The Low Volatility Anomaly Put To The Test

Erik Mulder

432751

The volatility anomaly is present in the time period 1963-2019 and robust over time and robust to methodological choices and assumptions. Using a methodology with value-weighted returns and NYSE breakpoints to form portfolios, the low volatility portfolio performs equally well in terms of return and outperforms the high volatility portfolio in terms of Sharpe Ratio. This result shows up in all size groups except for the biggest 20% of stocks, where the low volatility portfolio underperforms in terms of returns but still outperforms in terms of Sharpe Ratio. The anomaly has gotten stronger over time, with the anomaly first appearing in the 1980s. Although certain methodological choices affect the results, the anomaly is still present in different settings. The anomaly is robust to choice of portfolio size (10 or 5 portfolios), is present whether the returns are winsorized or not, becomes stronger after including penny stocks but is also very much present without these penny stocks, is robust to choice of volatility measure (past 3-year volatility or IVOL) and is a distinct effect from Beta.

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Financial Economics

Supervisor: P.J.P.M. Versijp

Second assessor: J.J.G. Lemmen

Date final version: November 15th 2020

Contents

1 Introduction

High risk, higher return. That statement is one that most people will know about and believe in. But is this statement true? This paper aims to answer that question. One way to measure risk is to use the volatility of returns of a stock or portfolio of stocks. And the literature has shown that low volatility stocks can outperform high volatility stocks. This phenomenon is called the Low Volatility anomaly. There has been some critique on the used methodologies however, so this paper compares results of different methodologies to see if the anomaly is a result of one particular chosen methodology, or whether it is robust to certain methodological choices. Next to that, the body of literature on the anomaly over time is small. This paper aims to contribute to that. Understanding the behavior of the anomaly over time and using different methodologies will help the overall understanding of the anomaly and can help investors make a more informed decision about the construction of their portfolios. Also, an updated data set is used with monthly data from 1963 to 2019, analyzing the anomaly in recent time periods in the US. Broadly speaking, the paper will first show that the anomaly exists in the data set, then analyze separate time periods and finally compare the results using different methodologies. The paper contributes to the literature in 3 ways. First, new evidence is presented that the anomaly exists within the latest data set. From 1963 to 2019, the anomaly exists in the US market. Second, the strength of the anomaly is analyzed over time. Third, the paper tries to refute criticism on the anomaly by showing that the anomaly is robust to certain methodological choices, and not limited to small stocks only.

Together, this will answer the research question: Does the anomaly survive after taking into account different methodologies and in all time periods?

The paper proceeds as follows. Section 2 summarizes the literature on the anomaly. Section 3 introduces the Data and Section 4 covers the Methodology. Section 5 presents the main results. Section 6 shows the robustness tests. Section 7 concludes.

2 Literature

When the CAPM was introduced by Sharpe (1964) and Lintner (1965), beta was widely interpreted as the measure of risk. The higher a portfolio's beta, the higher the risk but also the return. Soon after, evidence that this might not be the case was found. For example, Black (1972) and Fama & MacBeth (1973) both found evidence that high beta stocks earn too little compared to low-beta stocks. This was confirmed in further studies like Black, Jensen & Scholes (1972) and Miller & Scholes (1972). Haugen & Heins (1975) then drew the conclusion that portfolios with less variance have greater average returns than the riskier portfolios.

The CAPM was further challenged when Banz (1981) found that small stocks outperform larger stocks. This was extended by Fama & French (1992), who found that when taking into account size, market beta is unpriced in the cross-section of stock returns.

Then, Blitz & van Vliet (2007) provided a fresh look at this phenomenon and found that when sorting stocks on their past volatility, low volatility portfolios outperformed high volatility portfolios by a wide margin. This finding is not limited to the US market, but also present in international markets. These findings were confirmed in multiple studies such as Baker, Bradley & Wurgler (2011), Baker & Haugen (2012) and Frazzini & Pedersen (2014). However, various ways of measuring risk or volatility have been used throughout these studies.

For example, Blitz & van Vliet (2007) construct decile portfolios by ranking stocks on the past 3-year volatility of weekly returns. Baker, Bradley & Wurgler (2011) sort stocks based on their 5-year trailing total volatility or trailing beta. Baker & Haugen (2012) compute the volatility of total return for each company in each country over the previous 24 months, which they grouped into deciles, quintiles and halves.

Another popular way of measuring volatility is the IVOL measure introduced in Ang, Hodrick, Xing & Zhang (2006), who use daily return data from the past month to estimate idiosyncratic volatility (the IVOL effect). Fama & French (2016) then used this IVOL measure to test their 5-factor Model from Fama & French (2015). This methodology for IVOL was also used in Hsu & Chen (2017).

Beta is also still used as a measure of risk. For example, Garcia, Kochard, Sullivan & Wang (2015) specifically use beta as the risk measure to talk about the low-volatility anomaly. Frazzini & Pedersen (2014) use a slightly modified approach with the Betting-Against-Beta factor (BAB), which they define as a portfolio that holds low-beta assets, leveraged to a beta of one, and shorts high-beta assets, de-leveraged to a beta of one, creating a market neutral factor. However, Liu, Stambaugh & Yuan (2018) compare the Betting-Against Beta factor as constructed in Frazzini & Pedersen (2014) with the IVOL factor as mentioned before, from Ang et al. (2006) and find that beta shows slightly weaker results. Specifically, they show that the beta anomaly arises from beta's positive correlation with idiosyncratic volatility (IVOL). When controlling for either IVOL or simply excluding overpriced stocks with high IVOL, the beta anomaly becomes insignificant. This issue was then addressed by Asness, Frazzini, Gormsen & Pedersen (2020), by decomposing the BAB factor into two factors; betting against volatility (BAV) and betting against correlation (BAC). The theory behind lies in the definition of beta: the beta of a stock to the market index is equal to its volatility times its correlation with the market index, divided by the volatility of the market (Blitz, van Vliet & Baltussen (2019a). The results of Asness et al. (2020) are that volatility is the main driver of the low-risk effect, and that correlation only matters within buckets of stocks with similar levels of volatility. To calculate idiosyncratic volatility, they refer to Ang et al. (2006), and take the monthly residual volatility in the Fama-French 3-factor Model.

Next to the various ways to measure volatility and define it, there has also been critique. Interestingly, Bali & Cakici (2008) find contrasting results. They argue that the chosen methodology explains the results found in the literature, concluding that 'analyses based on two different measures of idiosyncratic volatility (estimated using daily and monthly data), three weighting schemes (value-weighted, equal-weighted and inverse-volatility weighted), three breakpoints (CRSP, NYSE , equal market share) and two different samples (NYSE/AMEX/NASDAQ and NYSE) indicate that no robustly significant relation exists between idiosyncratic volatility and expected returns'. They first replicate the methodology of Ang et al. (2006) and find similar results. They then change the methodology by using equal-weights instead of value-weights and find a small positive relationship return for the low-high idiosyncratic difference portfolio (although statistically insignificant). Weighting stocks by their inverse-volatility also gives results contrary to the findings of Ang et al. (2006). The results indicate that the weighting scheme used to compute average portfolio returns affects the cross-sectional relation between idiosyncratic risk and expected returns. This paper aims to extensively analyze the anomaly under different methodologies and tries to find out whether the criticism of Bali & Cakici (2008) is justified, and to what extent.

When analyzing the stocks inside the quintile portfolios of Ang et al. (2006) using CRSP breakpoints, they find that when portfolios are formed based on the CRSP breakpoints, quintile 1 (low idiosyncratic risk) contains mostly large stocks whereas quintile 5 (high idiosyncratic risk) contains much smaller stocks. Using NYSE breakpoints following Fama & French (1992) to generate quintiles with a more balanced average market share, they find the return differential between the low and high idiosyncratic risk portfolios to be very low and statistically insignificant, for the value-weighted, equal-weighted, and inverse-volatility weighting schemes. Thus, portfolios formed on the basis of NYSE breakpoints provide no evidence for the IVOL anomaly. Another important conclusion they draw is that small and illiquid stocks drive the results of Ang et al. (2006), which becomes apparent after excluding those stocks and seeing the results disappear.

Another point of critique came from Novy-Marx (2014), who criticizes the low-volatility anomaly by arguing that size is the main driver of the anomaly and that after controlling for size, profitability is the driver of the returns. This study estimates beta and volatility each month with the use of daily data. Quintiles are formed each month using NYSE breakpoints. Portfolio returns are value-weighted, and ignore transaction costs. Following Baker, Bradley & Wurgler (2011), the sample begins in January 1968. This start date coincides with the date at which high-volatility stocks began to underperform, thus creating a bias toward finding impressive results in favor of the low-volatility anomaly. Which break-points to use is also important for the results. Baker, Bradley & Wurgler (2008) use CRSP breaks, but this yields extreme tilts, with an end of the sample average market cap in the high volatility quintile of only 33 million, and the entire quintile making up on average less than 1.2% of total market capitalization. It thus creates a strong size bias in the volatility portfolios. Novy-Marx therefore uses NYSE breaks. See Bali & Cakici (2008) for an analysis of the differences in volatility strategy performance using CRSP and NYSE breaks.

In Fama & French (2015), the profitability factor (and an investment factor) is added to their three-factor Model (1993), to make it a five-factor Model. After Novy-Marx (2014) argued that profitability was able to explain the low-volatility effect, Fama & French (2016) tested this as well using their five-factor Model. Here, they argue that the low average returns associated with high beta, and high return volatility that are left unexplained by the three-factor Model are substantially captured by negative five-factor exposures to RMW and CMA, typical of less profitable firms that invest aggressively.

Blitz & Vidojevic (2017) acknowledge that the low-risk effect appears to be subsumed by the profitability factor in time-series regressions. However, based on a Fama-Macbeth test (1973) they conclude that it is premature to assume that the low-risk effect is explained by profitability or other factors. Additionally, Blitz, Baltussen & van Vliet (2019b) find that while the short side of a low-risk strategy (high risk) can be explained by the short side of the new Fama-French factors (poor profitability), the long side is not. In other words, they argue that the results of Novy-Marx (2014) and Fama & French (2016) are entirely driven by the short sides of factors. In a long-only setting, low-volatility certainly survives as a distinct factor.

Finally, a more general critique comes from Harvey (2017) who criticizes the finance community for p-hacking. For example, it is argued that publications are biased to the positive outcomes, and other areas like which results are chosen to be presented or the statistical methods can possibly be prone to manipulation. Moreover, Harvey, Liu & Zhu (2016) actually find that out of more than 300 anomalies analyzed (which they call the factor-zoo), many become insignificant after using a meticulous and diligent testing framework.

This paper will contribute to the existing literature in the following ways. First, it will use a recent data set to confirm the anomaly exists. Secondly, the anomaly will be challenged by taking into account different methodologies and robustness tests to tackle the anomaly from different angles, and conclude whether the anomaly still persists after doing this. Finally, the paper will show the time variation of the anomaly over the sample to see whether it has stayed consistent in strength, or whether it has increased or decreased.

2.1 Research Question

Does the anomaly survive after taking into account different methodologies and in all time periods?

The motivations for this research question are presented in the section below. From the literature, it shows that the recent research, as in Novy-Marx (2014), Fama & French (2016) and Blitz & Vidojevic (2017), but also Liu et al. (2018) and Asness et al. (2020), spark a discussion on the low volatility anomaly. A more detailed elaboration is in the section below, however, these developments create a need to further investigate the anomaly and take into account these recent findings.

2.2 Motivation

The paper uses roughly three motivations to study the low volatility anomaly.

First, Bali & Cakici (2008) have criticized some of the literature by breaking down their methodologies. They show that the chosen breakpoints, weighting schemes and different measures of volatility greatly influence the results. They provide details of the consequences of certain decisions regarding methodologies and this information should be taken into account when constructing portfolios. Also taking into account Harvey, Liu & Zhu (2016), it is important to test the anomaly in different settings to confirm it is not a data fluke. Also, low t-statistics will not be accepted, as proposed by Harvey et al. (2016).

Second, there is little research into the anomaly over time. This paper aims to analyze the anomaly per decade to gain insights into how the anomaly behaves over time. Generally speaking, factors are sometimes thought to have decreased over time after more research about it gets published. Therefore, it is also expected that the Low Volatility anomaly has decreased over time, as quite some time has passed since it gained popularity with the paper of Blitz & Van Vliet (2007).

Third, throughout the literature low beta and low volatility have been used interchangeably. However, as described in Liu, Stambaugh & Yuan (2018), the beta anomaly as in Frazzini & Pedersen (2014), also known as betting against beta (BAB), is a distinct anomaly. See also Asness, Frazzini, Gormsen & Pedersen (2020) who address this issue further. They agree with the criticism and split the BAB factor into 2: Betting against Volatility (BAV) and Betting against Correlation (BAC). They find that correlation matters, but only within volatility buckets. Also taking into account Blitz & Vidojevic (2017), it is useful to keep making the distinction between low beta and low volatility, and examine the results of papers where the terms are used interchangeably. Also, the paper will use another method to calculate volatility (idiosyncratic volatility, or IVOL) as a robustness test.

2.3 Hypothesis

Based on the literature, the following hypotheses are formulated based on the research question.

The low volatility anomaly is expected to be statistically significant in this dataset.

The volatility anomaly is expected to decrease in magnitude after using different methodologies.

The low volatility is expected to have decreased in magnitude over time.

3 Data

In this section, the dataset that is used is introduced and some characteristics of the dataset are highlighted. The section after this will describe the methodology used in this paper.

The sample in this study includes all common stocks (share codes 10 and 11) in the CRSP (Center for Research in Security Prices) database traded on NYSE, AMEX and NASDAQ exchanges, except those with an average price below 5 dollars. Excluding stocks with an average price below 5 dollars is done to exclude the typically illiquid penny stocks. Simply excluding all stocks below 5 dollars will serve a problem because a stock that goes bankrupt will always at some point drop below that 5 dollar mark. Thus, doing so will exclude bankrupt stocks from the sample, making the sample subject to survivorship bias. Survivorship bias refers to the tendency to look at a sample where the losers that have gone bankrupt are excluded from the sample. Survivorship bias can therefore lead to an overestimation of (in this case) returns by excluding the (often) negative returns of losers. To avoid excluding these bankrupt stocks and only exclude the illiquid penny stocks, stocks with an average price (over the sample) of below 5 dollars are excluded. To illustrate the effect of using the average price to drop stocks instead of just dropping all stocks with a price of $5, see the following example. The average monthly return when dropping stocks with an average price below $5 (before winsorization, and not in excess of the risk free rate) is 1.37%. However, if all stocks below $5 are dropped (instead of the average), the average return is 1.91%. However, as a robustness test these stocks will be included again to see the effect of excluding them. Still, the main analysis will exclude them.

The CRSP database delivers some prices as negative; this negative (-) sign is actually a symbol, instead of a negative indicator. It represents prices which are bid/ask averages instead of closing prices. Therefore, all prices with a symbol (-) are replaced by their absolute value. See Appendix section 8.2.4 for further definitions of price.

Another feature of the CRSP database is that holding returns are reported with some special return codes for a reason as to why the return is missing. These are indicated with -66.0, -77.0, -88.0 and -99.0 These are therefore not valid returns, so to deal with this, stocks with these return codes will be removed from the dataset.

Throughout this paper, there is use of and reference to factors. These factors are taken directly from the Ken French Data Library. The factors used are MktRf, SMB, HML, RMW, CMA and MOM. Below the factors are briefly introduced. Further definitions of how they are constructed can be found in the Appendix. The market return in excess of the risk-free rate is abbreviated as MktRf. SMB is the small minus big factor, also known as the size factor. HML is the high minus low factor, also known as the value factor. RMW is the robust minus weak factor, also known as the profitability factor. CMA is the conservative minus aggressive factor, also known as the investment factor. MOM is the momentum factor, which buys past winners and sells past losers.

3.1 Descriptive Statistics

This section tries to summarize some basic characteristics of the data to obtain an overview of what the sample looks like. Some statistics such as the number of stocks and the distribution of prices are presented and the correlations between the portfolios are highlighted.

First, the data is analyzed to get an idea of what it looks like To start with the raw dataset, Table 3.1 shows that there are on average 5195 stocks in the sample. The data is from 1963 to 2019, and at the lowest point there were 1977 stocks in the sample (March 1963), and at the peak 7615 (July 1997). Because small stocks are often more volatile than larger stocks and often have a large bid/ask spread and their illiquid nature, it is useful to look at the data without these penny stocks. Simply dropping stocks below $5 would make the sample suffer from survivorship bias. To avoid this, the average price can be used. If a stock in the sample has an average price of below $5, the stock is excluded. In this way, dropping stocks in distress that later bounce back up again is avoided, as well as (non-penny) stocks that go bankrupt and thus drop below $5, to $0. After doing this, there are on average 4170 stocks in the sample, with 1804 at the lowest point (February 1963) and 6069 stocks at the highest point (in both June and November 1997).

Table 3.1: Average Number of Firms in the Sample
The Table reports the average, minimum and maximum number of firms per month in the sample over the entire period. The sample starts in January 1963 and ends in December 2019, and includes all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. The numbers for the original, untouched dataset are reported, and below that the adjusted dataset. This is after excluding stocks that are on average below $5 in the sample.

	Average	Lowest	Highest
Original	5195	1977	7615
After adjustments	4170	1804	6069

If the exclusion of penny stocks as mentioned above is applied, the distribution of the prices changes as expected. In Table 3.2, the 10th percentile now has a price of 4.125$. The average price is now $38.13.

To further analyze the formed portfolios, a correlation matrix can be used.

Table 3.2: Distribution of Prices
The Table shows the distribution of the prices in the sample. The sample starts in January 1963 and ends in December 2019, and includes all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. Stocks with an average price below $5 in the sample are excluded.

Price or Bid/Ask Average	
1%	$0.86
5%	$2.63
10%	$4.13
25%	$8.25
50%	$16.13
75%	$28.63
90%	$45.88
95%	$61.07
99%	$115.44
Average	$38.13

Table 3.3: Correlation Matrix
The Table shows the correlations between the volatility portfolios. Portfolios are formed every month by sorting stocks on their past 3-year return volatility. The sample used runs from January 1963 to December 2019 and includes all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. Stocks are sorted into portfolios each month using CRSP breakpoints, and returns are equal-weighted. Stocks with an average price below $5 in the sample are excluded. Returns are winsorized at the 1% level.

	Volatility Portfolios									
	Low	2	3	4	5	6	7	8	9	High
Low	1.000									
2	0.934	1.000								
3	0.894	0.979	1.000							
4	0.867	0.964	0.986	1.000						
5	0.839	0.945	0.974	0.985	1.000					
6	0.816	0.924	0.955	0.973	0.982	1.000				
7	0.781	0.890	0.928	0.953	0.969	0.982	1.000			
8	0.739	0.851	0.894	0.924	0.945	0.967	0.983	1.000		
9	0.692	0.804	0.851	0.885	0.912	0.941	0.966	0.979	1.000	
High	0.624	0.736	0.785	0.824	0.858	0.894	0.928	0.952	0.974	1.000

Table 3.3 shows the return correlations between the portfolios. Comparing them, the Low and the High Volatility portfolio have a return correlation of 0.624. At the same time, portfolios Low and 2 have a correlation of 0.934. Portfolios 9 and High have a correlation of 0.974. This shows that the portfolios close

together have somewhat similar returns, but the portfolios at the ends of the sorts, so portfolio Low and High, have different returns as shown by the correlation of 0.624. This indicates the different portfolios have different return patterns.

4 Methodology

The main results of this paper are described in Section 5. These results rely on carefully constructed methodologies to derive these results. In this section, it is outlined and explained how these measures and results are derived, and which methodologies have been used. Subsection A will describe the formation of the portfolios. Subsection B will outline how the performances between portfolios are compared. Subsection C describes how the performance over time will be measured and compared. Finally, subsection D describes the robustness tests.

In the literature, there are multiple ways to construct the volatility sorted portfolios. The literature uses CAPM beta, past 3-year volatility of weekly returns, 5-year trailing total volatility, idiosyncratic volatility obtained by taking the standard deviation of the residual of the Fama French 3-factor Model on daily returns, past month daily data, volatility of total return of the past 24 months, 1 year of daily data, variance of daily returns and variance of daily residuals from the Fama French 3-factor Model (both) using 60 days of lagged returns. All these different methods can be used, but they produce slightly different results because their focus differs. As will be outlined later in the paper, choosing beta as a measure of volatility captures the beta effect and not the volatility effect, so beta is not the measure of choice in this paper. However, there is a section dedicated to the differences between the two in Section 6. The frequency of returns can produce different results as daily of weekly returns will be more volatile by definition than monthly returns. Comparing the differences in results between the data frequency is interesting but beyond the scope of this paper. Using either idiosyncratic volatility or volatility of returns is also an important difference. A section is dedicated to the impact of this choice in Section 6. In this paper, Blitz, Pang & Van Vliet (2013) is chosen as the main methodology. That means that past 3-year volatility of monthly returns will be used to test the main Model. Below, the used methods are explained in detail.

It should be noted that in the main Model, returns will be winsorized at the 1% level and stocks with an average price of below $5 will be excluded from the sample. Both are done to prevent the impact of extreme outliers or abnormally volatile or illiquid stocks. In section 6 an analysis is done on the impact of these two methodological choices on the results. Throughout the paper, returns are measured in excess of the risk free rate. Where appropriate, Newey-West t-statistics are reported with a lag of 6 months.

4.1 Forming portfolios on volatility

First, the volatility of each stock is calculated by taking the standard deviation of the past 3-year monthly returns, following Blitz, Pang & van Vliet (2013). Next, at the end of each month, stocks are ranked based on their calculated volatility and categorized into 10 equal weighted decile portfolios and value-weighted portfolios. As a robustness check, they are also divided into quintile portfolios. The top portfolio will contain the stocks with the highest past volatility, and the bottom portfolio the stocks with the lowest past volatility. Furthermore, the stocks are sorted into the portfolios using both CRSP breakpoints and NYSE only breakpoints. For the main Model, the latter will be used together with value-weighted returns.

4.2 Performance Comparison

For each portfolio, the average monthly return, standard deviation of returns and Sharpe ratio is reported. The returns are excess return, meaning the return in excess of the risk free rate, which can be taken from the Ken French Data Library, just as the return on the market. The returns presented are 1-month holding returns.

To talk about the performance of the volatility portfolios, the returns and the standard deviation of returns will be compared. To capture this in a single number, the Sharpe ratios of the portfolios can be used to give insights into the differences in the risk/reward relationship of the portfolios. To compare two Sharpe ratios, the Jobson & Korkie (1981) test with the Memmel (2003) correction can be used, to test for the statistical difference between two Sharpe ratios, following Blitz & Van Vliet (2007). The following formula will be used to calculate the z test statistic:

$$z = \frac{SR_1 - SR_2}{\sqrt{\frac{1}{T}[2(1-\rho_{1,2}) + \frac{1}{2}(SR_1^2 + SR_2^2 - SR_1 SR_2(1+\rho_{1,2}^2))]}}$$

Where SR_i will reflect the Sharpe ratios of portfolio i, $\rho_{i,j}$ the correlation between portfolios i and j, and T the number of observations.

Next to that, to further compare the variances of two portfolios, a test of equal variances will be used with a 95% confidence level. Based on the F-statistic and the P-value, a conclusion will be drawn whether the two portfolios have equal or unequal variance. With that information, a two-sample t-test of equal means to compare the averages of the two portfolios will be performed using either equal or unequal variances, again with a 95% confidence level. The calculated t-statistic and p-value resulting from this test will then give an indication whether the average returns of the two portfolios are different or not. This approach will be used in various sections of this paper.

Next to comparing the performances of the volatility portfolios, they will also be compared to the returns on other factors. The paper will use the Size (SMB), Value (HML), Profitability (RMW), Investment (CMA) and Momentum (MOM) factor returns taken from the Ken French Data Library to get an idea of the magnitude and the performance of the low volatility anomaly compared to other factors.

4.3 Performance over Time

After having formed the portfolios, the monthly returns are obtained for the 10 volatility sorted portfolios and compared over time. The entire sample will be split up per decade, and the returns, standard deviation of returns and the Sharpe Ratios of the portfolios will be compared over time. Analyzing the differences between the low volatility portfolio and the high volatility portfolio will give insights into how the anomaly has evolved over time.

A regression based methodology will be used to complete the analysis into the time periods. Specifically, regression (1) will be used to analyze the returns over time. Here, the beta will indicate how much a portfolio will move in relation to the excess market return, and alpha will indicate how much of the portfolio's movement is not explained by the excess market return. To try to approach why the anomaly evolves over time, market beta could prove to be an explanatory variable in the portfolio's return. The beta of the Low and High portfolios will be compared over the time periods, to see if any patterns appear. For example, whether the High Volatility portfolio always has a higher beta than the Low Volatility portfolio. Similarly, the alpha of the portfolios will be compared to see if any patterns become apparent over time, such as that the Low Volatility portfolio always has a low alpha whereas the High portfolio always has a high alpha.

The following regression will be used:

$$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + \epsilon_i \tag{1}$$

4.4 Robustness Tests

In this subsection, the methodologies used in section 6 will be introduced. The robustness tests will serve as an analysis as to whether the anomaly is dependent on a particular methodology or whether it is robust to particular choices and assumptions in the construction of the anomaly.

4.4.1 Decile vs. quintile portfolios

When forming portfolios, the choice can be made on how many portfolios to divide the stocks into. In the literature, using 5 or 10 portfolios is the most common choice. In the main analysis, each month stocks are divided into 10 portfolios. In this robustness test, the stocks will instead be divided into 5 portfolios each month and the performances will be compared by looking at differences in return patterns, Sharpe ratio patterns and statistical significance. For both methods, NYSE breakpoints will be used to form portfolios and returns are value-weighted. As in the main Model, returns are winsorized and stocks with an average price below \$5 are excluded. The results can be found in section 6.1.

4.4.2 Winsorized vs. non-winsorized

This section will explore how winsorizing the returns at the 1% level will affect the results. In the main Model, before portfolios are formed and even before past 3-year volatility is calculated, the returns are winsorized at the 1% level, diminishing the effects of extreme outliers. Skipping this step will provide the results of the non-winsorized data. The performances will be compared by looking at differences in return patterns, Sharpe ratio patterns and statistical significance. For both methods, NYSE breakpoints will be used to form portfolios and returns are value-weighted. As in the main Model, stocks with an average price below \$5 are excluded. The results can be found in section 6.2.

4.4.3 Dropping stocks below \$5 or not

This section will explore how dropping stocks with an average price below \$5 will affect the results. In the main Model, before portfolios are formed and even before past 3-year volatility is calculated, the stocks with an average price of below \$5 in the sample are excluded. Simply not doing this will provide the results including these so-called penny stocks. The performances will be compared by looking at differences in return patterns, Sharpe ratio patterns and statistical significance. For both methods, NYSE breakpoints will be used to form portfolios and returns are value-weighted. As in the main Model, returns are winsorized. The results can be found in section 6.3.

4.4.4 Volatility vs. idiosyncratic volatility

This section will perform the analysis of the main Model except with a different measure of volatility. As described above, this paper measures volatility as the standard deviation of past 3-year monthly returns. A different, often used measure is idiosyncratic volatility (IVOL, as in Ang et al. (2006)), which can be measured as the standard deviation of residuals, obtained by regressing returns on the Fama-French 3-factor Model. Often, daily returns are used to do this. Because monthly returns are used in this paper, here IVOL is calculated using monthly returns. In Liu et al. (2018) it is reported that there is little difference between using the residuals from a 1-factor regression or a 3-factor regression (they find a correlation of 0.99), but this paper will use the 3-factor regression anyways for the sake of completeness. So, each month, stock returns are regressed on the Fama-French 3-factor Model. The standard deviation of the residuals is then taken and used to sort stocks into 10 portfolios. The performances will be compared by looking at differences in return patterns, Sharpe ratio patterns and statistical significance. For both methods, NYSE breakpoints will be used to form portfolios and returns are value-weighted. As in the main Model, returns are winsorized and stocks with an average price below \$5 are excluded. The results can be found in section 6.4.

As mentioned, a regression based methodology will be used to compute the IVOL measure. To do this, the residuals from a regression on the Fama-French 3-factor Model are needed. The residuals are captured by the epsilon in the regression (2), and arise by regressing the constructed portfolios on the Fama-French 3-factor Model, i.e. on Size (SMB) and Value (HML). The factors mentioned above against which the volatility portfolios will be regressed will be taken from the Ken French data library. As mentioned, the standard deviation of these residuals will then be calculated to compute the IVOL measure.

The following regression will be used:

$$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_i SMB + h_i HML + \epsilon_i \tag{2}$$

4.4.5 Low Volatility & Beta

In the literature, volatility and beta are sometimes used interchangeably. In recent studies, Liu et al. (2018) make a clear distinction between the two and show that the two are different, and do not capture the same. This finding is confirmed in Asness et al. (2020), who find that splitting their Betting Against Beta factor as in Frazzini & Pedersen (2014) into Betting Against Correlation and Betting Against Volatility provides better explanatory value. This section will aim to highlight the differences between volatility and beta. To do this, the double-sorting technique is applied, by first ranking stocks on beta and subsequently on volatility within the beta buckets. A 5x5 double sort is used.

To construct beta, the methods of Liu, Stambaugh & Yuan (2018) are followed. They provide an extensive overview of the different ways to construct beta and compare the outcomes. To follow their methodology, a regression is used that accounts for lagged market return to accommodate non-synchronous trading effects. This gives us the following eq. (3):

$$R_{i,t} = \alpha_i + \beta_{i,0} r_{m,t} + \beta_{i,1} r_{m,t-1} + \epsilon_{i,t} \tag{3}$$

This regression will be run each month over a moving 60-month window. The beta estimate will be as eq. (4), applying summed slopes as in Dimson (1979).

$$\hat{\beta}_i^{ts} = \hat{\beta}_{i,0} + \hat{\beta}_{i,1} \tag{4}$$

The next most common step in the literature to estimating beta is to follow Vasicek (1973) to increase precision, by shrinking the beta towards 1.
See eq. (5).

$$\hat{\beta}_i = \omega_i \hat{\beta}_i^{ts} + (1 - \omega_i) \tag{5}$$

Note that Frazzini & Pedersen (2014) also use the Vasicek correction, and choose 0.6 as the weight. However, using the Vasicek correction here does not add any value because shrinking the beta does not change anything in the sorting process of the stocks into the portfolios. Because beta is in this paper only used for sorting purposes and not further regression analysis, correcting the beta has no effect.

To compare Volatility and Beta, a simple correlation is first calculated. Liu et al. (2018) find a correlation of 0.33. Then the double-sorting technique as mentioned above can be used, using a 5x5 or a 10x5 sort. As in Liu et al. (2018), little evidence for a beta effect is expected once we control for volatility, thus proving that they are distinct effects. The performances will further be compared by looking at differences in return patterns, Sharpe ratio patterns and statistical significance. For both methods, NYSE breakpoints will be used to form portfolios and returns are value-weighted. As in the main Model, returns are winsorized and stocks with an average price below $5 are excluded. The results can be found in section 6.5.

To summarize, the research question this paper aims to answer is whether the anomaly survives after taking into account different methodologies. Also, the evolution of the anomaly over time will be analyzed. To do this, the paper will first form portfolios on Volatiliy to construct the main Model. This Model will first be analyzed by looking at its performance per decade. This will answer the question of how the anomaly evolves and behaves over time. Then, to investigate whether the anomaly will survive after taking into account different methodologies, a number of robustness tests will be done. Here, each of the assumptions and methodological choices in constructing the portfolios will be let loose (ceteris paribus), to see if the anomaly is dependent on a particular methodological choice or if it is truly robust. After gaining insights into both, conclusions will be drawn about whether the anomaly is robust to methodological choices and time, and with that the research question can be answered.

5 Empirical Results

The Empirical Results section is structured as follows. First, the very basic equal weighted portfolios formed on volatility are presented. Then, these are compared against some other known factors. The critique of the impact of chosen methodologies is then immediately tackled by comparing equal weighted and value-weighted portfolios as well as the choice of breakpoints. The main methodology this paper will look at is then chosen and this methodology will be used throughout the rest of the paper to perform robustness tests against. Second, a double-sort on Volatility and the Size factor will be performed to see how Market Equity (ME) affects the magnitude and presence of the Low Volatility anomaly. And third, the anomaly will be analyzed over time. The section after that will test the robustness of the anomaly (Section 6).

5.1 Volatility Portfolios

To start, 10 equal-weighted portfolios are formed each month sorted on volatility. For each portfolio, the average 1-month holding return and its Sharpe Ratio is calculated. This simple overview will help us determine if the anomaly exists; if return or Sharpe ratio does not become greater moving from portfolio 1 to 10, then that means that higher risk is not rewarded with a higher return, or at least higher risk adjusted return. This is the idea of the low volatility anomaly. The results are published in table 5.1.

Table 5.1: **Portfolio Performance**
The Table shows the average 1-month holding returns in excess of the risk free rate, the standard deviation of returns and the Sharpe Ratios for portfolios formed every month by sorting stocks on their past 3-year return volatility. The sample runs from January 1963 to December 2019 and includes all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. Stocks are sorted into portfolios each month using CRSP breakpoints, and returns are equal-weighted. Stocks with an average price below $5 in the sample are excluded. Returns are winsorized at the 1% level. Reported t-statistics are simple t-statistics from a t-test whether the returns are significantly different from zero.

Volatility	Low	2	3	4	5	6	7	8	9	High
Return	.749	.808	.823	.858	.869	.884	.895	.830	.772	.513
t-statistic	(6.29)	(5.37)	(4.82)	(4.57)	(4.30)	(3.95)	(3.71)	(3.15)	(2.64)	(1.54)
SD (return)	3.03	3.83	4.34	4.77	5.13	5.68	6.13	6.70	7.44	8.45
Sharpe Ratio	0.248	0.211	0.190	0.180	0.169	0.156	0.146	0.124	0.104	0.061

We can see in the Table that going from the Low Volatility portfolio to the High, the returns slightly increase but the Sharpe Ratio definitely declines.

Next, the returns can be compared to those of some other well-known factors, to get an idea about the size and magnitude of the returns. The same decile returns are reported, and next to that the market return, and the factor returns of HML (High Minus Low), SMB (Small Minus Big), RMW (Robust Minus Weak) and CMA (Conservative Minus Aggressive) as in Fama & French (2016).

Table 5.2: Factor performance

Average returns, standard deviations of returns, and t-statistics for monthly factor returns, July 1963 to December 2019 (678 months) for all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. Reported t-statistics are simple t-statistics from a t-test whether the returns are significantly different from zero. Factor returns are taken directly from the Ken French Data Library.

	Rm-Rf	SMB	HML	RMW	CMA	MOM
Return	.54	.23	.30	.26	.27	.65
SD (return)	4.38	3.01	2.81	2.15	1.99	4.18
t-statistic	(3.21)	(1.97)	(2.82)	(3.13)	(3.58)	(4.02)

We see in Table 5.2 what the magnitude Fama-French and momentum factors are. This can be compared to Table 5.1 and it shows that the returns found from the decile portfolios are quite high. However, Table 5.1 shows equal-weighted returns with simple t-statistics and therefore not too many conclusions should be drawn from this. Also, perhaps more importantly, the factor returns are long-short portfolios, and comparing long-short portfolio returns with long-only portfolios is actually not comparable as they are inherently different. The next sections will provide further analysis with somewhat more sophisticated ways of measuring returns and significance levels.

To summarize, using a simple sort on past 3-year volatility of returns with equal weighted returns shows that the low volatility portfolio performs equally well as the high volatility portfolio in terms of return but does so with less fluctuations in the returns.

5.2 Comparison Different Sorting & Weighting Schemes

To address the concerns of Bali & Cakici (2008), in this section the performance of the anomaly is compared by using different methodologies. First, the results from Table 5.3 will be analyzed. Then, the results will be put into context by comparing them to the results found in the literature. Then the results will be extended by performing a double-sort on size and volatility. Finally, the section will conclude.

Talking about Bali & Cakici (2008) specifically, they find that the choice of breakpoints and the chosen weighting scheme affects the results. So their methodology is replicated and portfolios are formed using CRSP breakpoints, and NYSE breakpoints. Then, for both methods, equal-weighted and value-weighted returns are calculated. This gives us 4 different views, which are presented in Table 5.3.

Table 5.3 presents the equal-weighted and value-weighted returns of 10 decile portfolios formed by sorting stocks on their past 3-year volatility. Market Equity is used as a proxy for size, which is obtained by multiplying the price and shares outstanding of each stock every month. In Panel A, the portfolios are formed using the entire CRSP universe of stocks (as specified in section 3) as breakpoints, while Panel B shows the portfolios formed using only NYSE stocks as breakpoints. The reason for this is that Bali &

Table 5.3: Comparison sorting and weighting scheme

The Table shows the average 1-month holding returns in excess of the risk free rate, the standard deviation of returns and the Sharpe Ratios for portfolios formed every month by sorting stocks on their past 3-year return volatility. The sample runs from January 1963 to December 2019 and includes all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. Panel A shows the equal-weighted portfolios and Panel B shows the value-weighted portfolios. Market Equity (ME) is used to determine size. The left columns show these results for portfolios formed using CRSP breakpoints. The right columns show the results when using NYSE breakpoints only. Alphas are computed with respect to the Fama-French 3-factor Model, using a regression with the High-Low portfolio. T-statistics are reported between brackets, and are from a simple t-test whether the variable is significantly different from zero. Stocks with an average price below $5 in the sample are excluded. Returns are winsorized at the 1% level.

	CRSP Breakpoints			NYSE Breakpoints		
Decile	Return	Mkt. Share	Standard Deviation (return)	Return	Mkt. Share	Standard Deviation (return)
Panel A: Equal-weighted portfolios						
Low	.749	46,6%	3.026	.719	46,1%	2.936
2	.808	12,5%	3.825	.776	12,0%	3.576
3	.823	10,0%	4.337	.840	9,4%	4.062
4	.858	6,8%	4.770	.813	7,8%	4.440
5	.869	6,0%	5.130	.882	6,3%	4.743
6	.884	5,1%	5.681	.838	5,2%	5.039
7	.895	4,3%	6.131	.921	4,4%	5.469
8	.830	3,5%	6.704	.857	3,7%	5.929
9	.772	2,9%	7.435	.870	3,0%	6.464
High	.513	2,3%	8.452	.656	2,2%	7.755
High-Low	-0.236			-0.063		
	(-0.858)			(-0.255)		
Alpha	-0.341			-0.178		
	(-1.28)			(-0.75)		
Panel B: Value-weighted portfolios						
Low	.719	46,6%	3.662	.678	46,1%	3.587
2	.783	12,5%	4.300	.735	12,0%	3.853
3	.765	10,0%	4.434	.820	9,4%	4.445
4	.760	6,8%	4.757	.758	7,8%	4.519
5	.817	6,0%	5.064	.790	6,3%	4.762
6	.779	5,1%	5.562	.766	5,2%	4.992
7	.786	4,3%	6.005	.831	4,4%	5.361
8	.755	3,5%	6.623	.758	3,7%	5.742
9	.700	2,9%	7.429	.785	3,0%	6.348
High	.539	2,3%	8.436	.639	2,2%	7.655
High-Low	-0.180			-0.039		
	(-0.624)			(-0.151)		
Alpha	-0.222			-0.085		
	(-0.78)			(-0.34)		

Cakici (2008) raised concerns about common methodologies in the literature that use CRSP breakpoints. Their argument is that findings are biased by size and exchange-specific characteristics. To address this, this paper follows their methodology by using the NYSE breakpoints next to the CRSP breakpoints. We can see in Panel A that the return difference for the equal weighted decile sorted on CRSP breakpoints is -0.236% with a Newey West t-statistic of -.858. The value weighted returns for the CRSP breakpoints sorted difference portfolio is -0.180% with a Newey West t-statistic of -0.624. If these results are compared to the literature, there are slightly higher returns here. For example, Ang et al. (2006) find a result of -0.97% for their portfolios sorted by total volatility. Bali & Cakici (2008) find a result of -0.93% for portfolios sorted on idiosyncratic volatility. The reason my returns are higher could be because of the longer sample, or because they sort the stocks into quintiles instead of deciles. From the Table, portfolio 10 has substantially lower returns than portfolio 9, respectively 0.513% and 0.772%. Because the difference portfolio is calculated as portfolio 10 minus portfolio 1, the result is less negative than if quintiles would have been used. Another possible explanation is the different methodology this paper uses, while both Ang et al. (2006) and Bali et al. (2008) use idiosyncratic volatility, this paper uses past 3-year return volatility as in Blitz & van Vliet (2007). They use equal-weighted returns over the period 1986 to 2006, sorting stocks based on past 3-year volatility of weekly returns in decile portfolios. Interestingly, the returns in the deciles show the same behavior as in Blitz et al. (2007): They find that decile 9 delivers 7% return while decile 10 only 3.8%. Their returns do seem to be substantially higher than the findings in this paper. This could be due to the relatively small time period they use. This is confirmed by the high monthly return they find for the entire universe of US stocks: 8.1%. So for portfolio 1 they find a return of 6.9% and portfolio 10 a return of 3.8%. The difference is then -3.1%. Again, the results in this paper are less contrasting.

In Table 5.3 the market shares of the portfolios are also reported. When CRSP breakpoints are used to form decile portfolios, the average market share of decile 1 is 46.6% while decile 10 is only 2.3%. This shows us the negative correlation that exists between volatility and size. Apparently, the more volatile stocks are also the smaller sized stocks (measured by Market Equity). This effect can be further explored with a double sort on Volatility and Size. We see that the portfolios that are formed using NYSE breakpoints are somewhat better divided over the deciles. Now, portfolio 1 contains 46.1% and portfolio 2 contains 12% (compared to 46.6% and 12.5%, respectively). At the same time, portfolio 10 has an average market share of 2.2% (compared to 2.3%), meaning the changes happened in the middle portfolios rather than directly from the lower portfolios to the higher portfolios. Furthermore, in Table 8.3 (see Appendix), the market shares are compared when penny stocks are included or excluded. From this analysis it becomes evident that excluding penny stocks leads to a more balanced distribution of stocks amongst the portfolios.

Also, to further compare the results of using CRSP breakpoints versus NYSE breakpoints, the alphas (form the High-Low portfolio) as computed with the respect to the Fama & French 3-factor Model, just as in Bali et. al (2008), increase in Panel A (equal-weighted returns) from -0.341 (CRSP breakpoints) to -0.178 (NYSE breakpoints). However, the Newey West t-statistic becomes less, it decreases from -1.28 to -0.75. Doing the same comparison in Panel B(value-weighted returns), the alpha again increases from -0.222 (CRSP breakpoints) to -0.085 (NYSE breakpoints). But again the t-statistic becomes less significant as goes from -0.78 to -0.34.

Bali et. al (2008) also present the same results when using volatilities based on 2 to 5 years of data instead of daily return data. Since stocks are sorted based on past 3-year volatility, the results might be more comparable. It should be noted that they use idiosyncratic volatility, computed based on the Fama French three-factor Model whereas this paper computes simple volatilities of return. Interestingly, the results are more comparable than before. For example, they find that the value-weight difference (10 minus 1) portfolio produces a return of -0.39 (t-statistic -1.08) with an alpha (Fama French 3-factor Model) of -0.64 (t-statistic of -3.26). The Table shows a return of -0.180 (t-statistic of -0.624) with an alpha of -0.222 (t-statistic -0.78). The results could be more comparable because stocks can be less volatile on a yearly basis than on a daily basis. However, it is hard to say as the sample is also shorter and the way of measuring volatility also differs.

Because it shows from the Table above that size apparently plays a role in the returns of volatility portfolios, using value-weighted returns with portfolios formed using NYSE breakpoints gives us a fairer view of the anomaly. Therefore, the rest of the paper will use this methodology to form portfolios and compute returns. This is the same methodology that Fama & French (2016) use, which is value-weighted portfolios formed monthly using NYSE breakpoints.

To put the results of the value-weighted, NYSE breakpoints portfolios into context, the performance of these portfolios will be compared to other factors. This will give some insight into the magnitude and the performance of the anomaly. However, it should immediately be noted that the following Table will compare portfolios with factors that are constructed with a long-short portfolio. Because this is not very applicable to the Low-Volatility anomaly, long-only returns will here be compared with the factor returns. The results are presented in Table 5.4.

Table 5.4: Portfolio and Factor Performance
The Table shows the average 1-month holding returns in excess of the risk free rate, the standard deviation of returns (SD) and the Sharpe Ratios for portfolios formed every month by sorting stocks on their past 3-year return volatility & the performance of factor returns. Panel A shows the performance of the volatility portfolios and Panel B shows the performance of the factors. The sample runs from January 1963 to December 2019 and includes all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. Stocks are sorted into portfolios each month using NYSE breakpoints, and returns are value-weighted by size. Market Equity (ME) is used to measure size. Stocks with an average price below $5 in the sample are excluded. Returns are winsorized at the 1% level. Reported t-statistics are Newey-West t-statistics, and show whether the returns are significantly different from zero.

Decile	Low	2	3	4	5	6	7	8	9	High
Panel A: Portfolio Performance										
Return	.678	.735	.820	.758	.790	.766	.831	.758	.785	.639
t-statistic	(4.24)	(4.53)	(4.50)	(4.01)	(3.89)	(3.62)	(3.64)	(3.13)	(2.90)	(1.99)
SD (return)	3.59	3.85	4.44	4.52	4.76	4.99	5.36	5.74	6.35	7.66
Sharpe Ratio	.189	.191	.185	.168	.166	.154	.155	.132	.124	.083

Factor	Mktrf	SMB	HML	CMA	RMW	MOM
Panel B: Factor Performance						
Return	.541	.228	.304	.274	.259	.646
t-statistic	(3.06)	(1.86)	(2.40)	(3.14)	(2.73)	(3.93)
SD (return)	4.39	3.01	2.81	2.16	1.99	4.18
Sharpe Ratio	.123	.076	.108	.127	.130	.155

5.3 Double-sort on Volatility and Size

To further explore the relationship between Size and Volatility as shown in the previous section, the returns are presented, with standard deviations of return and Sharpe Ratios for 25 double sorted portfolios.

Table 5.5: Double sort on size and volatility
The Table shows the average 1-month holding returns in excess of the risk free rate, the standard deviation of returns (SD) and the Sharpe Ratios for 25 portfolios double sorted on size and volatility formed every month. The sample runs from January 1963 to December 2019 and includes all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. The portfolios are first sorted into quintiles on Market Equity, which is used to proxy for size. Then, conditional on this sort, the stocks are further sorted on past 3-year return volatility. The intersection of the sorts produces 25 Size-Volatility portfolios. Stocks with an average price below $5 in the sample are excluded. Returns are winsorized at the 1% level. Reported t-statistics are Newey-West t-statistics. Panel A shows the average return and standard deviation of 1-month holding returns of value-weighted portfolios. Panel B shows the same returns, but now with the calculated Sharpe Ratio in the right section.

Vol	Low	2	3	4	High	Low	2	3	4	High
Panel A: Returns and SD of portfolio excess returns										
	Return					SD (return)				
Small	.860	.907	.978	.963	.752	3.359	4.501	5.100	5.987	7.375
2	.816	.978	.893	.843	.752	3.301	4.429	5.068	5.823	7.243
3	.756	.865	.878	.832	.634	3.383	4.496	5.198	5.812	7.174
4	.678	.723	.734	.744	.643	3.423	4.427	5.102	5.665	7.198
Big	.681	.739	.680	.748	.865	4.185	4.791	5.003	5.642	7.553
Panel B: Returns and Sharpe Ratio of portfolio excess returns										
	Low	2	3	4	High	Low	2	3	4	High
	Return					Sharpe Ratio				
Small	.860	.907	.978	.963	.752	.256	.202	.192	.161	.102
2	.816	.978	.893	.843	.752	.247	.221	.176	.145	.104
3	.756	.865	.878	.832	.634	.223	.192	.169	.143	.088
4	.678	.723	.734	.744	.643	.198	.163	.144	.131	.089
Big	.681	.739	.680	.748	.865	.163	.154	.136	.133	.115

As presented in Table 5.5, Panel A, within each size quintile there is some variation in return, looking along the volatility quintiles. For example, it shows that small, low volatility stocks produce a return of 0.860% per month while the small, high volatility stocks deliver a return of 0.752%. Actually, in every size quintile the return for the low volatility portfolio is higher than the high volatility stocks, except for the Big stocks. Here, the low Volatility portfolio delivers a return of 0.681% and the high Volatility portfolio a return of 0.865%. But, looking at the right side of Panel A it shows that the standard deviation of returns of the High Volatility portfolios are much higher than the Low Volatility portfolios, indicating they are riskier. While this is expected because stocks are sorted on volatility, it is noteworthy because investors often look not just for return but return per unit of risk as well. This is a metric captured by the Sharpe Ratio, which calculates the return per unit of risk. Panel B confirms that the High Volatility portfolio returns are less compensated for their risk than the Low Volatility portfolios. For example, the Small, Low Volatility portfolio has a Sharpe Ratio of 0.256 while the Small, High Volatility portfolio a lower Sharpe Ratio of 0.102. The Sharpe ratio declines for the Low Volatility portfolios moving down to the bigger Size portfolios, as the Big, Low Volatility portfolio now only has a Sharpe ratio of 0.163. However, this is still higher than the Big, High Volatility portfolio which has a Sharpe Ratio of 0.115.

It becomes evident that within Volatility quintiles, Sharpe Ratios generally decline moving from Small to Big portfolios. In the High Volatility quintile, this relationship seems less evident as the Sharpe Ratio first increases from 0.102 to 0.104 moving from Small to quintile 2 (Size), then decreases again to 0.088 in quintile 3 and 0.089 in quintile 4, before increasing again in the Big quintile.

Also, the Size effect is confirmed in this Table. The returns are higher for small stocks than for big stocks, in every Volatility quintile except the High Volatility quintile. For example, in the 3rd Volatility quintile the Small portfolio has a return of 0.978% per month while the Big portfolio only has a return of 0.68%. The High Volatility quintile is the only one with different results; here the Small portfolio has a return of 0.752% while the Big portfolio outperforms with a return of 0.865%.

To summarize this subsection, the results show that using value-weighted returns and NYSE breakpoints to form the portfolios decrease the performance of the low volatility portfolio and increase the size of the high volatility portfolio (slightly). Thus, it somewhat suppresses the low volatility anomaly. However, not enough to make the effect disappear. To illustrate, the average monthly return of the low volatility portfolio is (still) slightly higher than the high volatility portfolio, and besides that is has a lower standard deviation of returns, so together the low volatility portfolio performs better than the high volatility portfolio, thus confirming the anomaly. Furthermore, this section also took into account the role of size. Specifically, these results indicate that the low volatility portfolio outperforms the high volatility portfolios for all size groups in terms of Sharpe Ratio, and for all except the Big size group in terms of returns.

5.4 Performance over time

In this section, the performance of the anomaly is examined over time. To analyze the evolution of the anomaly over time, the sample will be split up per decade. A consequence of this is that the first decade (1960 to 1969) will contain be smaller as the data set only starts in 1963 and volatility is calculated based on past 3-year returns. This provides 6 decades to analyze. The results are published in Table 5.6. In Panel A, the monthly excess returns and the t-statistics will be presented. In Panel B, the standard deviation of returns and the Sharpe Ratios will be reported. Again, the portfolios are formed using NYSE breakpoints and returns are value weighted. Stocks with an average price below $5 are dropped and returns are winsorized.

First, the results of Panel A will be discussed (returns), and then Panel B will be discussed (performance). Then, the results will be summarized and concluded.

As can be seen in Table 5.6, the return patterns are different per time period. The highest returns can be found in the period of 1990-1999, where the low volatility portfolio has an average monthly return of 1.098% and the highest volatility portfolio an average monthly return of 1.736%. Furthermore, the Low portfolio outperformed the High portfolio in terms of returns in 3 of the 6 periods, in 1980-1989, 2000-2009 and 2010-2019.

Starting at the earliest time period, 1966-1969, the high volatility portfolio outperformed the low volatility portfolio, showing a monthly average return of 0.522% (an insignificant t-statistic of 0.38) whereas the low volatility portfolio shows a monthly average return of -0.142%. Portfolio 2 already shows a higher return of 0.214% (although insignificant with a t-statistic of 0.38). Returns then almost monotonically increase with higher volatility portfolios with exceptions of portfolios 3 and 9. For example, portfolio 8 has an average monthly return of 0.370% with a t-statistic of 0.39. However, none of the returns are significant.

In the time period 1970-1979, the lower volatility portfolios deliver lower returns than the other portfolios. Portfolios 5 up to the High portfolio deliver an average monthly return of between 0.6% and 0.7% (with insignificant t-statistics of around 1.00). Moving down to the lower volatility portfolios, returns decrease as volatility decreases, with the Low Volatility portfolio only showing a return of 0.226% (insignificant t-statistic 0.52). Clearly, for the first 2 decades in the sample (so from 1966 to 1979), the Low Volatility portfolios did not outperform the High Volatility portfolios in terms of returns (although all returns are insignificant).

In the time period 1980-1989, the results look different. The Low Volatility portfolio outperforms the High Volatility portfolio, with a monthly average return of 0.987% (t-statistic 2.99) and a return of -0.008 (t-statistic -0.01) respectively. Also taking into account the second most (or least) volatile portfolios, it becomes evident that the Low volatility anomaly is still present here. Portfolio 2 shows a return of 1.131% (so a higher return than the Low portfolio) with a t-statistic of 2.61 and portfolio 9 shows a return of 0.410% with a t-statistic of 0.62. Portfolio 2 actually shows the highest average monthly return in this decade, higher than portfolio 3 (with a return of 1.120 and t-statistic 2.42). Returns actually decrease from portfolio 5 to higher volatility portfolios (and t-statistics decrease too). In this decade, it seems that higher volatility delivered lower returns than lower volatility.

In the period 1990-1999, returns overall were high. And the returns are very significant as well. The Low Volatility portfolio has a return of 1.098% (t-statistic of 3.07) and the High Volatility portfolio has a (higher) return of 1.736% per month (on average) with a t-statistic of 2.95. All the returns of the portfolios are around 1.1% on average per month, except for portfolios 2 (.841%, t-statistic 2.29) and portfolios 9 and the High portfolio. Portfolio 9 has a return of 1.436 (significant t-statistic of 3.37), which indicates that in this time period, high volatility portfolios outperformed the lower volatility portfolios in terms of return.

Table 5.6: Performance over time

The Table shows the average 1-month holding returns in excess of the risk free rate, the standard deviation of returns (SD) and the Sharpe Ratios for portfolios formed every month by sorting stocks on their past 3-year return volatility. The sample runs from January 1963 to December 2019 and includes all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. Stocks are sorted into portfolios each month using NYSE breakpoints, and returns are value-weighted by size. Market Equity (ME) is used to measure size. Stocks with an average price below $5 in the sample are excluded. Returns are winsorized at the 1% level. Reported t-statistics are Newey-West t-statistics. The sample is divided into decades. Pane A shows the returns and the t-statistic whether the returns are significantly different from zero, per decade. Panel B shows the standard deviation of returns (SD) and the Sharpe ratio of the portfolios, per decade.

Time period	Decile	Low	2	3	4	5	6	7	8	9	High
Panel A: Returns											
1966-1969	Return	-.142	.214	.039	.214	.242	.302	.351	.370	.250	.522
	t-statistic	(-0.26)	(0.38)	(0.06)	(0.28)	(0.29)	(0.33)	(0.36)	(0.39)	(0.20)	(0.38)
1970-1979	Return	.226	.361	.466	.484	.637	.592	.698	.660	.666	.612
	t-statistic	(0.52)	(0.82)	(1.02)	(1.00)	(1.20)	(1.01)	(1.07)	(0.95)	(0.94)	(0.73)
1980-1989	Return	.987	1.131	1.120	.897	.899	.691	.641	.479	.410	-.008
	t-statistic	(2.99)	(2.61)	(2.42)	(1.79)	(1.74)	(1.31)	(1.15)	(0.81)	(0.62)	(-0.01)
1990-1999	Return	1.098	.841	1.174	.956	1.107	1.049	1.110	1.194	1.436	1.736
	t-statistic	(3.07)	(2.29)	(3.14)	(2.93)	(2.94)	(3.11)	(3.52)	(3.35)	(3.37)	(2.95)
2000-2009	Return	.567	.322	.323	.508	.539	.588	.644	.574	.634	.142
	t-statistic	(1.21)	(0.81)	(0.64)	(0.95)	(0.97)	(1.01)	(0.97)	(0.85)	(0.85)	(0.16)
2010-2019	Return	.989	1.209	1.177	1.125	.926	.979	1.065	.869	.831	.733
	t-statistic	(3.38)	(5.05)	(4.55)	(4.16)	(2.77)	(2.84)	(3.24)	(2.20)	(1.94)	(1.39)

Time period	Decile	Low	2	3	4	5	6	7	8	9	High
Panel B: Sharpe Ratio											
1966-1969	SD (return)	.550	.559	.630	.756	.845	.914	.979	.959	1.239	1.381
	Sharpe Ratio	-0.258	0.383	0.062	0.283	0.286	0.330	0.359	0.386	0.202	0.378
1970-1979	SD (return)	.433	.438	.454	.482	.531	.587	.655	.696	.711	.837
	Sharpe Ratio	0.522	0.824	1.026	1.004	1.200	1.009	1.066	0.948	0.937	0.731
1980-1989	SD (return)	.330	.434	.462	.502	.517	.525	.555	.592	.663	.724
	Sharpe Ratio	2.991	2.606	2.424	1.787	1.739	1.316	1.155	0.809	0.618	-0.011
1990-1999	SD (return)	.357	.368	.374	.326	.377	.337	.315	.357	.426	.589
	Sharpe Ratio	3.076	2.285	3.139	2.933	2.936	3.113	3.524	3.345	3.371	2.947
2000-2009	SD (return)	.469	.397	.504	.534	.557	.581	.664	.675	.747	.901
	Sharpe Ratio	1.209	0.811	0.641	0.951	0.968	1.012	0.970	0.850	0.849	0.158
2010-2019	SD (return)	.292	.239	.258	.270	.334	.345	.329	.394	.428	.527
	Sharpe Ratio	3.387	5.059	4.562	4.167	2.772	2.838	3.237	2.206	1.942	1.391

The period 2000-2009 shows lower overall returns than the previous period. Also, the returns are insignificant with the highest t-statistic in the Low Volatility portfolio (t-statistic 1.21, so still insignificant). The returns seem to be the same for the volatility sorted portfolios, except for portfolios 2, 3 and the High portfolio. For the other portfolios, the returns fluctuate around 0.550%. For portfolios 2 and 3, the average monthly return is 0.322 and 0.323% respectively, and for the High Volatility portfolio the return is the lowest of the portfolios in this time period with an average monthly return of 0.142% (insignificant t-statistic of 0.16). Because the returns are all very insignificant, no hard conclusions can be drawn but it does seem to be the case that the High Volatility portfolio underperforms compared to the other portfolios in this time period.

For the most recent period of 2010-2019, the returns look more like those observed in the period 1990-1999. First of all, the highest return in this time period is achieved by portfolio 2 (and after that portfolios 3 and 4). Portfolio 2 achieves an average monthly return of 1.209% with a t-statistic of 5.05, making it significant even at the 1% level. Portfolios 3 and 4 show a return of 1.177% (t-statistic 4.55) and 1.125 (t-statistic 4.16) respectively. Comparing the Low and High Volatility portfolios, the Low portfolio shows a return of 0.989% (significant t-statistic of 3.38) while the High Volatility portfolio shows a return of 0.733% (insignificant t-statistic of 1.39). To put the returns of the higher volatility portfolios into context, only portfolio 7 shows a higher return than the Low portfolio, indicating that lower volatility portfolios clearly outperformed the higher volatility portfolios in terms of return in this decade.

Analyzing the t-statistics, it becomes clear that the first time period does not show significant results. This can be due to the smaller sample size, as volatility is calculated over the prior 3 years, and the entire sample only starts in 1964. Therefore for this decade, the results are possibly not very reliable. Further looking at the t-statistics, the results become more significant by the decade, with the exception of 2000-2009. For example, looking at the Low Volatility portfolio, the returns first become significant in the period 1980-1989, with a t-statistic of 2.99 (average monthly return 0.987). In the period 2010-2019, the low volatility portfolio delivered an average monthly return of 0.989% with a t-statistic of 3.38.

Panel B of Table 5.6 shows the standard deviation of returns (SD) of the returns in Panel A. In the second row, the Sharpe Ratio is calculated (see section 4 to see how it is calculated). In the previous paragraphs the returns were analyzed, now the performance measured in terms of Sharpe Ratio will be analyzed. To do this, the z-score will be used. This will answer the question of whether the Low Volatility portfolios perform better in terms of risk/reward over the time periods.

Because the portfolios are sorted on volatility, naturally the standard deviation of return is increasing with each portfolio. Strictly looking at the evolution of standard deviation of returns over time for the High Volatility portfolio, it becomes evident that some decades were more volatile than others. As mentioned before, the period 1966-1969 is somewhat unreliable because of the small sample size (47 months in this decade). Disregarding this period, the standard deviation of returns for the High portfolio dropped from the 1970s to the 1990s (from 0.837 to 0.589). The period 2000-2009 seems to be the most volatile with a standard deviation of returns of 0.901 for the High portfolio. On the other end, for the Low Volatility portfolio, returns became more stable moving from the 1970s to the 1990s (standard deviation of return of 0.433 to 0.357), and became slightly more volatile in the period 2000-2009 (standard deviation of return of 0.469). The period 2010-2019 was the least volatile for the Low portfolio with a standard deviation of monthly average returns of 0.292.

In terms of Sharpe Ratio, which captures the returns and its standard deviation in one number, the highest Sharpe Ratio can be found in the most recent period, 2010-2019. For portfolio 2, the Sharpe Ratio was 5.059. This is in line with the returns that were described in the previous paragraphs. Furthermore, the Low portfolio outperformed the High portfolio in terms of Sharpe Ratio in 4 of the 6 periods, that is in the periods 1980-1989 (z-score 14.006), 1990-1999 (z-score 0.597, insignificant even at the 10% level), 2000-2009 (z-score 8.029) and 2010-2019 (z-score 9.434). In the 1960s the z-score for the Low portfolio is -6.254, and for the 1970s -2.949. The negative z-score implies the Sharpe ratio of the Low portfolio was lower than that of the High portfolio. All z-scores, except for the period 1990-1999 are significant even at the 1% level.

In the period 1966-1969, the Low portfolio has a negative Sharpe Ratio but as mentioned before, might be unreliable due to the small sample size of this time period. In the 1970s, the Low portfolio showed a Sharpe Ratio of 0.522 while the High portfolio outperformed with a Sharpe Ratio of 0.731. This is because the returns were almost 3 times as high while the standard deviation of return was only two times as high (approximately). This is confirmed by the z-score of -2.949 (p-value 0.000) which implies that the Sharpe of the Low portfolio is almost 3 standard deviations lower than the Sharpe of the High portfolio. Portfolio 2 shows a somewhat higher Sharpe Ratio but this is still lower than that of portfolio 9. The highest Sharpe Ratio in this time period can be found in portfolio 5, right in the middle. It seems that in this time period, the Low Volatility anomaly was not present in terms of Sharpe Ratio.

In the period 1970-1979, the Sharpe ratio of the Low portfolio is still lower than the Sharpe ratio of the High portfolio. The Low portfolio has a Sharpe ratio of 0.522 where the High portfolio has a Sharpe ratio of 0.731. Comparing them, the z-score is -2.949 (p-value 0.002) implying that the Sharpe ratio of the Low portfolio is almost 3 standard deviations lower than that of the High portfolio, which is significant at the 1% level. Furthermore, the Sharpe ratios seem to increase the more volatile the portfolios get. For example, portfolio 5 has a Sharpe ratio of 1.200. In this period, the low volatility anomaly does not seem to exist in terms of Sharpe ratio (and as pointed out before, neither in terms of returns).

In the next time period, 1980-1989, the Low Volatility anomaly is very much present. The Low portfolio obtains the highest Sharpe Ratio of all portfolios in this time period with a Sharpe Ratio of 2.991. Sharpe Ratios then monotonically decrease with each portfolio, reaching the lowest in the High portfolio with a negative Sharpe Ratio of -0.011. The presence of the anomaly in this period seems to be due to the relatively low standard deviations of returns for the lower-end portfolios paired with the relatively higher returns these portfolios generated. The anomaly thus prevailed in terms of return and performance. Comparing the Sharpe ratio of the Low portfolio with that of the High portfolio, the z-score is 14.006 (p-value 0.000) implying that the Sharpe is 14 standard deviations higher (significantly) than the High portfolio. The Sharpe of the High portfolio is -0.011.

The period 1990-1999 shows somewhat mixed results. First of all, the highest Sharpe Ratio in this period can be observed in portfolio 9 (Sharpe Ratio of 3.371). The Low portfolio does not underperform by a large margin (Sharpe Ratio of 3.076), but this is not a result that is consistent with the idea of the Low Volatility anomaly. The Low portfolio still performs better than the High portfolio (which has a Sharpe Ratio of 2.947) but only slightly. As mentioned before, the z-score is 0.597 (insignificant) implying that statistically, the Sharpe ratios are not different.

Analyzing further, it becomes clear that the standard deviations of returns are quite close together for portfolios 1 through 8. To illustrate, portfolio 1 has a standard deviation of returns of 0.357 while portfolio 8 has the same (!) standard deviation. Portfolio 9 and the High portfolio then show a slightly higher standard deviation of returns (0.426 and 0.589 respectively) but their returns compensate for the increase in risk, which is what the Sharpe Ratio captures. In terms of reward for risk, all portfolios seem to be very equal. So although the Low portfolio (slightly) outperforms, there is no clear cut evidence that says the Low Volatility anomaly was present in this time period in terms of performance.

In the period 2000-2009, the anomaly strongly appears again. Just as in the period 1980-1989, the Low portfolio shows the highest Sharpe Ratio (1.209), outperforming the High portfolio (0.158). The z-score is 8.029 (p-value 0.000), implying that the Sharpe ratio of the Low portfolio is 8 standard deviations higher than the Sharpe ratio of the High portfolio, which is significant even at the 1% level. Looking at the second-most extreme portfolios, the performance of portfolio 9 is slightly better than that of portfolio 2, with Sharpe Ratios of 0.849 and 0.811 respectively. Here, the standard deviation of returns is higher for portfolio 9 (0.747) than portfolio 2 (0.397), but as becomes clear from the Sharpe Ratio, the returns compensate for this higher risk. Apparently, for this time period, the anomaly was present in the 2 extreme portfolios but somewhat disappeared when looking at the next 2 portfolios.

In the most recent time period, 2010-2019, the anomaly again prevailed. The Low portfolio outperformed the High portfolio in terms of Sharpe Ratio with 3.387 against 1.391 respectively. The z-score is 9.434 (p-value 0.000) implying the Sharpe ratio of the Low portfolio is more than 9 standard deviations higher than the High portfolio, which is significant even at the 1% level. The adjacent portfolios also show the same results. The superior performance of the lower-end portfolios seems to be due to the relatively low standard deviations of returns, with the standard deviation of returns of portfolios 2, 3 and 4 actually being slightly lower than that of the Low portfolio. Clearly, the Low Volatility anomaly was present in the most recent time period in terms of performance.

Finally, the evolution of the Sharpe Ratio of the Low and High portfolios can be compared over the time periods (by looking down in the Low or High Column). For the Low portfolio, the Sharpe Ratios increased every decade with a dip in the period 2000-2009, where it dropped from 3.076 to 1.209, increasing again in the next period (2010-2019) to 3.387. The adjacent portfolio 2 shows almost the same evolution except for the period 1990-1999 where it was slightly lower than 1980-1989. The evolution of the Sharpe Ratio of the High portfolio is less linear. After a slight increase from the 1960s to the 1970s to 0.731, it decreases sharply in the 1980s, before showing a turnaround in the 1990s to 2.947, and then dipping down in the period 2000-2009 to 0.158. It picks up again in the recent period 2010-2019 to 1.391. The adjacent portfolio 9 shows the same evolution. The z-scores can be found in table 8.6 in the Appendix. The z-scores confirm the patterns described above. For the Low portfolio, the Sharpe ratio in the 1990s in not significantly different from the Sharpe ratio in the 1980s, as indicated by the z-score 0.374 (p-value 0.708, which is not significant at any level). The other z-scores are significant at the 1% level, indicating that the evolution of the Sharpe ratio of the Low portfolio over time is significant at every step. For the High portfolio, the evolution of the Sharpe ratio is also significant at every step (so from decade to decade).

To summarize and answer the research question (How does the anomaly perform over time?), in terms of returns the Low portfolio outperformed the High portfolio in terms of returns in 3 of the 6 periods, in 1980-1989, 2000-2009 and 2010-2019. In terms of Sharpe Ratio the Low portfolio outperformed the High portfolio in 4 of the 6 periods, (in the periods 1980-1989, 1990-1999, 2000-2009 and 2010-2019). Thus, except for the 1990s, the Low Volatility strategy outperformed in terms of returns and Sharpe Ratio. In the 1990s the Low portfolio only provided a better risk/reward ratio, but not actual higher returns. It seems that for the period 1960-1969 and 1970-1979, investing in low volatility portfolios did not pay off in terms of returns and performance. Over time, the Sharpe Ratio of the Low portfolio increased while that of the High portfolio shows no clear pattern. Thus, to conclude, the anomaly seems to have gotten stronger over time.

5.5 Regressions on the market portfolio

To make an attempt to explain why the anomaly behaves like described above, a regression of the Low and High portfolios can be done in each decade, to see if either the alphas or market betas change over time, and whether the Low or High portfolio display any patterns in coefficients over time. For example, Blitz & van Vliet (2007) find that the low volatility portfolio has a low beta (0.58) and a positive alpha (4% per year), whereas the high volatility portfolio shows a somewhat higher beta (1.58) with a negative alpha (8% per year). However, this result is over their entire sample which runs from 1986-2006. Over time, they find that the low volatility portfolios generally underperform the market during bull markets (which suggests a beta lower than 1 during these periods) while it outperforms during down markets (also suggesting a beta lower than 1 during these periods). The high volatility portfolio then shows the exact opposite behavior.

In Table 8.8, (stored in section 8), the regression coefficients are presented from the regression of the Low and High portfolio on the excess market return in the 1960s. In this period, for both the Low and High portfolio, the alpha and beta are insignificantly different from zero. In this period the average monthly return of the market portfolio was 0.374% while the Low portfolio delivered a return of -0.142%. The standard deviation of return is about the same for these two portfolios (3.320 and 3.167, respectively). The High portfolio showed an average monthly return of 0.522%, but this is accompanied by a higher standard deviation of returns of 8.622.

The results for the 1970s are presented in Table 8.9. Here, the beta and alpha coefficients for the Low and High portfolio are not significantly different from zero either. The Low portfolio shows a return of 0.226% (standard deviation of returns of 3.765) while the market portfolio has an average monthly return of 0.148 with a higher standard deviation of return (4.765). The High portfolio has a return of 0.612% again with a higher standard deviation of return (8.109). Comparing the lowest monthly return achieved by the 3 portfolios in this time period, it seems that the low portfolio performs better than the market portfolio (-9.671% compared to -12.750%, respectively).

In Table 8.10 the results for the 1980s are presented. The alpha of the Low portfolio obtained from regressing this portfolio on the excess market return is 0.938% and is significant at the 1% level (t-statistic 3.23). This means that this portion of the returns is not explained by the market return in this time period. The Low portfolio seems to have higher returns than the market portfolio in this period (0.987% compared to 0.600%, respectively). Also, the max drawdown of the Low portfolio is less negative than that of the market portfolio (-15.426% compared to -23.240%). The highest monthly returns do however seem lower for the Low portfolio (8.002%) compared to the market portfolio (12.470%) and the High portfolio (16.393%).

The results for the 1990s are presented in Table 8.11. Here, again the alpha (1.209%) for the Low portfolio is significant at the 1% level (t-statistic 4.24). The beta is negative but insignificantly different from zero. The High portfolio has an alpha (1.498%) that is significant at the 5% level (t-statistic 2.27). The beta is positive but insignificant. Looking at returns, the Low portfolio seems to have lower average monthly returns (1.098%) than the market portfolio (1.209%) and the High portfolio (1.736%). The standard deviation of returns of the Low portfolio does however seem to be lower than that of the market portfolio (2.806 compared to 3.753).

In Table 8.12 the results for the 2000s are presented. In Panel B, the alphas of the Low and High portfolios are now insignificantly different from zero. However, the Low portfolio has a beta of 3.71, significant at the 1% level. The High portfolio has a beta of 3.14 significant at the 1% level. Apparently, it seems that the portfolio returns behave very similarly to that of the market in this time period. The Low portfolio seems to have higher average monthly returns in this period (0.567%) than both the High portfolio (0.142%) and the market portfolio (-0.010%). It also seems to have a lower standard deviation of returns than the market portfolio (4.443 compared to 4.743, respectively). The max drawdown for the Low portfolio is -12.086% whereas the market portfolio has a max drawdown of -17.230%. Furthermore, the highest monthly return achieved in this period is 14.317% for the Low portfolio whereas for the market portfolio it is 10.190%.

The results for the 2010s are presented in Table 8.13. Here, in Panel B the betas for the Low and High portfolio are both negative. The beta for the High portfolio is insignificantly different from zero (t-statistic -1.20) but the beta for the Low portfolio is significant at the 5% level (t-statistic -1.99). The alpha for the High portfolio is insignificantly different from zero (t-statistic 1.49) whereas the Low portfolio has a significant alpha of 1.191% (t-statistic 3.31, significant at the 1% level). In this period, the market portfolio seems to have higher returns (1.047% than both the Low portfolio (0.989%) and the High portfolio (0.733%). In terms of standard deviation of returns, the market portfolio seems to perform slightly better than the Low portfolio (3.498 compared to 3.626, respectively). Analyzing max drawdown, the market portfolio seems to perform slightly worse (-9.550%) than the Low portfolio (-8.521%). In terms of the highest returns in this period, the market portfolio seems to perform better (11.350%) than the Low portfolio (9.509%).

To summarize the results of the regression tests per time period, for 4 out of 6 periods the Low portfolio seems to have a positive beta, although the beta is only significantly different from zero for the 2 most recent time periods. The High portfolio seems to have a positive beta in 5 out of 6 time periods, although the beta is only significantly different from zero in 1 time period. The Low portfolio has a significant alpha in 3 out of the 6 periods and the High portfolio only has 1 significant alpha (at the 5% level). The alphas and betas do not seem to indicate any patterns that may suggest the behavior of the anomaly over time. The results do seem to indicate that the maximum drawdown of the Low portfolio is lower than that of the market portfolio (in 6 of the 6 periods). The highest returns the Low portfolio achieves seem to be lower in 3 out of the 6 time periods (and seem to be higher in the other 3), so no conclusions can be drawn about the upward potential in this case. The lower maximum drawdown is in line with the findings of Blitz & van Vliet (2007).

Finally, to conclude the Empirical Results section, the low volatility anomaly is present in this dataset, even after using NYSE breakpoints and value-weighted returns. The anomaly is also present in all size-groups, although it does not outperform in the biggest 20% of stocks in terms of returns (but it does in terms of Sharpe Ratio). Over time, the anomaly has increased in performance. Analyzing the market betas and alphas from a 1 factor regression across 6 time periods does suggest why the anomaly changes in behavior over time.

6 Robustness of the Anomaly

The anomaly can be further tested by doing some robustness checks. The value-weighted, NYSE breakpoint portfolios will be used as the main Model, and the following robustness tests will be done against this Model: using decile or quintile portfolios, using winsorized compared non-winsorized returns, dropping stocks below $5 or not, using volatility or idiosyncratic volatility and finally a comparison will be made between Volatility and Beta. The aim of these robustness checks is to gain insights into whether the results that appear in the main section are limited to the chosen methodologies. In the construction of the volatility portfolios, certain assumptions or choices are made. By relaxing the assumptions or choices one by one, and comparing the before and after results, insights can be given into the effects of each of these choices. If the performance only slightly changes, it means that the choice made has little impact. If the results are results are truly affected by a particular choice then indicates that the results of the anomaly are (to some degree) due to methodological choices. For the anomaly to be robust, the results should not be greatly impacted by the methodological choices.

6.1 Decile vs. quintile portfolios

The following section presents the portfolio performance comparison between using deciles or quintiles. To do this, the results as seen before for the 10 value-weighted portfolios sorted on volatility using NYSE breakpoints are presented in Table 6.1, Panel A. Then, this panel is compared to Panel B, where the same methodology is applied except that now stocks are sorted into 5 portfolios each month. The differences will be highlighted and then a conclusion will be drawn about this robustness test.

In Table 6.1, the difference in return between the lowest and highest portfolio becomes smaller if the stocks are sorted into quintiles instead of deciles. Panel A shows the performances of the decile portfolios and Panel B shows this for the quintile portfolios. In Panel A there is a relatively sharp decline in average returns moving from portfolio 9 to portfolio 10, where the return drops from 0.785% (t-statistic 3.14) to 0.639% (t-statistic 2.12) per month. Looking at the Newey-West t-statistic, it also becomes less significantly different from zero. Using quintiles (Panel B), the most volatile portfolio (quintile 5) has higher returns. Now, the return is 0.696% with a t-statistic of 2.51. This is slightly higher than quintile 1, where there is a return of 0.692% per month with a t-statistic of 4.58. The difference is highlighted in the last Column, where the return difference between quintile 5 and quintile 1 is 0.004% per month, although this is very insignificantly different from zero (t-statistic .02). In comparison, the last Column in Panel A shows that the most volatile portfolio delivers lower returns than the low volatility portfolio as the return difference between portfolio 10 and 1 is -0.039%, although this too is insignificantly different from zero (t-statistic -0.039).

So, using quintiles does not significantly make a difference in the return difference of the Low and High

Table 6.1: Decile vs Quintile Performance

The Table shows the average 1-month holding returns in excess of the risk free rate, the standard deviation of returns (SD) and the Sharpe Ratios for portfolios formed every month by sorting stocks on their past 3-year return volatility. The sample runs from January 1963 to December 2019 and includes all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. Stocks are sorted into portfolios each month using NYSE breakpoints, and returns are value-weighted by size. Market Equity (ME) is used to measure size. Stocks with an average price below $5 in the sample are excluded. Returns are winsorized at the 1% level. Reported t-statistics are simple t-statistics and show whether the returns are significantly different from zero. Panel A shows the results when stocks are sorted into 10 portfolios (deciles), and Panel B shows the results when stocks are sorted into 5 portfolios (quintiles).

Decile	Low	2	3	4	5	6	7	8	9	High	10-1
Panel A: Decile Portfolio Performance											
Return	.678	.735	.820	.758	.790	.766	.831	.758	.785	.639	-.039
t-statistic	(4.81)	(4.85)	(4.69)	(4.26)	(4.21)	(3.90)	(3.94)	(3.35)	(3.14)	(2.12)	(-.15)
SE	.141	.152	.175	.178	.187	.196	.211	.226	.250	.301	.258
SD (return)	3.59	3.85	4.44	4.52	4.76	4.99	5.36	5.74	6.35	7.66	6.55
Sharpe Ratio	.189	.191	.185	.168	.166	.154	.155	.132	.124	.083	

Quintile	Low	2	3	4	High	5-1
Panel B: Quintile Portfolio Performance						
Return	.692	.815	.779	.797	.696	.004
t-statistic	(4.58)	(4.56)	(4.09)	(3.68)	(2.51)	(.02)
SE	.151	.179	.191	.217	.277	.227
SD (return)	3.84	4.54	4.84	5.50	7.05	5.77
Sharpe Ratio	.180	.180	.161	.145	.099	

portfolio. Next to the return, the standard deviation of return can be analyzed. In both Panel A and Panel B, the standard deviation (SD) of returns of the low volatility portfolio is much lower than the high volatility portfolio. In Panel A, decile 1 shows a standard deviation of returns of 3.59 while decile 10 a standard deviation of returns of 6.55. The Sharpe Ratio is a measure that captures the reward for risk in a single number. It is computed by dividing the return in excess of the risk free rate by the standard deviation of returns. For decile 1, the Sharpe Ratio is 0.189. Decile 10 shows a lower Sharpe Ratio of 0.083. Moving to Panel B, quintile 1 has a standard deviation of returns of 3.84 while quintile 5 a standard deviation of 7.05. The Sharpe Ratio is 0.180 for quintile 1 while it is 0.099 for quintile 5. Using the z-score to compare the Sharpe ratios as presented in table 8.5 (Appendix), it becomes evident that the Low portfolio for both methodologies does not significantly differ, as the z-score is -0.681 (quintile Low compared to decile Low), with a p-value of 0.496. The High portfolios do differ significantly, as the z-score is 3.685 (p-value 0.000) is significant even at the 1% level. This means that the Sharpe ratio of the quintile High portfolio is more than 3 standard deviations higher than the Sharpe ratio of the decile High portfolio.

Furthermore, the performances of portfolios are compared using a variance comparison test (see Table 8.1, Section 8). Here, in Table 8.1, Panel A, the decile Low portfolio is different in variance compared to the quintile Low portfolio, significant at the 10% level. The decile High portfolio is significantly different from the quintile High portfolio at the 5% level. Comparing the average returns of the portfolios using a two-sample t-test of equal means produces no significant results. This implicates that using the choice between decile or quintile portfolios has an effect on the variance of the returns (but not on the average return) of the Low and High portfolio in this sample.

To conclude this robustness test, in terms of performance as measured by the Sharpe Ratio, the low volatility portfolio outperforms the high volatility portfolio by having a higher Sharpe Ratio, which is achieved by having about the same average returns but a lower volatility of returns. This effect is robust to choice of portfolios, it happens in both the decile portfolios and quintile portfolios, although the differences become somewhat smaller. However, the methodological choice does affect the results. The test of equal variances confirms that using quintiles produces a lower variance of returns in the High portfolio. Using quintiles increases the Sharpe ratio of the High portfolio significantly, which indicates that the 20% most volatile stocks perform significantly better than the top 10%. The choice of using deciles or quintiles therefore affects the results. However, the anomaly still exists in the quintile portfolios (in terms of Sharpe ratio).

6.2 Winsorized vs. non-winsorized

In this section, it is explored how winsorizing the returns at the 1% level affects the results. The comparison is presented in Table 6.2. Here, Panel A contains the main Model that is used throughout the paper, and Panel B will contain the exact same setup except that now the returns are not winsorized. After a short introduction, the results will be analyzed and the differences will be highlighted. Then a conclusion will be drawn about this robustness test.

First, the purpose and effect of winsorizing will be repeated. As only the top (and bottom) 1% is winsorized, not that much of an impact is expected unless the returns are skewed, and these extreme returns are very high. Apparently, this is the case in the sample. Because returns in decile 10 drop quite substantially (0.16% per month) after winsorizing, the returns that are in fact winsorized must be quite high to have such an impact. Because they are in the top 1% (and looking at the standard deviation of return, quite volatile as well), these stocks behave as lottery stocks, stocks that display unusual behavior by delivering extreme returns. To include them in an investment strategy is not realistic as they can be considered outliers, so winsorizing the returns gives us a more realistic view.

It becomes evident in Table 6.2 that winsorizing returns to some degree affect the results. From a return perspective, the higher portfolios have lower returns after winsorizing. For example, in Panel A, decile 6 has a return of 0.766% per month (t-statistic 3.90) while in Panel B (not winsorized) the portfolio has a return of 0.787% per month (t-statistic 4.00). Later in this section, it will be analyzed whether the returns are

Table 6.2: Winsorized vs Non-Winsorized Returns
The Table shows the average 1-month holding returns in excess of the risk free rate, the standard deviation of returns (SD) and the Sharpe Ratios for portfolios formed every month by sorting stocks on their past 3-year return volatility. The sample runs from January 1963 to December 2019 and includes all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. Stocks are sorted into portfolios each month using NYSE breakpoints, and returns are value-weighted by size. Market Equity (ME) is used to measure size. Stocks with an average price below $5 in the sample are excluded. Reported t-statistics are simple t-statistics and show whether the returns are significantly different from zero. Panel A shows the performance when using winsorized returns and Panel B shows the performance using non-winsorized returns.

Decile	Low	2	3	4	5	6	7	8	9	High
Panel A: Winsorized Returns, Portfolio Performance										
Return	.678	.735	.820	.758	.790	.766	.831	.758	.785	.639
t-statistic	(4.81)	(4.85)	(4.69)	(4.26)	(4.21)	(3.90)	(3.94)	(3.35)	(3.14)	(2.12)
SD (return)	3.59	3.85	4.44	4.52	4.76	4.99	5.36	5.74	6.35	7.66
Sharpe Ratio	.189	.191	.185	.168	.166	.154	.155	.132	.124	.083
Panel B: Non-Winsorized Returns, Portfolio Performance										
Return	.680	.735	.822	.753	.792	.787	.836	.780	.820	.793
t-statistic	(4.80)	(4.85)	(4.70)	(4.20)	(4.20)	(4.00)	(3.90)	(3.35)	(3.20)	(2.50)
SD (return)	3.59	3.86	4.46	4.54	4.80	5.03	5.48	5.95	6.53	8.15
Sharpe Ratio	.189	.190	.184	.166	.165	.156	.153	.131	.126	.097

significantly different from each other. Portfolios 7 through 10 display the same pattern of higher returns. The difference is especially prominent in decile 10. Here, the non-winsorized returns are 0.793% while the winsorized returns deliver us 0.639% per month (t-statistic 2.50 and 2.12, respectively). The higher return goes paired with a higher standard deviation of return of 8.15 as compared to 7.66 for the winsorized returns, but in terms of Sharpe Ratio the non-winsorized portfolio performs better than its winsorized counterpart (a Sharpe Ratio of 0.097 compared to a lower Sharpe Ratio of 0.083, respectively).

First of all, this means that decile 1 does not perform better anymore than decile 10 in terms of return. Whereas for the winsorized returns, decile 1 outperforms decile 10 by delivering a return of 0.678% per month versus a return of 0.639% per month, a different result appears in Panel B. Here, decile 1 underperforms decile 10 with a return of 0.680% per month versus the higher 0.793% per month. However, it should be noted that in both cases, the returns for the lowest decile are much more significantly different from zero than the high decile. In Panel A, returns of decile 1 have a t-statistic of 4.81 while returns of decile 10 have a t-statistic of 2.12. This is about the same in Panel B (t-statistic of 4.80 and 2.50, respectively). However, as analyzed in the paragraph below, this does not translate into a significant difference between the returns of the portfolios. In terms of Sharpe Ratio, decile 1 still performs better than decile 10 even though the returns are lower when the returns are not winsorized. This is because the standard deviation of returns has increased in decile 10 from 7.66 to 8.15. In Panel A, decile 1 has a Sharpe Ratio of 0.189 while decile 10 has a Sharpe Ratio of 0.083. In Panel B, the difference is slightly smaller. Decile 1 still has a Sharpe Ratio of

0.189 and decile 10 has a Sharpe Ratio of 0.097. The z-score that compares the Non-winsorized Model with the Main Model (using winsorized returns) shows that the Sharpe ratio of the Low portfolio is statistically speaking not different (z-score 0.000). Comparing the High portfolios produces a z-score of 3.256 (p-value 0.001), meaning the Sharpe ratio of the High portfolio without winsorizing returns has a Sharpe ratio that is more than 3 standard deviations higher than the Sharpe ratio of the High portfolio with winsorized returns, significant at the 1% level.

Furthermore, the comparison of variance test (see Table 8.1, Section 8) shows there is no significant difference in variance between the Low or High portfolios between the two methodologies. Also, the two-sample t-test of equal means (see 8.2, Section 8) indicates there are no significant differences in the average returns between the methodologies (when comparing the Low or High portfolios). This means that the choice to winsorize the returns or not does not lead to a significant difference in variance or average return in the Low or High portfolios in this sample, but the Sharpe ratio of the High portfolio without winsorizing returns is (statistically) significantly higher than the Sharpe ratio of the High portfolio when winsorizing returns.

To conclude this robustness test, whether returns are winsorized or not, the Low Volatility anomaly still exists in the sample. Even though the most volatile returns are more affected than the least volatile returns by winsorizing (on both sides, 1% level), the idea that low volatility portfolios perform better than high volatility portfolios is still true in the sample as confirmed by the test of equal variance and equal means, which indicate that winsorizing does not lead to a statistically significant effect on variance or average returns in the Low or High Volatility portfolio. Winsorizing does however lead to a significant decrease in Sharpe ratio in the High portfolio. Still, the anomaly exists in both methodological settings.

6.3 Dropping stocks below $5 or not

In this section, it is explored how dropping stocks with an average price below $5 affects the results. The comparison is presented in Table 6.3. Here, Panel A contains the main Model that is used throughout the paper, and Panel B will contain the exact same setup except that now these stocks are not excluded from the sample. The results will be analyzed and the differences will be highlighted. Then a conclusion will be drawn about this robustness test.

Table 6.3 presents the performance of the portfolios when including or excluding stocks with an average price (in the sample) below $5. In Panel A, the results from the main Model are presented. Panel B offers a comparison view. Analyzing the returns first, it becomes clear that for the lower volatility portfolios the results are not much different. Only the High Volatility portfolio shows a different result. To illustrate, the biggest difference in average monthly return for the volatility portfolios (portfolios 1 through 9) is in portfolio 8, where the difference is 0.021 percentage points. For the High Volatility portfolio, the average

Table 6.3: Excluding vs Including Penny Stocks
The Table shows the average 1-month holding returns in excess of the risk free rate, the standard deviation of returns (SD) and the Sharpe Ratios for portfolios formed every month by sorting stocks on their past 3-year return volatility. The sample runs from January 1963 to December 2019 and includes all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. Stocks are sorted into portfolios each month using NYSE breakpoints, and returns are value-weighted by size. Market Equity (ME) is used to measure size. Returns are winsorized at the 1% level. Reported t-statistics are simple t-statistics and show whether the returns are significantly different from zero. Panel A shows the performance of the portfolios when excluding stocks with an average price during the sample below $5 and Panel B shows the performance of the portfolios without excluding any stocks.

Decile	Low	2	3	4	5	6	7	8	9	High
Panel A: Excluding stocks below $5										
Return	.678	.735	.820	.758	.790	.766	.831	.758	.785	.639
t-statistic	(4.81)	(4.85)	(4.69)	(4.26)	(4.21)	(3.90)	(3.94)	(3.35)	(3.14)	(2.12)
SD (return)	3.59	3.85	4.44	4.52	4.76	4.99	5.36	5.74	6.35	7.66
Sharpe Ratio	.189	.191	.185	.168	.166	.154	.155	.132	.124	.083
Panel B: Not excluding any stocks										
Return	.677	.755	.802	.750	.788	.776	.826	.737	.747	.432
t-statistic	(4.80)	(4.90)	(4.62)	(4.22)	(4.18)	(3.94)	(3.88)	(3.21)	(2.99)	(1.43)
SD (return)	3.58	3.91	4.42	4.52	4.79	5.00	5.42	5.83	6.35	7.69
Sharpe Ratio	.189	.193	.181	.166	.165	.155	.152	.126	.118	.056

monthly return drops to 0.432%, a difference of 0.207 percentage points. So, it seems that in terms of return, penny stocks negatively affect the High Volatility portfolio in this sample. However, as discussed below, this difference is not statistically significant. So in terms of returns, excluding these penny stocks actually presents a more conservative view of the anomaly, although statistically the difference is not significant.

The standard deviations of the returns (SD) have also remained fairly unchanged. For the High portfolio, the difference is also negligible. To illustrate, in Panel A the High portfolio has a standard deviation of returns of 7.66 while in Panel B the number is 7.69. So, although penny stocks seem to hurt the returns of the most volatile portfolio (although not significantly), it does not increase its volatility.

As can be expected, the Sharpe Ratios do not differ that much either. Because the returns and standard deviations of returns have stayed mostly the same, the Sharpe Ratio will remain mostly the same by definition. Only the High portfolio shows a somewhat larger difference, where including penny stocks drops the Sharpe Ratio from 0.083 to 0.056. As already mentioned above, this is due to the decrease in return. The standard deviation of returns has stayed about the same. The difference in Sharpe ratio in the High portfolio is confirmed by the z score, which is 12.538 (p-value 0.000), which is significant at the 1% level.

Furthermore, the comparison of variance test (see Table 8.1, Section 8) shows there is no significant difference in variance between the Low or High portfolios between the two methodologies. Also, the two-sample t-test of equal means (see Table 8.2, Section 8) indicates there are no significant differences in the average returns between the methodologies (when comparing the Low or High portfolios). This means that the choice to include or exclude penny stocks does not lead to a significant difference in variance or average return in the Low or High portfolios in this sample. However, the Sharpe Ratios do significantly differ; not excluding these penny stocks causes a significantly lower Sharpe Ratio of the High portfolio.

Finally, the market share of each portfolio is compared when penny stocks are included or excluded from the sample in Table 8.3, stored in the Appendix. This topic is also briefly discussed in Section 5.2. Excluding penny stocks from the sample improves the distribution of stocks amongst the portfolios. For example, the Low portfolio now contains 46.7% of the stocks in terms of market share instead of 46.1%. Similarly, portfolio 2 now contains 12.3% instead of 12.0%. The more volatile portfolios contain smaller stocks in terms of market share when including penny stocks, where portfolio 9 only contains 2.7% instead of 3.0% and the High portfolio now contains 1.6% instead of 2.2%.

To conclude this robustness test, including penny stocks in the portfolios does not affect the result of the anomaly. In the setting where penny stocks are included, the anomaly still exists. The test of equal variances shows the variance of the Low or High portfolio is not affected, and the average returns neither. Including penny stocks decreases the Sharpe ratio of the High portfolio significantly, so the choice of excluding them in the main Model actually presents a more conservative view of the anomaly.

6.4 Volatility vs. idiosyncratic volatility

In this section, a simple overview of the anomaly is given when the past 3-year volatility of monthly returns is used (which is what is done throughout the paper), and what happens if a different methodology that is often seen in the literature is used, namely using the standard deviation of residuals of regressions on the Fama-French 3-factor Model. To compare this, a simple Table with 2 panels using value-weighted returns and portfolios formed using NYSE breakpoints is presented, with Newey-West t-statistics. Here, one panel will show the performance of the Main Model and the second panel will show the results for IVOL. Then a simple correlation is included to see how much the two measures are correlated, and possibly say something about whether the choice of how to measure volatility matters. It should be noted that the literature often uses daily returns when working with IVOL, but for simplicity, monthly returns are used.

The IVOL measure is constructed using the standard deviation of the residuals when regressing stock returns on the Fama-French three-factor Model. The correlation between IVOL and Volatility in the sample is 0.76. Although the correlation is quite high, there still are some differences between the two measures. That is reflected in the results in Table 6.4. Two panels are presented for comparison, with both panels using value-weighted returns and portfolios formed using NYSE breakpoints. In Panel A, the results as seen before are presented. These are the return averages, Newey-West t-statistics, standard deviation of returns and

Table 6.4: Past 3-year volatility vs IVOL

The Table shows the average 1-month holding returns in excess of the risk free rate, the standard deviation of returns (SD) and the Sharpe Ratios for portfolios formed every month by sorting stocks on their past 3-year return volatility or using the IVOL measure, calculated by taking the standard deviation of residuals obtained from the Fama-French 3-factor Model. The sample runs from January 1963 to December 2019 and includes all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. Stocks are sorted into portfolios each month using NYSE breakpoints, and returns are value-weighted by size. Market Equity (ME) is used to measure size. Stocks with an average price below $5 in the sample are excluded. Returns are winsorized at the 1% level. Reported t-statistics are Newey-West t-statistics and show whether the returns are significantly different from zero. Panel A shows the performance when using the past 3-year volatility of monthly returns to sort on and Panel B shows the performance using the IVOL measure.

Decile	Low	2	3	4	5	6	7	8	9	High
Panel A: Performance of portfolios sorted on past 3-year volatility										
Return	.678	.735	.820	.758	.790	.766	.831	.758	.785	.639
t-statistic	(4.24)	(4.53)	(4.50)	(4.01)	(3.89)	(3.62)	(3.64)	(3.13)	(2.90)	(1.99)
SD (return)	3.59	3.85	4.44	4.52	4.76	4.99	5.36	5.74	6.35	7.66
Sharpe Ratio	.189	.191	.185	.168	.166	.154	.155	.132	.124	.083

Decile	Low	2	3	4	5	6	7	8	9	High
Panel B: Performance of portfolios sorted on IVOL										
Return	.671	.854	.797	.863	.851	.816	.873	.869	.753	.475
t-statistic	(4.24)	(5.16)	(4.61)	(4.81)	(4.52)	(3.95)	(4.01)	(3.71)	(3.01)	(1.58)
SD (return)	4.02	4.01	4.15	4.36	4.49	4.89	5.18	5.58	6.02	7.04
Sharpe Ratio	.170	.213	.192	.198	.190	.167	.169	.156	.125	.067

Sharpe Ratio of 10 portfolios sorted on past 3-year volatility of monthly returns. For comparison, Panel B shows the same setup but now for portfolios formed by sorting on IVOL, the standard deviation of residuals obtained by regressing stock returns on the Fama-French three-factor Model. Comparing the Low and High deciles of both panels, both have a lower return in Panel B. For decile 1, this is 0.671% (t-statistic 4.24) against 0.678% which is quite similar. For decile 10 however, the results are more contrasting. Where the return for the past 3-year volatility methodology was 0.639%, with IVOL it is 0.475% per month. Next to the returns being lower, they are also less significantly different from zero (t-statistic drops from 1.99 to 1.58). In terms of Sharpe Ratio, both also perform worse. Decile 1 has a slightly lower standard deviation (4.02) so the Sharpe Ratio drops from 0.189 to 0.170. For decile 10, the standard deviation of return decreases to 7.04 (from 7.66), but the Sharpe Ratio declines (because of the much lower returns) to 0.067 (from 0.083). Interestingly, besides the two outer portfolios, the IVOL portfolios perform better in terms of returns in 6 (of the other 8) portfolios, and in terms of Sharpe Ratio in all of the remaining portfolios. For example, the return of decile 2 increases from 0.735 to 0.854 (t-statistic of 4.53 and 5.16, respectively) and the Sharpe Ratio increases from 0.191 to 0.213.

Because the returns drop sharply from decile 9 to decile 10 when using IVOL (decile 9 has a return of 0.753% while decile 10 a return of 0.475%, t-statistics 3.01 and 1.58 respectively), not only deciles 1 and 10 should be compared to confirm that the anomaly exists. Looking at the bottom 3 portfolios and the top 3 portfolios, the worst performing decile of the bottom 3 in terms of Sharpe Ratio (which is decile 1) still performs better than the best performing decile in the top 3 (decile 8). The highest Sharpe Ratio is obtained with decile 2. Here, the portfolio has a return of 0.854% per month (t-statistic of 5.16, the most significant decile) with a standard deviation of returns of 4.01. This results in the highest Sharpe Ratio (of all IVOL portfolios) of 0.213.

Finally, the comparison of variance test (see table 8.1, stored in the Appendix) shows there are significant differences in variance between the Low or High portfolios between the two methodologies. The difference in variance between the Low portfolios is significant even at the 1% level. The difference in variance between the High portfolios is significant at the 5% level. Next to that, the two-sample t-test of equal means (see 8.2, stored in the Appendix) indicate there are no significant differences in the average returns between the methodologies (when comparing the Low or High portfolios). This means that the measure of volatility that is used leads to a significant difference in variance (but not in average return) in the Low or High portfolios.

To conclude this robustness test, in both methodologies, the Low Volatility anomaly exists. That is, low volatility stocks outperform high volatility stocks in terms of Sharpe Ratio (and sometimes returns as well). Furthermore, there are significant differences in variance between the two measures. For the Low portfolio, using IVOL significantly increases variance of returns. For the High portfolio, using IVOL significantly decreases variance of returns, thus making the Low Volatility portfolio appear worse, and the High Volatility portfolio appears better.

6.5 Low Volatility & Beta

This section will go in-depth on the differences and similarities between Low Volatility and Low Beta. This will be done with the help of Table 6.5. Again, the same main Model is chosen as a base to compare and test against. First, some further background information into the issue will be given and then the results from the aforementioned Table 6.5 will be analyzed. Then a conclusion will be drawn about this robustness test.

In the literature, the low beta and low volatility are sometimes confused or used interchangeably. There is however a clear distinction between the two as recent literature shows. As highlighted in the methodology section of this paper, the methods of Liu et al. (2018) and Frazzini et al. (2014) will be combined, using parts of the methodology of Blitz et al. (2007) as well. To be precise, first the differences and similarities between Low Volatility and Beta are presented using the Volatility measure as in Blitz et al. (2007), which means measuring volatility as the standard deviation of past 3-year monthly returns. Liu et al. (2018) on the other hand use the standard deviation of the residuals on the Fama-French three-factor Model as their measure for volatility (IVOL, as in Ang et al. (2006)). Beta is measured following Liu et al. (2018) but instead of using dynamic weights for the Vasicek (1973) correction, this correction is not used at all as shrinking the betas does not affect the sort for portfolio formation. This step is therefore redundant. See also Section 4. The different measure of volatility in this paper could be a reason why the results might differ from the literature. Theoretically, the choice of dataset might also be a reason for differences, because market beta itself can be decomposed into volatility and market correlation, and correlations in my dataset can possibly be different. However, as my dataset covers 1963 to 2019 and Liu et al. (2018) cover 1963 to 2013, this is unlikely as the difference is only 6 years.

First of all, Liu et al. (2018) find a positive correlation between IVOL and Beta of 0.33. The volatility that is calculated in this paper and the Beta calculated as described earlier in this paper show a correlation of 0.52. As already explained, the different methodology could be the reason why a higher correlation is found. That would mean that the methodology used to calculate Volatility is relatively more related to Beta than the IVOL methodology, which focuses on residual volatility.

Now that a basic relationship between Volatility and Beta is shown with the correlation, the two can be further explored in a double sort. By independently sorting stocks into quintiles on Beta and Volatility and presenting the results in a 5x5 table, the returns and standard deviation of returns can be analyzed. This will help understand us if there is a beta anomaly once we control for volatility, and vice versa.

Looking at Table 6.5, in Panel A, in the Low Beta bucket, the Low Volatility portfolio's return is 0.656% per month while the High Volatility portfolio delivers a return of 0.674%, slightly higher. However, in the right section, the standard deviation of return seems substantially higher: the Low Volatility portfolio has a standard deviation of returns of 3.291 while the High Volatility portfolio has a standard deviation of returns of 6.314. The better performance is captured by the Sharpe Ratio in Panel B. Still in the Low Beta bucket, the Sharpe ratio for the Low Volatility portfolio is 0.199 while for the High Volatility portfolio it is 0.107. So within the Low Beta bucket, the Volatility anomaly seems to exist.

Looking at other Beta buckets, in Panel B, in Beta bucket 2, Low Volatility again outperforms High Volatility with a Sharpe Ratio of 0.200 versus a Sharpe Ratio of 0.130. In Beta bucket 3, the same shows: the Low Volatility portfolio outperforms with a Sharpe Ratio of 0.192 against a Sharpe Ratio of 0.123 of the High Volatility portfolio. However, in Beta buckets 4 and 5 something else happens. In bucket 4, the Sharpe Ratio of the Low Volatility portfolio is 0.093 while the High Volatility has a Sharpe Ratio of 0.102, which is slightly higher. Looking at Panel A again, it becomes clear where this comes from. In bucket 4 the Low Volatility portfolio has a monthly return of 0.516%, while the High Volatility portfolio has a return of 0.699% per month, which is higher. This is paired with a standard deviation of return of 5.527 and 6.820, respectively. In Panel B, Beta bucket 5 the results become even more contrasting. Low Volatility has a Sharpe Ratio of 0.001 while High Volatility has a Sharpe Ratio of 0.092. Analyzing these numbers further,

Table 6.5: Double Sort on Volatility and Beta

The Table shows the average 1-month holding returns in excess of the risk free rate, the standard deviation of returns (SD) and the Sharpe Ratios for 25 portfolios independently sorted every month on past 3-year return volatility and on Beta. The intersection of these sorts produces 25 double-sort portfolios. Beta is measured using a regression with a 60-month moving window on the excess market return and the lagged excess market return, summing these slopes and shrinking the beta to one. A stock needs a minimum of 36 months of non-missing returns to be assigned a beta. The sample runs from January 1963 to December 2019 and includes all common stocks (share codes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. Stocks are sorted into portfolios each month using NYSE breakpoints, and returns are value-weighted by size. Market Equity (ME) is used to measure size. Stocks with an average price below $5 in the sample are excluded. Returns are winsorized at the 1% level.

Vol	Low	2	3	4	High	Low	2	3	4	High
Panel A: Returns and Standard Deviation of returns										
Beta	Return					SD (return)				
Low	.656	.757	.771	.726	.674	3.291	3.907	4.155	4.919	6.314
2	.791	.890	.847	.857	.843	3.950	4.289	4.601	5.258	6.463
3	.908	.786	.780	.884	.806	4.740	4.917	5.077	5.367	6.551
4	.516	.726	.758	.791	.699	5.527	5.203	5.400	5.813	6.820
High	.009	.823	.982	.822	.717	6.568	6.697	6.276	6.584	7.824

Vol	Low	2	3	4	High	Low	2	3	4	High
Panel B: Returns and Sharpe Ratio of portfolio										
Beta	Return					Sharpe Ratio				
1	.656	.757	.771	.726	.674	.199	.194	.186	.148	.107
2	.791	.890	.847	.857	.843	.200	.208	.184	.163	.130
3	.908	.786	.780	.884	.806	.192	.160	.154	.165	.123
4	.516	.726	.758	.791	.699	.093	.140	.140	.136	.102
Big	.009	.823	.982	.822	.717	.001	.123	.156	.125	.092

Panel A shows that extremely low returns in the High Beta, Low Volatility portfolio are the main driver of this low Sharpe Ratio. Particularly, the Low Volatility portfolio shows a return of 0.009% per month while the High portfolio delivers a return of 0.717 per month. This is combined with a standard deviation of returns of 6.568 and 7.824, respectively.

The low returns in the high beta bucket seem to be limited to the lowest volatility quintile. Moving from volatility quintile 1 to quintile 2, the returns increase from 0.009% to 0.823%. An explanation could be the distribution of stocks per portfolio; there might be very few stocks that are in this category, which makes sense since first of all this is an independent sort, and secondly, because volatility and beta are somewhat similar (correlation of 0.52), there are simply not that many stocks that fall in the outermost sorts of these two characteristics. Further research into this is necessary to say something meaningful. It seems highly likely that the number of stocks that fall into this category is simply so low that the returns in the Table are not representative.

So for 3 out of 5 Beta buckets, the Low Volatility portfolio outperforms the High Volatility portfolio in terms of Sharpe Ratio. Decomposing this into returns and standard deviation of return, it shows that in only 1 of the Beta buckets, Low Volatility outperforms High Volatility in terms of average 1-month holding returns (Beta bucket 3, with a return of 0.908% against a return of 0.806%). In terms of standard deviation of return, Low Volatility scores better in all 5 Beta buckets than High Volatility but this is almost per definition the case. However, the differences in standard deviation of returns decrease, moving from lower to higher Beta buckets. For example, the standard deviations in Low and High Volatility in Beta bucket 1 are 3.291 and 6.314, in Beta bucket 3 this is 4.740 and 6.551 and in Beta bucket 4 this is 6.568 and 7.824. Clearly, the differences become smaller. Apparently, higher beta stocks are more volatile; they experience a higher standard deviation of return. This is confirmed by the pattern in Beta portfolios looking per Volatility bucket. In all 5 Volatility buckets, standard deviation of return is higher for High Beta portfolios than for Low Beta portfolios. For example, in Volatility bucket Low there is a standard deviation of return of 3.291 for the low beta portfolio and 6.568 for the high beta portfolio. In Volatility bucket 3, this is 4.155 compared to 6.276 and in the High Volatility bucket, it is 6.314 compared to 7.824, respectively.

Although this paper is not necessarily about the Beta Anomaly, the Beta Anomaly can shortly be explored. In short, the beta anomaly says that low beta stocks perform better than high beta stocks (Frazzini & Pedersen (2014), Liu et al. (2018), Asness et al. (2020)). Analyzing first the returns in Panel A, in the Low Volatility Column, Low beta has higher 1-month holding returns than High Beta (0.656% vs 0.009%). However, in the next 4 Volatility buckets, this pattern does not appear. In fact, returns are lower in these 4 volatility columns. For example, in Volatility Column 3 the low beta portfolio delivers a return of 0.771% while the high beta portfolio shows a return of 0.982%. Combining returns with its standard deviation to say something about performance measured by Sharpe Ratio, Low Beta outperforms High Beta in every Volatility bucket. For example, in Volatility bucket 2 the Low Beta portfolio has a Sharpe Ratio of 0.194 while the High Beta portfolio has a Sharpe Ratio of 0.123.

To conclude this robustness test, it seems that the Volatility Anomaly and the Beta Effect are two distinct anomalies. The Volatility effect survives in almost all of the Beta sorts, but so does the Beta effect, although the latter only survives in terms of Sharpe Ratio. The Volatility effect holds up a bit better in terms of returns as well.

Finally, to summarize and conclude the entire Section 6 Robustness of the Anomaly, relaxing the assumptions and methodological choices (ceteris paribus) slightly suppresses the anomaly to some degree, but overall the anomaly is resistant to these robustness tests. The choice of using deciles or quintiles significantly affects the results. However, the anomaly still exists when using quintile portfolios (in terms of Sharpe ratio). Winsorizing leads to a significant decrease in Sharpe ratio in the High portfolio. Still, the anomaly exists in both methodological settings. Including penny stocks decreases the Sharpe ratio of the High portfolio significantly, so the choice of excluding them in the main Model actually presents a more conservative view of the anomaly. For the High portfolio, using IVOL significantly decreases variance of returns in the High portfolio and increases variance in the Low portfolio, thus making the Low Volatility portfolio appear worse, and the High Volatility portfolio appears better.

7 Conclusion

The volatility anomaly is present in the dataset and robust over time and robust to methodological choices and assumptions. Using a methodology with value-weighted returns and NYSE breakpoints to form portfolios, the low volatility portfolio outperforms the high volatility portfolio in terms of return and Sharpe Ratio. This result shows up in all size groups except for the biggest 20% of stocks, where the low volatility portfolio underperforms in returns but outperforms in terms of Sharpe Ratio. The anomaly has gotten stronger over time, with the anomaly first appearing in the 1980s. The anomaly is robust to choice of portfolio size (10 or 5 portfolios), is present whether the returns are winsorized or not, becomes stronger after including penny stocks but is also very much present without these penny stocks, is robust to choice of volatility measure (past 3-year volatility or IVOL) and is a distinct effect from Beta.

The paper contributes to the literature in 3 ways. First, new evidence is presented that the anomaly exists within the latest data set. From 1963 to 2019, the anomaly exists in the US market. Second, the strength of the anomaly is analyzed over time. Third, the paper tries to refute criticism on the anomaly by showing that the anomaly is robust to certain methodological choices, and not limited to small stocks only. The paper does not provide further analysis on the interaction of the anomaly with the Fama-French factors or factors such as Momentum or Quality. A detailed analysis of this would be helpful to investors for the construction of their portfolio. Also, the paper does not analyze why the anomaly was not present in the first two decades of the sample, or why the anomaly increased in strength over time. Further research into this can provide further insights into the characteristics of the anomaly, and whether its performance is better in certain macroeconomic settings. Next to this, this paper focuses on US stocks. An analysis of the anomaly in different countries would further test the robustness of the anomaly out of setting.

The following hypotheses were formulated based on the research question and the literature. They will now be accepted/rejected.

The low volatility anomaly is expected to be statistically significant in this dataset. This hypothesis can not be rejected. For multiple time periods and multiple methodologies, the anomaly is statistically significant.

The low volatility anomaly is expected to decrease in magnitude after using different methodologies. This hypothesis can not be rejected. After putting the anomaly to the test with multiple robustness tests, some (more than others) of these tests decreased the magnitude of the anomaly. However, the anomaly remained statistically significant. Certain methodological choices do have effects on the output, but in all tested settings, the anomaly is present.

The low volatility anomaly is expected to have decreased in magnitude over time. This hypothesis can be rejected. The results indicate that the anomaly was not present in the first 2 decades of the sample, but then increased in magnitude over time. The most recent decade (2010-2019) was the strongest yet for the anomaly.

In conclusion, the results suggest that the volatility anomaly exists in the dataset. The results are robust to different methodologies and time periods, and not exclusive to small caps. To answer the research question, the low volatility anomaly survives after taking into account different methodologies and it has gotten stronger over time.

8 Appendix

8.1 Factor Definitions

Because the factors used come directly from the Ken French Data Library, below the definitions are directly quoted from his website. I do not claim that these are my own words, they are directly copied from the website for the sake of completeness and information.

Description of Fama/French 5 Factors (2x3)

Monthly Returns: July 1963 - December 2019

Construction:
The Fama/French 5 factors (2x3) are constructed using the 6 value-weight portfolios formed on size and book-to-market, the 6 value-weight portfolios formed on size and operating profitability, and the 6 value-weight portfolios formed on size and investment. (See the description of the 6 size/book-to-market, size/operating profitability, size/investment portfolios.)

SMB (Small Minus Big) is the average return on the nine small stock portfolios minus the average return on the nine big stock portfolios.

SMB(B/M) = 1/3 (Small Value + Small Neutral + Small Growth) - 1/3 (Big Value + Big Neutral + Big Growth).

SMB(OP) = 1/3 (Small Robust + Small Neutral + Small Weak) - 1/3 (Big Robust + Big Neutral + Big Weak).

SMB(INV) = 1/3 (Small Conservative + Small Neutral + Small Aggressive) - 1/3 (Big Conservative + Big Neutral + Big Aggressive).

SMB = 1/3 (SMB(B/M) + SMB(OP) + SMB(INV)).

HML (High Minus Low) is the average return on the two value portfolios minus the average return on the two growth portfolios.

HML = 1/2 (Small Value + Big Value) - 1/2 (Small Growth + Big Growth).

RMW (Robust Minus Weak) is the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios.

RMW = 1/2 (Small Robust + Big Robust) - 1/2 (Small Weak + Big Weak).

CMA (Conservative Minus Aggressive) is the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios.

CMA = 1/2 (Small Conservative + Big Conservative) - 1/2 (Small Aggressive + Big Aggressive).

Rm-Rf, the excess return on the market, value-weight return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11 at the beginning of month t, good shares and price data at the beginning of t, and good return data for t minus the one-month Treasury bill rate (from Ibbotson Associates).

Stocks: Rm-Rf includes all NYSE, AMEX, and NASDAQ firms. SMB, HML, RMW, and CMA for July of year t to June of t+1 include all NYSE, AMEX, and NASDAQ stocks for which we have market equity data for December of t-1 and June of t, (positive) book equity data for t-1 (for SMB, HML, and RMW), non-missing revenues and at least one of the following: cost of goods sold, selling, general and administrative expenses, or interest expense for t-1 (for SMB and RMW), and total assets data for t-2 and t-1 (for SMB and CMA).

8.2 CRSP Definitions

In the section below, the data definitions as provided by the CRSP database are collected. I do not claim that these are my own words, they are directly copied from the website for the sake of completeness and information.

8.2.1 Number of Shares Outstanding, SHROUT.

SHROUT is the number of publicly held shares, recorded in thousands.

8.2.2 Share code, SHRCD

SHRCD is a two-digit code describing the type of shares traded. The first digit describes the type of security traded.

First Digit - Share Code - Security Type

Code Definition

1 Ordinary Common Shares

2 Certificates

3 ADRs (American Depository Receipts)

4 SBIs (Shares of Beneficial Interest)

7 Units (Depository Units, Units of Beneficial

Interest, Units of Limited Partnership Interest, Depository Receipts, etc.)

Note: "Units" (code 7) does not represent combinations of common stock and anything else, such as warrants. The second digit gives more detailed information about the type of security traded.

Second Digit - Share Code - Security Type

Code Definition

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 Components (Primes and Scores).

4 Closed-end funds.

5 Closed-end fund companies incorporated outside the US

8 REIT's (Real Estate Investment Trusts).

For example: A SHRCD of 14 would represent ordinary common shares of a closed-end fund.

8.2.3 Holding Period Return, ret.

A return is the change in the total value of an investment in a common stock over some period of time per dollar of initial investment. RET(I) is the return for a sale on day I. It is based on a purchase on the most recent time previous to I when the security had a valid price. Usually, this time is I - 1. Returns are calculated as follows:

For time t (a holding period), let:

t = time of last available price ¡ t

r(t) = return on purchase at t?, sale at t

p(t) = last sale price or closing bid/ask average at time t

d(t) = cash adjustment for t

f(t) = price adjustment factor for t

p(t) = last sale price or closing bid/ask average at time of last available price ¡ t.

then r(t) = [(p(t)f(t)+d(t))/p(t')]-1

t is usually one period before t, but t can be up to ten periods before t if there are no valid prices in the interval. A series of special return codes specify the reason a return is missing.

Missing Return Codes

RET(t) Reason For Missing Return

-66.0 more than 10 periods between time t and the time of the preceding price t?

-77.0 not trading on the current exchange at time t

-88.0 no return, array index t not within range of BEGRET and ENDRET

-99.0 missing return due to missing price at time t

8.2.4 Closing Price or Big/Ask Average, PRC.

Prc is the closing price or the negative bid/ask average for a trading day. If the closing price is not available on any given trading day, the number in the price field has a negative sign to indicate that it is a bid/ask average and not an actual closing price. Please note that in this field the negative sign is a symbol and that the value of the bid/ask average is not negative.

If neither closing price nor bid/ask average is available on a date, prc is set to zero. In a monthly database, prc is the price on the last trading date of the month. The price series begins the first month-end after the security begins trading and ends the last complete month of trading.

If the security of a company is included in the Composite Pricing network, the closing price while listed on NYSE or AMEX on a trading date is the last trading price for that day on the exchange that the security last traded.

Similarly, highs, lows, and volumes include trades on all exchanges on which that security traded. For example, if a stock trades on both the NYSE and the PACX (Pacific Stock Exchange), and the last trade occurs on the PACX, the closing price on that day represents the closing price on the PACX, not the NYSE. Price data for Nasdaq securities comes directly from the NASD with the close of the day at 4:00 p.m. Eastern Time. Automated trades after hours on Nasdaq are counted on the next trading date, although the volumes are applied to the current date. Daily trading prices for The Nasdaq National Market securities were first reported November 1, 1982. Daily trading prices for The Nasdaq Small Cap Market were first reported June 15, 1992. prc for Nasdaq securities is always a negative bid/ask average before this time. All prices are raw prices as they were reported at the time of trading.

8.2.5 Primary Exchange, primexch

Primary Exchange is a one-character code which identifies the primary exchange on which the security trades.

Code Exchange

N NYSE

A AMEX

Q NASDAQ

X Other Exchange

R ARCA

8.3 Tests for Equal Variance and Two-Sample T-Tests

In this section, the results for the tests of equal variance and the two-sample t-tests for equal means are presented. In Table 8.1, Panel A, the decile Low portfolio is different in variance than the quintile Low portfolio, significant at the 10% level. The decile High portfolio is significantly different from the quintile High portfolio at the 5% level. Finally, the quintile Low portfolio is significantly different from the quintile High portfolio, even at the 1% level.

In Panel B, there is only one significant difference in variances between portfolio returns. The Non-Winsorized Low portfolio is significantly different in variance from the Non-Winsorized High portfolio, even at the 1% level.

In Panel C, the same finding appears as in Panel B. Only the Including Penny Stocks Low portfolio compared to the Including Penny Stocks High portfolio is significantly different in variance, even at the 1% level.

In Panel D, the Volatility Low portfolio is different in variance compared to the IVOL Low portfolio at the 1% level. The difference between the Volatility High portfolio and the IVOL High portfolio is significant at the 5% level. Finally, the difference between the IVOL Low and the IVOL High portfolios are significantly different even at the 1% level.

To conclude this subsection, the two tables produce various results. The variance comparison test shows that especially the decile vs quintile portfolios produce different variances, and there also seem to be differences in variance between the Volatility sorted portfolios and the IVOL sorted portfolios. For the test of equal means, there are no significant differences between the portfolios that were compared.

Table 8.1: Variance Comparison Tests for the Robustness Tests

The Table shows the F-Statistic and the P-Values for the test of equal variances performed on the average returns of the Low and High portfolios produced in the Section Robustness Tests. The table is grouped into 4 Panels. Panel A shows the results for the decile compared to the quintile portfolio average returns. Panel B shows the results for the winsorized compared to the non-winsorized return portfolios. Panel C shows the results for the portfolios when including or excluding penny stocks in the sample. Panel D shows the results for the portfolios sorted in past 3-year return volatility compared against the portfolios sorted on the IVOL measure. Significant p-values are indicated as follows: *p<0.1, **p<0.05 & ***p<0.01.

Portfolio	**F-Statistic**	**P-Value**
Panel A: Decile Portfolio vs Quintile Portfolios		
Decile Low vs Quintile Low	0.874	0.088*
Decile High vs Quintile High	1.181	0.035**
Quintile Low vs Quintile High	0.297	0.000***
Panel B: Winsorized Returns vs Non-Winsorized Returns		
Winsorized Low vs Non-Winsorized Low	1.000	1.000
Winsorized High vs Non-Winsorized High	0.883	0.116
Non-Winsorized Low vs Non-Winsorized High	0.194	0.000***
Panel C: Excluding vs Including Penny Stocks		
Excluding Low vs Including Low	1.006	0.944
Excluding High vs Including High	0.992	0.921
Including Low vs Including High	0.217	0.000***
Panel D: Volatility Sorted Portfolios vs IVOL Sorted Portfolios		
Volatility Low vs IVOL Low	0.798	0.004***
Volatility High vs IVOL High	1.184	0.032**
IVOL Low vs IVOL High	0.326	0.000***

Table 8.2: Two-Sample T-Tests of Equal Means with Equal/Unequal variances for the Robustness Tests

The Table shows the T-Statistic and the P-Values for the test of equal means performed on the average returns of the Low and High portfolios produced in the Section Robustness Tests. The table is grouped into 4 Panels. Panel A shows the results for the decile compared to the quintile portfolio average returns. Panel B shows the results for the winsorized compared to the non-winsorized return portfolios. Panel C shows the results for the portfolios when including or excluding penny stocks in the sample. Panel D shows the results for the portfolios sorted in past 3-year return volatility compared against the portfolios sorted on the IVOL measure. The two-sample t-test for equal means with unequal variances is performed for those portfolios marked with an asterisk. For the others, the two-sample t-test for equal means with equal variances is performed. Significant p-values are indicated as follows: *p<0.1, **p<0.05 & ***p<0.01.

Portfolio	**T-Statistic**	**P-Value**
Decile Portfolio vs Quintile Portfolios		
Decile Low vs Quintile Low	-0.068	0.946
Decile High vs Quintile High*	-0.139	0.889
Quintile Low vs Quintile High*	-0.013	0.990
Winsorized Returns vs Non-Winsorized Returns		
Winsorized Low vs Non-Winsorized Low	-0.010	0.992
Winsorized High vs Non-Winsorized High	-0.350	0.726
Non-Winsorized Low vs Non-Winsorized High*	-0.323	0.747
Excluding vs Including Penny Stocks		
Excluding Low vs Including Low	0.005	0.996
Excluding High vs Including High	0.485	0.628
Including Low vs Including High*	0.734	0.463
Volatility Sorted Portfolios vs IVOL Sorted Portfolios		
Volatility Low vs IVOL Low*	0.033	0.974
Volatility High vs IVOL High*	0.401	0.689
IVOL Low vs IVOL High*	0.615	0.539

8.4 Comparison Market Shares of Portfolios

In this section, the results for the market shares comparison between including and excluding penny stocks will be presented in Table 8.3. In this table, it becomes evident that including penny stocks leads to a more imbalanced distribution of stocks amongst the portfolios. The lower portfolios contain more (percentage points) market share than when excluding penny stocks. For example, the Low portfolio now contains 46.7% of the stocks in terms of market share instead of 46.1%. Similarly, portfolio 2 now contains 12.3% instead of 12.0%. The more volatile portfolios contain smaller stocks in terms of market share when including penny stocks, where portfolio 9 only contains 2.7% instead of 3.0% and the High portfolio now contains 1.6% instead of 2.2%.

To conclude, excluding penny stocks from the sample increases the stocks in the more volatile portfolios (in terms of market share, measured by market equity) and decreases the stocks in the lower volatility portfolios.

Table 8.3: Comparison of market shares between including or excluding penny stocks
The Table shows the market shares (in %) per portfolio. The column Excluding Penny Stocks shows the market shares when stocks with an average price in the sample below $5 are dropped. The column Including Penny Stocks shows the market shares these stocks are not excluded. Portfolios are formed every month by sorting stocks on their past 3-year return volatility. The sample runs from January 1963 to December 2019 and includes all common stocks (sharecodes 10 and 11) in the CRSP database traded on NYSE, AMEX and NASDAQ exchanges. Stocks are sorted into portfolios each month using NYSE breakpoints, and returns are value-weighted by size. Market Equity (ME) is used to measure size.

	Excluding Penny Stocks	Including Penny Stocks
Portfolio	Market Share %	Market Share %
Low	46.1%	46.7%
2	12.0%	12.3%
3	9.4%	9.6%
4	7.8%	7.9%
5	6.3%	6.3%
6	5.2%	5.1%
7	4.4%	4.3%
8	3.7%	3.6%
9	3.0%	2.7%
High	2.2%	1.6%

8.5 Z-scores for Sharpe Ratio Comparison

In this section, some tables will be presented that contain z-scores to compare the Sharpe ratios of two portfolios.

Table 8.4: Z-scores for comparison of Sharpe Ratios of the Low and High Portfolios within Time Periods. The table shows the z-score that compares the Low with the High portfolio within each decade. A positive score therefore indicates that the Low portfolio has a higher Sharpe ratio than the High portfolio, whereas a negative score indicates that the High portfolio has a higher Sharpe ratio. The P-value shows the (two-tailed) probability that the null is correct, and the Sharpe ratios are not different from another.

Decade	Z-score	P-value
1960s	-6.254	0.000
1970s	-2.949	0.002
1980s	14.006	0.000
1990s	0.597	0.551
2000s	8.029	0.000
2010s	9.434	0.000

Table 8.5: Z-scores for comparison of Sharpe ratios between Models The table shows the z-score that compares the Low and the High portfolios between models. A positive score therefore indicates that the Low (or High) portfolio has a higher Sharpe ratio than the Low portfolio in the other model, whereas a negative score indicates that the Low (or High) portfolio has a higher Sharpe ratio than the Low (or High) portfolio of the other model. The P-value shows the (two-tailed) probability that the null is correct, and the Sharpe ratios are not different from another.

Panel A: Decile vs Quintile		
	Low	High
Low	-0.681	
High		3.685

Panel B: Winsorized vs Non-Winsorized		
	Low	High
Low	0.000	
High		3.256

Panel C: Incl. vs Excl. Penny Stocks		
	Low	High
Low	0.000	
High		12.538

8.6 Regressions for the volatility portfolios through the time periods

To complete the analysis into the time periods, each decade the Low and High portfolios will be regressed on the market portfolio to compare the alpha and beta of the portfolios over time. Also, a summary of the returns will be presented per decade, where the average return of the Low and High portfolio as well as the market portfolio is presented, the standard deviation of these returns, and the lowest and highest monthly return these portfolios achieved in this decade.

Table 8.6: Z-scores for comparison of Sharpe ratio of the Low Volatility portfolio across time periods. The table presents the z-scores for the comparison of Sharpe ratio of the Low Volatility portfolio across time periods.

	1960s	1970s	1980s	1990s	2000s	2010s
1960s	0.000	-4.673	-10.962	-10.974	-7.196	-10.937
1970s	5.487	0.000	-11.407	-11.873	-4.606	-11.326
1980s	12.855	11.396	0.000	-0.467	9.038	-2.067
1990s	13.026	11.467	0.374	0.000	8.254	-1.262
2000s	8.533	4.475	-8.290	-8.250	0.000	-9.160
2010s	12.947	11.066	1.712	1.303	9.285	0.000

Table 8.7: Z-scores for comparison of Sharpe ratio of the High Volatility portfolio across time periods. The table presents the z-scores for the comparison of Sharpe ratio of the High Volatility portfolio across time periods.

	1960s	1970s	1980s	1990s	2000s	2010s
1960s	0.000	-2.164	2.396	-9.429	1.315	-5.444
1970s	2.557	0.000	5.432	-10.668	4.028	-4.319
1980s	-2.820	-5.432	0.000	-12.448	-1.277	-8.535
1990s	10.932	10.155	12.455	0.000	11.789	7.368
2000s	-1.544	-4.004	1.277	-11.788	0.000	-7.313
2010s	6.333	4.166	8.539	-7.491	7.324	0.000

Table 8.8: Summary of returns and regression coefficients for the volatility portfolios in the 1960s Panel A shows the average monthly return, standard deviation (SD) of returns, the lowest monthly return (Min. Return) and the highest monthly return (Max Return) achieved in the 1960s. Panel B shows the results from a regression of the Low portfolio and the High portfolio (sorted on past 3-year volatility) on the excess market return (MktRf). The coefficients are presented (alpha and beta), the t-statistic whether the coefficients are significantly different from zero, and the p-value accompanying the result.

Panel A	Return	SD (returns)	Min. Return	Max Return
Low Portfolio	-.142	3.167	-6.490	7.103
High Portfolio	.522	8.622	-15.386	21.200
Market Portfolio	.374	3.320	-7.910	9.050

Panel B		Coefficient	t-statistic	p-value
Low Portfolio				
	β	.039	(.33)	.742
	α	-.142	(-.30)	.762
High Portfolio				
	β	.233	(.72)	.475
	α	.521	(.41)	.682

Table 8.9: Summary of returns and regression coefficients for the volatility portfolios in the 1970s Panel A shows the average monthly return, standard deviation (SD) of returns, the lowest monthly return (Min. Return) and the highest monthly return (Max Return) achieved in the 1970s. Panel B shows the results from a regression of the Low portfolio and the High portfolio (sorted on past 3-year volatility) on the excess market return (MktRf). The coefficients are presented (alpha and beta), the t-statistic whether the coefficients are significantly different from zero, and the p-value accompanying the result.

Panel A	Return	SD (returns)	Min. Return	Max Return
Low Portfolio	.226	3.765	-9.671	18.485
High Portfolio	.612	8.109	-26.344	23.461
Market Portfolio	.148	4.765	-12.750	16.100

Panel B	Coefficient	t-statistic	p-value
Low Portfolio			
β	.025	(.33)	.745
α	-.223	(-.61)	.542
High Portfolio			
β	.215	(1.31)	.193
α	.581	(.75)	.457

Table 8.10: Summary of returns and regression coefficients for the volatility portfolios in the 1980s Panel A shows the average monthly return, standard deviation (SD) of returns, the lowest monthly return (Min. Return) and the highest monthly return (Max Return) achieved in the 1980s. Panel B shows the results from a regression of the Low portfolio and the High portfolio (sorted on past 3-year volatility) on the excess market return (MktRf). The coefficients are presented (alpha and beta), the t-statistic whether the coefficients are significantly different from zero, and the p-value accompanying the result.

Panel A	Return	SD (returns)	Min. Return	Max Return
Low Portfolio	.987	3.000	-15.426	8.002
High Portfolio	-.008	6.621	-30.423	16.393
Market Portfolio	.600	4.775	-23.240	12.470

Panel B	Coefficient	t-statistic	p-value
Low Portfolio			
β	.081	(1.34)	.183
α	.938	(3.23)	.002
High Portfolio			
β	.233	(1.75)	.082
α	-.147	(-.23)	.817

Table 8.11: Summary of returns and regression coefficients for the volatility portfolios in the 1990s Panel A shows the average monthly return, standard deviation (SD) of returns, the lowest monthly return (Min. Return) and the highest monthly return (Max Return) achieved in the 1990s. Panel B shows the results from a regression of the Low portfolio and the High portfolio (sorted on past 3-year volatility) on the excess market return (MktRf). The coefficients are presented (alpha and beta), the t-statistic whether the coefficients are significantly different from zero, and the p-value accompanying the result.

Panel A	Return	SD (returns)	Min. Return	Max Return
Low Portfolio	1.098	2.806	-8.174	11.902
High Portfolio	1.736	6.490	-23.644	18.876
Market Portfolio	1.209	3.753	-16.080	10.840

Panel B	Coefficient	t-statistic	p-value	
Low Portfolio				
β	-.086	(-1.19)	.238	
α	1.209	(4.24)	.000	
High Portfolio				
β	.184	(1.10)	.274	
α	1.498	(2.27)	.025	

Table 8.12: Summary of returns and regression coefficients for the volatility portfolios in the 2000s Panel A shows the average monthly return, standard deviation (SD) of returns, the lowest monthly return (Min. Return) and the highest monthly return (Max Return) achieved in the 2000s. Panel B shows the results from a regression of the Low portfolio and the High portfolio (sorted on past 3-year volatility) on the excess market return (MktRf). The coefficients are presented (alpha and beta), the t-statistic whether the coefficients are significantly different from zero, and the p-value accompanying the result.

Panel A	Return	SD (returns)	Min. Return	Max Return
Low Portfolio	.567	4.443	-12.086	14.317
High Portfolio	.142	8.014	-22.321	19.167
Market Portfolio	-.010	4.743	-17.230	10.190

Panel B	Coefficient	t-statistic	p-value	
Low Portfolio				
β	.317	(3.71)	.000	
α	.570	(1.41)	.161	
High Portfolio				
β	.493	(3.14)	.002	
α	.147	(.20)	.843	

Table 8.13: Summary of returns and regression coefficients for the volatility portfolios in the 2010s Panel A shows the average monthly return, standard deviation (SD) of returns, the lowest monthly return (Min. Return) and the highest monthly return (Max Return) achieved in the 2010s. Panel B shows the results from a regression of the Low portfolio and the High portfolio (sorted on past 3-year volatility) on the excess market return (MktRf). The coefficients are presented (alpha and beta), the t-statistic whether the coefficients are significantly different from zero, and the p-value accompanying the result.

Panel A	Return	SD (returns)	Min. Return	Max Return
Low Portfolio	0.989	3.626	-8.521	9.509
High Portfolio	0.733	6.338	-16.990	20.267
Market Portfolio	1.047	3.498	-9.550	11.350

Panel B	Coefficient	t-statistic	p-value
Low Portfolio			
β	-.197	(-1.99)	.050
α	1.191	(3.31)	.001
High Portfolio			
β	-.210	(-1.20)	.232
α	.950	(1.49)	.139

9 References

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. Journal of Finance, 61(1), 259-299.

Asness, C.S., Frazzini, A., Gormsen, N. J., and Pedersen, L. H. (2020). Betting against correlation. Journal of Financial Economics, 135(1), 629-652.

Baker, M., Bradley, B., and Wurgler, J. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. Financial Analysts Journal, 67(1), 40-54.

Baker, N. L., and Haugen, R. A. (2012). Low risk stocks outperform within all observable markets of the world. SSRN working paper no. 2055431.

Bali, T. G., Cakici, N. (2008). Idiosyncratic Volatility and the Cross Section of Expected Returns. The Journal of Financial and Quantitative Analysis, 43 (1), 29-58.

Banz, R.W. (1981). The relationship between return and market value of common stocks. Journal of Financial Economics, 9(1), 3-18.

Black, F. (1972). Capital market equilibrium with restricted borrowing. Journal of Business, 45(3), 444-455.

Black, F., Jensen, M.C., and Scholes, M. (1972). The capital asset pricing model: Some empirical tests. Studies in the Theory of Capital Markets, 81(3), 79-121.

Blitz, D., Baltussen, G., and van Vliet, P. (2019a). When equity factors drop their shorts. Working paper. Available online: https://ssrn.com/abstract=3493305.

Blitz, D., Baltussen, G., and van Vliet, P. (2019b). The Low Volatility Effect Revisited. Available online: https://ssrn.com/abstract=3442749.

Blitz, D., Pang, J., and van Vliet, P. (2013). The volatility effect in emerging markets. Emerging Markets Review, 16, 31-45.

Blitz, D., and Vidojevic, M. (2017). The profitability of low-volatility. Journal of Empirical Finance, 43(1), 33-42.

Blitz, D., and van Vliet, P. (2007). The volatility effect. Journal of Portfolio Management, 34(1), 102-113.

Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. Journal of Financial Economics, 7(1), 197-226.

Fama, E. F., and French, K. R. (1992). The cross-section of expected stock returns. Journal of Finance, 47(2), 427-465.

Fama, E. F., and French, K. R. (2015). A five-factor asset pricing model. Journal of Financial Economics, 116(1), 1-22.

Fama, E. F., and French, K. R. (2016). Dissecting anomalies with a five-factor model. The Review of Financial Studies, 29(1), 69-103.

Fama, E. F., and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. Journal of Political Economy, 81(3), 607-636.

Frazzini, A., and Pedersen, L. H. (2014). Betting against beta. Journal of Financial Economics, 111(1), 1-25.

Garcia-Feijo, L., Kochard, L., Sullivan, R. N., and Wang, P. (2015). Low-volatility cycles: The influence of valuation and momentum on low-volatility portfolios. Financial Analysts Journal, 71(3), 47-60.

Harvey, C.R. (2017). Presidential Address: The scientific outlook in financial economics, Journal of Finance, 72(4), 1399-1440.

Harvey, C.R., Liu, Y. and Zhu, H. (2016). ... and the cross-section of expected returns. The Review of Financial Studies, 29(1), 5-68.

Haugen, R.A. and Heins A.J. (1975). Risk and the rate of return of financial assets: Some old wine in new bottles. Journal of Financial and Quantitative Analysis, 10(5), 775-784.

Hsu, C.C. and Chen, M.L. (2017). The timing of low-volatility strategy. Finance Research Letters, 23, 114-120.

Jobson, J.D. and Korkie, B.M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. The Journal of Finance, 36(4), 889-908.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics, 47(1), 13-37.

Liu, J., Stambaugh, R. F., and Yuan, Y. (2018). Absolving beta of volatility's effects. Journal of Financial Economics, 128(1), 1-15.

Memmel, C. (2003). Performance hypothesis testing with the Sharpe ratio. Finance Letters, 1(1), 21-23.

Miller, M.H. and Scholes, M. (1972). Rates of return in relation to risk: a reexamination of some recent findings. Studies in Theory of Capital Markets, Praeger, New York, 47-78.

Novy-Marx, R. (2014). Understanding defensive equity. NBER working paper no. 20591.

Sharpe, W.F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance, 19(3), 425-442.