Returns around Earnings Announcements for Companies with Seasonality in Earnings

Ritika Dokania

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Statistics

Vijay Singal, Chair
Inyoung Kim
Xinwei Deng

May 9, 2018
Blacksburg, VA

Keywords: Abnormal Returns, Seasonality

Returns around Earnings Announcements for Companies with Seasonality in Earnings

Ritika Dokania

(ABSTRACT)

This thesis examines returns around earnings announcements for companies with seasonality in earnings.*Earnrank* is used as a measure of seasonality where *earnrank* for a company is calculated quarterly by taking last five years of earnings data, ranking them and taking the average of the ranks for the respective quarter. For seasonal firms, we find robust evidence that abnormal returns are created when such firms announce their earnings for the highest seasonality quarter as measured by their *earnrank*. Additionally, the results were consistent for different time periods and abnormal returns were found to increase over time. We also performed the analysis industry-wise and found significant difference in returns for most and least seasonal firms in Manufacturing, Financial and Construction sectors. The results for Construction sector is in conflict to our hypothesis and require further exploration. We also study which kind of firms exhibit seasonality and found evidence for high seasonality in large firms, value firms, old firms, firms with lower turnover and firms with lower accruals. Lastly, we studied factors determining abnormal returns relative to the four-factor model and found size to be a significant explanatory variable. The long-short portfolio based on seasonality generated an alpha of 62 basis points per month.

Returns around Earnings Announcements for Companies with Seasonality in Earnings

Ritika Dokania

(GENERAL AUDIENCE ABSTRACT)

This thesis examines returns around earnings announcements for companies with seasonality in earnings. Earning Seasonality is a phenomenon wherein firms show predictably higher earnings in one quarter of the year due to the underlying cyclical nature of the firms business. The quarter with the highest earnings is termed as positive seasonality quarter. *Earnrank* is used as a measure of seasonality where *earnrank* for a company is calculated quarterly by taking last five years of earnings data, ranking them and taking the average of the ranks for the respective quarter. For seasonal firms, we find robust evidence that abnormal returns are created when such firms announce their earnings for the highest seasonality quarter as measured by their *earnrank*. Additionally, the results were consistent for different time periods and abnormal returns were found to increase over time. We also performed the analysis industry-wise and found significant difference in returns for most and least seasonal firms in Manufacturing, Financial and Construction sectors. The results for Construction sector is in conflict to our hypothesis and require further exploration. We also study which kind of firms exhibit seasonality and found evidence for high seasonality in large firms, value firms, old firms, firms with lower turnover and firms with lower accruals. Lastly, we studied factors determining abnormal returns relative to the four-factor model and found size to be a significant explanatory variable. The long-short portfolio based on seasonality generated an alpha of 62 basis points per month.

Dedicated to:

My Husband, Parang Saraf
for being my pillar of support,
and encouraging me to scale bigger heights,

My Mother, Usha Dokania
who never let me be weak

My Father, Sajan Kumar Dokania
who believed in me more than anyone

And

My Daughter, Diya Saraf
who was my partner in achieving this feat

Acknowledgments

Doing Masters with a thesis option has not been a norm in the Statistics Department at Virginia Tech. When I joined the program, I went through the graduate handbook and found that there is an option to do Masters with thesis. Since my goal was to apply Statistics in Finance, I knew doing thesis was the appropriate way forward.

In this regards, I consider myself fortunate that Dr. Vijay Singal showed confidence in me and agreed to guide me through this process and did so in a highly systematic manner. Therefore, I would like to begin my acknowledgements by profusely thanking Dr. Vijay Singal and my committee members, Dr. Kim Inyoung and Dr. Xinwei Deng for guiding me through the process and helping me apply statistical concepts in finance. It would not have been possible for me to achieve what I have done without their help and support.

I would especially like to thank Dr. Singal, who has been a mentor and guide to me in this process. I want to thank him for giving me time whenever I requested, explaining me the technical details and resolving my queries in a very seamless fashion. I couldn't have asked for a better advisor than Dr. Singal.

Finally, I would like to thank my family and friends for supporting me through the process. I had my baby during my graduate program and there were many a time when it got very stressful to handle multiple obligations all at once. I would like to thank my mother, Usha Dokania, who kept me going strong, my father, Dr. Sajan Kumar Dokania, who asked me to keep working hard and not worry about the results, my husband, Dr. Parang Saraf, who showed immense faith and confidence in my abilities, my sisters Reema, Gunjan and Shashi, who kept my life entertaining, my brother, Neeraj for always having a word or two of encouragement and my daughter, Diya, who showed unconditional love. I am grateful to each one of them from the bottom of my heart. Their love and support has greatly helped me in successfully completing my Master's in Statistics program at Virginia Tech.

Ritika Dokania

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Seasonality is a phenomenon in which earnings in a particular quarter is higher as compared to earnings in other quarters. Retailers such as Macy's witness higher earnings in the December quarter as compared to other quarters. For an ice-selling company, June quarter would be the highest earning quarter. This trend is repeated year after year, thereby establishing seasonality for that firm or industry. Figure 1.1 explains the concept of seasonality through a use case. Quarterly earnings of Macy's is plotted from 2000-2005. Quarters 1, 2 and 3 are shown in green bars where quarter 4 (Q4) is represented in orange bar. Q4 earnings are found to peak every year from 2000-2005 emphasizing the existence of seasonality in earnings of Macys. Q4 would then be termed as positive seasonality quarter for Macys.

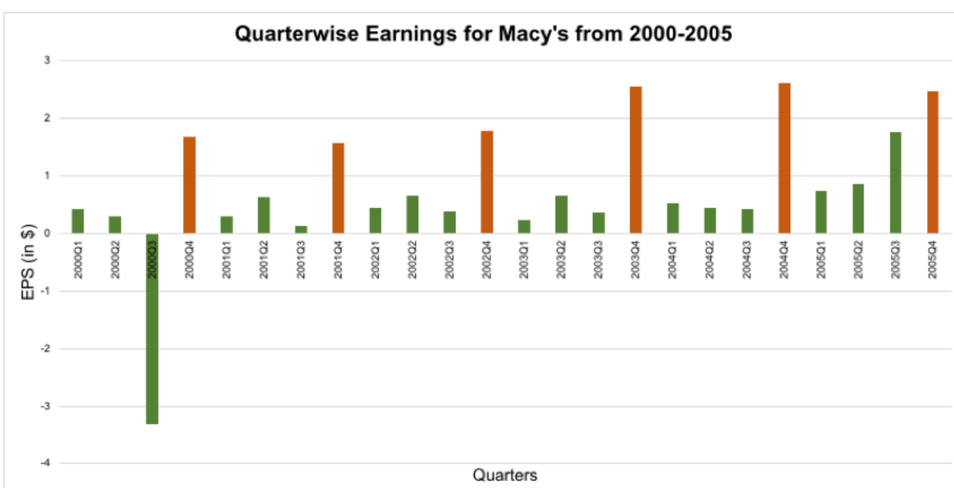

Figure 1.1: Example of Seasonality in Earnings for Macy's Inc.

Seasonality is persistent and repetitive in nature. It occurs mainly due to the underlying cyclical nature of the industry to which the company belongs. Holidays, climatic conditions, religion, customs, traditions are some of the factors that cause seasonality. For example, in the December quarter, in US, two big festivals – Christmas and Thanksgiving are celebrated in which people traditionally give gifts to each other, thereby increasing retail sales. Mega sale events such as 'Black Friday', 'Boxing Day' are also planned around this quarter for which people reserve their shopping for the whole year and this further adds to sales momentum in Q4.

Seasonality is not a stationary phenomenon and it varies by geography and demographic factors. As different festivals are celebrated in different parts of the world, the spending behavior of people change accordingly. Again, climatic conditions vary with countries and this leads to different production and consumption pattern in that area. Therefore, phenomenon of seasonality is highly dependent on geography, demography, diversity, culture and surrounding environmental factors.

Seasonality is widespread and pervasive. It is not limited to any particular industry or company but is seen across several industries, the timing of which depends on the demand factors driving the industry. A good example to mention here are Ski resorts, the demand of which is highest in Winter season as that is the time when the resort has adequate snow to facilitate skiing. For home 'Do It Yourself' stores like Lowes and Home Depot, Q2 may be the best when homeowners start buying items for the summer. Similarly, gas stations, airlines, and hotels will probably do well in the summer. Electric utilities may do better in the winter; and grocery stores may not exhibit any meaningful seasonality. For H&R Block, it may be Q1. Table 1.1 contains examples of some of the industry groups displaying seasonality.

Table 1.1: Example of Seasonality in Earnings across different Industry Groups

Industry	**Example Companies**	**Seasonality quarter**
Retail	Macys, Target	December quarter, higher sales due to Christmas, Thanksgiving
Hospitality	Priceline, Hilton Hotels	June quarter: higher sales due to more trips in holiday season
Accounting	H&R Block	March quarter: as the deadline for tax filing is in April, demand for tax accountants surge in the first quarter

As different industries are closely linked with one another, seasonality in one may lead to seasonality in another. For example, in order to meet the high retail sales demand in the holiday season, there is a spur in manufacturing activity in the months leading to holiday season with manufacturing rising sharply around August [2]. Further, in order to meet the increasing production demand, manufacturing firms need financing and the demand for capital increases. Loan demand peaks in June causing high seasonal sales for the financial sector in this quarter [14]. Another related example is the seasonality seen in the advertising industry. Ad rates are the highest in Q4, where companies have the maximum ad budget to encourage sales in the holiday season with budgets dwindling substantially in January. The Online Ad Revenue Index shown in figure 1.2 depicts a clear pattern of seasonality in sales for the advertising industry where peaks of revenues can be seen in Nov-Dec months followed by sharp decline in January [4]. For this reason, people sharing videos on YouTube for earnings, are advised to reserve some video posts for September – December time period to maximize their earnings [19].

In nutshell, different industries experience positive seasonal quarters in different months depending on factors governing sales for that industry.

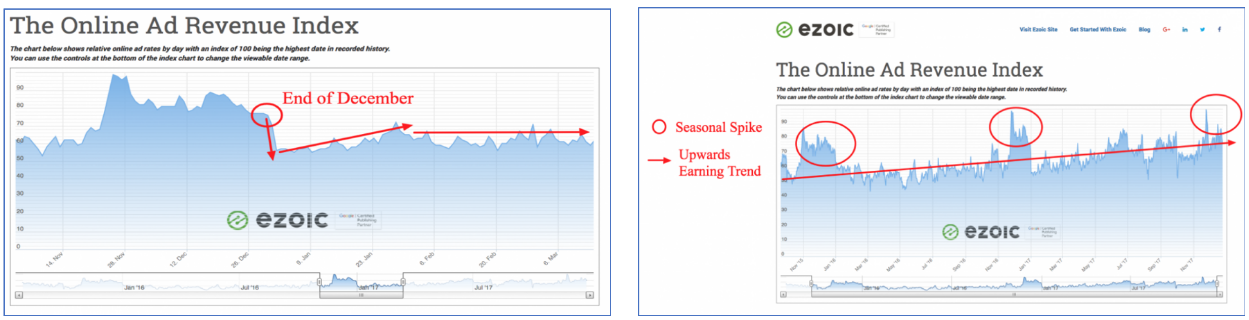

Figure 1.2: Seasonality in Advertising Industry depicted by Ad Revenue Index

1.1 Seasonality and Abnormal returns

One of the notable features of Modern Portfolio Theory is Efficient Market Hypothesis proposed by Eugene F Fama in 1970 [7]. Fama proposed that market is informationally efficient as stock prices fully reflect all the information available in the market. Stocks trade at their fair value in an efficient market at all times. This eliminates the scope of making abnormal returns in the market by buying low and selling high as market adjust itself to new information as soon as it arrives giving no time to investors to place their bet. The stock prices follow random walk and there is no way to predict changes in stock market prices based on the past returns.

On the contrary, financial research is filled with bountiful information challenging the Efficient Market Theory. The January Effect identified by Rozeff and Kinney (1976) [22] is one such anomaly which defies the Efficient Market Hypothesis. For an equal-weighted index of NYSE price, Rozeff and Kinney presented evidence that an average stock return in the month of January between 1904 and 1974 is found to be 3.5% as compared to 0.5% in other months. More than one-third of the annual returns accrued in January and it was found to be eight times higher than other months [28, 12]. The puzzling question is 'if we know this, why does this happen?'. Richard Roll (1983) asserts that some investors buy more in January in anticipation of price increases due to January Effect therefore contributing to the rise in prices [20]. Heston and Sadka (2007) found that stocks tend to have higher and lower returns in the same month of the calendar year [13]. Abnormal returns are found to be observed in recurring firm events such as stock splits, dividend announcement, earnings announcements, stock dividend etc. Grinblatt et al.(1984) presented empirical evidence of positive excess returns in the period surrounding stock dividends and stock splits [11]. Fama et al (1969) shows that stock splits are preceded by a period of unusually high returns [8]. In this context, we are interested in looking into whether there are any abnormal returns associated with seasonality, which is yet another recurring firm event. Seasonality of earnings is not news as per Samuelson (1965) but a phenomenon that is well-known and much anticipated [25]. Investors are aware of seasonal earnings much in advance and as per efficient market theory, such information must already be reflected in stock prices eliminating any scope of abnormal returns. However, Chang et al (2016) in their paper present evidence that market does not properly account for the information contained in the seasonality in earnings, which leads to significant abnormal returns in the months when companies announce their earnings for the positive seasonality quarter [6]. Salamon and Stober (1994) examine the effect of seasonality in earnings on the stock prices when such earnings were announced and provided evidence that excess returns are created for seasonal companies announcing results for their peak sales quarter [24].

Consider the example of Macys retail chain for the trading period 1993 to 2013 (see Table 1.2). Macys has a highly seasonal business with most of its earnings concentrated in the fourth quarter. The average monthly market-adjusted return for Macys fourth- quarter announcements was 3.72%, compared with 1.95% for all other quarters. Higher earnings in the fourth quarter is an inherent part of the Macys business and it is well-known to the investor community. Yet, analysts tend to underestimate the earnings in the fourth seasonality quarter leading to abnormal returns. The evidence was presented by Chang et al (2016) where analysts were found to make forecast errors due to recency bias [6].

Table 1.2: Comparing returns during earnings announcement for positive seasonality quarter with other quarters for Macy's Inc.

Company Name: Macy's	
Average monthly returns for positive seasonality quarter	3.72%
Average monthly returns of other quarters	1.95%

In this thesis, we replicate the empirical evidence of abnormal returns for companies announcing earnings for positive seasonality quarter as shown by Chang et al (2016) and extend their work by performing the analysis of abnormal returns for different industries and for different time periods. In addition, we will evaluate if there exist significant difference in abnormal returns between industries. We will further analyze which kind of firms exhibit seasonality and what are the factors determining abnormal returns. We believe that this thesis will provide a deep insight into the concept of seasonality, seasonality measures and abnormal returns associated with announcement of seasonal earnings. As mentioned in the previous section, seasonality is industry-specific and quarterly earning peak at different times for companies in different industries. Therefore, this thesis will provide an interesting insight into whether abnormal returns are created due to seasonality of earnings in different industry groups. Based on this, we define our problem statement and research questions as follows.

1.2 Problem Statement

To determine whether positive seasonality companies experience abnormal returns as compared to companies that do not exhibit seasonality?

- Is this phenomenon same across different time periods?
- Is this phenomenon same across industries?

1.3 Research Questions

In this thesis, based on the above mentioned problem statement, we examine the following three research questions:

- Research Question 1: Which kind of firms exhibit seasonality?
- Research Question 2: How do we quantify the existence of abnormal returns for most and least seasonal companies?
- Research Question 3: What are the factors determining abnormal returns?

1.4 Data description

For our analysis, we primarily used the following two databases:

1. CRSP [18]: The Center for Research in Security Prices, or more commonly known as CRSP, provides one of the most comprehensive databases for historical security prices and returns information for the stocks listed in United States. CRSP has stock market data for NYSE since 1926, AMEX data since 1962 and NASDAQ data since 1972. NYSE acquired AMEX in 2008 and the AMEX Exchange was renamed to NYSE MKT. The Exchange Code 2 or the Primary Exchange Code A refers to AMEX/NYSE MKT data.

2. Compustat Database [26]: Compustat is a financial database providing information on the annual and quarterly income statement, balance sheet, cash flow, pension and other supplemental data for various companies in the world. Compustat maintains data for both active and inactive companies. It has coverage of approximately 99,000 global securities, covering close to 99% of the market capitalization with annual company data available back to 1950 and quarterly data available back to 1962. This database is widely used by institutional investors, universities, bankers, financial advisors and multi-asset class asset managers for their research and analysis.

For our analysis, we used the data for the period dated October 1^{st}, 1972 – September 30^{th}, 2013. We include the common stock with CRSP share codes 10 and 11, listed on NYSE, AMEX and NASDAQ (Exchange Code 1, 2 and 3 respectively). This resulted in 2.7M rows corresponding to 22,330 companies in CRSP and 1.3M rows corresponding to 30,231 companies in Compustat. The data was further cleaned in an iterative manner to take care of any duplicate entries, missing data and noise. Common companies were then identified between CRSP and Compustat and the final dataset comprised of 1.7M rows in CRSP, 662K rows in Compustat and 10,493 unique companies.

1.5 Outline

The rest of the thesis is organized as follows: Chapter 2 describes CRSP and Compustat data, and the steps we took to clean and organize the data for our analysis. Chapters 3, 4, and 5 describe in detail each of the three research questions : 1) identifying the type of firms that exhibit seasonality, 2) whether abnormal returns exist for positive seasonality companies, and perform similar analysis industry-wise and for different time periods, and lastly, 3) what are the factors determining abnormal returns. We finally conclude in chapter 6.

Chapter 2

Working with CRSP and Compustat Data

Center for Research in Security Prices (CRSP) [18], is one of the most comprehensive databases for historical security prices and returns information for the stocks listed in United States. The database is widely used by the finance industry; academia and research institutions alike in United States and across the world for academic research and back-testing applications. CRSP has stock market data for companies listed on New York Stock Exchange (NYSE) since 1926, on American Stock Exchange (AMEX) since 1962 and on NASDAQ since 1972. NYSE acquired AMEX in 2008 and the AMEX Exchange was renamed to NYSE MKT. The Exchange code 2 or the Primary Exchange Code A refers to AMEX/NYSE MKT data.

Compustat [26] is a financial database providing information on the annual and quarterly income statement, balance sheet, cash flow, pension and other supplemental data for various companies in the world. Compustat maintains data for both active and inactive companies. It has coverage of approximately 99,000 global securities, covering close to 99% of the market capitalization with annual company data available back to 1950 and quarterly data available back to 1962. This database is widely used by institutional investors, universities, bankers, financial advisors and asset managers dealing in both fixed income and equity markets.

CRSP and Compustat, although considered gold standard data in finance, are found to contain various noise and missing values which require attention before performing the analysis. It is also important to understand the precise data definitions and organization of data in order to put the data to proper use. We used a number of iterative steps to clean the data before performing our analysis. [21]

2.1 CRSP

2.1.1 Data Definitions

We downloaded the following data from CRSP for each stock. Table 2.1 provides a brief description of each of the data columns. The detailed description of data is provided later.

Table 2.1: CRSP Data Definitions

CRSP Column Name	**Column Description**
Date	Date
PERMNO	Permanent Number
COMNAM	Company Name
TICKER	Ticker
SHRCD	Share Code
EXCHCD	Exchange Code
SICCD	SIC Code
PERMCO	CRSP Permanent Company Number
CUSIP	CUSIP
PRC	Price
VOL	Share Volume
RET	Holding Period Return
SHROUT	Number of Shares Outstanding
DIVAMT	Dividend Amount
DLSTCD	Delisting Code
DLPRC	Delisting Price
DLRET	Delisting Return

Detailed description of data:

Date: refers to the date for which stock price is reported

PERMNO: refers to the unique permanent identification number assigned by CRSP to each security. Unlike TICKER and company name, the PERMNO neither changes during an issues trading history, nor is it re-assigned after an issue ceases trading. The user may track a security through its entire trading history in CRSPs files with one PERMNO, regardless of name or capital structure changes. PERMNO is a 5–digit integer for all common securities in the CRSP files. It was observed that the same company with different share class has different PERMNO. Therefore, PERMNO is security-specific and not company-specific as shown through an example in Table 2.2

Table 2.2: Example of a company (Alphabet Inc.) with different PERMNOs corresponding to different Share Classes

PERMNO: **14542** TICKER: GOOG Company Name: Alphabet Inc Share Class: C	PERMNO: **90319** TICKER: GOOGL Company Name: Alphabet Inc Share Class: A

COMNAM: refers to the name of the company

TICKER: refers to the alphabetic symbol assigned to a security by an exchange. Tickers can be reused over time. So, a TICKER does not uniquely identify a security. PERMNO does.

SHRCD: refers to 2-digit Share Code. First digit describes the type of security, second digit provides further detail on security or company type. Table 2.3 lists the different types for Share Code.

EXCHCD: refers to Integer code indicating the exchange on which the security is listed. Table 2.4 lists various types of Exchange Codes.

SICCDrefers to SIC Code to classify companies into different industry groups.

PERMCO: refers to unique permanent company identification number assigned by CRSP to all companies with issue on CRSP files. This number is permanent for all securities issued by a company regardless of name changes. Please note that PERMCO is different from PERMNO, which is a unique identifier for each security on CRSP file. An interesting example of PERMCO is given in figure 2.1. In this figure, it can be seen that for the same company name and PERMCO, there are different TICKER, PERMNO and CUSIP.

CUSIP: like PERMNO, uniquely identifies a security. It consists of nine characters al-

Table 2.3: Description of Share Codes

First Digit

Code	Share Type
1	Ordinary Common Share
2	Certificates
3	ADRs
4	Share of Beneficial Interest
5	Units / ETF

Second Digit

Code	Security / Company Detail
0	Securities which have not been further defined
1	Securities which need not be further defined
2	Companies incorporated outside the US
3	Americus Trust Component, ETF
4	Closed ended funds
5	Closed-end fund companies incorporated outside the US
8	REITs

Table 2.4: Description of Exchange Codes

Code	Exchange
1	NYSE
2	NYSE MKT / AMEX
3	NASDAQ
4	ARCA
-1	Suspended Trading by Primary listing Exchange

PERMNO	date	SHRCD	EXCHCD	TICKER	COMNAM	SHRCLS	PERMCO	CUSIP
10001	10/29/93	11	3	GFGC	GREAT FALLS GAS CO		7953	36720410
10001	11/30/93	11	3	EWST	ENERGY WEST INC		7953	36720410
10001	8/31/09	11	3	EGAS	ENERGY INC		7953	36720410
10001	7/30/10	11	2	EGAS	GAS NATURAL INC		7953	36720410

Figure 2.1: Example of a company that has evolved over time. It has the same PERMNO, PERMCO and CUSIP but different TICKER and Company Name

phanumeric code which contains information about the issuer, the type of financial instrument offered and a validation digit. CRSP database normally truncates the last character from CUSIP making it an eight-letter identifier. CUSIP has been used to connect CRSP data to COMPUSTAT data. Figure 2.2 breaks down the CUSIP code for Amazon. Figure 2.3 gives an example of how CUSIP changes with different issues for the same issuer. Finally, figure 2.4 given an example of how CUSIP changed for the same company that went through multiple organizational changes such as name change, reorganization and acquisition.

PRC: refers to the last trading price of the security for the reported date. If closing price is

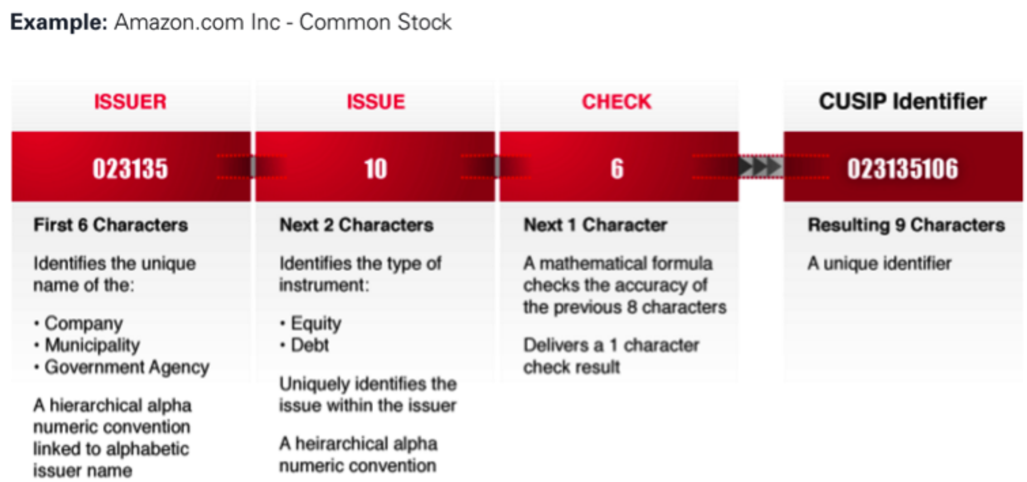

Figure 2.2: Breakdown of Amazon's CUSIP code

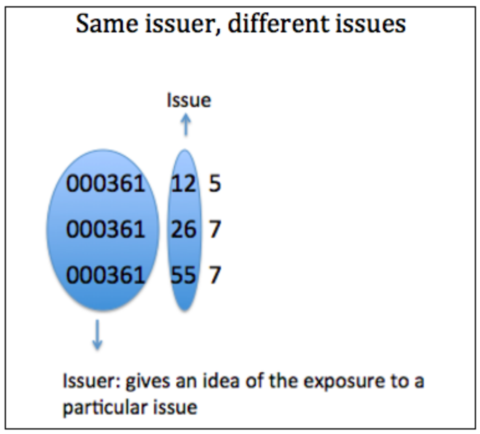

Figure 2.3: CUSIP example of a company with different issues

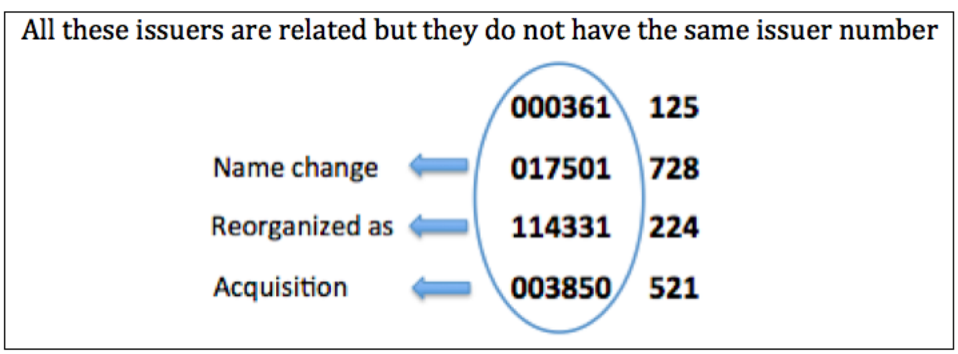

Figure 2.4: Change in CUSIP for an example company that evolved over time – name change, reorganization, and acquisition

not available, average of bid-ask is used with a leading dash for that date. Absolute value of price is used for our analysis in order to avoid any misunderstanding of dash as a negative sign.

VOL: refers to average daily volume of the security traded on the Exchange

RET: refers to Holding Period Return (HPR), which includes dividend and capital gains. HPR is calculated as follows:

$$\text{HPR} = \frac{\text{Income (dividend)} + \text{End of period value (price)} - \text{Initial value (price)}}{\text{Initial Value}} * 100$$

SHROUT: refers to shares outstanding for the company in thousands

DIVAMT: refers to ordinary cash dividends paid

DLSTCD: When a stock is delisted from exchange, a delisting code is assigned to the security indicating the reason for delisting. Table 2.5 lists the different reasons for delisting present in CRSP

Table 2.5: Description of Delisting Codes

Code	**Delisting Reason**
1	Active
2	Mergers
3	Exchange
4	Liquidations
5	Dropped
6	Expirations
9	Domestic become foreign

DLPRC: refers to Delisting Price, a trade price or a price quote (given as the average of bid and ask quotes) on another exchange or over-the-counter.

DLRET: refers to return of a security after it has been delisted from NYSE, NYSE MKT or NASDAQ. The Delisting Return is calculated by comparing the security′s amount after delisting with its price on the last day of trading. The amount after delisting can be either an off-exchange price, an off-exchange price quote, or the sum of a series of distribution payments.

2.1.2 Data Analysis and Noise Removal

For our analysis, data must meet the following four requirements:

1. Monthly stock data for the time period: October 1972 – September 2013 (∼40 years)
2. Stocks, with Common Stock Share Code – 10, 11
3. Stocks trading on NYSE, NYSE MKT/AMEX or NASDAQ, i.e. Exchange Code 1, 2 and 3
4. Stocks with no missing market capitalization for the listing period

Applying the first three filters in the downloaded data resulted in 2.7M rows and 22,339 unique companies as shown in Table 2.6.

At this step, it was found that there is a lot of noise present in the data and if the market cap filter is applied directly to the above subset of data, it would lead to a loss of large number of rows. Hence, using an iterative process, first various noise was identified and cleaned after which the market cap filter was applied. [3] Mentioned below is a list of all the noise that was removed:

1. Removal of rows with duplicate months
2. Removal of rows after the company has been delisted
3. Removal of top rows with missing price
4. Removal of bottom rows with no delisting code
5. Removal of stocks with price missing in the middle
6. Removal of top rows for missing months
7. Removal of bottom rows for missing months
8. Removal of stocks with months missing in the middle

Table 2.6 lists the number of data points left after applying all the noise removal filters.

Table 2.6: Summary of all the filters applied to CRSP data for data cleaning and noise removal. The table also lists the number of rows and unique stocks remaining after the application of each filter

Description	**Filters**	**Number of Rows**	**Number of Unique Stocks**
Downloaded Data	Oct 1, 1972 – Sep 30, 2013	3,419,425	28,531
Sharecode Filter	Share Codes: 10, 11	2,737,097	22,343
Exchange Code Filter	Exchange Codes: 1, 2, 3	2,686,160	22,339
Noise Removal	Removing rows with duplicated months. For such rows, only the dividend amount and record date were different. Price, HPR, etc. stayed the same	2,673,230	22,339
	Noise relating to Delisting: a) Removing companies where the first row has delist code b) Removing rows after the company gets delisted	2,670,071	22,331
Market cap filter: Missing prices	Removing top rows with missing prices	2,669,437	22,321
	Removing bottom rows with missing prices and no delist code	2,669,291	22,321
	Removing companies with missing prices in the middle	2,385,543	20,618
Market cap filter: Missing months	Removing top rows with missing months	2,385,528	20,618
	Removing bottom rows with missing months	2,385,519	20,618
	Removing companies with missing months in the middle	2,247,425	19,665

Table 2.7: Noise removal led to a saving of 841k rows corresponding to 4700 stocks

Description	**Total number of rows**	**Total number of unique stocks**
Total data downloaded	3,419,425	28,531
Data left after the application of filters: *After Noise Removal*	2,247,425	19,665
Data left after the application of filters: *Without Noise Removal*	1,405,714	14,965
Savings	**841,711**	**4,700**

Table 2.7 compares the effect of noise removal on the final data. As can be seen in the table 2.7, after applying all the filters, we are left with 2,247,425 rows (out of 3,419,425 rows) and 19,665 unique stocks (out of 28,531 stocks). However, if we apply the filters directly to our downloaded data without removing any noise, we are left with only 1,405,714 rows and 14,965 stocks. Therefore, we saved 841,711 rows corresponding to 4,700 stocks through an efficient data cleaning process.

2.2 Compustat

2.2.1 Data Description

Table 2.8 lists all the columns that were downloaded along with the description of each of those columns.

Table 2.8: Compustat Data Description

Column Name	**Description**
GVKEY	GVKEY or Global company key, is a unique 6-digit number key assigned to each company in the Capital IQ Compustat database
Datadate	Date for which the information is being provided
FYR	Month in which fiscal year ends. When a company changes it's FYR during the year, Compustat keeps the old and new values
FYEARQ	Information on fiscal year
FQTR	Information on fiscal quarter
DATACQTR	Information on calendar year and calendar quarter
DATAFQTR	Information on fiscal year and fiscal quarter
TIC	Ticker
CONM	Company Name
RDQ	Report date of quarterly earnings
CSHOQ	Common Shares Outstanding
EPSPXQ	EPS basic (quarterly) excluding extra-ordinary items
EPSPXY	EPS basic (annually) excluding extra-ordinary items
NIQ	Net Income
EXCHG	Exchange. '11' for NYSE, '12' for AMEX and '14' for Nasdaq-NMS stock market
MKVALTQ	Total market value
SIC	Industry Classification code

2.2.2 Adjusting EPS for Stock Splits

Compustat reports quarterly EPS data as reported by the company. No adjustments are made for stock splits. As most of our analysis would involve comparison of returns over time with respect to EPS, it is important that EPS is normalized for a fair comparison. In order to adjust EPS for stock splits, Factor to Adjust Price (FACPR) was used from CRSP data.

FACPR values are assigned by CRSP as shown in table 2.9:

Table 2.9: FACPR values assigned by CRSP

FACPR Value	**Description**
0	in case of cash dividends
-1	when a company is delisted
> 0	in case of stock split
< 0	in case of reverse stock split

For our purpose, we did not make adjustments in the following cases:

- when FACPR was equal to 0,
- when FACPR was equal to -1 and
- when $-0.1 <$ FACPR < 0.1 (as the change in shares is minimal)

For all other cases, EPS was adjusted for stock splits. Figure 2.5 shows EPS adjustments for an example company, Rocky Mountains, Ticker: RMCF.

Please note the day FACPR data is released, EPS is already adjusted, so no adjustment of EPSPXQ is done for that date. In CRSP, holding period returns are already adjusted for stock splits and therefore do not require any adjustment. In our analysis, FACPR was adjusted for 5302 companies. Since no rows were deleted in this step, the number of rows and unique companies remained intact after EPS adjustment.

PERMNO	date	TICKER	FACPR
10044	8/31/88	RMCF	-0.6667
10044	3/28/02	RMCF	0.33333
10044	2/27/04	RMCF	0.5
10044	5/28/04	RMCF	0.1
10044	6/30/05	RMCF	0.33333

(a) FACPR data taken from CRSP for Rocky Mountains

gvkey	datadate	tic	epspxq	adj facpr1	adj facpr2	adj facpr3	adj facpr4	adj facpr 5	net adj	adj EPS
.	.	.	.	.	.	.	.	.	.	.
11976	2/29/88	RMCF	0.01	0.33	1.33	1.50	1.10	1.33	0.98	0.01
11976	5/31/88	RMCF	-0.05	0.33	1.33	1.50	1.10	1.33	0.98	-0.05
11976	8/31/88	RMCF	0.01		1.33	1.50	1.10	1.33	2.93	0.00
11976	11/30/88	RMCF	-0.67		1.33	1.50	1.10	1.33	2.93	-0.23
.	.	.	.	.	.	.	.	.	.	.
11976	2/28/02	RMCF	0.23		1.33	1.50	1.10	1.33	2.93	0.08
11976	5/31/02	RMCF	0.18			1.50	1.10	1.33	2.20	0.08
11976	8/31/02	RMCF	0.25			1.50	1.10	1.33	2.20	0.11
.	.	.	.	.	.	.	.	.	.	.
11976	11/30/03	RMCF	0.25			1.50	1.10	1.33	2.20	0.11
11976	2/29/04	RMCF	0.15				1.10	1.33	1.47	0.10
11976	5/31/04	RMCF	0.14					1.33	1.33	0.11
11976	8/31/04	RMCF	0.24					1.33	1.33	0.18
11976	2/28/05	RMCF	0.19					.	.	.
11976	5/31/05	RMCF	0.16					1.33	1.33	0.12
11976	8/31/05	RMCF	0.18							0.18
11976	11/30/05	RMCF	0.18							0.18
11976	2/28/06	RMCF	0.17							0.17

(b) EPSPXQ adjustment in Compustat data. The highlighted rows in yellow indicate the stock split events

Figure 2.5: Example of EPS adjustment for stock splits for a sample company – Rocky Mountains, Ticker: RMCF

2.2.3 Noise Removal

Table 2.10 lists number of data points left in Compustat data after applying filters for noise removal and identifying common companies between CRSP and Compustat.

Table 2.10: Summary of all the filters applied to Compustat data for data cleaning and noise removal. The table also lists the number of rows and unique stocks remaining after the application of each filter

<table>
<tr><th>Description</th><th>Filters</th><th>Number of Rows</th><th>Number of Unique Stocks</th></tr>
<tr><td>Downloaded Data</td><td>Oct 1, 1972 – Sep 30, 2013</td><td>1,318,631</td><td>30,231</td></tr>
<tr><td>Common Companies between CRSP and Compustat</td><td></td><td>735,691</td><td>15,093</td></tr>
<tr><td rowspan="2">Noise Removal</td><td>Removing rows with duplicated months</td><td>735,211</td><td>15,093</td></tr>
<tr><td>Removing companies with less than 24 rows</td><td>663,074</td><td>10,493</td></tr>
<tr><td rowspan="2">Ready to use Data</td><td>Number of rows</td><td colspan="2">663,074</td></tr>
<tr><td>Number of Unique Stocks</td><td colspan="2">10,493</td></tr>
</table>

2.3 Data Summary: CRSP and Compustat

Tables 2.11 and 2.12 list summary statistics corresponding to final clean CRSP and Compustat data.

Table 2.11: Number of CRSP and Compustat data points used in our analysis

Description	**Statistics**
Number of Rows CRSP	1,708,819
Number of Rows Compustat	662,720
Number of Unique Stocks	10,493
Number of Portfolio Months	420

Table 2.12: Summary statistics of clean CRSP and Compustat data used in our analysis

Description	**Average**	**Standard Deviation**
Returns	1.43%	17.50%
Market Capitalization	$1766.M	$10,904M
Earnrank	10.4	2.8

Chapter 3

Firms that exhibit Seasonality

In this chapter, we are interested in finding out what kind of firms exhibit seasonality. In order to do that we first explore methods to quantify seasonality. Thereafter, we check how the phenomenon of seasonality varies for different firm characteristics i.e. size, book to market value, share turnover, accruals and age.

3.1 Measuring seasonality

An important question to answer at this point is: how do we measure Seasonality?

Seasonality as a concept is known to mankind through experiences when a common pattern repeats year after year. A doctor knows when is she going to have the maximum flow of patients, a confectionery store owner is aware when to expect highest demand of candies while a tourist guide knows when will he have the maximum number of visitors and so on. It is easy to identify positive seasonality quarters and the period of high sales is implicitly known to the business-owner. However, what is not simple is to quantify seasonality.How to differentiate whether firm A exhibit higher seasonality as compared to firm B?

Why do we want to quantify seasonality? Seasonality tends to have important economic ramifications for various stakeholders [1]. It impacts pricing of goods and services, demand and supply of labor, production, distribution and marketing activities ultimately impacting stock prices. Chang et al [6] presented evidence that abnormal returns are created when a company announces earnings for its positive seasonality quarter. It is therefore important to quantify seasonality before we can understand its impact on returns.

Over the years, researchers have used various methodologies to quantify seasonality. One of the most common techniques used for seasonal adjustment is X-12-ARIMA software package developed by U.S. Census Bureau [10]. While the process has been standardized for general use, the aforementioned tool is a black-box for regular user and lacks interpretability as the model makes various assumptions to calculate seasonality factor. Another method commonly used to measure seasonality is the method of simple averages in which seasonality for a particular quarter is calculated by taking the ratio of earnings of that particular quarter to average of earnings for all the four quarters in the calendar year. While the method of simple averages is easy to interpret, it may not be most optimal to represent seasonality especially for a growing firm where earnings growth may be due to growth factors, which can be mistaken as seasonality. Salamon and Stober (1994) employs the method of sales ratio to calculate seasonality [24]. Given this method also use absolute values of sales to measure the seasonality factor, it would suffer from the same drawback as the method of simple averages.

For our analysis, we have used earnrank measure to calculate seasonality as formulated by Chang et al in their paper Being surprised by the Unsurprising [6]. Earnrank for a quarter t is calculated by taking earnings of the last 20 quarters starting from t-4 to t-23, sorting the earnings in ascending order and ranking them from 1 to 20. In order to calculate earnrank for quarter t, we would then take the average of ranks corresponding to the earnings of quarter t-4, t-8, t-12, t-16 and t-20. The higher the earnrank value, higher is the seasonality. Given below is an example of calculating earnrank in a step-wise manner. In this example, we are calculating earnrank for Apple stock for 2016 Quarter 1.

Step 1: Select the last 20 quarters of earnings starting from t-4 to t-23. Earnings data should be present for all 20 quarters in order to calculate earnrank. To calculate earnrank for 2016Q1 for Apple, earnings from 2010Q2 to 2015Q1 is selected

datadate	TICKER	year	quarter	adj eps	earnrank
31-Mar-10	AAPL	2010	Q1	3.39	
30-Jun-10	AAPL	2010	Q2	3.57	
30-Sep-10	AAPL	2010	Q3	4.71	
31-Dec-10	AAPL	2010	Q4	6.53	
31-Mar-11	AAPL	2011	Q1	6.49	
30-Jun-11	AAPL	2011	Q2	7.89	
30-Sep-11	AAPL	2011	Q3	7.13	
31-Dec-11	AAPL	2011	Q4	14.03	
31-Mar-12	AAPL	2012	Q1	12.45	
30-Jun-12	AAPL	2012	Q2	9.42	
30-Sep-12	AAPL	2012	Q3	8.76	
31-Dec-12	AAPL	2012	Q4	13.93	
31-Mar-13	AAPL	2013	Q1	10.16	
30-Jun-13	AAPL	2013	Q2	7.51	
30-Sep-13	AAPL	2013	Q3	8.31	
31-Dec-13	AAPL	2013	Q4	14.59	
31-Mar-14	AAPL	2014	Q1	11.69	
30-Jun-14	AAPL	2014	Q2	1.29	
30-Sep-14	AAPL	2014	Q3	1.43	
31-Dec-14	AAPL	2014	Q4	3.08	
31-Mar-15	AAPL	2015	Q1	2.34	
30-Jun-15	AAPL	2015	Q2	1.86	
30-Sep-15	AAPL	2015	Q3	1.97	
31-Dec-15	AAPL	2015	Q4	3.3	
31-Mar-16	**AAPL**	**2016**	**Q1**	**1.91**	**11.6**
30-Jun-16	AAPL	2016	Q2	1.43	8.2
30-Sep-16	AAPL	2016	Q3	1.68	8.4
31-Dec-16	AAPL	2016	Q4	3.38	13.8

Qtr t - 4 to t - 23

Figure 3.1: Calculating Earnrank (Step 1): select data from the last 20 quarters of earnings from t-4 to t-23

Step 2: Rank the 20 quarters of earnings data from smallest to largest

year	qtr	eps	rank
2014	Q2	1.29	1
2014	Q3	1.43	2
2015	Q1	2.34	3
2014	Q4	3.08	4
2010	Q2	3.57	5
2010	Q3	4.71	6
2011	Q1	6.49	7
2010	Q4	6.53	8
2011	Q3	7.13	9
2013	Q2	7.51	10
2011	Q2	7.89	11
2013	Q3	8.31	12
2012	Q3	8.76	13
2012	Q2	9.42	14
2013	Q1	10.16	15
2014	Q1	11.69	16
2012	Q1	12.45	17
2012	Q4	13.93	18
2011	Q4	14.03	19
2013	Q4	14.59	20

Figure 3.2: Calculating Earnrank (Step 2): rank earnings data from smallest to largest

Step 3: Earnrank, for quarter t is the average rank of quarters t − 4, t − 8, t − 12, t − 16, and t − 20. This is the average rank of the same fiscal quarter as the upcoming announcement from previous years.

Sorted by quarters

year	qtr	eps	rank
2015	Q1	2.34	3
2011	Q1	6.49	7
2013	Q1	10.16	15
2014	Q1	11.69	16
2012	Q1	12.45	17
2014	Q2	1.29	1
2010	Q2	3.57	5
2013	Q2	7.51	10
2011	Q2	7.89	11
2012	Q2	9.42	14
2014	Q3	1.43	2
2010	Q3	4.71	6
2011	Q3	7.13	9
2013	Q3	8.31	12
2012	Q3	8.76	13
2014	Q4	3.08	4
2010	Q4	6.53	8
2012	Q4	13.93	18
2011	Q4	14.03	19
2013	Q4	14.59	20

Earnrank for 2016Q1 = (3+7+15+16+17) / 5
= 11.6

Figure 3.3: Calculating Earnrank (Step 3): Average rank of the same fiscal quarter to calculate earnrank

Why Earnrank?: Earnrank as a measure of seasonality is conceptually easy to understand. It is intuitive and simple to construct. When we start ranking the earnings of last 20 quarters to calculate earnrank, we start getting the sense of seasonality as the quarters with traditionally higher earnings will start getting higher rank. Quarter with the highest earnrank value will be called positive seasonality quarter. Some other benefits of earnrank is that it is not affected by negative earnings unlike certain other seasonality measures which involves growth in earnings. Earnrank is less sensitive to earnings growth, negative earnings and is not much affected by the presence of outlier in data.

3.2 Type of firms exhibiting seasonality

In order to answer the research question: which kind of firms exhibit seasonality, we perform a regression analysis.

3.2.1 Regression Analysis

The dependent and independent variables for the regression analysis are defined as below:

Dependent variable

- Change in Earnrank between a firms highest and lowest announcements over the calendar year (for firms with four announcements)

  Change in earnrank is used as a measure because the difference between the highest and the lowest earnrank implies the magnitude of seasonality experienced by a company. Earnrank is an absolute measure and does not carry much meaning on its own unless compared to some baseline. For a particular company, in a given year, the quarter with highest earnrank points to the highest seasonality quarter and the quarter with the lowest earnrank refers to the least seasonal quarter. The question is: can we comment on seasonality by looking at a certain earnrank value in isolation? Lets consider the following two cases to explain this point further:

Table 3.1: An example case to explain difference in earnrank for measuring the magnitude of seasonality in a firm

Company	**Year**	**Lowest Earnrank**	**Highest Earnrank**	**Earnrank difference**
Company A	2015	8	9	1
Company B	2015	7	11	4

In the above example, the question is: which company exhibit higher seasonality? Can we answer this question by solely looking at a particular earnrank value? The answer would be no.

In this example, Company B exhibit higher seasonality as the difference between the highest and lowest earnranks is higher for company B as compared to Company A. Seasonality is defined as a phenomenon when earnings in one particular quarter is predictably higher as compared to other quarters. Difference between the highest and the lowest earnranks enables us to capture the essence of comparability between quarters that is embedded in the definition of seasonality. Hence, we feel that range will be an appropriate measure for seasonality for a firm.

Independent variables

- Five predictor variables are used to define the different firm types: log market capitalization, log book-to-market ratio,share turnover, accruals, log of firm age.

  Each of the predictor variables is described in detail in table 3.2

Table 3.2: Description of Independent Variables

Predictor variable	**Description**
log market capitalization	This predictor variable is used to capture the size of the firm whether the firm is a large-cap or a small-cap
log book-to-market ratio	This predictor variable is used to capture whether the firm is a value or a growth firm
share turnover	This predictor variable is used to capture whether the firm is high or low momentum
accruals	This predictor variable is used to capture whether the firm has higher or lower accruals. Accruals is calculated using the following formula: Accruals = Net Income – Cash flow from Operations – Cash flow from Investing
log of firm age	This predictor variable is used to capture whether the firm is an old or a young firm. Number of years since the IPO is used as a proxy for age

Results:

For regression analysis, the dependent variable is the difference between the highest and lowest earnrank for a company over next year is regressed on various types of firm characteristics from the previous year: log market capitalization, log book to market, share turnover, accruals and firm age. The results for the regression analysis is given in figure 3.4

```
Total Rows: 58540

                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.034
Model:                            OLS   Adj. R-squared:                  0.034
Method:                 Least Squares   F-statistic:                     414.5
Date:                Wed, 25 Apr 2018   Prob (F-statistic):               0.00
Time:                        04:40:22   Log-Likelihood:            -1.5133e+05
No. Observations:               58540   AIC:                         3.027e+05
Df Residuals:                   58534   BIC:                         3.027e+05
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.3492      0.053     82.164      0.000       4.245       4.453
x1             0.1534      0.016      9.369      0.000       0.121       0.186
x2             0.4138      0.040     10.408      0.000       0.336       0.492
x3         -1.893e-05   3.73e-06     -5.081      0.000   -2.62e-05   -1.16e-05
x4             1.3576      0.047     28.599      0.000       1.265       1.451
x5         -1.739e-05   6.99e-07    -24.883      0.000   -1.88e-05    -1.6e-05
==============================================================================
Omnibus:                     3916.046   Durbin-Watson:                   0.429
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             4722.322
Skew:                           0.691   Prob(JB):                         0.00
Kurtosis:                       2.841   Cond. No.                     1.26e+05
==============================================================================

x1: log(marketcap), x2: log(book_to_market), x3: accruals, x4: log(age), x5: share_turnover
```

Figure 3.4: Regression results: Type of firms exhibiting seasonality

From the results in Table 3.4, the seasonal shift in earnings is found to be significant for each firm type considered in problem. Seasonality in earnings is found to be prominent in large firms, value firms, old firms, firms with lower turnover and firms with lower accruals. All the values are found to be statistically significant at 1% level.

At this point, it would be appropriate to mention that the sample size used in the problem has a slight bent towards old firms, as the earnrank measure, which is used for the dependent variable, requires at least last five years of earnings data. The firms, which are relatively old and have a continuous flow of quarterly earnings information are more likely to be included in the analysis as compared to younger firms, which have shorter trading history and are more likely to be discarded due to insufficient data. The result obtained may not generalize well to the young firms.

3.3 Conclusion

Seasonal shift in earnings are more common for large firms, value firms, old firms, firms with lower turnover and firm with lower accruals.

Chapter 4

Quantifying Abnormal Returns

This chapter deals with a pertinent research question we are trying to answer in this thesis i.e. whether seasonality in earnings can lead to abnormal returns. In the previous chapters, we determined the measure of seasonality and the type of firms that exhibit seasonality. In this chapter, we are exploring the question: whether announcement of earnings for the positive seasonality quarter leads to abnormal returns? Such results have a potential to discover an anomaly in the market thereby challenging Efficient Market theory, which reasons that stocks trade at fair value at all times, therefore, there is no opportunity for investors at any time to generate abnormal returns. All the publicly available information is reflected in stock prices instantly as soon as it reaches market. Seasonality, which is a highly persistent phenomenon known to investors way in advance, must be completely reflected in stock price before the earnings from a high seasonality quarter is announced. In this thesis, we are challenging this hypothesis and using empirical analysis to test whether abnormal returns can be explained by the seasonality factor.

We further delve into three sub-problems:

1. Test the hypothesis of abnormal returns due to seasonality factor for different time periods
2. Test the hypothesis of abnormal returns due to seasonality factor for different industries
3. Compare the abnormal returns between industries and test whether the difference is significant or not.

4.1 Quantifying abnormal returns for overall data

4.1.1 Research Question

How do we quantify the existence of abnormal returns for most and least seasonal companies?

4.1.2 Methodology

In order to test our hypothesis, we form a portfolio every month and sort the companies by their seasonality factor, which is earnrank in our case, to calculate abnormal returns. For portfolio formation for the current month, we first identify the companies which announced their earnings in the same month in the previous year. We used 'rdq' i.e. reporting date of quarterly earnings data from Compustat to identify the companies.

In our portfolio, we exclude stocks with a price less than $5 or a missing market capitalization at the end of the previous month. We placed a limit of at least ten stocks to form a portfolio. For the months, where we have less than 10 stocks in a portfolio, such months are excluded from our analysis. Once we have formed the portfolio, we sort the companies by their earnrank in an ascending order and divide the portfolio into quintiles based on earnrank. By doing this, the companies exhibiting lowest seasonality move to quintile 1 and the companies reporting their most seasonal earnings fall in quintile 5. We calculate the average monthly returns of quintile 1 stocks and quintile 5 stocks and then take their difference to test whether the difference is statistically significant or not. Based on these criteria, we were able to form portfolio for 420 months.

4.1.3 Exploratory analysis and statistical tests

Before testing for statistical significance in difference between average monthly returns of quintile 1 and quintile 5 stocks, we use various exploratory techniques and statistical tests to check the distribution of data, i.e. whether data is normally distributed or not.

1. Box Plots:

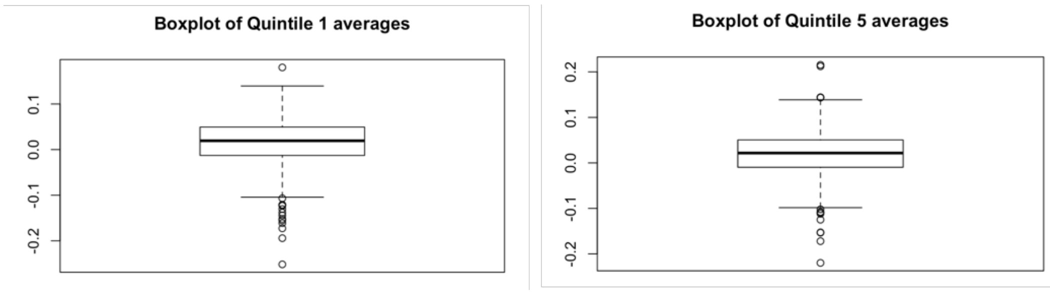

Figure 4.1: Box plots for Quintile 1 and Quintile 5 average returns

Both the datasets (Quintile 1 average stock returns and Quintile 5 average stocks returns, n = 420) are approximately balanced around zero; thereby pointing to the fact that the mean in both cases is close to zero. However, there are many outliers in both the datasets as evident from the circles outside the whiskers on either side for both the plots. While the presence of the outliers does not completely rule out the normal distribution, it does raise questions around skewness and kurtosis in data [27, 30]

2. Histograms:

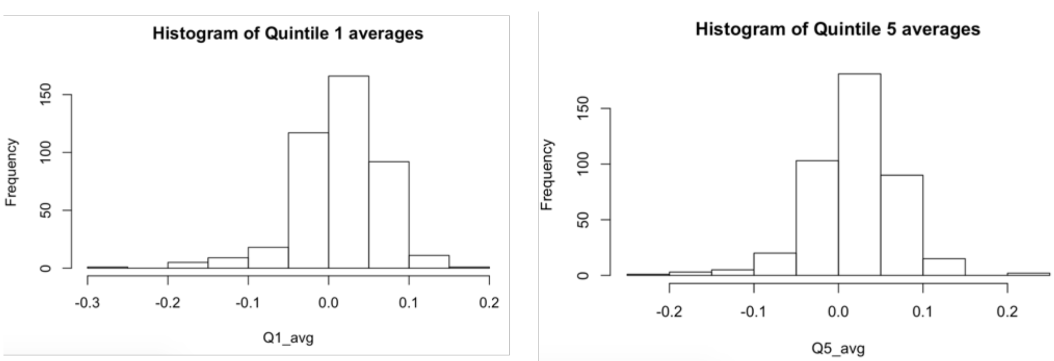

Figure 4.2: Histogram for Quintile 1 and Quintile 5 average returns

The histograms reveal that data is slightly left skewed. This is mainly due to the presence of several outlying negative returns stretching the curve to the left. The histograms suggest a more rigorous analysis of data to check the satisfaction of normality assumption.

3. Q-Q Plots:

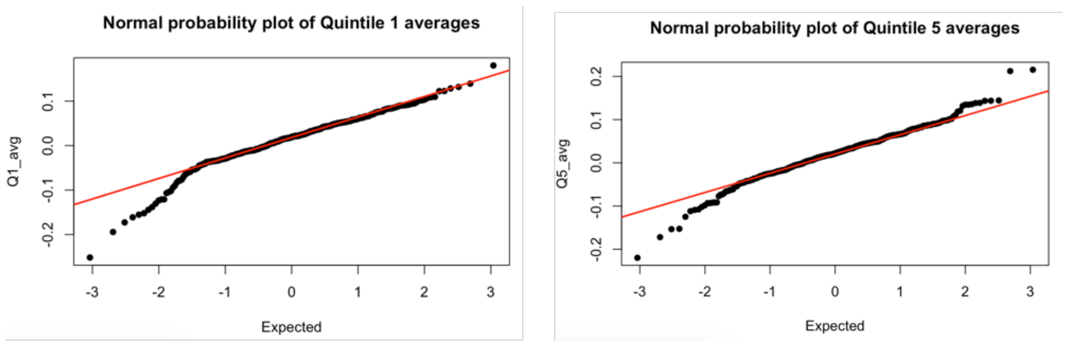

Figure 4.3: Q-Q plots for Quintile 1 and Quintile 5 average returns

Quantile-Quantile (Q-Q) plots are another graphical technique for comparing two probability distributions by plotting the quantiles of data against the quantiles of theoretical distribution. It helps in providing a graphical representation of goodness of fit as against a numerical summary. Q-Q plots are also suggestive of other properties of data such as skewness, kurtosis, outliers etc [15]. For our purpose, we compare whether our data follows a normal distribution or not using a Q-Q plot. If the two distributions are similar, i.e. our data follows a normal distribution, the points on the Q-Q plot would fall around the $y = x$ line. Deviation of data from $y = x$ line suggests a deviation from normality. As we can see in the above Q-Q plots that many data points fall apart from the $y = x$ line suggesting deviations away from normality and the presence of fat tails.

4. Statistical Tests for checking normality:

   (a) Shapiro Wilk Test:

   Hypothesis for Shapiro-Wilk test is as follows:
   H_o: Data is normal
   H_a: Data is not normal

```
> shapiro.test(q1)

        Shapiro-Wilk normality test

data:  q1
W = 0.9517, p-value = 1.742e-10
```

Figure 4.4: Shapiro Wilk Test of Normality for Quintile 1 returns

Conclusion: As p-value is very close to 0, null hypothesis is rejected concluding Quintile 1 average returns data is not normal

```
> shapiro.test(q5)

        Shapiro-Wilk normality test

data:  q5
W = 0.9751, p-value = 1.337e-06
```

Figure 4.5: Shapiro Wilk Test of Normality for Quintile 5 returns

Conclusion: As p-value is very close to 0, null hypothesis is rejected concluding Quintile 5 average returns data is not normal

(b) Andersen Darling Test:

Hypothesis for Andersen Darling test is as follows:
H_o: Data is normal
H_a: Data is not normal

```
> ad.test(q1)

        Anderson-Darling normality test

data:  q1
A = 3.4471, p-value = 1.248e-08
```

Figure 4.6: Andersen Darling Test of Normality for Quintile 1 returns

Conclusion: As p-value is very close to 0, null hypothesis is rejected concluding Quintile 1 average returns data is not normal

```
> ad.test(q5)

        Anderson-Darling normality test

data:  q5
A = 1.8313, p-value = 0.0001097
```

Figure 4.7: Andersen Darling Test of Normality for Quintile 5 returns

Conclusion: As p-value is very close to 0, null hypothesis is rejected concluding Quintile 5 average returns data is not normal

Based on above exploratory analysis and statistical tests, the data was found to be not normal. Hence, non parametric test, Sign Test was performed to check the significance in difference between quintile 1 and quintile 5 average returns.

4.1.4 Research Hypothesis

H_o: Median returns for companies experiencing positive seasonality quarter = Median returns for companies that do not experience seasonality

H_a: Median returns for companies experiencing positive seasonality quarter $\neq$ Median returns for companies that do not experience seasonality

4.1.5 Testing and Results

Paired test: Given that quintile1 and quintile5 data were formed based on a common attribute: earnrank, it therefore seemed most appropriate to use a paired test to test the difference in their returns. Figure 4.8 plots the quintile 1 and quintile 5 average returns month-wise for 420 months. The correlation between their returns is 0.855.

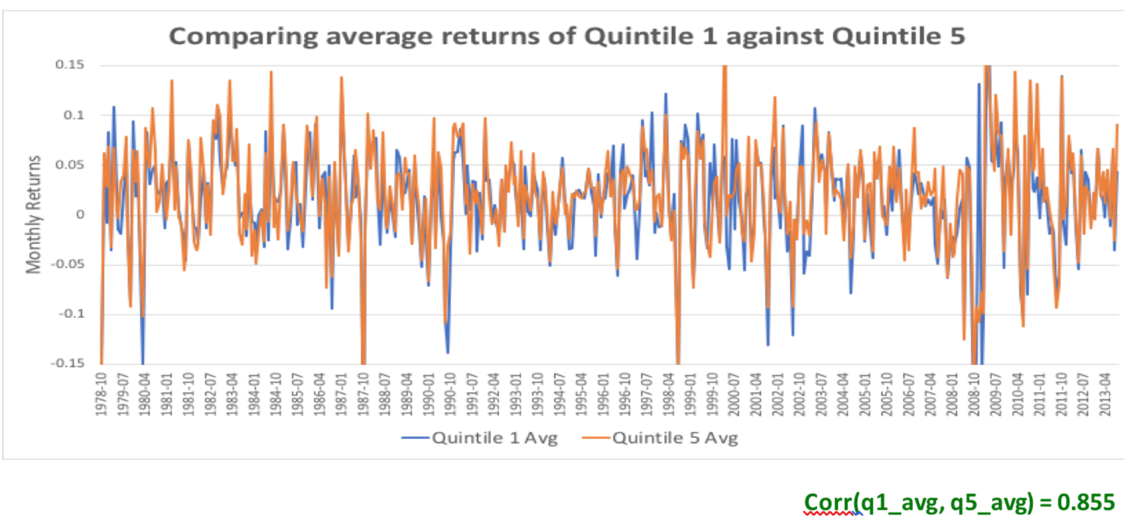

Figure 4.8: Trend analysis of Quintile 1 and Quintile 5 average return

Sign Test, a non-parametric test, which is an alternative to paired t-test for non-normal data, is used for testing whether the difference in returns between Quintile 5 and Quintile1 are significant or not. The result for Sign Test is shown in Table 4.1 and Figure 4.9

Table 4.1: Summary of result: Sign Test

Sign Test Statistic	247
p-value	0.00036
Significant?	Yes

```
> SIGN.test(q5,q1)

        Dependent-samples Sign-Test

data:  q5 and q1
S = 247, p-value = 0.0003565
alternative hypothesis: true median difference is not equal to 0
95 percent confidence interval:
 0.001910401 0.006063465
sample estimates:
median of x-y
  0.004209777

                  Conf.Level L.E.pt U.E.pt
Lower Achieved CI     0.9431 0.0019 0.0060
Interpolated CI       0.9500 0.0019 0.0061
Upper Achieved CI     0.9547 0.0019 0.0061
```

Figure 4.9: Detailed result for Sign Test for testing abonormal returns

The p-value for the Sign Test is found to be 0.00036, which indicates that the difference between Quintile 5 average returns and Quintile 1 average returns are statistically significant at 1% level. The result provides evidence that there exist abnormal returns for the stocks announcing earnings for the positive seasonality quarter. These abnormal returns cannot be explained by risk factors as the volatility of Quintile 5 returns is about the same as Quintile 1 returns as shown in Table 4.2. We will further explore the factors determining abnormal returns in Chapter 5. On average abnormal return of 0.59% and a median abnormal return of 0.42% is generated for the long-short portfolio.

Table 4.2: Summary of results: Mean, Median and Volatility of Abnormal Returns

	Mean	**Median**	**Standard Deviation**
Quintile 5 (Q5)	2.07%	2.16%	5.25%
Quintile 1 (Q1)	1.48%	1.93	5.23%
Abnormal Return (Q5 - Q1)	0.59%	0.42%	2.82%

4.2 Quantifying abnormal returns for different time periods

In order to perform the analysis for abnormal returns across different time periods, the total data for the period October 1972 – September 2013 was divided into three sub-periods as follows:

1. Sub-period 1: 1978 – 1989 (12 years)
2. Sub-period 2: 1990 – 2001 (12 years)
3. Sub-period 3: 2002 – 2013 (12 years)

As the earnrank calculations require a minimum of 5 years of data, we have divided the periods starting from 1978 as that is the first year we start forming the portfolios.

Research hypothesis

H_o: Median returns for the companies experiencing positive seasonality quarter for a given sub-period = Median returns for the companies that do not experience seasonality for the given sub-period

H_a: Median returns for the companies experiencing positive seasonality quarter for the given sub-period $\neq$ Median returns for the companies that do not experience seasonality for the given sub-period

Methodology

In order to test our hypothesis, the data is first divided into three different time periods: 1978 – 1989, 1990 – 2001 and 2002 – 2013. Thereafter, portfolios are formed and hypothesis testing is done for each of the three sub-periods.

For a given sub-period, in order to form a portfolio for the current month, such companies are identified which announced their earnings in the same month in the previous year. 'rdq', which stands for reporting date of quarterly earnings, is used to identify the companies which announced their results in a particular month. In our portfolio, we exclude stocks with a price less than $5 or a missing market capitalization at the end of the previous month. Only those portfolios are included which have at least 10 stocks in them.

After a portfolio is formed, the companies are sorted in an ascending order based on their earnrank and then the portfolio is divided into quintiles. By doing this, the companies exhibiting lowest seasonality move to quintile 1 and the companies exhibiting highest seasonality fall in quintile 5. Average monthly returns of quintile 1 stocks and quintile 5 stocks are calculated for the given sub-period and their difference is taken to find abnormal returns. It is further checked whether abnormal return is statistically significant or not.

Result and Analysis

Figure 4.10 summarizes the results for hypothesis testing for each of the three sub-periods:

Year	Q5_Mean	Q1_Mean	Q5 - Q1 Mean	Normal Assumption	Q5_Median	Q1_Median	Q5 - Q1 Median	Sign Test Statistic	p-value	Significant?
1978 - 1989	2.24%	1.85%	0.40%	No	2.15%	2.16%	0.34%	80	0.0385	Yes
1990 - 2001	2.10%	1.51%	0.59%	No	1.98%	2.03%	0.40%	116	0.0492	Yes
2002 - 2013	2.00%	1.28%	0.72%	No	2.19%	1.61%	0.51%	118	0.0131	Yes
Overall	2.07%	1.48%	0.59%	No	2.16%	1.93%	0.42%	247	0.0004	Yes

Figure 4.10: Hypothesis Testing Results for different Time-Periods

The difference in quintile5 average returns and quintile1 average returns is found to be statistically significant for each of the three sub periods at 5% level. As the data did not meet normal assumption in each of the three sub-periods, Sign Test, which is a non-parametric alternate to paired t-test is used to test the significance in difference. Following figure 4.11 shows the trend of abnormal returns in each of the three sub-periods and compared with the abnormal returns of the overall period. The figure shows that abnormal returns for positive seasonality companies have steadily increased over time.

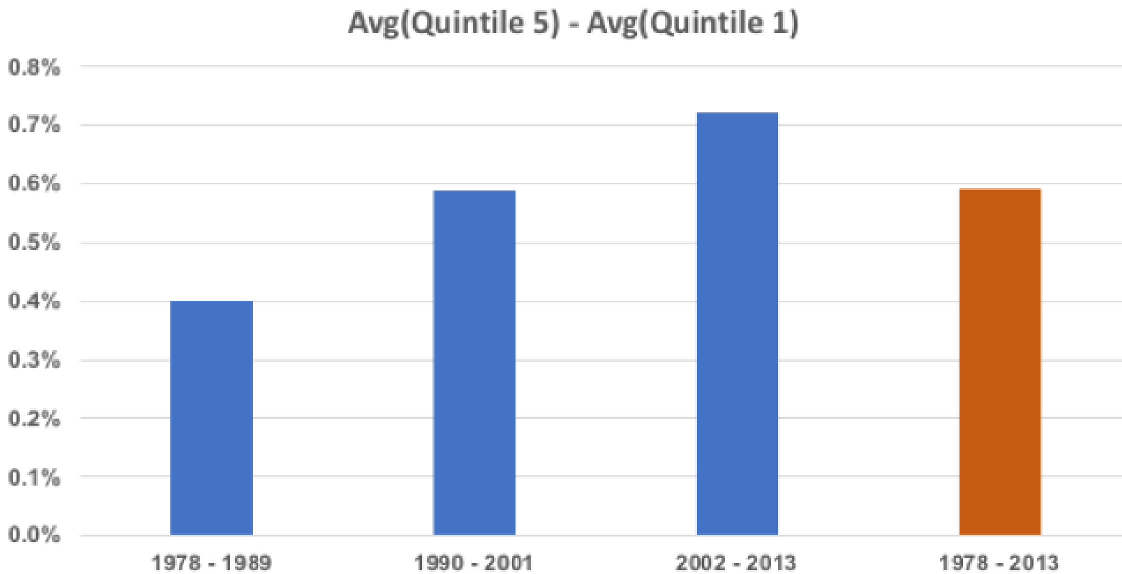

Figure 4.11: Abnormal Returns for different Time-Periods

4.2.1 Quantifying abnormal returns for different industries

In this sub-problem we evaluate how the abnormal returns for positive seasonality quarter vary by industry. The Standard Industrial Classification (SIC) Codes, a four-digit code, is used to classify the companies into major group, industry group and division. The first two digit of the SIC code indicate the major group, the first three indicate the industry group and each division encompasses a range of SIC codes. The companies were classified into different industries based on the industry classifications provided in Table 4.3:

Table 4.3: List of SIC Codes

Range of SIC Codes	**Division**
0100-0999	Agriculture, Forestry and Fishing
1000-1499	Mining
1500-1799	Construction
1800-1999	Not Used
2000-3999	Manufacturing
4000-4999	Transportation, Communications, Electric, Gas and Sanitary service
5000-5199	Wholesale Trade
5200-5999	Retail Trade
6000-6799	Finance, Insurance and Real Estate
7000-8999	Services
9100-9729	Public Administration
9900-9999	Nonclassifiable

Figure 4.12 represents the distribution of our data for the period October 1972 – September 2013 into different industry groups. Manufacturing, Finance and Services are the top three industry groups followed by Transportation, Retail, Mining, Construction and Agriculture.

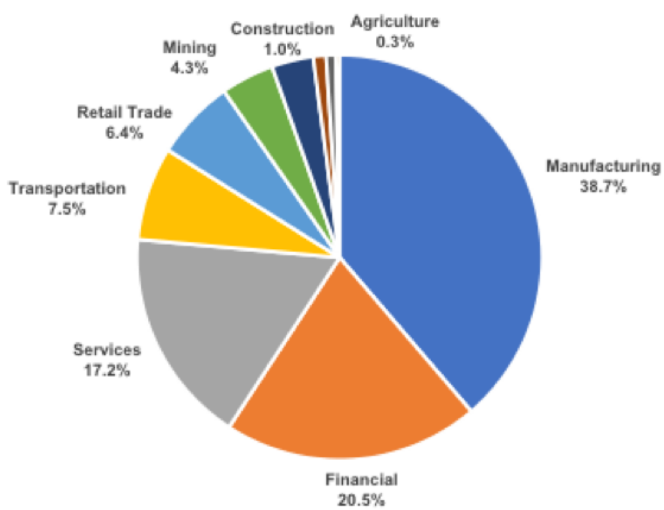

Figure 4.12: Distribution of Companies by Industry

Research hypothesis

H_o: Median returns for the companies experiencing positive seasonality quarter in a given industry group = Median returns for the companies that do not experience seasonality in a given industry group

H_a: Median returns for the companies experiencing positive seasonality quarter in a given industry group $\neq$ Median returns for the companies that do not experience seasonality in a given industry group

Methodology

In order to test our hypothesis, we first divide the data into 10 industry groups using the industry classification code provided in the previous section. We exclude 'Nonclassifiable' and 'Not used' classification codes for our analysis as there is no interpretability value for such codes.

For a particular industry, to form a portfolio for the current month, we first identify the companies which announced their earnings in the same month in the previous year. We used 'rdq' i.e. reporting date of quarterly earnings data from Compustat to identify the companies. In our portfolio, we exclude stocks with a price less than $5 or a missing market capitalization at the end of the previous month. We exclude the portfolios which have less than 10 stocks from our analysis. Based on industry data, different number of portfolios were formed for each industry.

Once we have formed the portfolio, we sort the companies by their earnrank in an ascending order and divide the portfolio into quintiles based on earnrank. By doing this, the companies exhibiting lowest seasonality move to quintile 1 and the companies exhibiting highest seasonality fall in quintile 5. We calculate the average monthly returns of quintile 1 stocks and quintile 5 stocks and take their difference to test whether the difference is statistically significant or not for a particular industry.

Results and Analysis

Figure 4.13 summarizes the results for hypothesis testing for each of the industry groups:

Industry	No. of Companies	#Months	Q5_Mean	Q1_Mean	Q5 - Q1 Mean	Normal Assumption	Paired t-stat	p_value	Significant?	Q5_Median	Q1_Median	Q5 - Q1 Median	Sign Test Statistic	p-value	Significant?
Construction	106	114	0.87%	3.12%	-2.25%	Yes	-2.049	0.0427	Yes	1.58%	1.92%	-1.11%	54	0.6398	No
Financial	2148	281	1.67%	1.16%	0.51%	No	NA	NA	NA	1.45%	1.37%	0.28%	159	0.0316	Yes
Manufacturing	4064	420	1.81%	1.55%	0.26%	No	NA	NA	NA	2.04%	1.70%	0.50%	238	0.0072	Yes
Mining	446	279	1.05%	1.52%	-0.48%	No	NA	NA	NA	1.17%	1.48%	-0.06%	138	0.9047	No
Retail Trade	675	420	2.47%	1.75%	0.72%	No	NA	NA	NA	2.43%	1.72%	0.55%	229	0.0709	No
Services	1808	414	1.99%	1.71%	0.28%	No	NA	NA	NA	2.06%	2.23%	0.43%	216	0.4035	No
Transportation	784	280	1.35%	1.00%	0.35%	No	NA	NA	NA	1.38%	1.44%	0.25%	155	0.0829	No
Wholesale Trade	349	271	1.51%	1.36%	0.16%	No	NA	NA	NA	2.01%	1.03%	0.56%	145	0.2742	No
Overall	**10493**	**420**	**2.07%**	**1.48%**	**0.59%**	No	NA	NA	NA	2.16%	1.93%	0.42%	247	0.0004	Yes

Figure 4.13: Hypothesis Testing Results for different Industries

The hypothesis testing is done for only 8 out of 10 industries. Industry groups – Agriculture, Forestry and Fishing and Public Administrations were excluded as no portfolios were formed for these groups because they did not meet the criteria of at least 10 companies in a portfolio for any month in the given period.

For each of the industry groups, we checked the normality assumption to help us choose the appropriate test. For the industry where data met the normality assumption paired t-test was used to check significance and where the data did not meet the normality assumption, a non-parametric test i.e. Sign Test was used to perform the hypothesis testing. Of all the industry groups, only Construction met the normality assumption for which paired t-test was done to check significance in difference. For all other industry groups, which did not meet the normality assumption, Sign Test was done to test the abnormal returns.

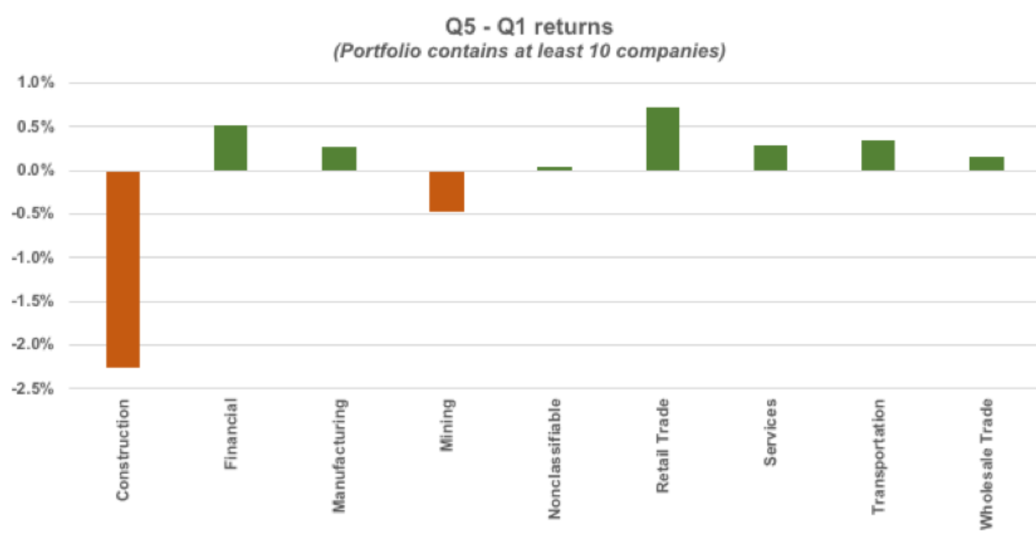

Figure 4.14: Abnormal Returns for different Industry Groups

The abnormal returns are found to be significant for three industry groups: Construction, Finance and Manufacturing. One puzzling result is that the abnormal return is negative for Construction sector as seen in Figure 4.14, implying average returns for positive seasonality stocks is lower as compared to average returns of stocks that do not exhibit seasonality. Myers and Swerdloff (1967) explains that seasonality in Construction sector is caused by slowdown in construction activity in the winter months rendering large population of construction workers unemployed [17]. The way we have conceptualized seasonality, we assume predictably higher earnings in a particular quarter as compared to other quarters in a calendar year. The subtle difference in the way we define seasonality and the manner it manifests in the Construction sector could be the reason behind the puzzling result of negative abnormal return. Another potential reason can be the number of portfolio months. For Construction, we were able to form the portfolio for only 114 months as opposed to 420 in Retail. This suggests that the data points may not be sufficient to represent the overall picture of the industry and more robust analysis can be performed to detect abnormal returns in the Construction sector.

Retail Trade is found to have the maximum abnormal returns, which is significant at 10% level. The sample size for the retail sector was relatively small: there were only 675 companies in Retail sector as compared to 2148 companies in Financial and 4064 companies in Manufacturing sector. However, Retail Trade appears to be an important sector driving abnormal returns and there is scope of further analysis to identify abnormal returns in this sector.

4.3 Comparing abnormal returns between industry groups

At this point it would be interesting to examine whether there is a significant difference in abnormal returns between different industry groups. Various exploratory analyses were performed to check the heteroscedasticity and distribution of data. Further, different tests were conducted in order to identify the appropriate statistical test that can be used to perform the comparative analysis. A parametric test such as Analysis of Variance (ANOVA), which is used to test the difference in means in more than one group, assumes that the data for each group is normally distributed and variance of data is equal across different groups. If our industry data satisfies this assumption, we will use ANOVA to check the difference in means else we will evaluate certain non-parametric techniques to perform our analysis.

4.3.1 Exploratory Analysis for Industry Groups

1. Testing homogeneous variance

	Construction	Finance	Manufacturing	Mining	Retail	Services	Transportation	Wholesale Trade
Mean	-2.25%	0.51%	0.26%	-0.48%	0.72%	0.28%	0.35%	0.16%
Variance	1.38%	0.23%	0.10%	0.49%	0.53%	0.45%	0.16%	0.60%

Figure 4.15: Mean and Variance of Abnormal Returns for different Industry Groups

Mean and Variance for the different industry groups can be seen in the Figure 4.15. In order to test the homogeneity of variance between the industry groups, a statistical test called Levene Test is performed. The hypothesis for the test is as follows:

H_o: Variance is homogeneous between groups H_a: Variance is not homogeneous

```
Levene's Test for Homogeneity of Variance (center = median)
        Df F value    Pr(>F)
group    7  51.562 < 2.2e-16 ***
      2471
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 4.16: Results for Levene Test for Homogeneous Variance between Industry Groups

As the p-value is close to zero, we reject the null-hypothesis and conclude that the variance between the groups is not homogeneous.

2. Box Plots

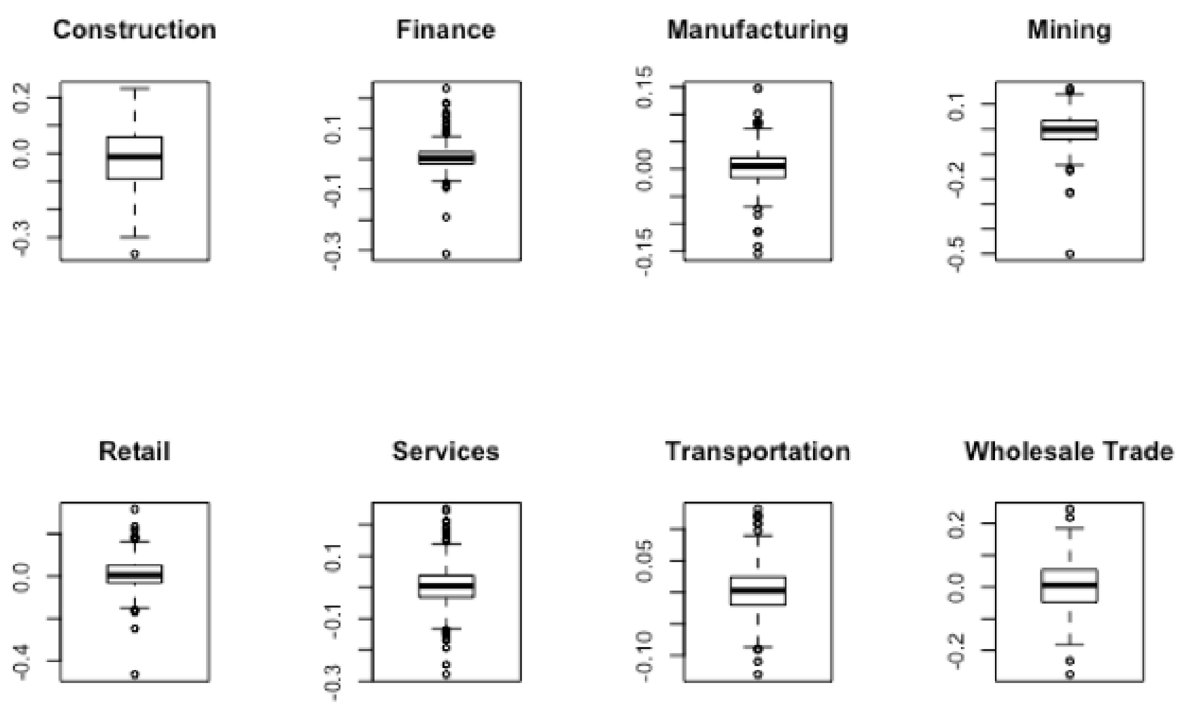

Figure 4.17: Box plots for different Industry Groups

Presence of outliers in boxplots for some of the industry groups question the normality assumption of data.

3. Histograms

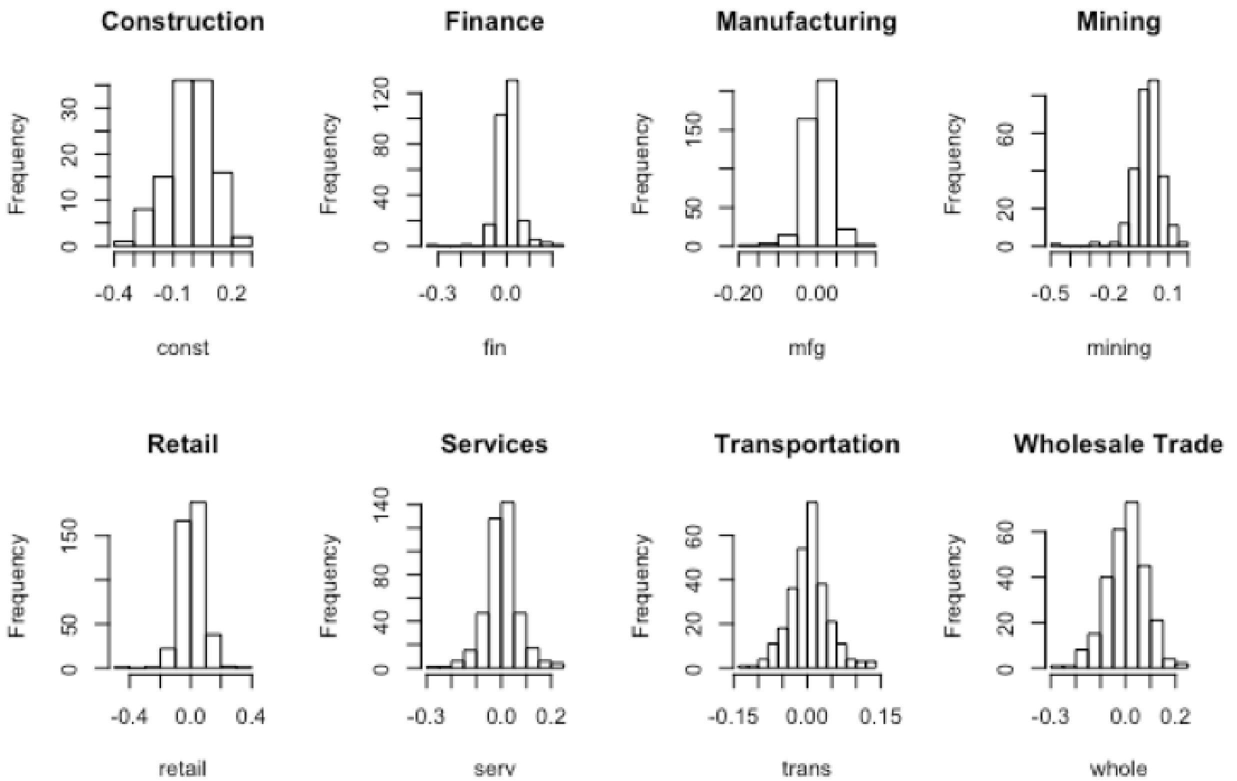

Figure 4.18: Histogram for different Industry Groups

The histograms for the different industry groups are approximately bell-shaped. However, skewness can be seen in certain industry groups.

4. Q-Q plots

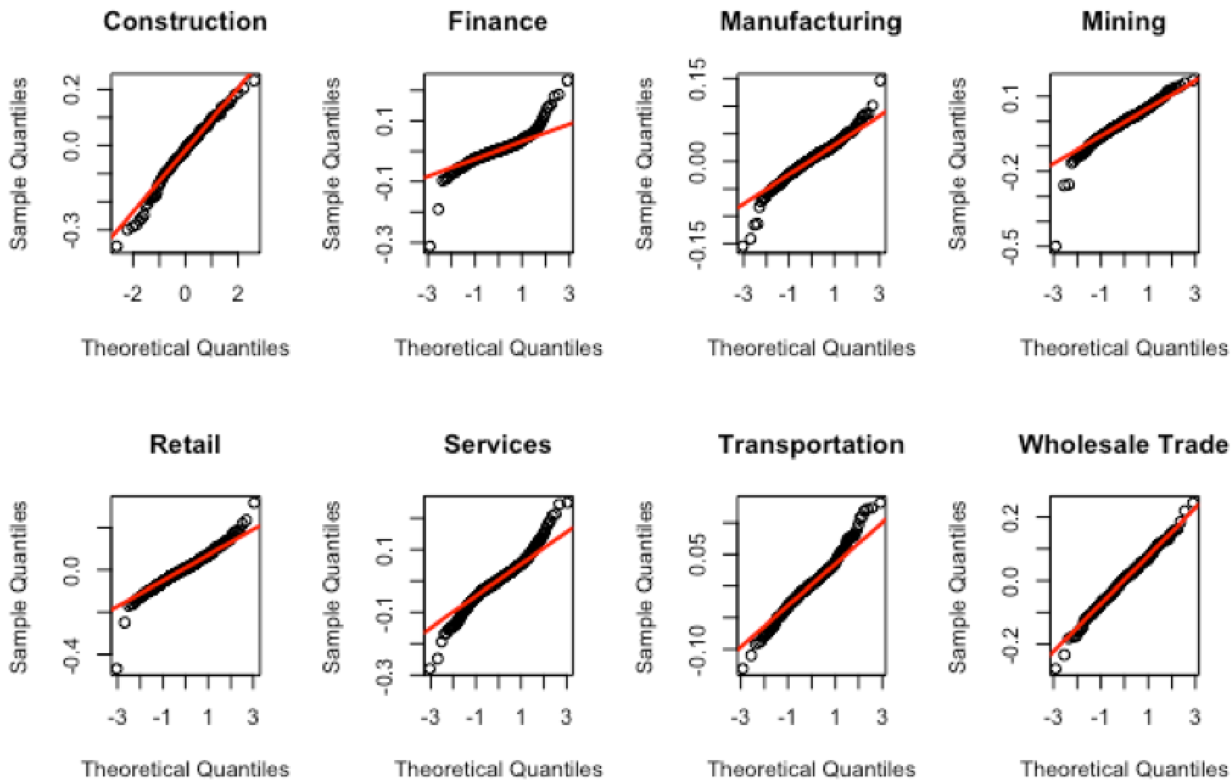

Figure 4.19: Q-Q plots for different Industry Groups

Q-Q plots for many industry groups are found to deviate from the y=x line, which signals data distribution away from normality for such industries.

5. Testing for Normality

   Shapiro Wilk Test is done to test the normality in data in each of the industry groups, the results of which is shown in figure 4.20. Only Construction and Wholesale trade data are found to be normally distributed. The data for the remaining industry groups fail the normality assumption as tested by Shapiro-Wilk test. This deviation from normality is supported by the exploratory analysis using boxplots, histograms and Q-Q plots.

```
> shapiro.test(construction) #yes

        Shapiro-Wilk normality test

data:  construction
W = 0.9829, p-value = 0.1558
```

```
> shapiro.test(finance) #no

        Shapiro-Wilk normality test

data:  finance
W = 0.8558, p-value = 1.654e-15
```

```
> shapiro.test(manufacturing) #no

        Shapiro-Wilk normality test

data:  manufacturing
W = 0.962, p-value = 5.961e-09
```

```
> shapiro.test(mining) # no

        Shapiro-Wilk normality test

data:  mining
W = 0.9159, p-value = 2.06e-11
```

```
> shapiro.test(retail) #no

        Shapiro-Wilk normality test

data:  retail
W = 0.9599, p-value = 2.795e-09
```

```
> shapiro.test(services) #no

        Shapiro-Wilk normality test

data:  services
W = 0.9687, p-value = 9.659e-08
```

```
> shapiro.test(transportation) #no

        Shapiro-Wilk normality test

data:  transportation
W = 0.9839, p-value = 0.003041
```

```
> shapiro.test(wholesale) #yes

        Shapiro-Wilk normality test

data:  wholesale
W = 0.9949, p-value = 0.5113
```

Figure 4.20: Shapiro Wilk Test for Normality Results for different Industry Groups

4.3.2 Statistical Analysis and Results

In order to find a suitable test to test whether the difference in abnormal returns between the industry groups is significant or not, we perform a comparative analysis of different parametric and non-parametric methods that are used to test equality of means in more than two samples. The comparative analysis is presented in Figure 4.21 [23]

ANOVA	Welch ANOVA	Kruskal Wallis Test
Normal Distribution	Normal Distribution	All groups should have the same shape distributions
Equal Variance	Unequal Variance	No variance assumption
H_0: $\mu_1 = \mu_2 = \mu_3 \ldots = \mu_k$ H_1: Means are not all equal. where k = the number of independent comparison groups.	H_0: $\mu_1 = \mu_2 = \mu_3 \ldots = \mu_k$ H_1: Means are not all equal. where k = the number of independent comparison groups.	H_0: population medians are equal / Samples (groups) are from identical populations. H_1: population medians are not equal / At least one of the samples (groups) comes from a different population than the others.

Figure 4.21: Comparison of Different Statistical Tests for Equality of Means in more than two Samples

As we do not meet the homogenous variance and normality assumption, ANOVA is ruled out. Data for only two industries, Construction and Wholesale trade follow normal distribution. Other industries do not satisfy the normality assumption therefore Welch ANOVA is also ruled out.

Kruskal Wallis, a non-parametric test, is found to be most suitable in this scenario as it does not require normality and equal variance assumption [29]. Based on this, we formulate the research hypothesis as follows:

Research Hypothesis

H_o: Industry medians for abnormal returns are same
H_a: At least one industry median for abnormal return is different

Test results:

```
	Kruskal-Wallis rank sum test

data:  values by groups
Kruskal-Wallis chi-squared = 7.2717, df = 7, p-value = 0.4012
```

Figure 4.22: Kruskal Wallis Test for testing equality of medians between industry groups

As the p-value for the test is much greater than 0.05, the test does not indicate significant difference between any two industry groups.

4.4 Conclusion

Abnormal returns are found to be significant for the stocks exhibiting positive seasonality. This phenomenon is found to be consistent when considered for different time periods. The rate of abnormal returns increased over time. In Construction, Manufacturing and Financial sectors abnormal returns due to seasonality factors were found to be significant at 5% level and in Retail sector, results were significant at 10% level. In Construction, however, abnormal returns were negative, which is in contrast to our hypothesis, and requires further attention. Lastly, abnormal returns between the industry groups were not found to be significant.

Chapter 5

Factors determining abnormal returns

In this chapter we examine if abnormal returns due to seasonality in earnings can be explained through the standard risk factors. We take the portfolios sorted into quintiles based on their earnrank and calculate abnormal return by taking the difference between quintile 5 and quintile 1 average returns as the dependent variable. For the independent variables, we use four factors proposed by Fama and French (1993) and Carhart (1997) [5, 9]. The four factors are: excess returns on the market, Small minus Big (SMB), High minus Low (HML) and Up minus Down (UMD). In addition to identifying factors determining abnormal returns, in this chapter, we also assess the returns of the positive seasonality portfolio relative to four-factor model and the returns of the least seasonality portfolio relative to four-factor model.

5.1 Regression Analysis

5.1.1 Abnormal Returns on Four-Factors

We perform regression analysis to identify the factors determining abnormal returns. For this purpose, we regress abnormal returns on 4-factor model variables. The dependent and independent variables for this regression are given below:

Dependent variable

- Abnormal return

  Abnormal return = Average return of Quintile 5 – Average return of Quintile 1 portfolio

Independent variables

- Market excess returns
- Small minus big (Size)
- High minus low (Value)
- Up minus Down (Momentum)

n = 420 (No. of portfolios formed in the period 1972-2013). The risk factors were downloaded from Ken French website [16]

```
lm(formula = abnormal_return ~ mkt_excess_return + SMB + HML +
    UMD)

Residuals:
      Min        1Q    Median        3Q       Max
-0.239883 -0.014347 -0.002398  0.012314  0.121908

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.006192   0.001414   4.379 1.51e-05 ***
mkt_excess_return  0.016386   0.032096   0.511 0.609951
SMB               -0.176878   0.046942  -3.768 0.000188 ***
HML               -0.035053   0.050522  -0.694 0.488186
UMD                0.002110   0.030702   0.069 0.945240
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02787 on 415 degrees of freedom
Multiple R-squared:  0.03337,   Adjusted R-squared:  0.02405
F-statistic: 3.581 on 4 and 415 DF,  p-value: 0.006928
```

Figure 5.1: Regression results for Abnormal Returns on 4-factors

For this regression, only SMB is found to be significant. Other variables are not significant. Abnormal returns, can therefore be explained by size effect to some extent. The R-square value for this result is 0.03. The negative coefficient of SMB factor variable indicates that the abnormal returns are more for large firms.The long-short portfolio is found to have an alpha of 62 basis points per month as measured from intercept value, which is significant.

5.1.2 Quintile 5 Excess Returns on Four-Factors

Dependent variable

- Quintile 5 excess returns

  Quintile 5 excess returns = Quintile 5 Average Returns – Risk free rate

Independent variables

- Market excess returns
- Small minus big (Size)
- High minus low (Value)
- Up minus Down (Momentum)

```
lm(formula = quintile5_excess_returns ~ mkt_excess_return + SMB +
    HML + UMD)

Residuals:
      Min        1Q    Median        3Q       Max
-0.150260 -0.014239 -0.002718  0.010335  0.188322

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.009422   0.001276   7.384 8.50e-13 ***
mkt_excess_return  0.961124   0.028959  33.189  < 2e-16 ***
SMB                0.391302   0.042355   9.239  < 2e-16 ***
HML                0.338035   0.045586   7.415 6.89e-13 ***
UMD               -0.058857   0.027702  -2.125   0.0342 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02515 on 415 degrees of freedom
Multiple R-squared:  0.7743,    Adjusted R-squared:  0.7721
F-statistic: 355.9 on 4 and 415 DF,  p-value: < 2.2e-16
```

Figure 5.2: Regression results for Quintile 5 Excess Returns on 4-factors

All the four factors market excess returns, SMB, HML and UMD are significant at 5% level. The highest seasonality quintile portfolio has an alpha of 94 basis points per month as indicated from significant intercept value. Therefore, we conclude that Market risk, Size, Value and Momentum explain 77% (R-square) of the variation in the positive seasonality portfolio returns.

5.1.3 Quintile 1 Excess Returns on Four-Factors

Dependent variable

- Quintile1 excess returns

  Quintile 1 excess returns = Quintile 1 Average Returns – Risk free rate

Independent variables

- Market excess returns
- Small minus big (Size)
- High minus low (Value)
- Up minus Down (Momentum)

```
lm(formula = quintile1_exceess_returns ~ mkt_excess_return +
    SMB + HML + UMD)

Residuals:
      Min        1Q    Median        3Q       Max
-0.077972 -0.010818 -0.001074  0.011850  0.089596

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.003226   0.001070   3.015  0.00273 **
mkt_excess_return  0.944788   0.024281  38.910  < 2e-16 ***
SMB                0.568229   0.035513  16.001  < 2e-16 ***
HML                0.373048   0.038221   9.760  < 2e-16 ***
UMD               -0.060990   0.023227  -2.626  0.00896 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02109 on 415 degrees of freedom
Multiple R-squared:  0.8402,    Adjusted R-squared:  0.8386
F-statistic: 545.3 on 4 and 415 DF,  p-value: < 2.2e-16
```

Figure 5.3: Regression results for Quintile 1 Excess Returns on 4-factors

All the four factors: market excess returns, SMB, HML and UMD are significant at 1% level. The lowest seasonality quintile portfolio has an alpha of 32 basis points per month as compared to 94 basis points for the highest seasonality quintile portfolio. Therefore, we conclude that Market risk, Size, Value and Momentum factors explain 84% (R-square) of the variation in the lowest seasonality portfolio returns.

5.2 Conclusion

Regression results based on Fama four factor model showed that abnormal returns can be explained by size factor to some extent where return increased with size. Other than this, the highest seasonality quintile portfolio had an alpha of 94 bps per month as compared to an alpha of 32 bps per month for the lowest seasonality portfolio. The long-short portfolio was found to have an alpha of 62 bps per month.

Chapter 6

Conclusion

In this Master's thesis, we tested the hypothesis whether seasonality can explain abnormal returns in stocks. We found strong evidence that abnormal returns existed for firms announcing earnings for the positive seasonality quarter. We used *earnrank* as the seasonality measure as *earnrank* was an easy to construct and simple to understand measure, which made minimal assumptions unlike traditional X-12 seasonality measure based on time series models. *Earnrank* was also least affected by negative earnings, growth in earnings and outliers in data. Portfolios sorted by earnrank was divided into quintiles, where quintile 5 contained the most seasonal companies and quintile 1, the least seasonal companies. Difference between quintile 5 and quintile 1 returns was used to calculate abnormal returns. These abnormal returns were found to be significant at 1% level

The hypothesis for abnormal returns associated with seasonality was tested for different time periods and for different industry groups. For time-wise analysis, overall data was divided into three equal periods. Abnormal returns were found to be significant in each of the time periods and they increased over time.

For the industry analysis, abnormal returns were found to be significant for three sectors: Construction, Manufacturing and Financial at 5% level and for Retail Trade at 10% level. In the Construction sector, abnormal returns were found to be negative, which is in contrast to our hypothesis and requires further attention. This opens an area of potential future research.

In general, seasonal shift in earnings were found to be more common for large firms, value firms, old firms, firms with lower turnover and firms with lower accruals.

Lastly, we also examined factors determining abnormal returns relative to standard risk factors proposed by Fama four factor model and found SMB to be a significant variable in explaining abnormal returns. The long-short portfolio based on seasonality generated an alpha of 62 basis points per month.

In conclusion, the result that abnormal returns can be explained by seasonality, which is a persistent and recurring firm event is particularly interesting as it represents a market anomaly that challenges the Efficient Market Theory.

Bibliography

[1] Tom Baum and Svend Lundtorp. *Seasonality in tourism*. Elsevier, 2001.

[2] J Joseph Beaulieu and Jeffrey A Miron. The seasonal cycle in us manufacturing. Technical report, National Bureau of Economic Research, 1990.

[3] Robert Bennin. Error rates in crsp and compustat: A second look. *The Journal of Finance*, 35(5):1267–1271, 1980.

[4] Bishop, Tylor. Why ad earnings are lower in january for most publishers.

[5] Mark M Carhart. On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82, 1997.

[6] Tom Y. Chang, Samuel M. Hartzmark, David H. Solomon, and Eugene F. Soltes. Being surprised by the unsurprising: Earnings seasonality and stock returns. *The Review of Financial Studies*, 30(1):281–323, 2017.

[7] Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970.

[8] Eugene F Fama, Lawrence Fisher, Michael C Jensen, and Richard Roll. The adjustment of stock prices to new information. *International economic review*, 10(1):1–21, 1969.

[9] Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

[10] David F Findley, Brian C Monsell, William R Bell, Mark C Otto, and Bor-Chung Chen. New capabilities and methods of the x-12-arima seasonal-adjustment program. *Journal of Business & Economic Statistics*, 16(2):127–152, 1998.

[11] Mark S Grinblatt, Ronald W Masulis, and Sheridan Titman. The valuation effects of stock splits and stock dividends. *Journal of financial economics*, 13(4):461–490, 1984.

[12] Mark Haug and Mark Hirschey. The january effect. *Financial Analysts Journal*, 62(5):78–88, 2006.

[13] Steven L Heston and Ronnie Sadka. Seasonality in the cross-section of stock returns. *Journal of Financial Economics*, 87(2):418–445, 2008.

[14] Investopedia. Is the banking sector subject to any seasonal trends?

[15] Chambers John, William Cleveland, Beat Kleiner, and Paul Tukey. Graphical methods for data analysis. *Wadsworth, Ohio*, pages 128–129, 1983.

[16] Kenneth R French. U.s. research returns data.

[17] Robert J Myers and Sol Swerdloff. Seasonality and construction. *Monthly Lab. Rev.*, 90:1, 1967.

[18] University of Chicago. Center for Research in Security Prices, 1994.

[19] Quarter Lab. Advertising seasonality: Earn more money on youtube.

[20] Richard Roll. Vas ist das? *Journal of Portfolio management*, 9(2):18–28, 1983.

[21] Barr Rosenberg and Michel Houglet. Error rates in crsp and compustat data bases and their implications. *The Journal of finance*, 29(4):1303–1310, 1974.

[22] Michael S Rozeff and William R Kinney Jr. Capital market seasonality: The case of stock returns. *Journal of financial economics*, 3(4):379–402, 1976.

[23] Kristin L Sainani. Dealing with non-normal data. *PM&R*, 4(12):1001–1005, 2012.

[24] Gerald L Salamon and Thomas L Stober. Cross-quarter differences in stock price responses to earnings announcements: Fourth-quarter and seasonality influences. *Contemporary Accounting Research*, 11(1):297–330, 1994.

[25] Paul A Samuelson. Proof that properly anticipated prices fluctuate randomly. *IMR; Industrial Management Review (pre-1986)*, 6(2):41, 1965.

[26] S&P Global Market Intelligence. Fundamental Data, 2018.

[27] Michaela Spitzer, Jan Wildenhain, Juri Rappsilber, and Mike Tyers. Boxplotr: a web tool for generation of box plots. *Nature methods*, 11(2):121, 2014.

[28] Richard H Thaler. Anomalies: the january effect. *Journal of Economic Perspectives*, 1(1):197–201, 1987.

[29] Andrew J Tomarken and Ronald C Serlin. Comparison of anova alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin*, 99(1):90, 1986.

[30] John W Tukey. *Exploratory data analysis*, volume 2. Reading, Mass., 1977.