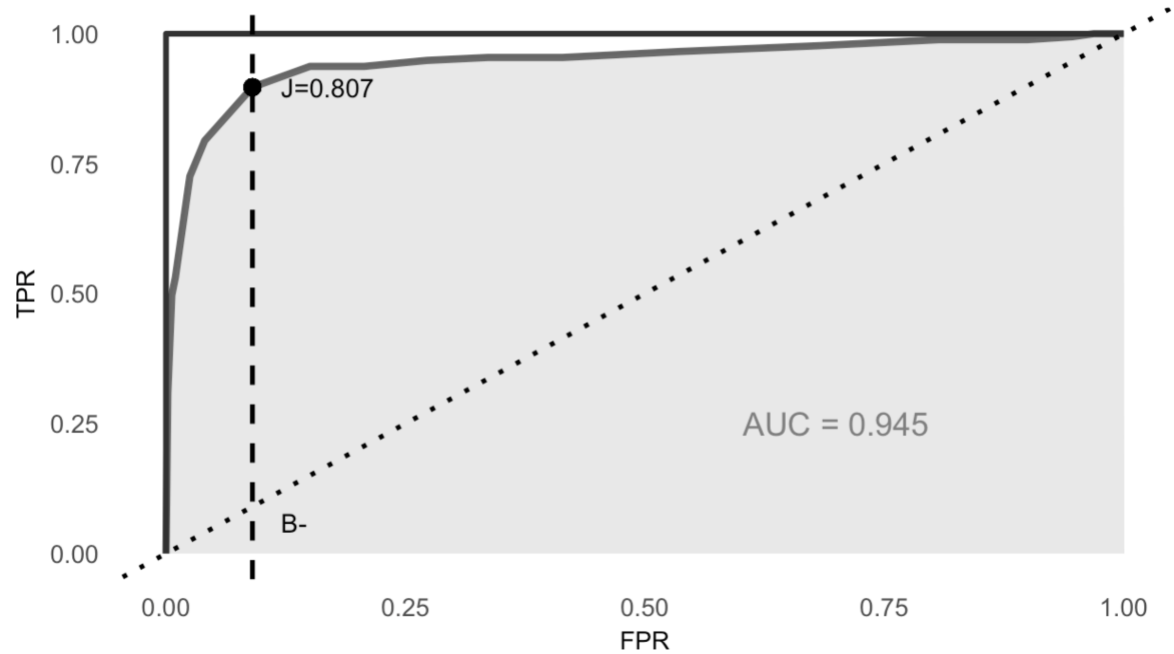


Figure 6.5.4: *ROC Curve Fitch*



The ROC curves for both Fitch and S&P demonstrate excellent predictive performance, as reflected in their respective AUC values. The calculated AUC value for S&P is 0.923, and 0.945 for Fitch. According to Çorbacıoğlu et al. (2023), AUC values between 0.9 and 1.0 are considered excellent, signifying a high level of model discrimination.

The ROC curve for Fitch (Figure 6.5.4) shows a steep initial rise, reflecting its effectiveness in identifying a large proportion of defaulting entities early, with a high HR, contributing to its larger AUC. Fitch's strength lies in concentrating defaults within its highest-risk categories (C and CC), achieving a TPR of 0.4971 by rating CC compared to S&P's 0.2358 (Tables A.3 and A.4). This contrasts with S&P (Figure 6.5.3), which distributes defaults more broadly across lower ratings, resulting in a slightly less pronounced initial rise and a lower, but still excellent, AUC (Çorbacıoğlu et al., 2023). By rating BB-, S&P captures 99% of all defaulting companies, whereas Fitch achieves this by rating BBB+, indicating S&P's greater ability to avoid classifying defaults within its investment-grade ratings. These findings align with the observed results in Chapter 6.4. Our findings extend the work of Engelmann et al. (2003), who validated the utility of ROC curves for assessing credit rating accuracy but did not compare agencies but rather used aggregated ratings.

While Fitch has a larger AUC, its AR value is lower than S&P's. This can be explained by the fact that the incline in the CAP curve, for the perfect model, arises because the x-axis represents the cumulative proportion of all issuers (including non-defaulters), while the y-axis represents the cumulative proportion of defaults. As the curve progresses, even a perfect model incorporates non-defaulting issuers along with defaults, leading to an incline rather than a vertical rise. For S&P, this incline favors its AR because S&P distributes defaults more gradually across its ratings, meaning the CAP curve remains closer to the perfect model in the start. This minimizes the area lost between the S&P CAP curve and the perfect model, increasing its AR. This distinction highlights the importance of using both the CAP and ROC models, as their focus on different aspects can lead to varying results.

6.6 Testing for Discriminiative Power of AUC

In this section, we assess the performance of the ROC by analyzing the AUC. To quantify the uncertainty in the AUC estimates, confidence intervals were calculated using the asymptotic normal approximation of the Mann-Whitney (1947) U-statistic. Additionally, hypothesis testing was performed to evaluate whether the observed AUC values significantly exceed predefined thresholds. These thresholds, 0.5, 0.7, 0.8, and 0.9, align with the standards established (Çorbacıoğlu et al., 2023). for interpreting AUC’s strength. The robustness analysis provide a framework for assessing the ROC’s predictive capabilities and their effectiveness in distinguishing between different outcomes.

Table 6.6.1: *Summary of Classification Results and AUC for S&P and Fitch*

		S&P	**Fitch**
N_D	*True Positives + False Negative*	*1 141*	*175*
N_{ND}	*True Negatives + False Positive*	*9 224*	*2 648*
AUC		*0.923*	*0.945*

Using the Mann-Whitney (1947) U-statistic, we calculated the following values:

Table 6.6.2: *Discriminative Power AUC*

AUC	*Sigma*	*95% Conf. Int.*	*99% Conf. Int.*	*Z-Value*
S&P				
0.5	0.0025	[0.919, 0.928]	[0.917, 0.930]	169.59
0.7	0.0116	[0.901, 0.946]	[0.894, 0.953]	19.32
0.8	0.0130	[0.898, 0.949]	[0.890, 0.957]	9.53
0.9	0.0135	[0.897, 0.950]	[0.889, 0.958]	1.74
Fitch				
0.5	0.0080	[0.930, 0.961]	[0.925, 0.966]	55.41
0.7	0.0302	[0.886, 1.005]	[0.868, 1.023]	8.13
0.8	0.0339	[0.879, 1.012]	[0.858, 1.033]	4.29
0.9	0.0356	[0.876, 1.092]	[0.854, 1.037]	1.28

Both agencies have AUC values significantly exceeding the random guessing threshold of 0.5, with exceptionally high Z-values (S&P: 169.59, Fitch: 55.41), confirming their discriminatory power is statistically significant at this basic benchmark. This indicates both models can reliably distinguish between defaulters and non-defaulters. Similarly, both surpass the "poor" and "fair" thresholds of 0.7 and 0.8 with substantial Z-values (19.32 and 9.53 for S&P, and 8.13 and 4.29 for Fitch). These results confirm statistically significant performance at these benchmarks, as defined by Çorbacıoğlu et al. (2023), underlining the robust discriminatory capabilities of both agencies.

At the "excellent" (Çorbacıoğlu et al., 2023), threshold of 0.9 however, the results become more nuanced. While Fitch achieves a slightly higher AUC of 0.945 compared to S&P's 0.923, the associated Z-values drop considerably. For Fitch, the Z-value of 1.28 falls short of both the 95% and 90% confidence levels, indicating that its performance cannot be deemed statistically significant for the "excellent" threshold. In contrast, S&P's Z-value of 1.74, while not meeting the 95% threshold (Z = 1.96), surpasses the critical value of 1.645 required for statistical significance at the 90% confidence level. This suggests that S&P's discriminatory power achieves "excellent" classification at the less stringent 90% confidence level, reflecting its slightly stronger and more consistent ability to identify defaults relative to Fitch.

These results support failing to reject *H2*, indicating that both Fitch and S&P demonstrate predictive accuracy significantly better than random guessing.

7.0 Conclusion

This thesis has explored the differences between Fitch and S&P in their ability to classify companies approaching default in the poorest rating categories, thereby signaling credit risk effectively. The importance of this topic lies in its contribution to understanding how credit rating agencies influence financial decision-making, market stability, and investor confidence. By examining inter-agency differences, this study addresses a gap in existing research, making it relevant to investors, and financial institutions seeking to assess credit risk reliably.

Using a dataset of 13,983 rating updates from 2014 to 2023, this study employed transition matrices, calculated default probabilities for all rating categories, and compared the AR and ROC curves of the two agencies. The analysis revealed differences in how Fitch and S&P utilize their different rating categories. Fitch's inclusion of a C rating creates a more normalized distribution of default probabilities across ratings. S&P, while having the C rating included within its rating scale, did not issue it for any US companies within our analyzed timeframe. As a result, default risk is spread more across S&P's lower speculative-grade ratings.

Both agencies demonstrated excellent discriminatory power through AR and ROC analyses. However, Fitch have a larger proportion of companies approaching default within investment-grade categories, contrasting with S&P, which categorized almost all defaulting companies to their speculative-grade ratings. Additionally, the analysis of downgrade patterns revealed that S&P applied gradually larger downgrades closing in on default. In contrast, Fitch issued larger downgrades earlier, resulting in less pronounced increases in downgrade size as defaults approached compared to S&P.

Overall, the thesis highlights the importance of understanding the distinct methodologies of CRAs and their implications for the financial market. The findings emphasize that differences in the use of ratings, can significantly impact the perception of credit risk. By addressing a gap in previous research, this study offers valuable insights into the complexities of credit rating practices and their impact on financial systems.

7.1 Further Research

Future research could expand on this study by exploring the role of Moody's alongside Fitch and S&P to provide a more comprehensive inter-agency comparison of credit rating practices. Additionally, further research could uncover potential cases of leniency in specific agencies, which may suggest the presence of conflicts of interest inherent in the issuer-pays model. Such an investigation could support the findings about conflicts of interest highlighted by Partnoy (1999). In particular, examining the 2008 financial crisis would provide valuable insights into whether certain agencies demonstrated greater leniency during this period. Furthermore, incorporating company-specific financial metrics, such as debt-to-equity ratios, liquidity indicators, and market volatility into machine learning models could significantly enhance the precision of predicting rating changes and defaults. Additionally, stress-testing credit ratings under simulated macroeconomic scenarios, such as sharp increases in interest rates or economic recessions, could offer valuable insights into how agencies adapt to global market shifts and refine the accuracy of their rating methodologies.

Bibliography

Altman, E. I., & Herbert A. Rijken. (2006). A Point-in-Time Perspective on Through-the-Cycle Ratings. Financial Analysts Journal, 62(1), 54–70. Retrieved from: http://www.jstor.org/stable/4480730

Basel Committee on Banking Supervision. (2010). Basel III: A global regulatory framework for more resilient banks and banking systems. Bank for International Settlements. Retrieved from: https://www.bis.org/publ/bcbs189.pdf

Belkin, B., Forest, L., & Suchower, S. (1998). A one-parameter representation of credit risk transition matrices. CreditMetrics Monitor, 44–56.

Bhalla, D. (2019). Gini, cumulative accuracy profile (CAP), and AUC. ListenData. Retrieved from: https://www.listendata.com/2019/09/gini-cumulative-accuracy-profile-auc.html

Bhatia, A. V. (2002). Sovereign credit ratings methodology: An evaluation (IMF Working Paper No. WP/02/170). International Monetary Fund. Retrieved from: https://papers.ssrn.com/Sol3/papers.cfm?abstract_id=880211

Binici, M., Hutchison, M. M., & Schindler, M. (2018). Are credit rating agencies discredited? Measuring market price effects from agency sovereign debt announcements (No. 3125010). Social Science Research Network. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3125010

Bloomberg L.P. (2024). Credit ratings data for our analysis. Retrieved from: Bloomberg Terminal at NHH.

Bobbitt, Z. (2020). *Two proportion Z-test: Definition, formula, and example*. Statology. Retrieved from: https://www.statology.org/two-proportion-z-test/

Bush, C. (2022). Dealing with the conflicts of interest of credit rating agencies: A balanced cure for the disease. Capital Markets Law Journal, 17(3), 334–364. Retrieved from: https://doi.org/10.1093/cmlj/kmac012

Cantor, R., & Mann, C. (2003). Measuring the performance of corporate bond ratings. Moody's Investors Service. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=996025

Claessens, S., Law, A., & Wang, T. (2018). How do credit ratings affect bank lending under capital constraints? BIS Working Papers No. 747. Bank for International Settlements. Retrieved from: https://www.bis.org/publ/work747.pdf

Çorbacıoğlu, Ş. K., & Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. Turkish Journal of Emergency Medicine, 23(4), 195–198. Retrieved from: https://pmc.ncbi.nlm.nih.gov/articles/PMC10664195/pdf/TJEM-23-195.pdf

Curtis, T. (2015). US shale oil dynamics in a low price environment (OIES Paper: WPM No. 62). The Oxford Institute for Energy Studies. Retrieved from: https://www.econstor.eu/bitstream/10419/246535/1/978-1-78467-043-6.pdf

Engelmann, B., Hayden, E., & Tasche, D. (2003). Testing rating accuracy. Risk, 16(1), 82–86. Retrieved from: https://www.bundesbank.de/resource/blob/704150/b9fa10a16dfff3c98842581253f6d141/mL/2003-10-01-dkp-01-data.pdf

European Securities and Markets Authority (ESMA). (2021). COVID-19 and credit ratings. Retrieved from: https://www.esma.europa.eu/sites/default/files/trv_2021_2-covid-19_and_credit_ratings.pdf

Fitch Ratings. (2020). Rating through cycle: What does this mean during the coronavirus pandemic? Video. Retrieved from: https://www.fitchratings.com/videos/rating-through-cycle-what-does-this-mean-during-coronavirus-pandemic-29-06-2020?autoplay=true

Fitch Ratings. (n.d.). Rating definitions. Retrieved from: https://www.fitchratings.com/products/rating-definitions

Grinder, B., & Cooper, D. (2017). Spies, abolitionists and the origins of credit rating agencies. Financial History Magazine, 120. Museum of American Finance. Retrieved from: https://www.moaf.org/publications-collections/financial-history-magazine/120/_res/id=Attachments/index=0/Eds%20Perspective.pdf

Hathaway, M., & Seitzinger, M. V. (2010). Credit rating agencies and their regulation (CRS Report No. R40613). Congressional Research Service. Retrieved from: https://crsreports.congress.gov/product/pdf/R/R40613/12

Irwin, R. J., & Irwin, T. C. (2013). Appraising credit ratings: Does the CAP fit better than the ROC? International Journal of Finance & Economics, 18(4), 396–408. Retrieved from: https://onlinelibrary.wiley.com/doi/abs/10.1002/ijfe.1471

Kiefer, N., & Larson, E. (2004). Testing simple Markov structures for credit rating transition. Economic and Policy Analysis Working Paper.

Kiff, J., Kisser, M., & Schumacher, L. (2013). Rating through-the-cycle: What does the concept imply for rating stability and accuracy? (200 Working Paper No. WP/13/64). International Monetary Fund. Retrieved from: https://www.imf.org/external/pubs/ft/wp/2013/wp1364.pdf

Krämer, W., & Güttler, A. (2008). On comparing the accuracy of default predictions in the rating industry. Social Science Research Network.

Listokin, Y., & Taibleson, B. (2010). If you misrate, then you lose: Improving credit rating accuracy through incentive compensation. Yale Journal on Regulation. Yale Law School, Public Law Working Paper No. 203; Yale Law & Economics Research Paper No. 402. Retrieved from: https://ssrn.com/abstract=1535514

Machek, O., & Hnilica, J. (2013). A stochastic model of corporate lifespan based on corporate credit ratings. International Journal of Engineering Business Management, 5(45). Retrieved from: https://doi.org/10.5772/56918

Mann, H., and D. Whitney, 1947, On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other, Annals of Mathematical Statistics 18, 50–60.

Moody's Investors Service. (2003). Moody's rating symbols and definitions. Retrieved from: https://www.moodys.com/sites/products/DefaultResearch/2003000000434370.pdf

Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and practical use for clinicians. Korean Journal of Anesthesiology, 75(1), 25–36. Retrieved from: https://pmc.ncbi.nlm.nih.gov/articles/PMC8831439/pdf/kja-21Nahm209.pdf

Partnoy, F. (1999). The Siskel and Ebert of financial markets?: Two thumbs down for the credit rating agencies. Washington University Law Quarterly, 77(3), 619–712.

S&P Global Ratings. (2020). Top 10 investor questions on our ratings process. Retrieved from: https://www.spglobal.com/ratings/en/research/articles/200604-top-10-investor-questions-on-our-ratings-process-11518157

S&P Global Ratings. (2023). Default, transition, and recovery: 2023 annual global financial services default and rating transition study. Retrieved from: https://www.spglobal.com/ratings/en/research/articles/240624-default-transition-and-recovery-2023-annual-global-financial-services-default-and-rating-transition-study-13137806

S&P Global Ratings. (2024a). *Default, transition, and recovery: 2023 annual global corporate default and rating transition study*. S&P Global. Retrieved from: https://www.spglobal.com/ratings/en/research/articles/240328-default-transition-and-recovery-2023-annual-global-corporate-default-and-rating-transition-study-13047827

S&P Global Ratings. (2024b). Default, transition, and recovery: Corporate defaults jumped 80% in 2023 to 153. S&P Global. Retrieved from: https://www.spglobal.com/ratings/en/research/articles/240116-default-transition-and-recovery-corporate-defaults-jumped-80-in-2023-to-153-12968832

S&P Global Ratings. (n.d.). Understanding credit ratings. Retrieved from: https://www.spglobal.com/ratings/en/about/understanding-credit-ratings

Seneta, E. (2006). Markov and the creation of Markov chains. University of Sydney. Retrieved from: https://www.maths.usyd.edu.au/u/eseneta/senetamcfinal.pdf

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. Science, 240(4857), 1285–1293. Retrieved from: https://basin.earth.ncu.edu.tw/download/courses/seminar_MSc/2008/0319-2_Measuring%20the%20Accuracy%20of%20Diagnostic%20Systems.pdf

Sy, A. N. R. (2009). The systemic regulation of credit rating agencies and rated markets (IMF Working Paper No. WP/09/129). International Monetary Fund. Retrieved from: https://doi.org/10.2139/ssrn.1426448

U.S. Securities and Exchange Commission. (2002). *Testimony concerning the roles of the SEC and FASB in financial reporting*. Retrieved from: https://www.sec.gov/news/testimony/051402tsrkh.htm

Vipond, T. (n.d.). Rating agency. Corporate Finance Institute. Retrieved from: https://corporatefinanceinstitute.com/resources/fixed-income/rating-agency/

White, L. J. (2010). The credit rating agencies. The Journal of Economic Perspectives, 24(2), 211-226. Retrieved from: https://www.aeaweb.org/articles?id=10.1257/jep.24.2.211

Yin, J., & Tian, L. (2014). Joint inference about sensitivity and specificity at the optimal cut-off point associated with Youden index. Computational Statistics & Data Analysis, 77, 1–13. Retrieved from: https://www.sciencedirect.com/science/article/pii/S0167947314000450?casa_token=0eVNJesvgE0AAAAA:hrDPzeYoB06ia3cur6p3Vglkd7fEyINAZivutx1OXWbHFJYyJCzgMb8UEpPMI0v07KAst5HI

Youden, W. J. (1950). Index for rating diagnostic tests. Cancer, 3(1), 32–35. Retrieved from: https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.1002/1097-0142(1950)3%3A1%3C32%3A%3AAID-CNCR2820030106%3E3.0.CO%3B2-3

Z Table. (n.d.). Z table: Standard normal distribution table. Retrieved from: https://www.ztable.net

Appendix

Table A.1: *S&P Transition Matrix (with Default)*

[1] "S&P Transition Matrix (with Default):"

	AAA	AA+	AA	AA-	A+	A	A-	BBB+	BBB	BBB-	BB+	BB	BB-	B+	B	B-	CCC+	CCC	CCC-	CC	C	Default
AAA	0.400	0.400	0.200	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AA+	0.167	0.000	0.417	0.333	0.000	0.000	0.000	0.000	0.083	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AA	0.000	0.119	0.429	0.452	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AA-	0.000	0.014	0.149	0.338	0.419	0.081	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A+	0.000	0.000	0.000	0.081	0.306	0.540	0.048	0.008	0.008	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A	0.000	0.000	0.000	0.003	0.199	0.273	0.428	0.077	0.017	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A-	0.000	0.000	0.000	0.003	0.006	0.243	0.240	0.464	0.039	0.003	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BBB+	0.000	0.000	0.000	0.002	0.005	0.042	0.394	0.202	0.319	0.027	0.002	0.002	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000
BBB	0.000	0.000	0.000	0.000	0.002	0.002	0.029	0.369	0.204	0.325	0.053	0.002	0.007	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.002
BBB-	0.000	0.000	0.002	0.002	0.002	0.002	0.007	0.039	0.423	0.206	0.215	0.078	0.014	0.002	0.005	0.000	0.000	0.000	0.000	0.002	0.000	0.000
BB+	0.000	0.000	0.000	0.000	0.000	0.004	0.002	0.004	0.071	0.313	0.180	0.353	0.058	0.008	0.002	0.004	0.000	0.000	0.000	0.000	0.000	0.000
BB	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.003	0.012	0.018	0.383	0.193	0.303	0.064	0.013	0.004	0.000	0.000	0.000	0.000	0.000	0.000
BB-	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.005	0.009	0.021	0.366	0.150	0.358	0.064	0.014	0.010	0.001	0.000	0.000	0.000	0.000
B+	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.006	0.010	0.005	0.015	0.325	0.139	0.419	0.062	0.005	0.008	0.001	0.001	0.000	0.003
B	0.000	0.001	0.000	0.001	0.000	0.002	0.001	0.003	0.004	0.003	0.004	0.010	0.020	0.248	0.114	0.501	0.066	0.017	0.004	0.003	0.000	0.001
B-	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.004	0.002	0.001	0.001	0.007	0.024	0.284	0.076	0.396	0.148	0.017	0.015	0.000	0.022
CCC+	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.001	0.001	0.029	0.245	0.034	0.292	0.168	0.103	0.000	0.124
CCC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.010	0.082	0.173	0.027	0.237	0.165	0.000	0.300
CCC-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.053	0.076	0.000	0.306	0.000	0.558
CC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.029	0.025	0.014	0.000	0.000	0.925
C																						
Default																						

Table A.2: *Fitch Transition Matrix (with Default)*

[1] "Fitch Transition Matrix (with Default):"

	AAA	AA+	AA	AA-	A+	A	A-	BBB+	BBB	BBB-	BB+	BB	BB-	B+	B	B-	CCC+	CCC	CCC-	CC	C	Default
AAA	0.700	0.300	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AA+	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AA	0.000	0.167	0.000	0.833	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AA-	0.000	0.000	0.400	0.180	0.420	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A+	0.000	0.000	0.000	0.292	0.185	0.462	0.031	0.000	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015
A	0.000	0.000	0.000	0.013	0.131	0.331	0.450	0.062	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006
A-	0.000	0.000	0.000	0.015	0.005	0.188	0.257	0.455	0.054	0.000	0.000	0.020	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BBB+	0.000	0.000	0.000	0.000	0.007	0.014	0.389	0.186	0.325	0.061	0.007	0.004	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.004
BBB	0.000	0.000	0.000	0.000	0.006	0.003	0.011	0.444	0.203	0.306	0.025	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BBB-	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.033	0.529	0.154	0.208	0.043	0.010	0.003	0.003	0.000	0.000	0.000	0.000	0.000	0.005	0.000
BB+	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.022	0.489	0.143	0.251	0.091	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.073	0.391	0.089	0.292	0.089	0.047	0.005	0.005	0.000	0.000	0.000	0.000	0.000
BB-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.006	0.011	0.022	0.318	0.168	0.302	0.117	0.028	0.011	0.011	0.000	0.000	0.000	0.000
B+	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.011	0.011	0.071	0.290	0.115	0.361	0.109	0.000	0.022	0.005	0.000	0.000	0.000
B	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.005	0.010	0.047	0.257	0.084	0.466	0.037	0.063	0.000	0.010	0.005	0.005
B-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.000	0.000	0.000	0.034	0.283	0.069	0.221	0.234	0.048	0.028	0.062	0.000
CCC+	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.024	0.000	0.143	0.333	0.071	0.190	0.071	0.143	0.000	0.024
CCC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.027	0.027	0.135	0.041	0.054	0.041	0.297	0.270	0.108
CCC-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.600	0.400	0.000
CC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.548	0.452
C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.034	0.000	0.017	0.000	0.931
Default																						

Table A.3: *TPR and FPR S&P*

Rating	*Count D*	*Count ND*	*Cum D*	*Cum ND*	*TPR*	*FPR*	*J*
C	0	0	0	0	0.0000000	0.0000000	0.0000000
CC	269	23	269	23	0.2357681	0.0024935	0.2332646
CCC-	239	99	508	122	0.4452235	0.0132264	0.4319971
CCC	232	380	740	502	0.6485539	0.0544232	0.5941307
CCC+	223	756	964	1258	0.8448729	0.1363833	0.7084896
B-	99	1355	1063	2613	0.9316389	0.2832827	0.6483562
B	45	1147	1108	3760	0.9710780	0.4076323	0.5634457
B+	15	999	1123	4759	0.9842244	0.5159367	0.4682877
BB-	6	780	1129	5539	0.9894829	0.6004987	0.3889842
BB	2	673	1131	6212	0.9912358	0.6734605	0.3177752
BB+	2	527	1133	6739	0.9929886	0.7305941	0.2623945
BBB-	2	493	1135	7232	0.9947415	0.7840416	0.2106998
BBB	3	562	1138	7794	0.9973707	0.8449696	0.1524011
BBB+	2	503	1140	8297	0.9991236	0.8995013	0.0996223
A-	1	403	1141	8700	1.0000000	0.9431917	0.0568083
A	0	266	1141	8966	1.0000000	0.9720295	0.0279705
A+	0	136	1141	9102	1.0000000	0.9867736	0.0132264
AA-	0	66	1141	9168	1.0000000	0.9939289	0.0060711
AA	0	36	1141	9204	1.0000000	0.9978317	0.0021683
AA+	0	15	1141	9219	1.0000000	0.9994579	0.0005421
AAA	0	5	1141	9224	1.0000000	1.0000000	0.0000000

Table A.4: *TPR and FPR Fitch*

Rating	*Count D*	*Count ND*	*Cum D*	*Cum ND*	*TPR*	*FPR*	*J*
C	54	5	54	5	0.3085714	0.0018882	0.3066832
CC	33	11	87	16	0.4971429	0.0060423	0.4911006
CCC-	6	10	93	26	0.5314286	0.0098187	0.5216098
CCC	34	40	127	66	0.7257143	0.0249245	0.7007898
CCC+	12	41	139	107	0.7942857	0.0404079	0.7538779
B-	18	132	157	239	0.8971429	0.0902568	0.8068861
B	7	159	164	398	0.9371429	0.1503021	0.7868407
B+	0	150	164	548	0.9371429	0.2069486	0.7301942
BB-	2	172	166	720	0.9485714	0.2719033	0.6766681
BB	1	169	167	889	0.9542857	0.3357251	0.6185606
BB+	0	208	167	1097	0.9542857	0.4142749	0.5400108
BBB-	2	323	169	1420	0.9657143	0.5362538	0.4294605
BBB	2	391	171	1811	0.9771429	0.6839124	0.2932305
BBB+	2	326	173	2137	0.9885714	0.8070242	0.1815473
A-	0	245	173	2382	0.9885714	0.8995468	0.0890246
A	1	125	174	2507	0.9942857	0.9467523	0.0475334
A+	1	58	175	2565	1.0000000	0.9686556	0.0313444
AA-	0	38	175	2603	1.0000000	0.9830060	0.0169940
AA	0	22	175	2625	1.0000000	0.9913142	0.0086858
AA+	0	9	175	2634	1.0000000	0.9947130	0.0052870
AAA	0	14	175	2648	1.0000000	1.0000000	0.0000000

Table A.5: *Z-table negative*

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414
-0.1	.46017	.45620	.45224	.44828	.44433	.44034	.43640	.43251	.42858	.42465
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-1	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08692	.08534	.08379	.08226
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-2	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-3	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-4	.00003	.00003	.00003	.00003	.00003	.00003	.00002	.00002	.00002	.00002

Extracted from Z Table. (n.d.).

Table A.6: *Z-table positive*

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749	.57142	.57535
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
+1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
+1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
+1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
+1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
+2	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
+2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
+2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
+2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
+2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
+2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
+2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
+2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
+2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
+2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
+3	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
+3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
+3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
+3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
+3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
+3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
+3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
+3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
+3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
+3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997
+4	.99997	.99997	.99997	.99997	.99997	.99997	.99998	.99998	.99998	.99998

Extracted from Z Table. (n.d.).