UNIVERSITY OF CALIFORNIA

Los Angeles

Determinants of Returns in the US High Yield Bond Market from 2018 to 2023

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Applied Statistics and Data Science

by

Kanishka Malik

2024

ABSTRACT OF THE THESIS

Determinants of Returns
in the US High Yield Bond Market
from 2018 to 2023

by

Kanishka Malik
Master of Applied Statistics and Data Science
University of California, Los Angeles, 2024
Professor Hongquan Xu, Chair

High Yield bonds are a major component of the investment portfolios of institutions and individuals, and credit analysts use a wide range of variables to predict their returns. This study explores the nature of quarterly High Yield bond returns from June 30 2018 to June 30 2023 of a sample of High Yield bonds issued by US companies, and several of the variables commonly used by credit analysts. Using methods such as Bivariate Regression, Multiple Regression, and Random Forest, this research analyzes the relationship between Bond Returns and the independent variables, and uses different validation techniques to assess the quality of those models.

The thesis of Kanishka Malik is approved.

Dave Anthony Zes

Maryam Mahtash Esfandiari

Frederic R.P. Schoenberg

Hongquan Xu, Committee Chair

University of California, Los Angeles

2024

To my parents

TABLE OF CONTENTS

1 Introduction **1**

2 Data Collection & Analysis **2**

2.1 High Yield Bond - Definition 2

2.2 Data Collection 2

2.3 Dependent Variable: Quarterly Returns of Bonds 4

2.4 Independent Variables 6

2.4.1 Financial Leverage 7

2.4.2 Interest Coverage 9

2.4.3 Lagging Treasury Bond Returns 9

2.4.4 Lagging Bond Returns 9

2.4.5 Sector 10

2.4.6 Lagging Stock Returns of the Issuer 10

2.4.7 Modified Duration 11

2.4.8 Lagging Total Returns of the S&P 500 Index 11

2.4.9 Lagging Returns of the High Corporate Bond Index 11

2.4.10 Lagging GDP Growth 11

2.5 Exploratory Data Analysis 12

3 Forecasting Models **25**

3.1 Linear Regression 25

3.1.1 Relationship with Lagging Bond Returns 28

3.2 Time Series Cross Validation 32

3.3 Cyclic Permutation Test 34

3.4 Multiple Regression 36

3.5 Random Forest Analysis 41

4 Conclusion & Limitations 44

4.1 Conclusion 44

4.2 Limitations 45

A Appendix of Supplementary Figures and Tables 46

References 54

LIST OF FIGURES

2.1 Percentage of Bonds in Sample with no prices in the quarter 5
2.2 Distribution of Quarterly Bond Returns in Sample 13
2.3 Boxplot of Bond Returns by Quarter 14
2.4 Standard Deviation of Bond Returns by Quarter 14
2.5 Median Bond Returns by Quarter 15
2.6 Distribution of Sign Log of Net Leverage as Multiple of LTM EBITDA . . . 16
2.7 Distribution of Sign Log of Net Leverage as Multiple of Forward EBITDA . 17
2.8 Distribution of Sign Log of Net Debt to UFCF 17
2.9 Distribution of Sign Log of Interest Coverage as multiple of LTM EBITDA . 18
2.10 Distribution of Sign Log of Interest Coverage as multiple of Forward EBITDA 18
2.11 Distribution of Net Debt to Total Enterprise Value 19
2.12 Distribution of Stock Return of Underlying Issuers 20
2.13 Boxplot of Modified Duration of Bonds by Quarter 20
2.14 Quarterly Returns of 10y Treasury, 5y Treasury, Bloomberg HY Index and S&P 500 22
2.15 Quarterly-on-Quarter Nominal and Real GDP Growth by Quarter 22
2.16 Year-on-Year Nominal and Real GDP Growth by Quarter 23
2.17 Correlation Matrix 24

3.1 Distributions of Residuals of Bivariate Regressions 27
3.2 Distribution of ACF of Bond Returns at Lag of 1 29
3.3 Distribution of ACF of Bond Returns at Lag of 2 30
3.4 Scatterplot and Regression Line of Bond Returns vs Lagging Bond Returns . 31

3.5 Scatterplot and Regression Line of Bond Returns vs Lagging Bond Returns by Regime 31
3.6 Illustration of Evaluation on Rolling Forecasting Origin [HA21] 33
3.7 Test Correlations for Bivariate Regression from Simulated Datasets, with Correlation from Actual dataset in Red Line 36
3.8 Fitted vs Residuals 39
3.9 Quantile-Quantile Plot 40
3.10 Distribution of Residuals from Multiple Regression 40
3.11 Test Correlations for Multiple Regression from Simulated Datasets, with Correlation from Actual dataset in Red Line 41
3.12 Variable Importance in Random Forest Regression 43

A.1 Histogram of Quarterly Bond Returns excluding AMC Entertainment's bonds 46
A.2 Histogram of Residuals for Bivariate Regressions with Quantitative Variables excluding AMC Entertainment's bonds 47
A.3 Linear Relationship with each Independent Variable 48
A.4 Regression Lines for each Independent Variable by Regime 49
A.5 Quarterly Bond Returns Distributions per Quarter 50

LIST OF TABLES

3.1 Summary Statistics of Bivariate Regressions 26

3.2 Summary Statistics of Bivariate Regressions excluding AMC bonds 28

3.3 Cross Validation Results of Bivariate Regressions 34

3.4 Summary Statistics of Multiple Regression excluding highly correlated variables and GICS Industry 38

3.5 VIF of variables in Multiple Regression excluding highly correlated variables and GICS Industry 39

A.1 Summary Statistics of Multiple Regression Excluding AMC Entertainment's bonds 51

A.2 Summary Statistics of Multiple Regression using 10-year Treasury bond's returns instead of the 5-Year Treasury 51

A.3 VIF of Multiple Regression including all Quantitative Variables 52

A.4 Summary Statistics of Multiple Regression using all the Variables 53

ACKNOWLEDGMENTS

I would like to thank Professor Dave Zes, my Advisor, for his invaluable support in performing this research. I am also grateful to Oaktree Capital Management LP, for allowing me to access resources that made this research possible.

CHAPTER 1

Introduction

High Yield bonds, also known as 'Junk Bonds', are a major source of financing for corporations in the United States, estimated at $1.4 trillion outstanding in 2023 [Hor23]. While bonds are often held by investors from issuance to maturity, a large number of them are traded between investors in the secondary market through broker-dealers. As a result, it is possible to observe the statistical relationship between market returns of a High Yield bond and other variables, such as the financial leverage of the company, returns of a company's underlying stock, past returns of the bond itself, past returns of major stock or bond indices, or macroeconomic variables such as GDP growth.

Bond prices may change due to a multitude of factors such as changes in interest rates, financial conditions of a company or industry, economic conditions, regulations, and changes in expectations of such factors. This study uses data related to a portfolio of bonds that were rated high yield as of June 2018, to better understand what factors best explained the quarterly returns of the US High Yield bond market from mid-2018 to mid-2023, which saw a wide range of economic conditions. The research used bivariate regression, multiple regression, and Random Forest, to determine the relationship between the quarterly returns of the bonds and various lagging independent variables. It used time series cross validation with a rolling origin point and conducted cyclic permutation to understand the significance of these relationships.

CHAPTER 2

Data Collection & Analysis

2.1 High Yield Bond - Definition

A bond is classified as High Yield (also known as 'Junk' or 'Speculative Grade') if it is issued by a company that has a higher risk of default, and such companies typically have a S&P rating that is BB+ or worse [Fab12]. A company may have a higher risk of default due to a combination of factors such as higher financial leverage, deteriorating financial performance, elevated risks in its sector, or being an emerging company. This is in contrast with Investment Grade companies, which have a much lower risk of default and are rated BBB- or above. When a company issues a bond, it has the obligation to pay a coupon to the bondholder, which is usually a fixed percentage of the face value, and to repay the face value to the lenders at a pre-determined maturity date in the future, along with meeting certain requirements set forth in the bond indenture. The higher the default risk is for a borrower, the higher the coupon it needs to offer to generate sufficient demand from buyers for its bonds. Since the least risky bond is a Treasury bond, the bonds of all other issuers are issued at a premium to the yield of Treasury bond with the equivalent maturity. The riskier a bond is, the greater this premium needs to be to attract sufficient interest from investors.

2.2 Data Collection

The two primary sources of data for this research were Bloomberg and S&P Capital IQ, which are market and financial data platforms. The author obtained data related to the prices and returns of bonds through his access to Bloomberg Terminal, which has vast repositories of market data related to fixed income and other instruments. Financial Data related to the companies that issued these bonds was obtained using Capital IQ, which is owned and operated by S&P Global. Bloomberg also provided qualitative data related to the bonds and its issuer, and quantitative data related to the returns of US five-year and ten-year Treasuries, quarterly returns of the underlying issuer's stock. The "SRCH" function on Bloomberg allows the user to select bonds that were outstanding at a given point in time based on a range of criteria from the entire bond universe available in Bloomberg's database.

The criteria selection process was determined with the objective of building a dataset of High Yield borrowers for whom there would be sufficient pricing data and for whose issuers there would be publicly available financial data. The research selected USD-denominated High Yield bonds that had at least $250 million outstanding face value on June 30, 2018, since bonds with lower outstanding face value are less likely to trade frequently and would have less accurate pricing data. The study selected bonds whose maturity was beyond June 30, 2023, to increase the likelihood of capturing a large enough time period, with sufficient data across the different monetary regimes and economic conditions. As a result, all the bonds selected had an effective maturity of over five years. As of June 30, 2018, the median and mean maturity of the bonds was 6.9 years and 8.5 years respectively, with the longest maturity of 26 years. Also, bonds of the same bond issue are often issued as both, 144A and Reg S; the study only looks at the 144A bonds as pricing on Reg S bonds was not available on Bloomberg. It also excluded convertible bonds as they have a different risk profile than non-convertible High Yield bonds, and excluded private placement bonds as those do not typically trade and thus, have no actionable pricing.

Moreover, the study focused on bonds of US based companies that were publicly traded as of June 30, 2018. This is because financial data of public companies are available in their SEC filings and is aggregated on platforms like S&P Capital IQ, while data of private companies is usually not disclosed to the public. Also, this made it possible to use the returns of the underlying issuer's stock and Net Debt to Total Enterprise Value as independent variables. For financial data related to the issuers, such as financial leverage, the study relied on S&P's Capital IQ platform .

The final dataset was a combination of the datasets from Bloomberg and Capital IQ, comprising quarterly bond returns data for 222 bonds issued by 118 companies, with lagging quarterly financial leverage and interest coverage of the issuers, the issuers' lagging quarterly stock return, the lagging quarterly returns on US Treasury Bonds (labeled as 'Govt_Ret_5y' for 5-year Treasury Returns and 'Govt_Ret_10y' for the 10-year Treasury Returns), annual and quarterly GDP growth in the prior quarter. A company can have multiple issues of high yield bonds outstanding, and this data set included 60 companies that each had more than one high yield bond issue in the sample. The study also excluded bonds of financial companies, such as banks, insurers, and lending companies since these companies have very different business models and analysts typically rely on a different set of financial variables and frameworks to measure their creditworthiness. For instance, the core activity of a bank is to borrow from depositors and bond holders to lend to other borrowers – they are essentially in the business of managing financial assets and appear highly levered by the metrics of a non-financial company.

As the unit of observation is quarterly bond return, all the lagging returns are also quarterly, such as that of the underlying stock, S&P 500 index, Treasury Bonds. The objective of the study was to use the data available in one quarter to project a bond's returns in the subsequent quarter. For this study, quarters end on March 31, June 30, September 30, and December 31, and they are labeled as 'Q1', 'Q2', 'Q3' and 'Q4' respectively, for each calendar year. The paper focuses on quarterly bond returns as most companies report financial performance on a quarterly basis, enabling us to use updated lagging financial information for the analysis. This analysis was considered from the standpoint of an investor holding these bonds for the medium to long term, so the optimal unit of observation was quarterly returns, analyzed over a multi-year period.

2.3 Dependent Variable: Quarterly Returns of Bonds

The dependent variable is the quarterly total return of a bond, which is calculated as the change in price in the quarter plus the coupon accrued in the quarter for each bond, divided by the price of the bond at the end of the prior quarter, in each quarter where prices were available. To determine this, the quarter-end prices for each bond were captured from Bloomberg and the change in price from the last quarter's end was used. If there is no change in price, the total return would just be the coupon accrued in the quarter. The median coupon of the bonds in this dataset was 5.75%, the average was 5.98%, and the maximum coupon was 11%.

During this five-year period, some bonds would be refinanced or prepaid prior to their maturity date, and 21 bonds or 9.5% of the initial sample would default, due to which, pricing data was not available for those bonds for certain quarters. 119 of the 222 bonds or 53.6% had no pricing in Q2 of 2023, the final time period of this analysis (See Figure 2.1).

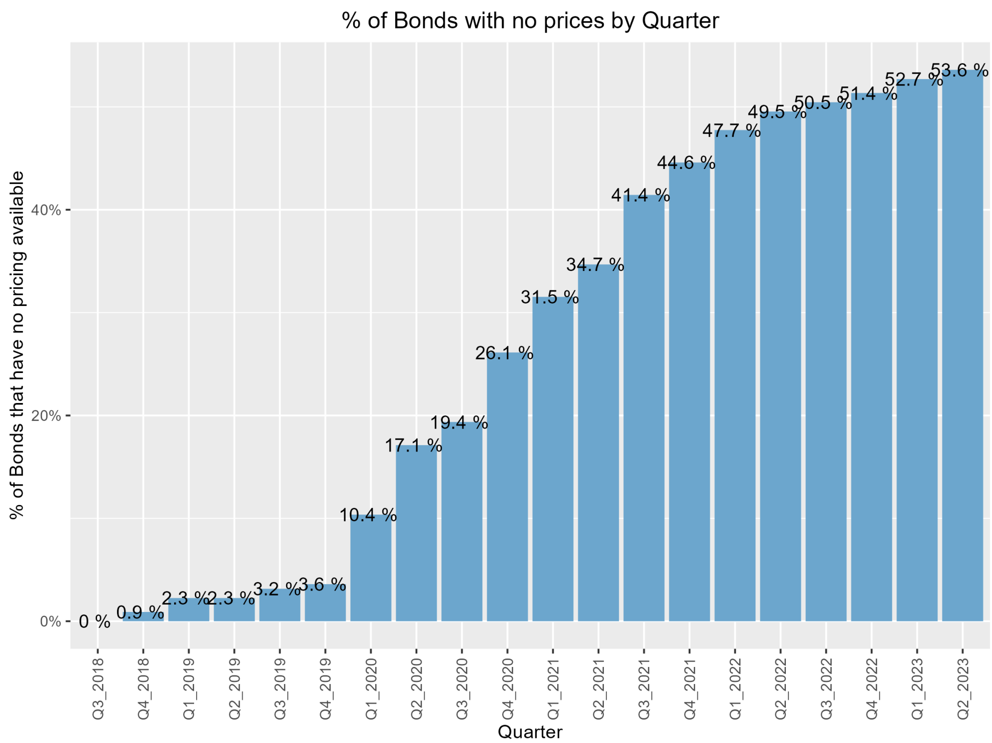

Figure 2.1: Percentage of Bonds in Sample with no prices in the quarter

For bonds that prepaid before mid-2023, the price used was the one at which the bond was repaid or called (usually anywhere from 100 to 102). For bonds that defaulted, the last traded price available on Bloomberg was used. While it is possible that after the issuer restructured its debt, the eventual recovery on the defaulted bond would be different than the last traded price, this research is focused on performing high yield bonds and determined that outcomes of restructuring or bankruptcies were outside the scope of this study. Once a bond was called or stopped trading after a default, the bond price and subsequent quarterly return was denoted as NA. There is a sharp increase in the percentage of the sample that had no pricing available during Covid, as some companies defaulted on their debt while others refinanced their bonds after interest rates were lowered.

2.4 Independent Variables

The final dataset included a wide range of variables, all of which were time variant, except for GICS Industry, which was fixed through time. The time-variant independent variables broadly fall into two buckets. First, there are variables that are specific to a bond or its issuer, such as lagging bond returns or net leverage of an issuer, and which vary each quarter. The second bucket are variables that are not specific to an issuer and are macroeconomic in nature, such as lagging GDP growth or lagging Treasury Bond returns. Not all of these variables would be statistically significant for explaining quarterly high yield bond returns, but the research considered all of them. The quantitative variables lag the dependent variable by 1 quarter as the research tried to use the latest data available in a given quarter to predict the next quarter's returns. For example, the return of a company's stock in Q4 2018 would be used to predict the quarterly returns of its bonds in Q1 2019, or the S&P 500 index return in Q2 2019 would be used to predict the bonds' returns in Q3 2019.

For financial data related to companies, the data would effectively be lagging by two quarters. Most companies take 30 to 60 days to report their financial results for a quarter, so the financial data of the issuers would, in almost all cases, effectively lag the dependent variable by two quarters. For instance, Acadia Healthcare reported its Q4 2018 results on Feb 28, 2019. If, at the end of Q4 2018, on Dec 31, 2018, the returns for Q1 2019 had to be projected using Acadia's financial performance, the latest available financial data for Acadia Healthcare would be from its fiscal quarter Q3 2018, ending September 30, 2018, since results for Q4 2018 would not be available as yet. Some companies have a different reporting calendar, with fiscal quarters ending on different dates, but the research used the most recent data that was available in a given quarter to make predictions for the next quarter.

For companies that defaulted, such as the McClatchy Company, or were acquired, such as CDK Global or Covanta Holdings, public financials were not available subsequent to the event and their financial data for those quarters was entered as NA. To the extent a company was acquired by a public company and its bonds remained outstanding in subsequent quarters, the research used the acquirer's financials for the remaining quarters of that bond's issuer since the acquiring company typically assumes the target's debt. The final data set had 17 independent variables, only one of which, the GICS industry, was categorical. The definitions of these variables and context for using them are laid out below.

2.4.1 Financial Leverage

The dataset included four measures of leverage for each company – the ratio of Net Debt to the Last Twelve Months ("LTM") Earnings Before Interest Tax and Depreciation ("EBITDA"), and Net Debt to Next Twelve Months (or Forward) EBITDA, Net Debt to LTM Unlevered Free-Cash-Flow ("UFCF"), Net Debt to Total Enterprise Value ("TEV"). All of the variables used to compute these ratios were obtained using the financial data of the issuers from Capital IQ. Net Debt is a measure of a company's indebtedness, measured as the face value of all debt minus cash & cash equivalents held on the company's balance sheet. The EBITDA of a company is a measure of economic profit and is often used to determine the level of debt that a company can service. Capital IQ also aggregates the forward expectations from various sources to determine an expected or Forward EBITDA level for the next twelve months, which is also used for measuring leverage since it is an indicator of expected economic profit.

Financials analysts also use the free-cash-flow of a business to determine the level of debt a business can support as this is a measure of the cash generated by a business after accounting for capital expenditures and net working capital requirements. Therefore, the ratio of Net Debt to the LTM UFCF ("UFCF") was used as a predictor. Due to the significant dispersion in the EBITDA and UFCF levels, as discussed in Chapter 2.5, the analysis used a sign log transformation for ratios using those metrics.

Another measure of leverage is the ratio of Net Debt to TEV. The TEV is typically measured as the sum of the Net Debt of a company and its market capitalization. As market capitalization changes in real time, so does the TEV. For any given quarter, the ratio uses the latest available Net Debt figure and the TEV. For instance, for Acadia Healthcare, the TEV as of March 31, 2020 would use the market capitalization on March 31, 2020 but the Net Debt and EBITDA data from the quarter ending Dec 2019, as data for Q1 2020 would not yet be available.

Market Capitalization changes every trading day, so the denominator in this ratio contains up-to-date information about the market's perception of future performance and risks of the business. Financial accounting information such as Net Debt and EBITDA is only updated quarterly, so the Net Debt to TEV may contain more up to date information on the default risk. For instance, in Q1 2020, as many companies experienced a severe decline in stock prices, the Net Debt to TEV ratio at the end of Q1 2020 would have been higher, which can be used to make projections for bond returns in Q2 2020. But the Net Debt to LTM EBITDA for Q1 2020 would not be reported until Q2 2020 for most companies, so any projection made using this data in Q1 2020 for Q2 2020 would be using data from Q4 2019, which would not reflect the rapidly changing outlook at the onset of Covid. Note that this ratio can be negative if a company has more cash than total debt, resulting in a negative Net Debt figure.

The higher a company's financial leverage, the greater its risk of default, and the higher the yield on the bonds are likely to be. However, what constitutes high financial leverage for a company depends on a range of factors, such as the cyclicality of earnings, the size of the company, secular trends of its industry, capital intensity, its market power relative to suppliers and customers.

2.4.2 Interest Coverage

The analysis also considered the Interest Coverage of the issuers, which is the measure of a company's capacity to meet the interest payments on its debt. There were two measures of Interest Coverage – the ratio of LTM EBITDA to LTM Interest Expense and Forward EBITDA to LTM Interest Expense. The higher the interest coverage, the more likely the issuer is to meet its interest payments. As Interest Coverage as a multiple of LTM EBITDA was highly correlated with Interest Coverage as a multiple of Forward EBITDA, only the former was used in the forecasting models.

2.4.3 Lagging Treasury Bond Returns

The coupons of High Yield bonds are priced at a premium to the yields of Treasury bonds. As interest rate expectations change, so do Treasury bond prices. As returns on corporate bonds are also affected by interest rates and are at least partially linked to the risk-free Treasury bonds, the research analyzed the relationship between the lagging returns of government bonds. There were multiple Treasury bonds of various maturities outstanding in June 2018, so two of those bonds were selected, with approximately five years and ten years remaining maturity as of June 30 2018, and the research considered their lagging quarterly returns to predict the quarterly returns of the High Yield bonds. Lin, Wang, and Wu (2014) [LWW14] attempt to predict corporate bond excess returns for a sample of corporate bonds using several variables, one of which is a factor composed using the linear combination of forward rates and the term spread between the 10-year and 1-year Treasury yields. This paper simply used lagging quarterly returns of the two Treasury bonds to predict subsequent returns of the High Yield bonds.

2.4.4 Lagging Bond Returns

The returns of a bond in a given quarter may contain predictive value for returns of the same bond in the next quarter. Chapter 3.1.1 discusses the autocorrelation in bond returns and how the relationship with lagging returns varied across different monetary regimes.

2.4.5 Sector

The returns and relationship of returns to other variables may vary by sector. MSCI determines the GICS or Global Industry Classification Standard which classifies companies across different sectors [MSC23]. The dataset had bonds across 10 different sectors with Consumer Discretionary comprising the largest number of bonds at 59.

2.4.6 Lagging Stock Returns of the Issuer

A company's common stock represents ownership in that business and is a residual claim on its cash flows, and is subordinated to all other obligations, including bonds, in the company's capital structure. Unlike bonds, the stock is a claim of ownership in a business, not a debt owed by the business, and the stock has no contractual maturity date or any obligation to pay interest. Bondholders may only receive the face value at maturity date, possibly with a small call premium if the indenture allows for it and assuming the bond matures within the required time frame, but a stock has no limit on its price. However, the bonds and stock of a company are both linked to a company's cash flows and assets. Therefore, lagging stock returns may contain predictive information for the bond's returns. As all companies in this sample were public companies, there was stock return data for every issuer in almost every time period. For instances where the company was acquired by another public company, the stock returns of the acquirer's stock were used.

For companies that defaulted or were acquired by a private company, there was no stock return data for the periods subsequent to the event of default. The research studies the relationship between the stock's quarterly return in a given quarter with the issuer's bond's return in the subsequent quarter. Hong, Lin, and Wu (2012) [HLW12] show evidence that returns in the Investment-Grade and High-Yield bond markets can be predicted by past stock returns.

2.4.7 Modified Duration

The modified duration of a stock measures the sensitivity of a bond's price to a change in the underlying yield [Fab12]. The modified duration of each bond was computed, for each quarter, using the *BondValuation* package in R, built using the methodology proposed by Wadim Djatschenko in 2019 [Dja19]. The modified duration generated by the *BondVal.Yield()* function from this package represents approximately how much the bond's price would change for a 100 basis points change in yield.

2.4.8 Lagging Total Returns of the S&P 500 Index

The S&P 500 Index is a market capitalization weighted index of the 500 largest companies listed on exchanges in the US [SPX23]. As it is among the most widely referenced indices that tracks the performance of the US corporate sector, past quarterly returns of the index may explain subsequent returns of the bonds in this sample.

2.4.9 Lagging Returns of the High Corporate Bond Index

The Bloomberg Corporate High Yield Index tracks the performance of a portfolio of USD denominated High Yield bonds issued by non-emerging market companies [Blo24]. While its composition is different from the sample in this portfolio, the companies are of similar credit quality and many of them were members of the index in this time period.

2.4.10 Lagging GDP Growth

The analysis included lagging Quarter-on-Quarter and Year-on-Year GDP growth by quarter, for both, nominal and real GDP, and all seasonally adjusted. GDP is a measure of economic output for a country and real GDP growth omits the impact of inflation from this figure [GDP24]. This data was obtained from Bloomberg, which aggregates it from the Bureau of Economic Analysis. Similar to the financial data, GDP growth data is typically reported with a lag, so the GDP return for Q3 will be reported in Q4. Therefore, if, at the end of

Q4, GDP growth is used to predict the returns on a bond in Q1, it would be using the GDP growth experienced in Q3, since that would be the latest data available. Similarly, if returns for Q2 2023 are being projected in Q1 2023, it will be using the GDP growth for Q4 2022 as that would be the latest quarterly data available.

2.5 Exploratory Data Analysis

The EDA was conducted in conjunction with the development of the dataset and selection of the independent variables. It relied extensively on the visualization of each variable to understand its distribution and identify outliers. For the dependent variable, the median and average total quarterly returns of the bonds were 1.6% and 1.9% respectively and ranged from -92% to 393% (see Figures 2.2 and 2.3).

There were quarters with relatively greater returns - in the first quarter of 2020, at the onset of the Covid-19 pandemic, there was a sharp downturn across equity and credit markets, during which the median return of the bonds in this data set were -11.9% for the quarter (see Figure 2.5). Markets recovered in the subsequent quarter, following monetary loosening and fiscal stimulus measures, when median returns were 10.8%. In Q1'2020, at the onset of Covid-19, there were 21 bonds, such as those of California Resources Corp and Carrizo Oil & Gas, that declined over 50%, 44 bonds returned lower than -25%, and 186 of the 199 bonds with returns data for that quarter had negative returns. In the subsequent quarter, many of the same bonds would recover in value so there were 5 bonds that returned over 100%, 15 bonds returned over 50%, and 173 of the 184 bonds with pricing data had positive returns. Due to the high volatility in this period, these quarters would be influential in driving the results of the analysis.

Also of note was the performance AMC Entertainment's bonds' returns in Q1 2021, which were outliers, returning in excess of 300%, as the bonds recovered substantially in value. Figures [A.1] illustrates the distribution of returns without AMC. The variance of the dependent variable was 2.96% and the standard deviation was 17.2% for the entire data set. Figure 2.4 illustrates the standard deviation of returns by quarter. The quarters Q1 2020 and Q2 2020 exhibit high standard deviations of 20.6% and 27.1% respectively. Q1 2021 has the highest standard deviation 51.7%, though this is primarily due to the bonds of AMC, without which the standard deviation for that quarter would be 7.9%. The distribution of returns for this sample for each quarter can be viewed in Figure A.5.

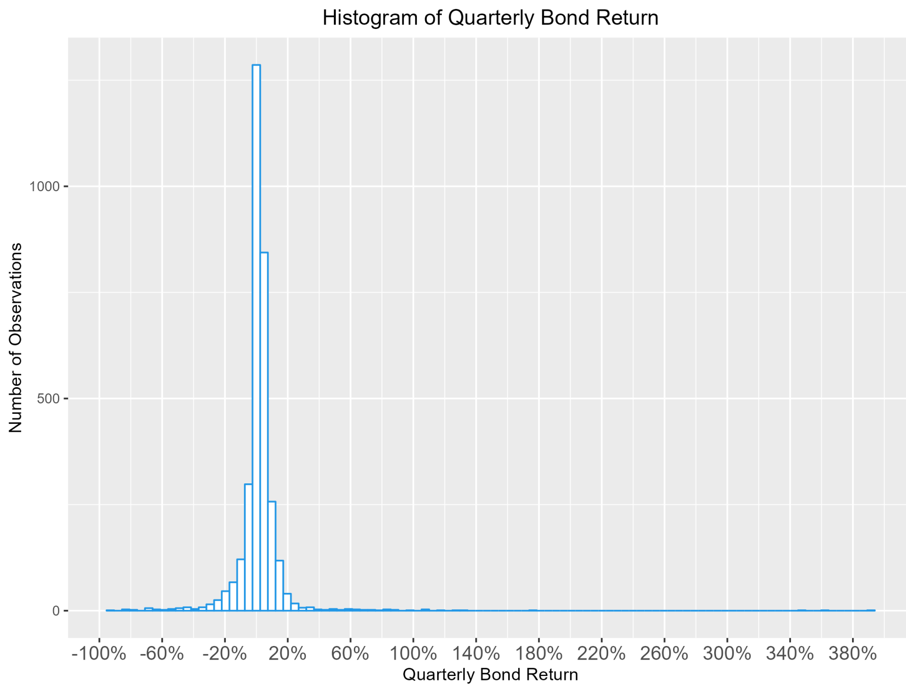

Figure 2.2: Distribution of Quarterly Bond Returns in Sample

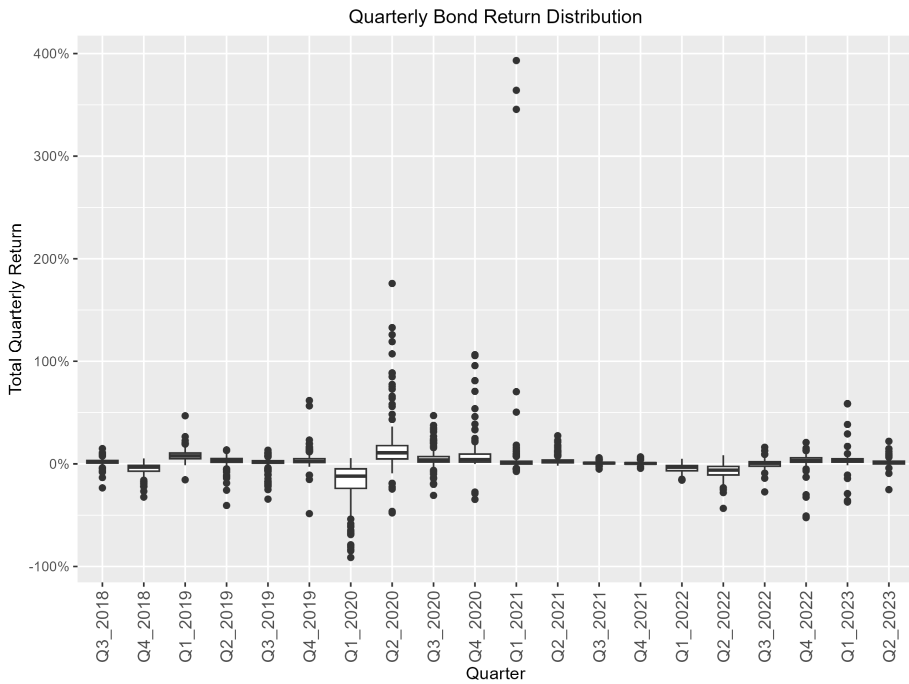

Figure 2.3: Boxplot of Bond Returns by Quarter

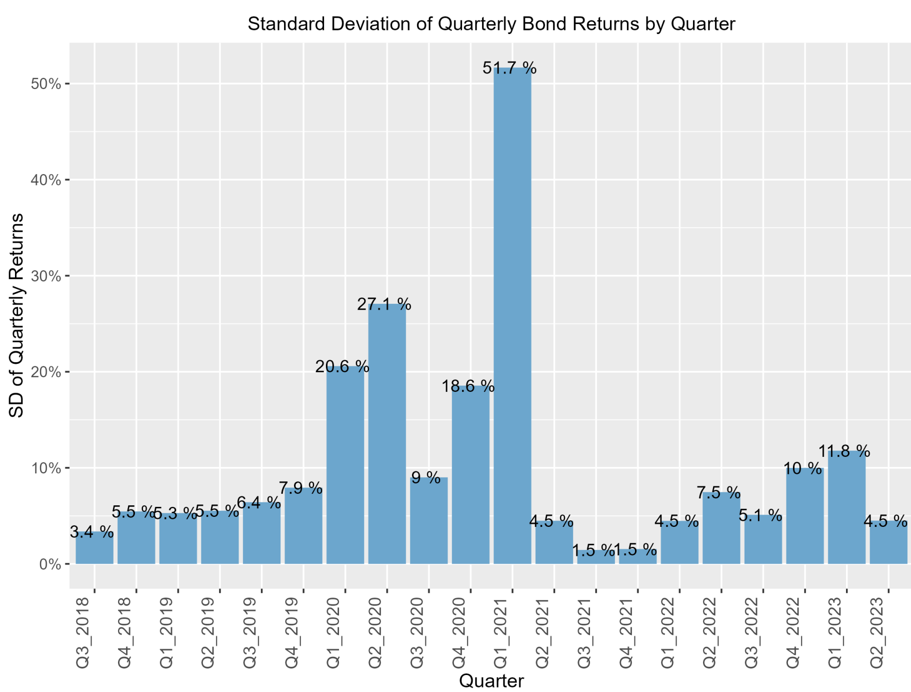

Figure 2.4: Standard Deviation of Bond Returns by Quarter

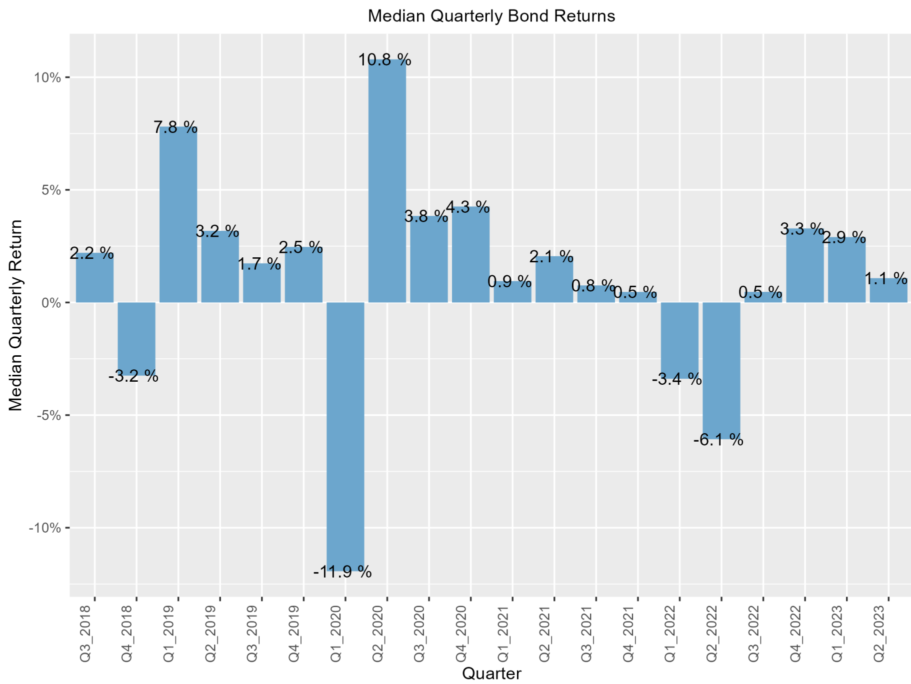

Figure 2.5: Median Bond Returns by Quarter

There were outliers in the net leverage and interest coverage, both as a multiple of LTM and Forward EBITDA. Several issuers experienced extreme volatility in their earnings due to the disruptions caused during the COVID-19 period or due to other factors. For instance, oil & gas companies experienced severe declines to their earnings through 2020 followed by a rebound in late 2021. Figures 2.6 and 2.7 illustrate the distribution of the sign log of net leverage as a multiple LTM EBITDA and multiple of Forward EBITDA, respectively. Due to the declines in EBITDA, these ratios were extremely high and the graphs use the sign log as the values without the transformation were too large to be reasonably illustrated.

Similarly, for Interest Coverage, there were outliers due to small LTM EBITDA or Interest Expense which resulted in extremely large values. A company's EBITDA or UFCF can also be negative or extremely low at a given point due to severe underperformance or one-time costs, which can result in abnormally high Debt to EBITDA or Debt to UFCF levels relative to the rest of the dataset, and therefore affect the results of a prediction model. During Covid-19, there were many companies in this dataset that reported negative EBITDA or UFCF due to a slowdown in their business. To account for the outliers, sign log transformation was done for Net Leverage, Net Debt to UFCF, and Interest Coverage (see Figures 2.6, 2.7, 2.8, 2.9 for distribution of Sign Log transformed variables). The distribution of lagging Net Debt to TEV had most of its values between 0% and 100% (see Figure 2.11).

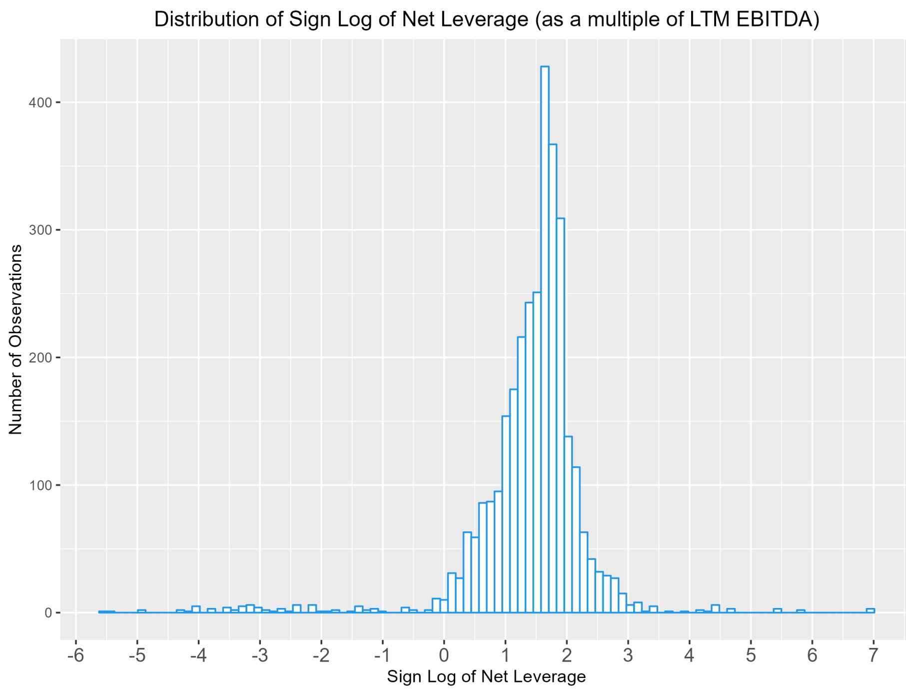

Figure 2.6: Distribution of Sign Log of Net Leverage as Multiple of LTM EBITDA

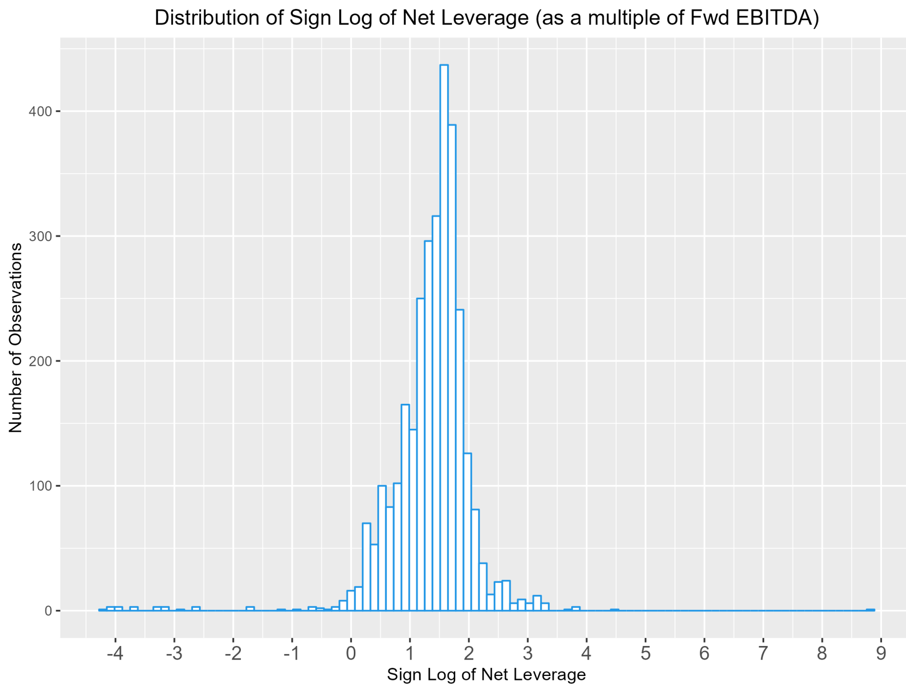

Figure 2.7: Distribution of Sign Log of Net Leverage as Multiple of Forward EBITDA

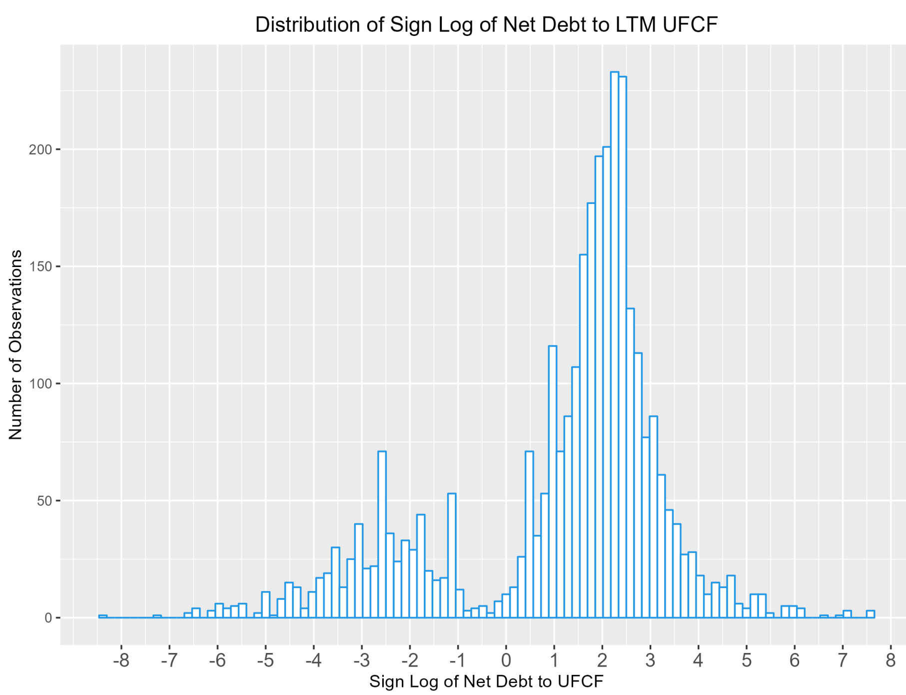

Figure 2.8: Distribution of Sign Log of Net Debt to UFCF

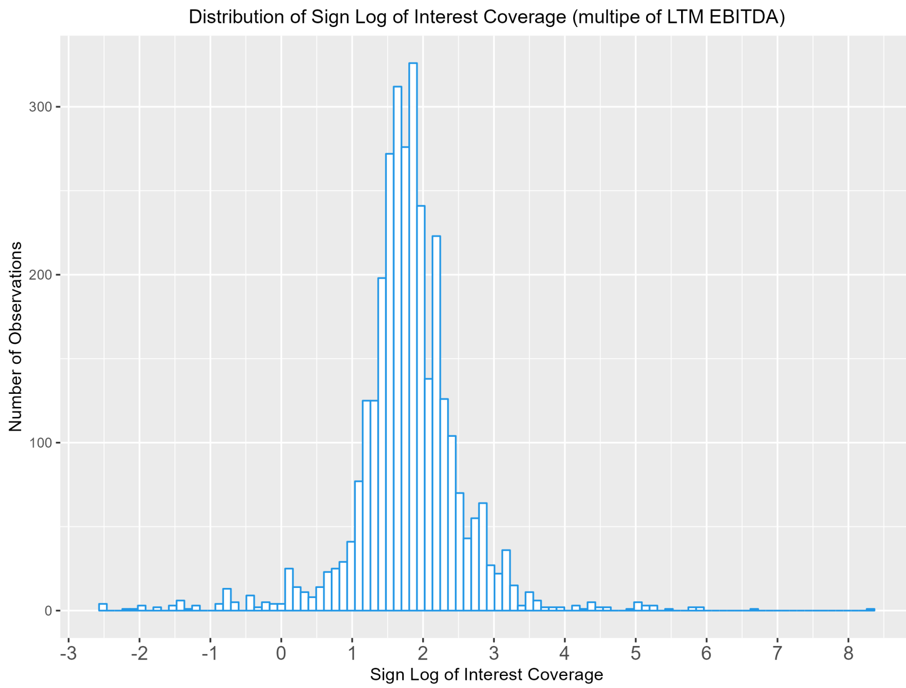

Figure 2.9: Distribution of Sign Log of Interest Coverage as multiple of LTM EBITDA

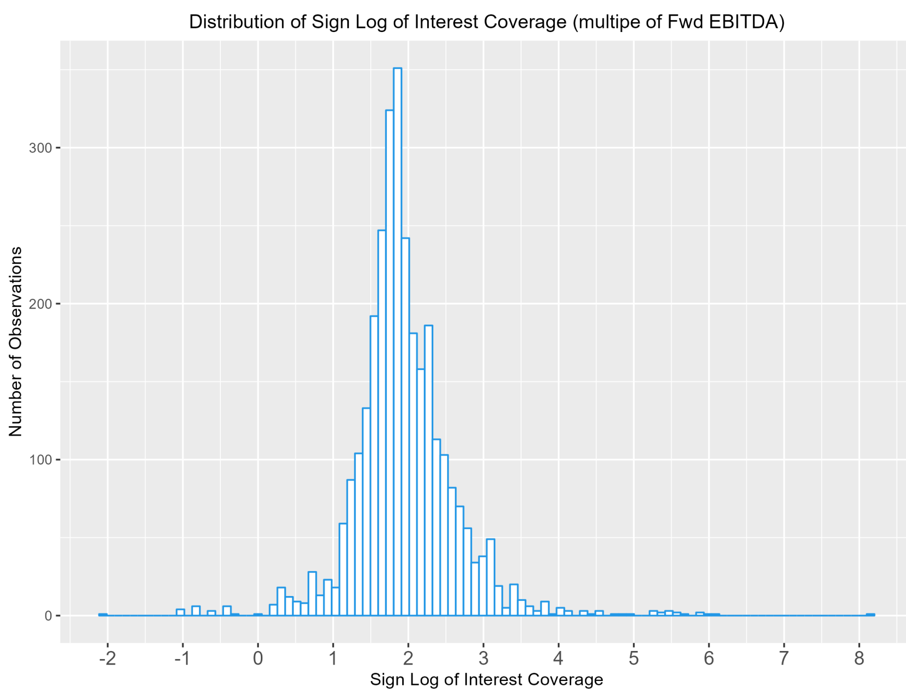

Figure 2.10: Distribution of Sign Log of Interest Coverage as multiple of Forward EBITDA

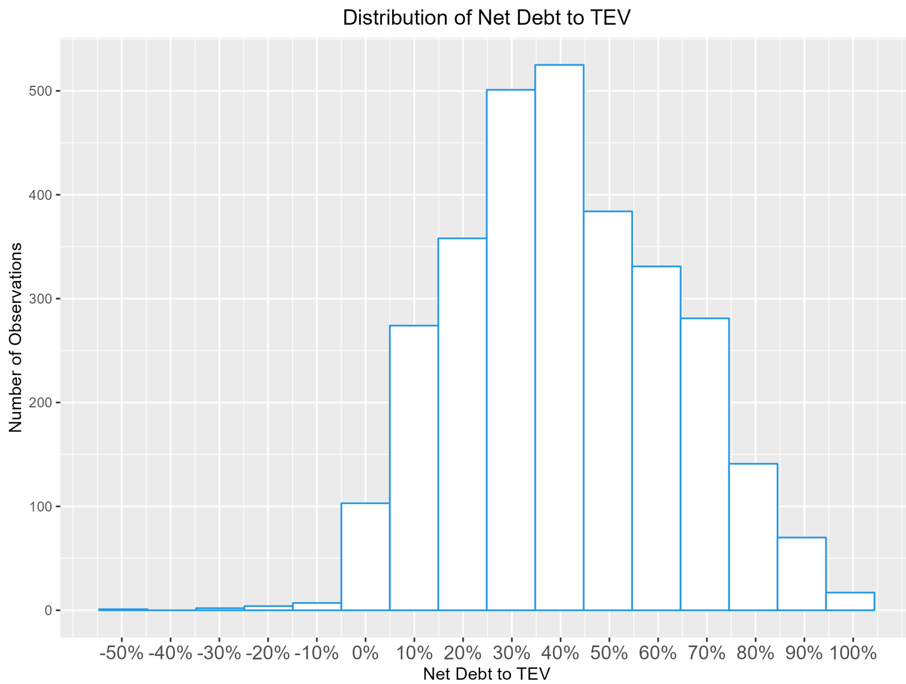

Figure 2.11: Distribution of Net Debt to Total Enterprise Value

The quarterly stock return of the underlying issuers was within a range of -93% and 455%, with the median of 2.9% and average of 5.5% (see Figure 2.12). Due to the presence of certain stocks, like AMC and SM Energy Co., which had large positive moves in certain quarters, the distribution is right skewed. Modified duration declines through time as it is directly proportional with the maturity of the bonds, and as the bond gets closer to maturity, the duration also declines (see Figure 2.13).

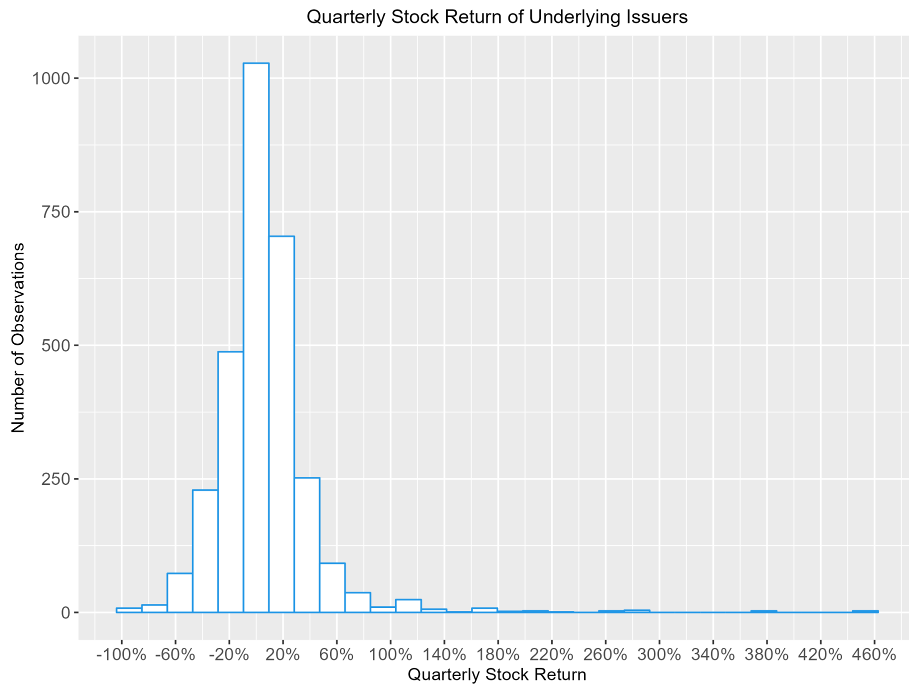

Figure 2.12: Distribution of Stock Return of Underlying Issuers

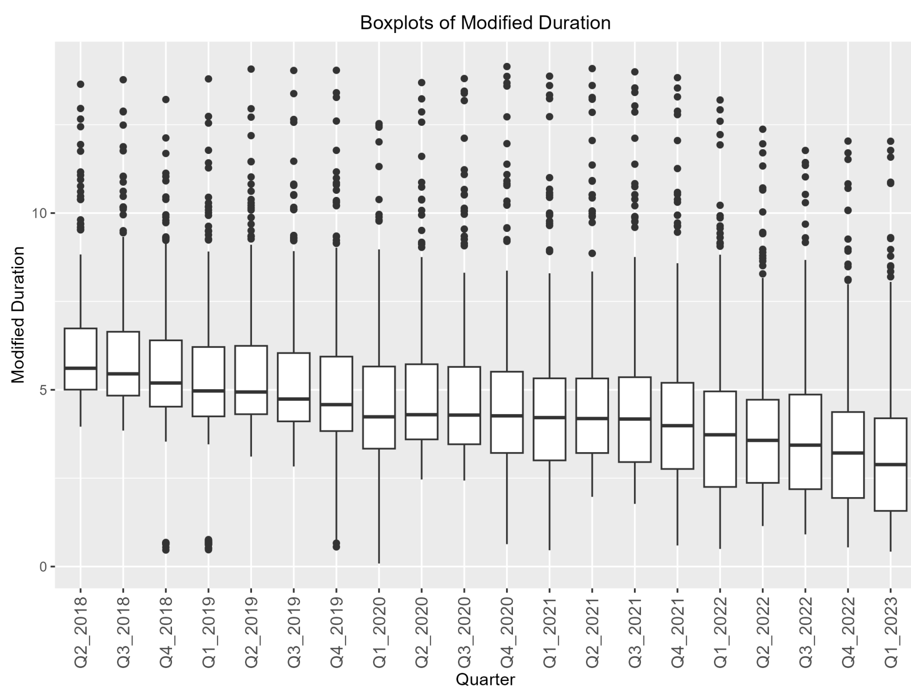

Figure 2.13: Boxplot of Modified Duration of Bonds by Quarter

Figure 2.14 illustrates that the returns of the S&P 500 and Bloomberg Corporate High Yield index are directionally the same as each other for each quarter, and directionally the same as the median returns of the bonds in this sample. The S&P 500 has more variance in its returns likely because stocks are more sensitive to the fluctuations of a company's performance since they are subordinate to the bonds in the company's capital structure, and unlike bonds, have no theoretical upper limit on their price.

The figure also shows how returns of 5-year and 10-year Treasuries are directionally similar to each other though not always the same. Their returns may also be directionally inverse to those of the stock or HY bond indices. For instance, in Q1 2020, as there was a flight to quality in the wake of the Covid pandemic, Treasuries performed well, in contrast to the S&P 500 and HY Bond Index. However, during the first three quarters of 2022, as inflation increased and interest rates increased, all of these assets had negative quarterly returns.

Figure 2.15 shows that Quarter-on-Quarter GDP growth were sharply negative in Q2 2020 due to the downturn from Covid-19 and positive in the subsequent quarter as the economy partially recovered. The research also considered Year-on-Year growth for GDP growth, which exhibited similar trends, as shown in Figure 2.16. Directionally, Real and Nominal GDP growth were the same in almost every quarter except Q1 2022 and Q2 2022 when there was a substantial increase in inflation. Due to the lag in reporting GDP data, the GDP growth used as a predictor in a given quarter would be the growth as of the prior quarter. So the GDP growth available in Q3 2020 to project Q4 2020 would be the growth experienced in Q2 2020.

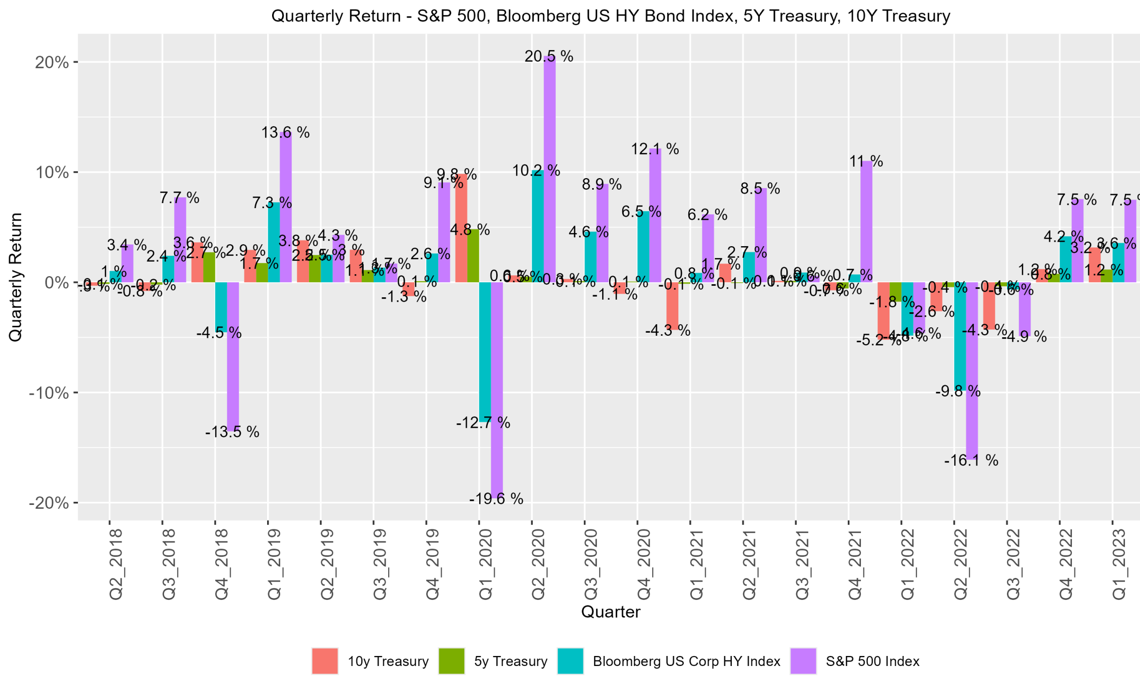

Figure 2.14: Quarterly Returns of 10y Treasury, 5y Treasury, Bloomberg HY Index and S&P 500

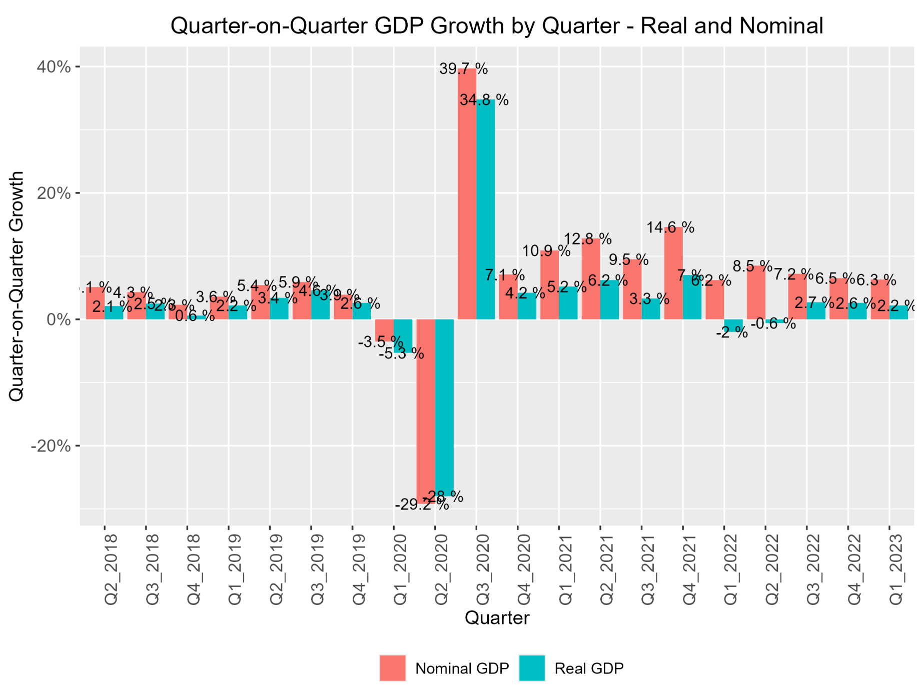

Figure 2.15: Quarterly-on-Quarter Nominal and Real GDP Growth by Quarter

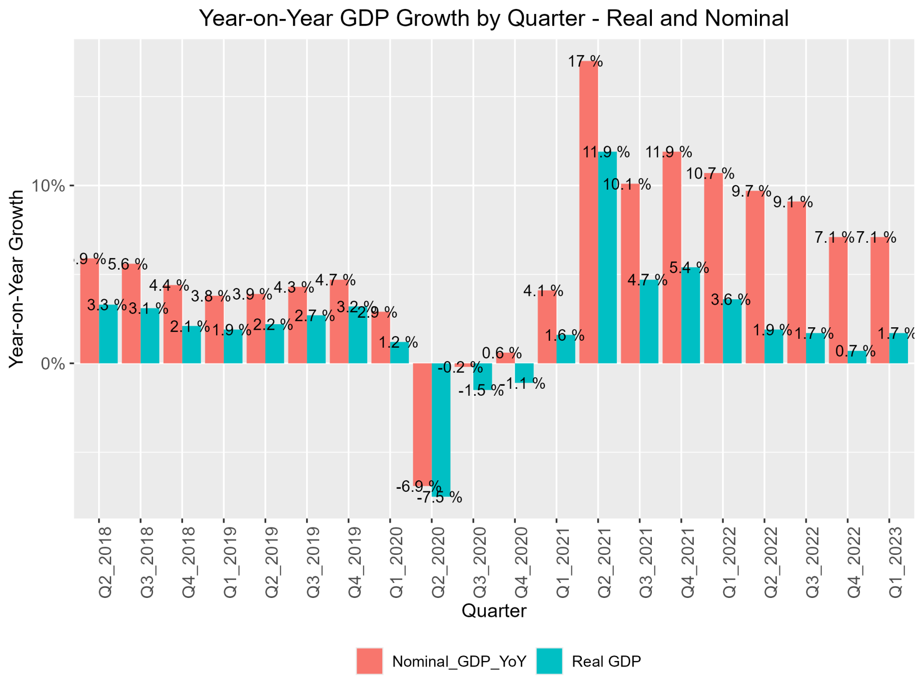

Figure 2.16: Year-on-Year Nominal and Real GDP Growth by Quarter

Figure [2.17] illustrates the correlation between all the quantitative variables in this research. Note that this correlation matrix only looks at correlations for those observations where the dependent variable was available, as those with NA values would not be used in the forecasting models. The dependent variable did not have very high correlations with most of the quantitative variables. The notable correlations were with the lagging 5-year and 10-year government bond returns, at 0.24 and 0.23 respectively, and with lagging Net Debt to TEV at 0.16. Sign Log of Interest Coverage as a multiple of LTM EBITDA was highly correlated with Sign Log Interest Coverage as a multiple of Forward EBITDA, due to which only the former was used in the forecasting models. Sign Log of Net Debt to LTM EBITDA had a correlation of 0.47 with Sign Log of Net Debt to Forward EBITDA.

The lagging quarterly returns for the High Yield Bond Index and the S&P 500 were highly correlated at 0.95. Similarly, there was a high correlation between 5-year and 10-year treasury returns, due to which only one of them would be used in the multiple regression. Real GDP Growth was also highly correlated with Nominal GDP growth, for both, quarter-on-quarter growth and year-on-year growth within a quarter. The modified duration of a bond was not highly correlated with any of the other variables.

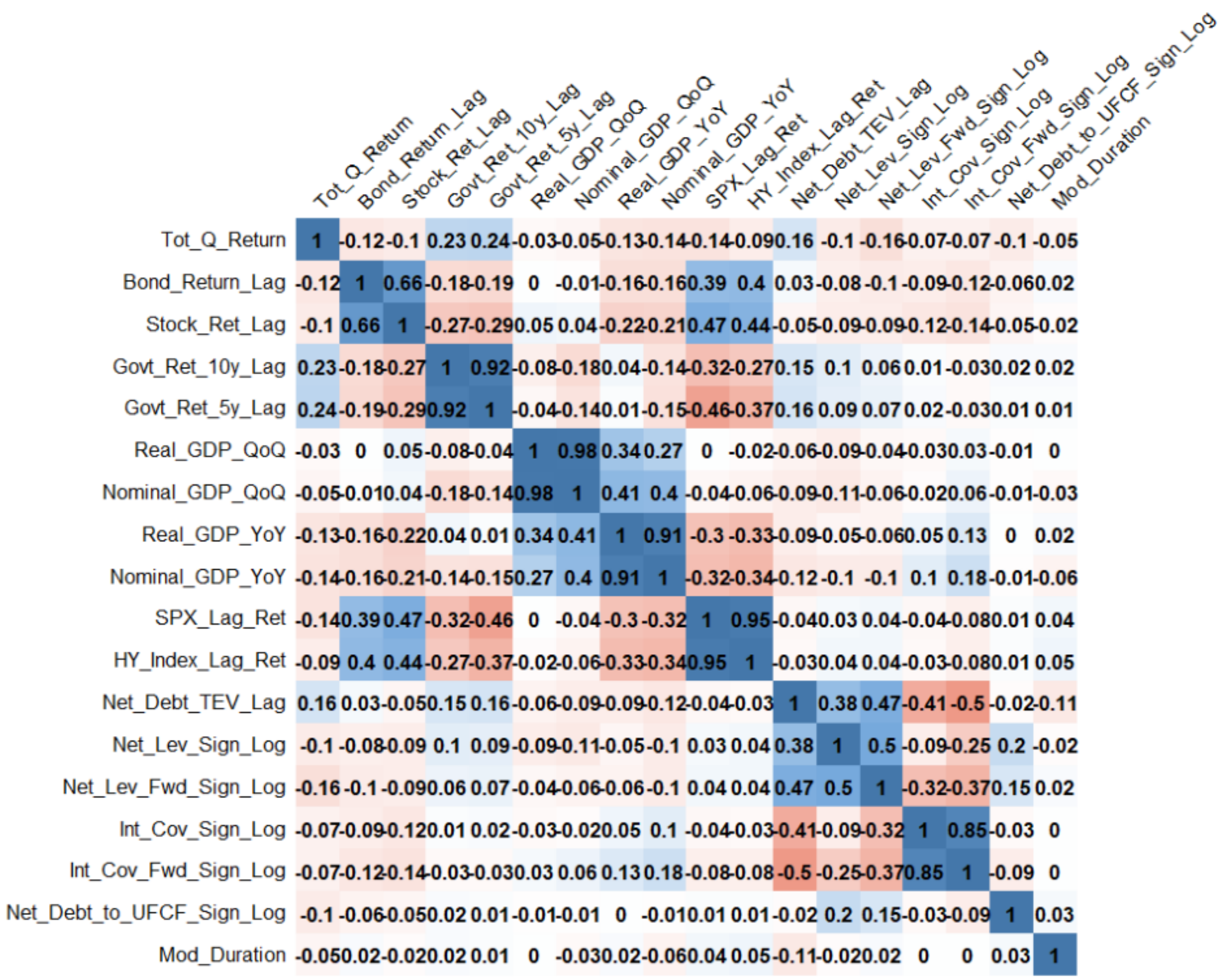

Figure 2.17: Correlation Matrix

CHAPTER 3

Forecasting Models

The research conducted different bivariate and multiple regressions using the variables explored in Chapter 2, and validated the models using windowing and cyclic permutation techniques. It also used Random Forest to determine the most important variables for predicting quarterly bond returns.

3.1 Linear Regression

Linear regressions were performed with each of the quantitative variables that lagged by one quarter. These are simple linear regressions that were repeated independent of each other. Table [3.1] below presents the output of these 16 different linear regressions, with the intercept, coefficient, p-value and R-Squared. Except for quarterly Real GDP growth, quarterly Nominal GDP growth and modified duration, all of the variables' coefficients were significant at the 1% level. All of them had an inverse relationship except for the lagging quarterly returns of 10-year and 5-year Treasury bonds (labeled as 'Govt_Ret_10y_Lag' and 'Govt_Ret_5y_Lag') and Net Debt to TEV, which have a positive relationship. The R-squared is indicative of how much of the variation in the quarterly returns is explained by the linear model. The R-squared values were very low, with the highest one of 5.9% for lagging 5-year Treasury bond returns.

The residuals in Figure [3.1] are right skewed with heavy tails and are very similar to the distribution of the dependent variable, partly because these models explain a very small percentage of the variation in the dependent variable. The returns of AMC bonds were major outliers which skewed the distribution of the residuals. The coefficients represent the average change in the quarterly return of the bonds for a unit change in the independent variable. For instance, if the returns on the 5-year Treasury increase by a given percentage, the returns on the High Yield bonds would increase, on average, by 2.873 times that percentage. These relationships can also be viewed graphically in Figure A.3.

Variable	**Interecept**	**Coefficient**	**P-Value**	**R-Squared**
Bond_Return_Lag	0.0216	-0.1258	0.0000	0.015
Stock_Ret_Lag	0.0256	-0.0429	0.0000	0.010
Net_Debt_TEV_Lag	-0.0247	0.1162	0.0000	0.025
Net_Lev_Sign_Log	0.0470	-0.0196	0.0000	0.011
Net_Lev_Fwd_Sign_Log	0.0751	-0.0408	0.0000	0.026
Net_Debt_to_UFCF_Sign_Log	0.0277	-0.0074	0.0000	0.009
Int_Cov_Sign_Log	0.0484	-0.0163	0.0000	0.005
Govt_Ret_10y_Lag	0.0084	1.2089	0.0000	0.055
Govt_Ret_5y_Lag	-0.0034	2.8733	0.0000	0.059
Real_GDP_QoQ	0.0202	-0.0525	0.0785	0.001
Nominal_GDP_QoQ	0.0229	-0.0704	0.0082	0.002
Real_GDP_YoY	0.0339	-0.7052	0.0000	0.018
Nominal_GDP_YoY	0.0456	-0.5223	0.0000	0.019
SPX_Lag_Ret	0.0271	-0.2372	0.0000	0.020
HY_Index_Lag_Ret	0.0224	-0.3015	0.0000	0.009
Mod_Duration	0.0347	-0.0031	0.0099	0.002

Table 3.1: Summary Statistics of Bivariate Regressions

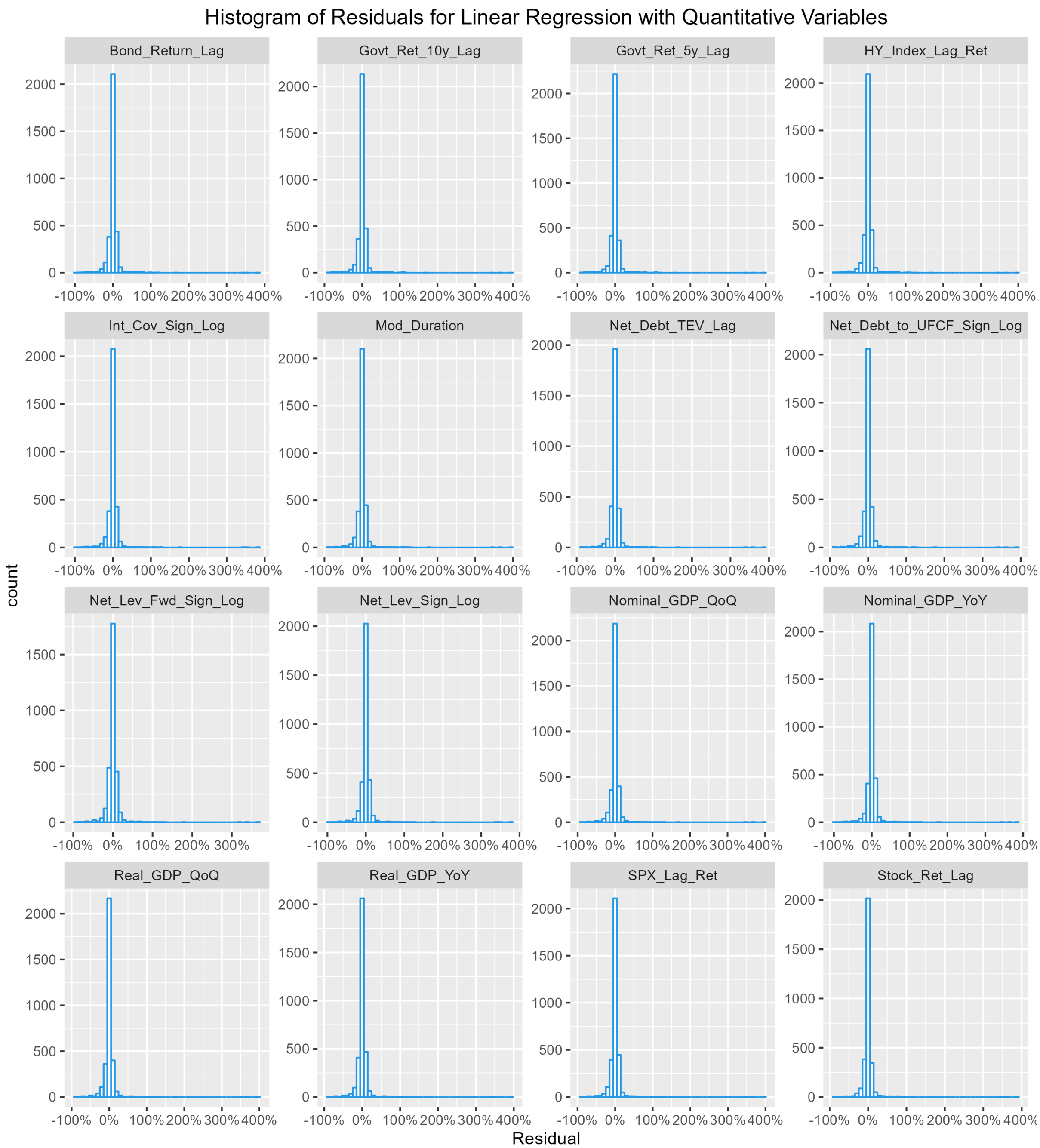

Figure 3.1: Distributions of Residuals of Bivariate Regressions

The model results (in Table [3.2]) and the tail of the residuals look different when AMC's bonds are excluded (see Figure [A.2]). In that regression, the Net Leverage variables, Net Debt to UFCF, Interest Coverage, and Modified Duration variables are not significant. The cross validation described in Chapter 3.2 partly helps mitigate the impact of these outliers.

Variable	Interecept	Coefficient	P-Value	R-Squared
Bond_Return_Lag	0.0208	-0.2208	0.0000	0.044
Stock_Ret_Lag	0.0231	-0.0436	0.0000	0.017
Net_Debt_TEV_Lag	-0.0159	0.0897	0.0000	0.032
Net_Lev_Sign_Log	0.0139	0.0020	0.4817	0.000
Net_Lev_Fwd_Sign_Log	0.0148	0.0016	0.6889	0.000
Net_Debt_to_UFCF_Sign_Log	0.0194	-0.0023	0.0283	0.002
Int_Cov_Sign_Log	0.0213	-0.0025	0.4188	0.000
Govt_Ret_10y_Lag	0.0055	1.2709	0.0000	0.109
Govt_Ret_5y_Lag	-0.0068	3.0030	0.0000	0.117
Real_GDP_QoQ	0.0208	-0.1725	0.0000	0.019
Nominal_GDP_QoQ	0.0263	-0.1715	0.0000	0.023
Real_GDP_YoY	0.0296	-0.6081	0.0000	0.024
Nominal_GDP_YoY	0.0399	-0.4560	0.0000	0.026
SPX_Lag_Ret	0.0260	-0.2731	0.0000	0.047
HY_Index_Lag_Ret	0.0210	-0.3783	0.0000	0.025
Mod_Duration	0.0253	-0.0017	0.0591	0.001

Table 3.2: Summary Statistics of Bivariate Regressions excluding AMC bonds

The relationship of returns with the GICS Industry classification was also tested but none of the coefficients for the different industries were significant so those results have been excluded.

3.1.1 Relationship with Lagging Bond Returns

The research further explored the relationship between bond returns and lagging returns. If Y_i represents the return of bond i, and X_i the return of the bond i in the prior time period, the regression equation is

$$Y_i = \beta_0 + \beta_1 \ X_i + \epsilon$$

In this model is X is just the back-shifted version of Y_i (returns from the previous time period for bond). β_0 is the intercept, representing the mean return when the lagging return is 0. β_1 is the coefficient for the lagging return and ϵ is the error term. The null hypothesis is

$$H_0\text{: } \beta_1 = 0$$

Which states that the lagged bond return has no effect on the present quarter's return for a given bond. The coefficient for the regression in Table [3.1] above suggests that there is a negative relationship between bond returns lagging by 1 quarter and the present quarter's bond returns. The ACF test for the returns of each bond showed that the ACF values at a lag of 1 had a mean of -0.22, and the histogram of ACFs is right skewed with a substantial number of bonds that have negative values. This suggests that on average, for this sample, bond returns are negatively autocorrelated with returns in the prior quarter. However, for a lag of 2, the mean of the ACFs is close to 0 and the distribution resembles a normal distribution with a mean of 0, suggesting little or no autocorrelation with the return's relationship with returns from 2 quarters ago.

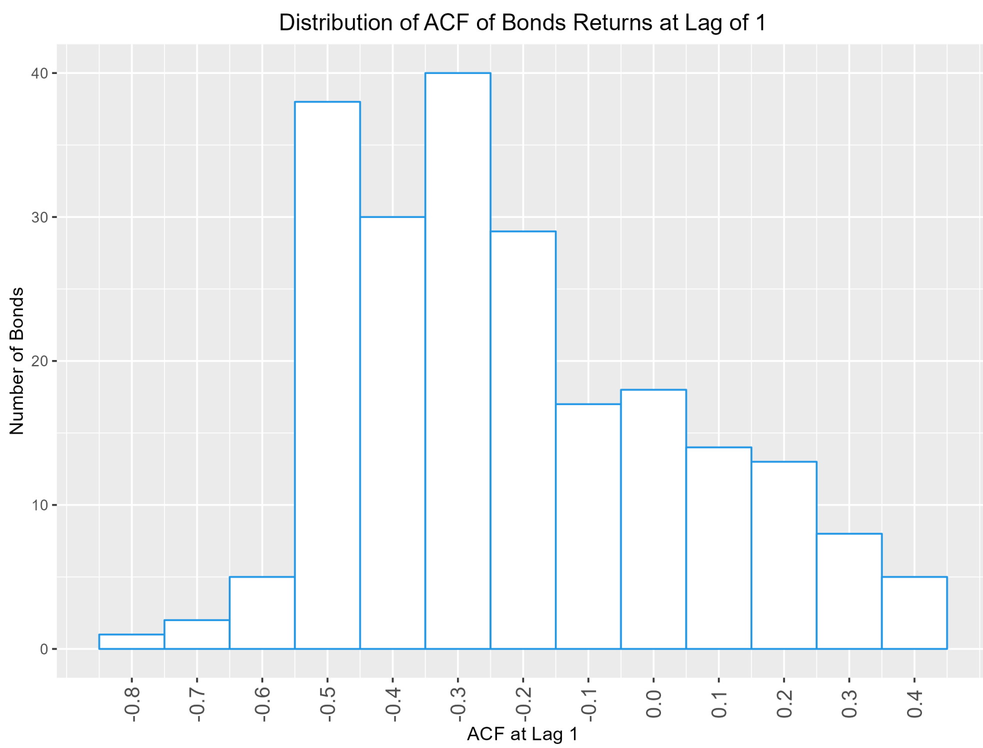

Figure 3.2: Distribution of ACF of Bond Returns at Lag of 1

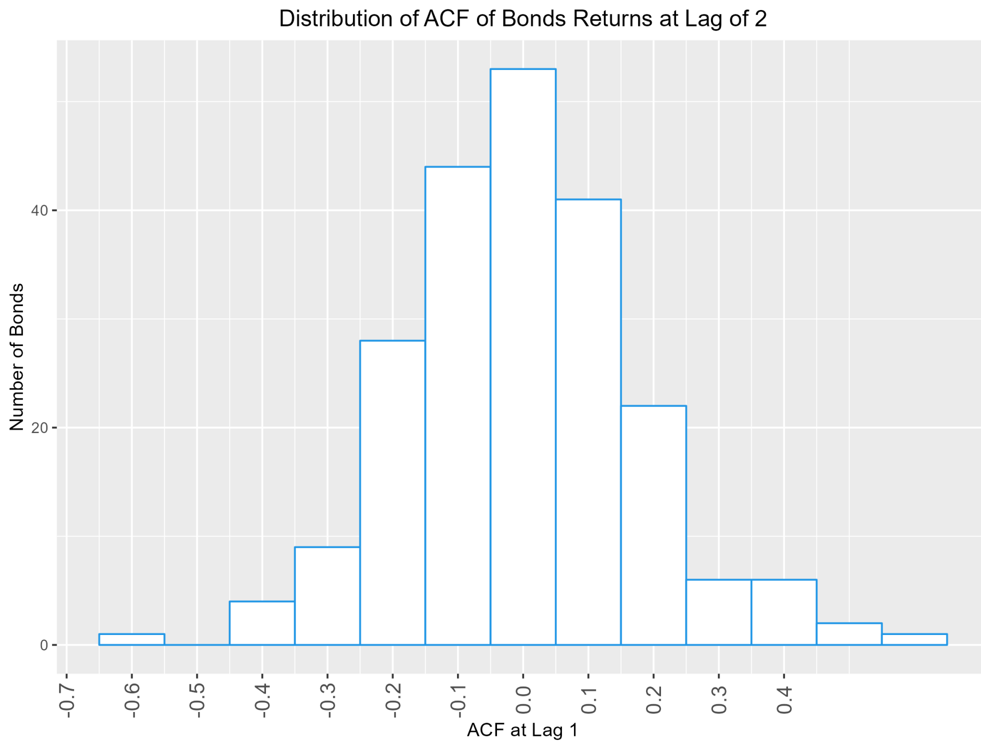

Figure 3.3: Distribution of ACF of Bond Returns at Lag of 2

During the time period for this analysis, three different monetary regimes were identified for the US, that were largely related to the economic effects from Covid but also other factors. During Covid, from Q1 2020 to Q4 2021 the Federal Reserved lowered the Federal Funds Rate and began an expansive Quantitative Easing Program [MW24]. In Q1 2022, as inflationary pressures mounted, the Federal Reserve began tightening interest rates, which continued till Q2 2023 [Fed24]. To observe the impact of these regimes, a new categorical variable called 'Regime' was defined, in which Q3 2018 to Q4 2019 was categorized as 'Pre-Covid', Q1 2020 to Q4 2021 as 'Covid', and Q1 2022 to Q2 2023 as 'Post-Covid'.

The relationship between bond returns and lagging returns varies across the three regimes. The relationship in the pre-Covid period appears to be negative though this partly because the final quarter of the pre-Covid period was Q4 2019, and bond returns were negative in Q1 2020 during the Covid downturn, so projections made in Q4 2019 for Q1 2020 are a major reason why the relationship between lagging return and current return are negative in that regime. The relationships in the 'Covid' and 'Post-Covid' regime are also negative but these are much weaker relationships. Similar differences across regimes for other variables can be observed in Figure A.4.

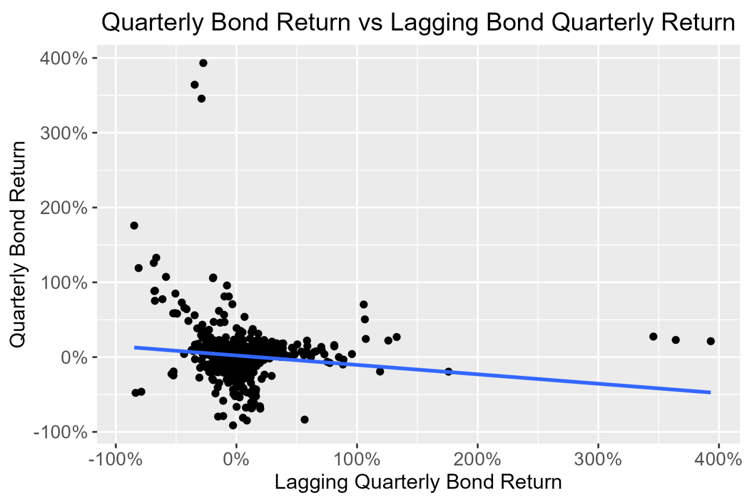

Figure 3.4: Scatterplot and Regression Line of Bond Returns vs Lagging Bond Returns

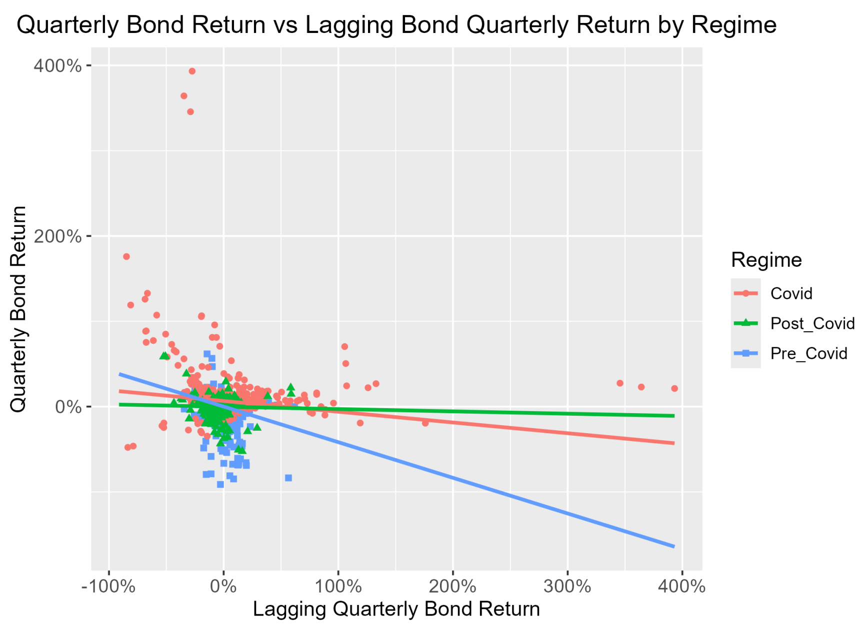

Figure 3.5: Scatterplot and Regression Line of Bond Returns vs Lagging Bond Returns by Regime

3.2 Time Series Cross Validation

The cross validation for these linear models uses a windowing technique, where the training model is updated by training it with the first k quarters of data and testing it on the k+1th quarter, and repeating this until k = N-1 or 19 quarters. The process starts with training a model using data from the first seven or k=7 quarters, and testing that model on data from the k+1th or eighth quarter to predict the returns for that quarter. It starts with seven quarters to give the training model sufficient data for a reasonably stable forecast. The predicted quarterly returns for the eighth quarter for each bond are stored in a vector. The training set moves forward by one quarter each time so in the next iteration, the model is training using the data from the first eight or k=8 quarters to generate new model coefficients. That model is used to predict the returns for the ninth quarter and the predicted values vector is updated using those figures. This continues till k=19 quarters, to predict the returns of the final quarter, Q2 2023.

This process returns a vector of predicted quarterly returns for all the quarters from k=8 to k=20 or from Q2 2020 to Q2 2023, computed using models that were trained with data from all the quarters immediately prior to the one which is being tested. The Figure [3.6], from Hyndman and Athansopoulos' book *Forecasting: Principles and Practice* [HA21], illustrates how the training tests (blue dots) and test sets (orange dots) are updated at each step. This process is known as 'evaluation on a rolling forecasting origin' [HA21] since the origin of the forecast changes by one step each time. This method resembles what would happen in practice for a bond investor, who would use new information from each period to make predictions for the next time period. The results of this validation technique can be viewed in Table [3.3]

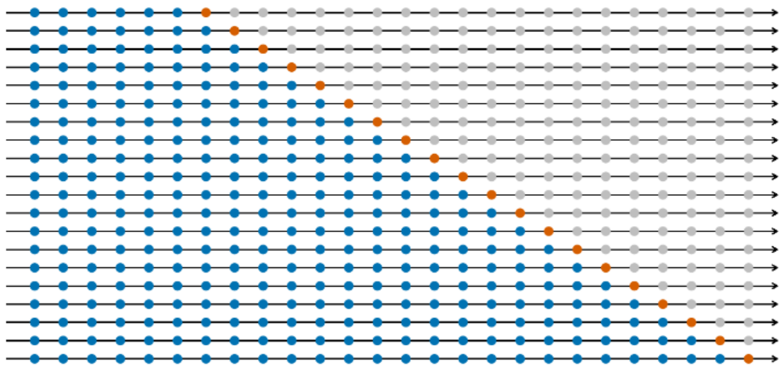

Figure 3.6: Illustration of Evaluation on Rolling Forecasting Origin [HA21]

The predicted values, $\hat{Y}_{it}$, and actual values, Y_{it}, for those quarters are used to obtain the Mean Squared Error (MSE), Root Mean Square Error (RMSE), and the correlation between the predicted returns and actual returns (Test Correlation). The residual is the difference between the actual return in a given quarter t for a given bond i and the return predicted by the model, for those observations where actual returns and predicted returns were available:

$$Y_{it} \text{ - } \hat{Y}_{it}$$

The MSE is the average of the sum of the errors squared for all 222 bonds across the 13 quarters. Since not all quarters' observations will have complete values due to missing prices or independent variables, n would not be the product of 222 and 13 quarters but just the number of complete observations:

$$\text{MSE} = \frac{\sum_{i,t}(Y_{it} - \hat{Y}_{it})^2}{n}$$

The RMSE is the square root of the MSE. The MSE indicates how much the model's projected return deviates from the actual return, on average. For instance, variance of the prediction error for lagging bond return is, on average, 4.92%. The variables with lowest MSE are Net Debt to TEV and 5-year Treasury bond returns. Test Correlation is the correlation between the predicted values and the actual values, where the predictions for each quarter k are made with a model trained with k-1 quarters. It is a measure of the quality of the model in practice and all the predictors appear to be weakly correlated.

Variable	**Test Correlation**	**MSE**	**RMSE**
Bond_Return_Lag	0.0822	0.0492	0.2218
Stock_Ret_Lag	0.0369	0.0432	0.2078
Net_Debt_TEV_Lag	0.0489	0.0405	0.2012
Net_Lev_Sign_Log	-0.0636	0.0444	0.2106
Net_Lev_Fwd_Sign_Log	-0.0632	0.0471	0.2171
Net_Debt_to_UFCF_Sign_Log	-0.0961	0.0436	0.2087
Int_Cov_Sign_Log	-0.1204	0.0440	0.2096
Govt_Ret_10y_Lag	0.1851	0.0435	0.2085
Govt_Ret_5y_Lag	0.1985	0.0417	0.2041
Real_GDP_QoQ	0.0383	0.0563	0.2373
Nominal_GDP_QoQ	0.0538	0.0567	0.2381
Real_GDP_YoY	0.0044	0.0432	0.2077
Nominal_GDP_YoY	0.0898	0.0427	0.2067
SPX_Lag_Ret	0.0395	0.0445	0.2109
HY_Index_Lag_Ret	0.0271	0.0440	0.2098
Mod_Duration	-0.1752	0.0432	0.2079

Table 3.3: Cross Validation Results of Bivariate Regressions

3.3 Cyclic Permutation Test

In the cyclic permutation tests a dataset is simulated by switching the quarterly returns data between different quarters. In the first cycle, the quarterly returns of Q4 2018 are assigned to Q3 2018, the quarterly returns of Q1 2019 are assigned to Q4 2018, so on and so forth, until the final quarter, where the quarterly returns of Q2 2023 are assigned to Q1 2023, and the quarterly returns of Q3 2018 are assigned to Q2 2023. Note that only the dependent variable for a quarter is changed and the independent variables are left unchanged. As a result, this simulated dataset has a dependent variable with quarterly returns that were not the actual ones for that quarter but from a different quarter. With this simulated dataset, the cross-validation process explained above is repeated, in which a model is trained on the first k quarters of the simulated data to test on the k+1th quarter to obtain a set of predicted values for that quarter, and then updating that training model with k+1 quarters to test it on data in the subsequent k+2th quarter. This will give a set of predicted values generated on this simulated dataset, and the correlation coefficient between the values predicted by the model and the 'actual' quarterly returns assigned for the quarter are stored in a vector.

In the next cycle, another dataset is simulated where the returns of Q1 2019 are assigned to Q3 2018, Q2 2019 returns are assigned to Q4 2018, so on and so forth, until Q1 2023, where the Q3 2018 returns assigned, and for Q2 2023, which is assigned the Q4 2018 results. The cross-validation process is repeated to determine a correlation between the predicted and actual returns for this cycle. Since there are 20 quarters of data, there are 19 such cycles, to ensure that each quarter is assigned the returns from every other quarter at least once. A distribution of correlations for these cycles is obtained and compared to test correlations that were obtained in the cross-validation for the actual dataset. When testing the models for lagging bond returns, there are only 18 cycles instead of 19, to avoid an iteration where the simulated dataset has a dependent variable that is identical to the independent variable.

The correlations obtained in these regressions from the simulated data can be compared to the test correlation obtained above using the actual data in Chapter 3.2. It helps determine if the test correlation observed with the actual data was random – if the test correlation was smaller or larger than all the simulated correlations then it was likely not a random result. Figure 3.7 below shows the histogram of the correlations from the simulated datasets, with the red vertical line indicating the test correlation from the actual dataset. For variables that are significant, such as 5-year or 10-year Treasury returns, lagging bond return, Year-on-Year Nominal and Real GDP Growth, lagging stock returns, the test correlations of the actual dataset is higher than all or almost all of the correlations from the simulated datasets, suggesting that the relationships revealed by the cross validation may not be owed to chance. For other variables, such as Modified Duration, the financial leverage or interest coverage variables, several cycles have lower correlations than the correlation from the actual dataset, suggesting that the correlation observed from the actual data set may have been a random occurrence.

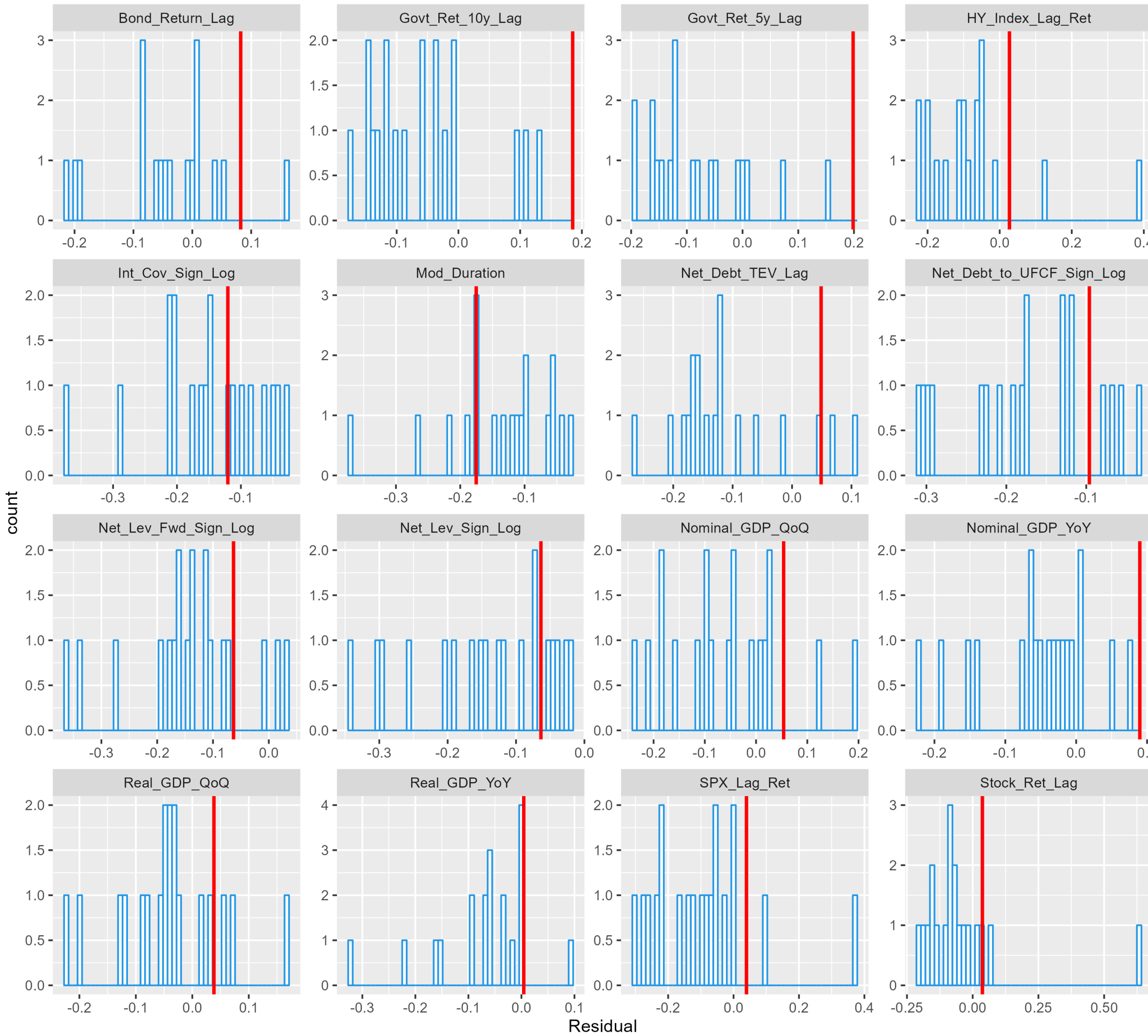

Figure 3.7: Test Correlations for Bivariate Regression from Simulated Datasets, with Correlation from Actual dataset in Red Line

3.4 Multiple Regression

Multiple regression was initially performed with all the variables in the dataset. However, when all of the variables were included, the VIF was very high for certain variables, such as the Treasury returns and GDP growth figures (see Table A.3). To avoid multicollinearity, the highly correlated variables were excluded from the multiple regression. Therefore, the 10-year Treasury bond return, which was highly correlated with the 5-year Treasury bond return, was omitted. The regression excluded the Nominal GDP growth variables, which were highly correlated with Real GDP growth variables, and excluded the S&P 500 Index returns, which were highly correlated with the High Yield Bond Index returns. The excluded variables contained similar information to ones included in the model. The final model also excluded the GICS Industry variable as none of the ten industry levels had a significant relationship at the 1% level, so it was omitted to avoid using more degrees of freedom.

After excluding GICS Industry and the four highly correlated variables, none of the variables had a VIF of over 2.1 (see Table [3.5] for the list of variables and VIF of each). The model had an R-squared of 0.22, and an F-statistic of 66.2 with a p-value well below 0.0001, suggesting that the joint null hypothesis that all of the coefficients are 0 is rejected. Table [3.4] shows that the variables with statistically significant coefficients are Lagging Bond Return, Net Debt to TEV, all three of the sign log of Net Leverage variables, Interest Coverage, lagging 5-year Treasury returns, and lagging YoY growth in Real GDP.

Of the variables that have significant coefficients, the Net Debt to TEV and 5y-Treasury Bond returns have a positive relationship while the rest have a negative relationship. The residuals have long tails due to the outliers but are distributed around 0 (Figure [3.10]). In the quantile-quantile plot (Figure [3.9]) they deviate significantly from the plotted line, suggesting they are not normally distributed. In Residuals vs Fitted values plot (Figure [3.8]), the residuals appear to be larger for larger fitted values, suggesting heteroskedasticity. The outliers are the bonds of AMC, which influence the residual plots and the regression results.

Multiple regression was also done with the 5-year Treasury returns swapped for the 10-year Treasury returns, and the results were similar (Table [A.2]). The independent variables had the same relationships directionally, and the same variables were significant in that model, except that the 10-year Treasury's relationship was not as strong. The R-Squared of that model was marginally lower at 0.21 and it had an F-Statistic that was statistically significant.

Table [A.4] shows the results of the multiple regression with all the variables included, including GICS Industry and the highly correlated variables. This model had an R-squared of 0.24, marginally higher, and directionally, the coefficients appear to be the same but the standard errors are larger.

Table [A.1] shows the results without the AMC bonds that were outliers - directionally, the relationships were the same, except for lagging stock returns and Real GDP QoQ, which were also statistically significant relationships in that regression, while Sign Log of Net Leverage as a multiple of Trailing EBITDA was not significant. That model had a slightly higher R-squared of 0.24 and an F-Statistic that was statistically significant.

Variable	**Estimate**	**Std Error**	**T-Value**	**P-Value**
(Intercept)	0.1286	0.0139	9.2574	0.0000
Bond_Return_Lag	-0.1765	0.0226	-7.8205	0.0000
Stock_Ret_Lag	-0.0109	0.0101	-1.0718	0.2839
Net_Debt_TEV_Lag	0.1818	0.0156	11.6416	0.0000
Net_Lev_Sign_Log	-0.0183	0.0037	-4.9612	0.0000
Net_Lev_Fwd_Sign_Log	-0.0712	0.0050	-14.3552	0.0000
Net_Debt_to_UFCF_Sign_Log	-0.0056	0.0013	-4.3458	0.0000
Int_Cov_Sign_Log	-0.0257	0.0040	-6.4724	0.0000
Govt_Ret_5y_Lag	2.4961	0.2099	11.8909	0.0000
Real_GDP_QoQ	0.0578	0.0285	2.0256	0.0429
Real_GDP_YoY	-0.8418	0.0955	-8.8110	0.0000
HY_Index_Lag_Ret	0.1674	0.0636	2.6314	0.0085
Mod_Duration	-0.0009	0.0011	-0.8184	0.4132

Table 3.4: Summary Statistics of Multiple Regression excluding highly correlated variables and GICS Industry

Variable	VIF
Bond_Return_Lag	1.868888
Stock_Ret_Lag	2.060620
Net_Debt_TEV_Lag	1.628740
Net_Lev_Sign_Log	1.462522
Net_Lev_Fwd_Sign_Log	1.624076
Net_Debt_to_UFCF_Sign_Log	1.071428
Int_Cov_Sign_Log	1.292903
Govt_Ret_5y_Lag	1.256619
Real_GDP_QoQ	1.172301
Real_GDP_YoY	1.341993
HY_Index_Lag_Ret	1.561665
Mod_Duration	1.031894

Table 3.5: VIF of variables in Multiple Regression excluding highly correlated variables and GICS Industry

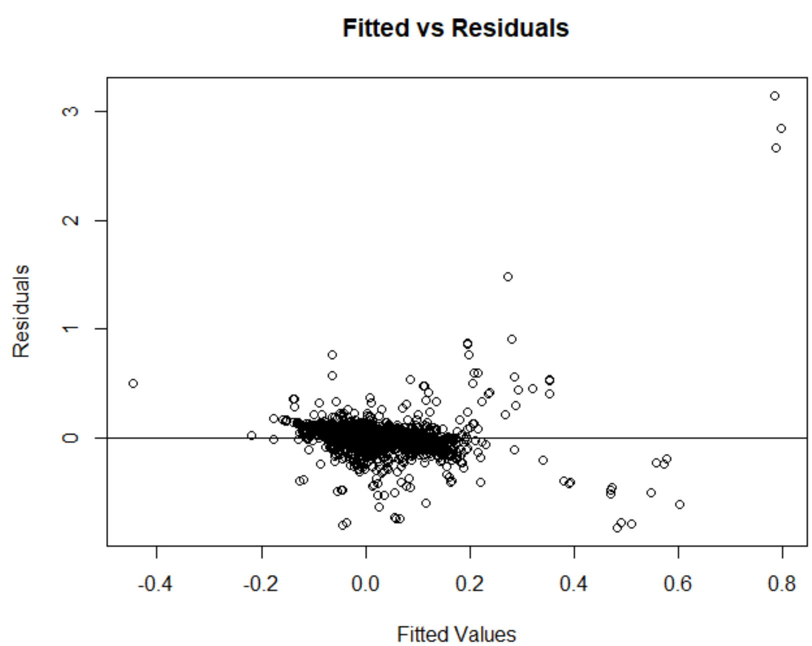

Figure 3.8: Fitted vs Residuals

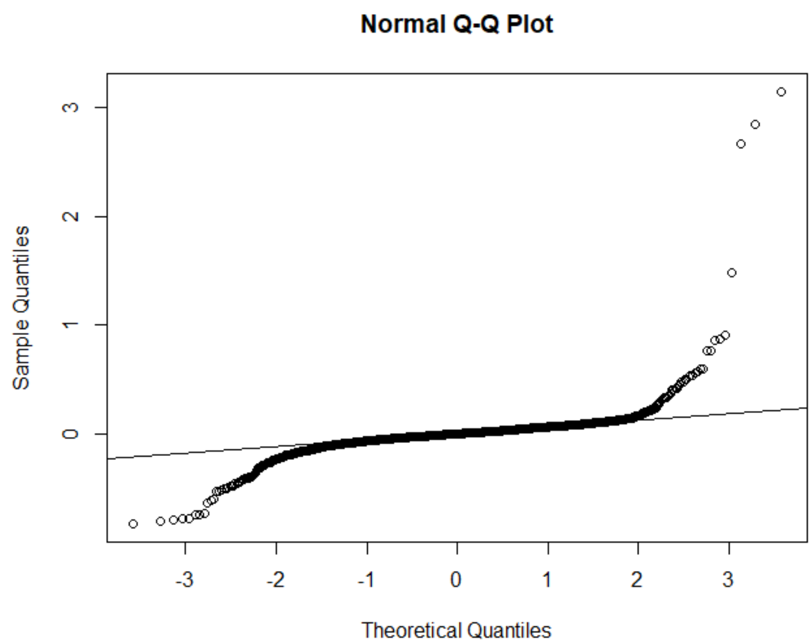

Figure 3.9: Quantile-Quantile Plot

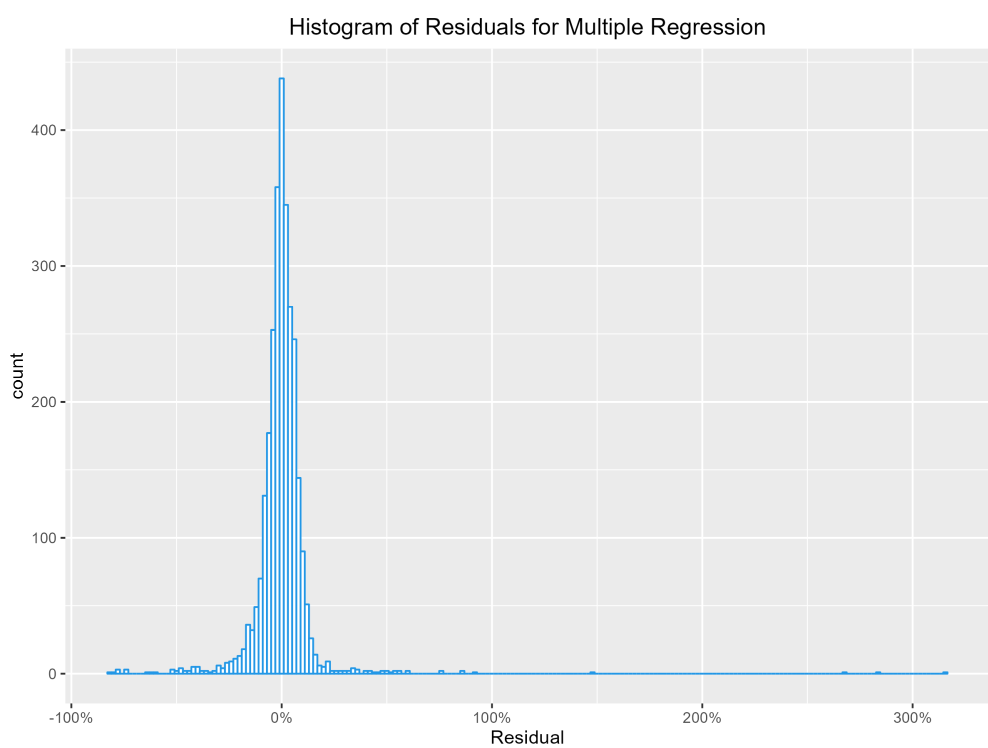

Figure 3.10: Distribution of Residuals from Multiple Regression

The cross validation for the multiple regression, using the same variables as the one used in Table [3.4], also used the windowing technique discussed in Chapter 3.2, where the training model is incrementally updated by an additional quarter of data, and tested on the subsequent quarter. The MSE is 10.12%, the RMSE is 31.81% and the test correlation is 0.05736. Cyclic permutation is performed across the simulated datasets and showed that two of the eighteen cycles (or 11% of total) have test correlations that are greater than the correlation from the actual result, suggesting the result from the actual dataset may have been a random occurrence, as illustrated in Figure [3.11].

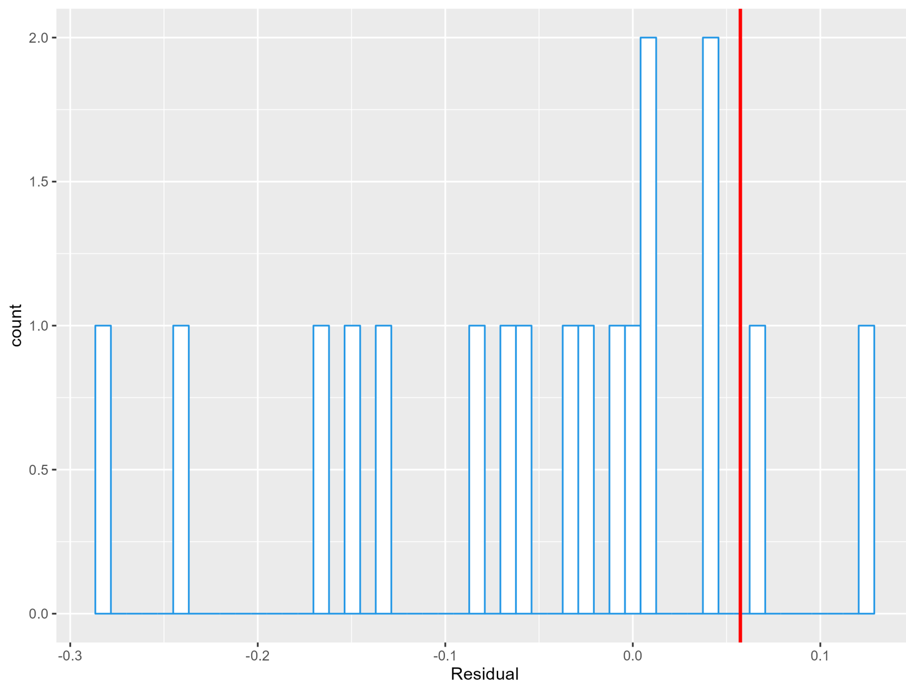

Figure 3.11: Test Correlations for Multiple Regression from Simulated Datasets, with Correlation from Actual dataset in Red Line

3.5 Random Forest Analysis

The *randomForest()* package in R was used to perform a Random Forest regression analysis of the data. In this technique, data is randomly sampled from the training set, with replacement, and it uses a subset of the sixteen independent quantitative variables at each split to generate a decision tree, which reduces the correlation between trees. The model outputs the average of these trees. The training set comprised the data from the first fourteen quarters, through Q4 2021, and the test set was from the remaining six quarters. The data was filtered for missing values for the dependent variable (quarterly bond return), and to manage the missing values within the independent variable, the *na.roughfix* option was used, which replaces the NA values with the median of the variable. As a result, these observations did not have to be omitted, and for most independent variables, there were not those many missing values, which was why this seemed the most efficient fix. The number of trees was set at 1000, and 5 variables were used at each split. The Importance Plot below shows the most important variables in the model from the training set.

The first four variables – Lagging Bond Return, Sign Log of Net Debt to UFCF, Sign Log of Net Leverage using Forward EBITDA, 10-year Treasury returns, were most important for the predicting the quarterly bond return, since they had the greatest impact in reducing MSE. The variables that increased the Node purity the most on aggregate were Sign Log of Net Leverage as a multiple of Forward EBITDA, and Lagging Bond Return. The R-squared of the training model was 0.77, which was higher than what was obtained in the multiple regression though this could be due to overfitting. The training model was then used on the test set and that produced an MSE of 0.629%, which was a lot lower than the R-squared of the multiple regression or any of the bivariate regressions. The R-squared of this model with the test set was 0.14, which was lower than the r-squared of the training model. For future research, it would be interesting to fine-tune the random forest model to obtain a more stable result and understand the variation in model outputs across time periods.

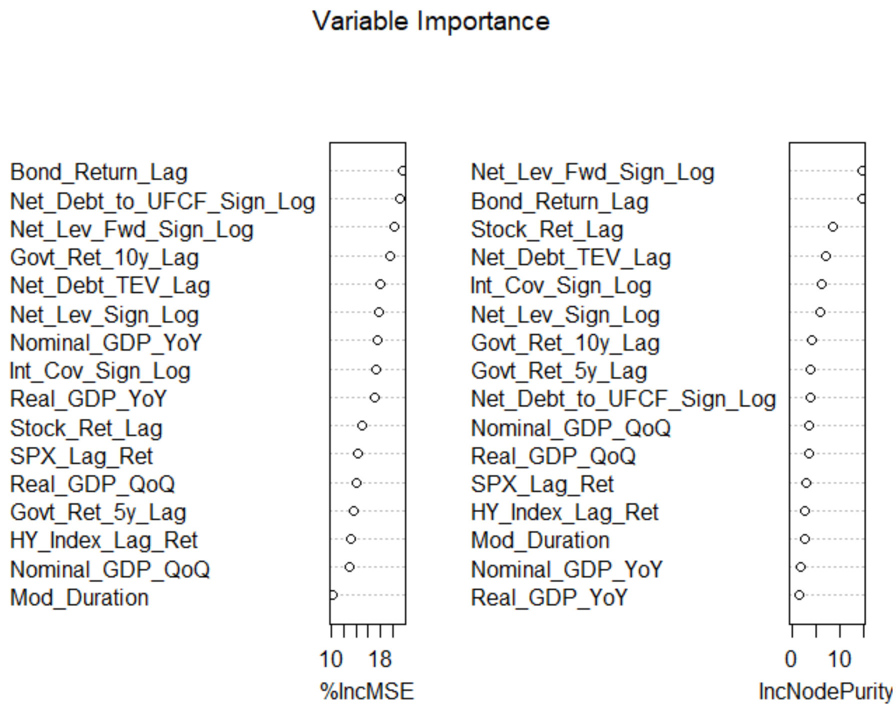

Figure 3.12: Variable Importance in Random Forest Regression

CHAPTER 4

Conclusion & Limitations

4.1 Conclusion

The bivariate regressions showed a significant relationship between the returns of the bonds in this sample and several of the lagging variables, but the low R-squared and test correlations indicate that these models did not have strong predictive power for quarterly returns. The multiple regression, after omitting the highly correlated variables, had an R-Squared of 0.22, and indicated a significant relationship for several of the relationships, but did not have a high test correlation, as discussed in Chapter 3.4. Most notably, the lagging returns of the 5-Year and 10-Year Treasury bonds had a positive and statistically significant relationship in the Bivariate and Multiple Regression, had low MSEs as per the cross validation, and appeared to have non-random relationships as per the cyclic permutation. Similarly, lagging Bond Returns had a negative and statistically significant relationship in both regressions, although the relationship varies depending on the monetary regime. Real GDP YoY Growth had a negative relationship though this could be affected by the fact that the data lagged the bond returns by more than a quarter so projections made for an influential quarter like Q1 2020 resulted in a negative relationship (see Figure A.4).

The relationships of other variables were less consistent across the two methods. For instance, Net Debt to TEV had a significant positive relationship in the multiple regression but a negative relationship in the bivariate regression. The Random Forest method had the highest R-squared for the training model and the lowest MSE, with the most influential variables being the Lagging Bond Return, Sign Log of Net Debt to UFCF, Sign Log of Net Leverage as a multiple of Forward EBITDA, and lagging returns of the 10-year Treasury bond. The returns of AMC's bonds were outliers and excluding them altered the results of the regressions. There were also influential quarters, such as Q1 2020 and Q2 2020, that experienced high volatility and affected the relationships with the quarterly bond returns.

For building more expansive forecasting models, the impact of influential quarters, such Q1 2020 and Q2 2020, need to be further explored, and if a longer time series or a wider set of independent variables werer available, it could provide a better understanding of how the relationships varied by economic conditions. Using other regression techniques or transformations of the variables may have also uncovered interesting relationships.

4.2 Limitations

The analysis excluded bonds of private companies, which are major issuers of High Yield bonds. It relied on Bloomberg Terminal to filter the initial list of bonds. An alternate data source, such as Thomson Reuters Eikon, or a combination of sources may have resulted in a different and more expansive dataset. Also, the High Yield universe was defined as those companies with an S&P Issuer Rating of BB+ or worse. An alternative criteria such as using a different rating agency's ratings may have resulted in a different universe with different results. Moreover, the data set was focused on bonds with a maturity of five years or more. Including bonds with a shorter maturity or a different maturity range could uncover different relationships.

For defaulted bonds, the final trading price was used as the final price, but it is possible that through the restructuring process the bondholders realized a different return. Even for performing bonds, the trading price on Bloomberg may not necessarily be actionable on a given day. Moreover, using daily or monthly bond returns could have resulted in different relationships. Lastly, the measure for EBITDA used in this analysis did not make any adjustments for non-recurring or non-cash costs that credit analysts typically take into consideration. If net leverage was computed after adjusting the EBITDA figure for non-recurring items, that may have been a more realistic measure of leverage.

APPENDIX A

Appendix of Supplementary Figures and Tables

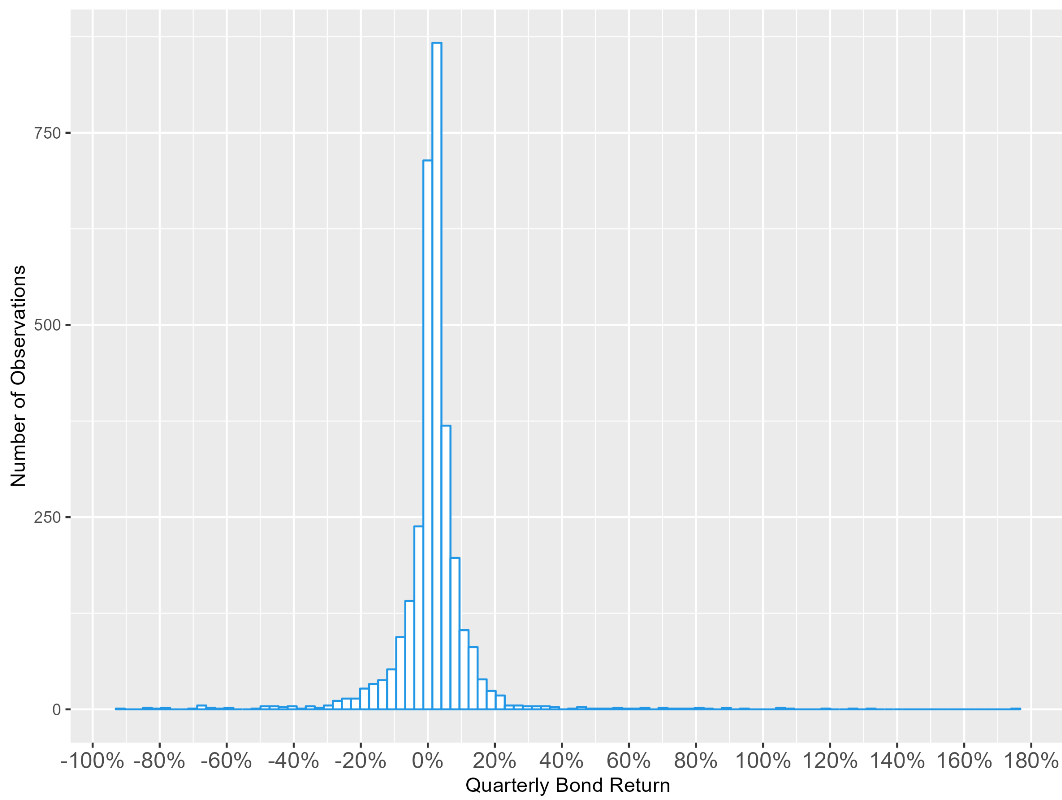

Figure A.1: Histogram of Quarterly Bond Returns excluding AMC Entertainment's bonds

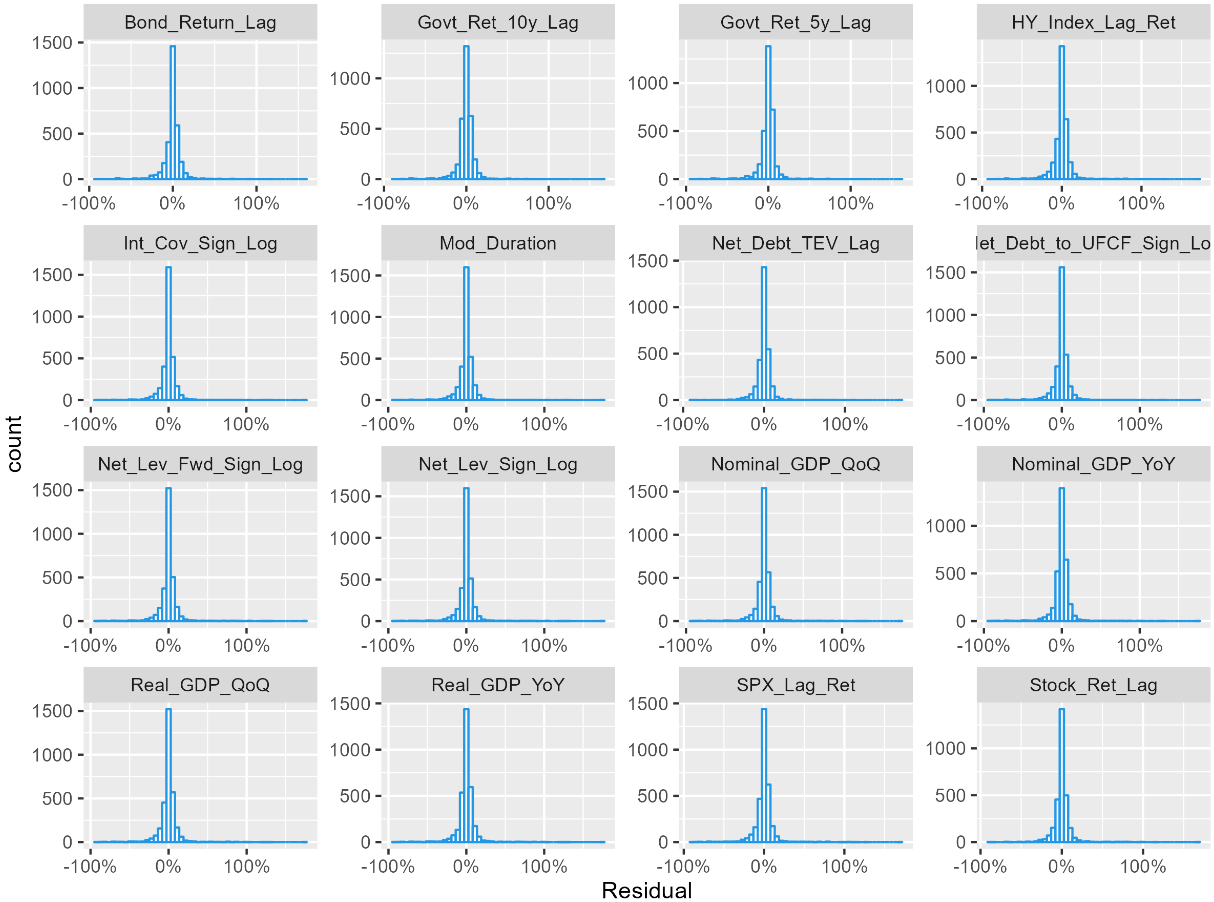

Figure A.2: Histogram of Residuals for Bivariate Regressions with Quantitative Variables excluding AMC Entertainment's bonds

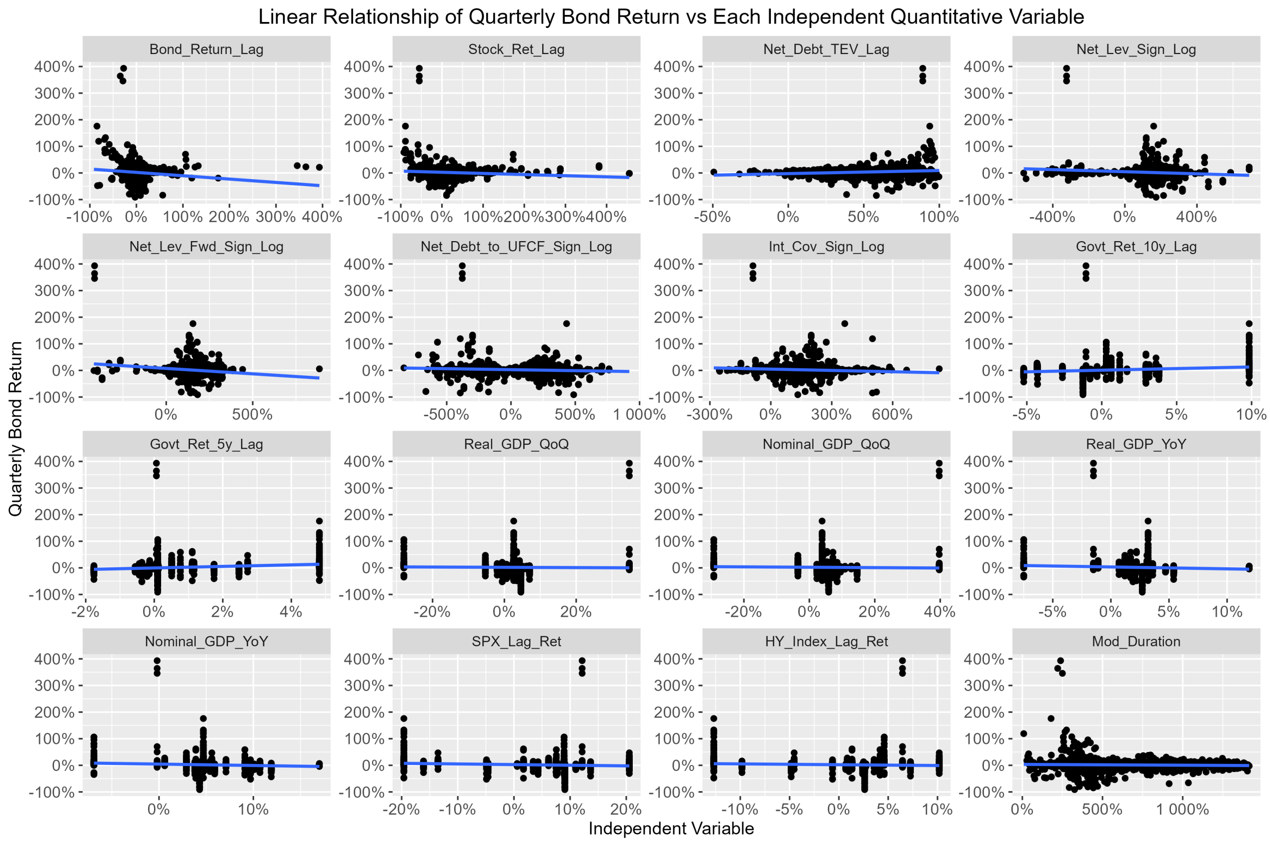

Figure A.3: Linear Relationship with each Independent Variable

Figure A.4: Regression Lines for each Independent Variable by Regime

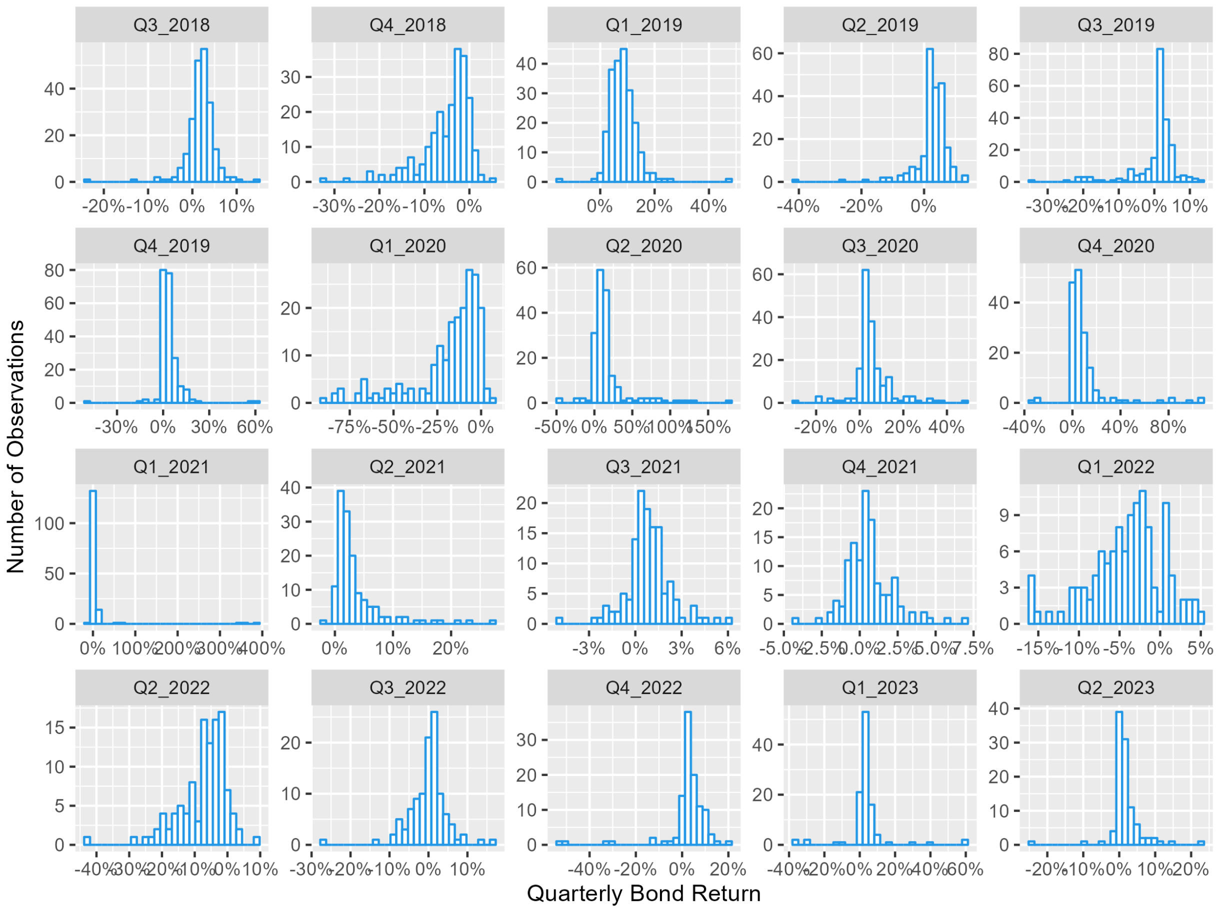

Figure A.5: Quarterly Bond Returns Distributions per Quarter

Variable	Estimate	Std Error	T-Value	P-Value
(Intercept)	0.0230	0.0104	2.2135	0.0269
Bond_Return_Lag	-0.3245	0.0248	-13.0855	0.0000
Stock_Ret_Lag	0.0556	0.0081	6.8520	0.0000
Net_Debt_TEV_Lag	0.0885	0.0110	8.0710	0.0000
Net_Lev_Sign_Log	-0.0023	0.0029	-0.7855	0.4322
Net_Lev_Fwd_Sign_Log	-0.0185	0.0047	-3.9144	0.0001
Net_Debt_to_UFCF_Sign_Log	-0.0029	0.0009	-3.2752	0.0011
Int_Cov_Sign_Log	-0.0041	0.0029	-1.4066	0.1597
Govt_Ret_5y_Lag	2.6498	0.1412	18.7646	0.0000
Real_GDP_QoQ	-0.0928	0.0192	-4.8427	0.0000
Real_GDP_YoY	-0.5548	0.0643	-8.6307	0.0000
HY_Index_Lag_Ret	0.0817	0.0459	1.7799	0.0752
Mod_Duration	-0.0005	0.0007	-0.7239	0.4692

Table A.1: Summary Statistics of Multiple Regression Excluding AMC Entertainment's bonds

Variable	Estimate	Std Error	T-Value	P-Value
(Intercept)	0.1374	0.0139	9.8692	0.0000
Bond_Return_Lag	-0.1680	0.0226	-7.4398	0.0000
Stock_Ret_Lag	-0.0120	0.0102	-1.1781	0.2389
Net_Debt_TEV_Lag	0.1838	0.0156	11.7533	0.0000
Net_Lev_Sign_Log	-0.0181	0.0037	-4.8966	0.0000
Net_Lev_Fwd_Sign_Log	-0.0702	0.0050	-14.1345	0.0000
Net_Debt_to_UFCF_Sign_Log	-0.0056	0.0013	-4.3689	0.0000
Int_Cov_Sign_Log	-0.0244	0.0040	-6.1305	0.0000
Govt_Ret_10y_Lag	0.9963	0.0880	11.3177	0.0000
Real_GDP_QoQ	0.0827	0.0286	2.8890	0.0039
Real_GDP_YoY	-0.9436	0.0951	-9.9263	0.0000
HY_Index_Lag_Ret	0.0562	0.0615	0.9128	0.3614
Mod_Duration	-0.0009	0.0011	-0.8402	0.4008

Table A.2: Summary Statistics of Multiple Regression using 10-year Treasury bond's returns instead of the 5-Year Treasury

Variable	**VIF**
Bond_Return_Lag	1.882065
Stock_Ret_Lag	2.127542
Net_Debt_TEV_Lag	1.630046
Net_Lev_Sign_Log	1.471654
Net_Lev_Fwd_Sign_Log	1.635287
Net_Debt_to_UFCF_Sign_Log	1.078544
Int_Cov_Sign_Log	1.324151
Govt_Ret_5y_Lag	12.037803
Govt_Ret_10y_Lag	8.518462
Real_GDP_QoQ	175.025811
Real_GDP_YoY	15.346422
HY_Index_Lag_Ret	13.081214
Mod_Duration	1.069105
SPX_Lag_Ret	15.690561
Nominal_GDP_QoQ	186.058688
Nominal_GDP_YoY	23.885378

Table A.3: VIF of Multiple Regression including all Quantitative Variables

Variable	Estimate	Std Error	T-Value	P-Value
(Intercept)	0.0833	0.0185	4.4983	0.0000
Bond_Return_Lag	-0.1803	0.0223	-8.0716	0.0000
Stock_Ret_Lag	-0.0086	0.0102	-0.8402	0.4008
Net_Debt_TEV_Lag	0.1964	0.0166	11.7991	0.0000
Net_Lev_Sign_Log	-0.0190	0.0037	-5.1977	0.0000
Net_Lev_Fwd_Sign_Log	-0.0706	0.0050	-14.2087	0.0000
Net_Debt_to_UFCF_Sign_Log	-0.0063	0.0013	-4.8797	0.0000
Int_Cov_Sign_Log	-0.0234	0.0041	-5.7752	0.0000
Govt_Ret_5y_Lag	2.4877	0.6407	3.8825	0.0001
Govt_Ret_10y_Lag	0.4516	0.2350	1.9215	0.0548
Real_GDP_QoQ	-2.3923	0.3443	-6.9476	0.0000
Real_GDP_YoY	0.5914	0.3192	1.8528	0.0640
HY_Index_Lag_Ret	1.1359	0.1816	6.2534	0.0000
Mod_Duration	-0.0007	0.0011	-0.6284	0.5298
SPX_Lag_Ret	-0.4648	0.1045	-4.4489	0.0000
Nominal_GDP_QoQ	2.2810	0.3173	7.1891	0.0000
Nominal_GDP_YoY	-1.5811	0.2867	-5.5154	0.0000
GICS_IndustryConsumer Discretionary	0.0238	0.0096	2.4956	0.0126
GICS_IndustryConsumer Staples	0.0188	0.0184	1.0237	0.3060
GICS_IndustryEnergy	0.0210	0.0104	2.0264	0.0428
GICS_IndustryHealth Care	0.0254	0.0141	1.7974	0.0724
GICS_IndustryIndustrials	0.0060	0.0119	0.5043	0.6141
GICS_IndustryInformation Technology	0.0198	0.0160	1.2374	0.2160
GICS_IndustryMaterials	0.0087	0.0113	0.7728	0.4397
GICS_IndustryReal Estate	0.0690	0.0214	3.2296	0.0013
GICS_IndustryUtilities	0.0081	0.0179	0.4540	0.6499

Table A.4: Summary Statistics of Multiple Regression using all the Variables

REFERENCES

[Blo24] "US Corporate High Yield Index." *Bloomberg*, 2024. https://assets.bbhub.io/professional/sites/27/US-Corporate-High-Yield-Index.pdf Accessed: Aug 5 2024.

[Dja19] Wadim Djatschenko. "The Nitty Gritty of Bond Valuation: A Generalized Methodology for Straight Bond Analysis." *SSRN*, 2019. http://dx.doi.org/10.2139/ssrn.3205167.

[Fab12] Frank J. Fabozzi. *The Handbook of Fixed Income Securities*. McGraw-Hill, New York, eighth edition, 2012.

[Fed24] "Effective Federal Funds Rate." *Federal Reserve Bank of New York*, 2024. https://www.newyorkfed.org/markets/reference-rates/effr Accessed: Aug 3 2024.

[GDP24] "Gross Domestic Product." *Bureau of Economic Analysis*, 2024. https://www.bea.gov/data/gdp/gross-domestic-product Accessed: Aug 3 2024.

[HA21] R.J Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Australia, third edition, 2021.

[HLW12] Yongmiao Hong, Hai Lin, and Chunchi Wu. "Are corporate bond market returns predictable." *Journal of Banking & Finance*, **36**:2216–2232, 2012.

[Hor23] Bill Hortz. "Revisiting the High Yield Bond Market." *Nasdaq*, 2023. https://www.nasdaq.com/articles/revisiting-the-high-yield-bond-market Accessed: Aug 1 2024.

[LWW14] Hao Lin, Junbo Wang, and Chunchi Wu. "Predictions of corporate bond excess returns." *Journal of Financial Markets*, **21**:123–152, 2014.

[MSC23] "Global Industry Classification Sector (GICS®)." *MSCI*, 2023. https://www.msci.com/documents/1296102/11185224/GICS+Sector+Definitions+2023.pdf Accessed: Aug 3 2024.

[MW24] Eric Milstein and David Wessel. "What did the Fed do in response to the COVID-19 crisis?" *Brookings Institution*, 2024. https://www.brookings.edu/articles/fed-response-to-covid19/ Accessed: Aug 6 2024.

[SPX23] "S&P 500: The Gauge of the U.S. Large-Cap Market." *S&P Global*, 2023. https://www.spglobal.com/spdji/en/documents/additional-material/sp-500-brochure.pdf Accessed: Aug 3 2024.