Master thesis: Idiosyncratic volatility anomaly magnitude and dissection

Michel Roos, 414648

Supervisor: Dr. P.J.P.M Versijp

Second assessor: Dr. Jan Lemmen

Date final version: 5-8-2020

Stocks with high idiosyncratic risk tend to have low future returns, as was found first in Ang, Hodrick, Xing and Zhang (2006). This has been named the idiosyncratic volatility anomaly in academic financial research. Multiple studies have since attempted to uncover explanations for the prevalence of the anomaly. Examples of papers are Chen et al (2009), Jiang et al (2009) and Baker et al (2009). The purpose of this study was to uncover the robustness of the idiosyncratic volatility anomaly, in particular with respect to the January effect and market cycles. This paper looked into the effect of using different systemic risk factors on the idiosyncratic volatility variable as well as the effect of a diversity of other anomalies on the idiosyncratic volatility anomaly. The research finds that the idiosyncratic risk anomaly is very robust and identifies several potential driving forces and constituents. The main factors that were identified as significant are Investment, sales growth, accruals and cash and equivalents. The paper concludes breaking the sample up in different stock sectors.

Keywords: Volatility, Idiosyncratic, Return, Stock, Index, CAPM

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

Erasmus University Rotterdam

Erasmus School of Economics

Table of Contents

Non-plagiarism

I condone that this report is fully based on my own personal study and research. I have referred properly to all sources used in my work, no matter the origin of the source. This assignment has not been previously assessed in any way or form. I have not copied or otherwise plagiarized work of other people and/or students.

Introduction

The capital asset pricing model or CAPM for short predicts a positive sign in the relationship between systemic risk exposure and expected returns in the future. This is the theoretical core that a lot of finance research builds upon. However, there are anomalies in empirical market behavior, where the CAPM predictions seem false or not as accurate as predicted. One of such anomalies is called the low volatility anomaly.

The low volatility anomaly is the phenomena that in general studied markets stocks with low volatility have higher returns than high volatility stocks. The volatility of a stock is generally broken up in a systemic risk part. This is the co-movement with the market also known as beta. The residual of the volatility that cannot be explained by systemic risk is known as the idiosyncratic risk. The mismatch between returns and volatility is known as a stock market anomaly and deviates from the predictions of common financial theory, which state that investors receive higher compensation for taking more risk. The low volatility anomaly is referred to in several other ways, namely the low-beta, minimum-volatility or minimum-variance anomaly. Some forms of this anomaly even show that absolute return for stocks with higher volatility is lower, when compared to low-volatility stocks.

The low volatility anomaly has been documented over extended periods of time and in many countries. For example, in the United States stock market the anomaly has been present for a 90 year period. Historical evidence can be found for this from 1929 in the paradox investing database (Paradox, 2020). The company has used US stock data, split it in volatility deciles and compared the returns of the portfolios. Here the portfolio with the lowest volatility has a higher past return then the highest volatility portfolio.

In this paper the focus is on the idiosyncratic volatility anomaly, one measure of volatility where systemic risk is taken out of the equation, instead of looking at the total volatility of stocks. Considerable attention has been given to this variant of the low volatility anomaly. This study looks into the anomaly magnitude and dissects it using data from to the CRSP stock universe. This data is reported in different time frequencies by the CRSP.

This study investigates the magnitude of the idiosyncratic volatility anomaly as well as the relationship to stock returns and alpha as calculated in Carhart(1997). Further looking into the influence of usage of equal and value-weighted index returns, market cycles, the January effect, investment, past sales growth and accruals on the magnitude and significance of the anomaly and previously described relationships.

The goal is to further clarify the presence of the anomaly in CRSP stock data as Chen et al(2009) shows that for a large sample of stocks have the magnitude of idiosyncratic volatility anomalies and respective returns. The idea is to see what results indicate when the full stock sample is analyzed and dissected, with several robustness checks.

January is known to be a month with odd stock return and volatility behavior it is therefore of particular interest to research, high volatility stocks appear to do considerably better in January. Market cycles are known to influence correlations amongst stocks and other markets, they can possibly be of importance to the behavior of stocks with different levels of idiosyncratic volatility. Many balance accounts are known to influence stock return and volatility in the long run and are therefore interesting to consider analyzing. Examples of these include cash (flow), R&D expenditure and Capital expenditure.

In this paper I show that certain sample selection parameters can lead to different conclusions about the idiosyncratic volatility anomaly. Such as identifying what cycle the market is in and the month that we are currently in. The dataset is then further filtered to look for potential causes of the idiosyncratic volatility anomaly and either confirm or reject the presence of them.

The focus of the analysis is on mapping the relationship between returns and idiosyncratic volatility by calculating the alpha, portfolio return and average idiosyncratic volatility for different decile portfolios. The same calculations are then made for Januarys and other months. Robustness analysis is then done for market cycles, different computational methods off the idiosyncratic volatility. Analysis is done for annual time horizon, where several robustness checks are done for company balance accounts.

The fact that the anomaly its magnitude and parameters are subject to the sample selection is interesting and surprising. The value of the analysis is to try to reveal the potential causes and figure out what is behind the anomaly, and further robustness analysis. This is found in many papers where the volatility is analyzed periodically over a significant time horizon, for example Bandyopadhyay et al(2010).

The argument is that sample selection criteria are important when identifying stock characteristics. The market sentiment can exert influence in the magnitude of the idiosyncratic volatility anomaly. The underlying investor behavior appears significantly different for the state of the market as well whether it is January or not. A selection of balance accounts taken as financial ratio can exert influence in the future stock returns and reduce the significance and magnitude of the idiosyncratic volatility anomaly.

The idiosyncratic volatility anomaly is robust to controlling for other anomalies. It is therefore not a manifestation of other known market anomalies. Most variables to appear to absorb a part of the coefficient of the idiosyncratic volatility, but these are not enough to influence the significance of the idiosyncratic volatility anomaly.

The implications of the results are that the idiosyncratic risk anomaly is very robust as well as identifying several potential driving forces and constituents. The main factors that were identified as important are Investment, sales growth, accruals and cash and equivalents. The paper concludes breaking the sample up in different stock sectors. The IVOL anomaly sign flips for January where the risk return relationship seems significantly better than in other months. The IVOL anomaly is strongly persistent in market bear cycles and appears to weaken in bull cycles.

The rest of the paper starts off with theoretical background surrounding the idiosyncratic volatility anomaly, followed by discussion of the hypothesis. Next the methodology and used data is discussed, before diving into the results based on analysis of a monthly and an annual dataset, following up with the conclusions.

Literature review

In what is often referred to as the low-volatility anomaly, researchers have shown that measures of prior stock price variability—including total return volatility, idiosyncratic volatility, and beta—relate to future performance but not necessarily in the way that theory suggests, which is that investors are compensated with higher returns for taking a larger amount of risk.

The academic evidence has been stacking up in recent decades. Since the discovery of the anomaly in the early 1970s a lot of additional research has been done on the topic. Examples of research papers are Baker and Haugen (1991), Chan et al (1999), Jangannathan and Ma (2003), Clarke De Silva and Thorley (2006), Blitz (2007) Baker et al (2011 and 2012) and many more. Some of these papers look into the international stock market.

In research several broad explanations have been proposed and/or proven. Many investors would like to beat the market and perform relatively well compared to others, discussed in detail by Baker et al (2011). Many institutional investors have different interests when managing clientele funds, for example a bonus when beating the market, proposed by Karceski (2001). In general investors have a so called skewness preference as proposed by Barberis and Huang (2008). Leverage and shortage constraints possibly strengthen the anomaly as proposed by Brennan (1971) and tested by Frazzini and Pederson (2014). Behavioral biases are among named causes. These include but are not limited to heuristics, overconfidence and limited attention, further causes have been proposed.

Classical CAPM

The classical CAPM short for capital asset pricing model is a theory that suggest the return of a given stock should be a function of its beta or market risk if you will. It was introduced by William F. Sharpe (1964), John Lintner (1965) and Jan Mossin (1966). The model entails that market volatility/beta should be proportional to the expected market return. Sharpe later developed a measure of risk to return of an asset referred to as the Sharpe ratio. Despite the numerous empirical tests the CAPM did not survive it remains a popular analysis vehicle because of high utility and simplicity in numerous situations.

The CAPM predicts that returns should be proportional to the riskiness of the stock relative to the market portfolio. However the CAPM does not hold in practice, low volatility portfolios have significantly outperformed over the last five decades (Ang, 2006). These portfolios showed high average returns and small drawdowns, leading to the thesis that risk is not related linearly by profits. The low-volatility anomaly shows that prior stock price total return volatility, beta and idiosyncratic volatility relate to future performance but not continuously in the way that is assumed by the CAPM. The deviation is so extensive that this anomaly has been proposed as “the greatest anomaly in finance” by Baker et al (2011).

Under the efficient markets hypothesis investors can only realize higher returns by taking extra risk. Risky stocks should have high average returns and safe stocks should have the opposite. The generally accepted theoretical framework of risk therefore points people in a different direction.

Research on volatility anomaly

Already in the 70s researchers were aware of the empirical risk return relationship being different than predicted by the CAPM. Black et al (1972) show that the relationship between risk and realized returns. Haugen and Heins (1975) uncovered deficiencies in earlier studies about the relationship between risk and realized returns.

Research conducted by Constantinides (1984) determined that there is an option that is tax-timing connected with stock return volatility. Stocks with higher idiosyncratic volatility tend to have a higher tax-timing option value. Thus, investors could require lower expected returns to hold them.

The paper from Miller (1977) reports that stocks are priced optimistically when short sales are constrained. When pessimistic investors are kept out of the market and the remaining investors having diverse beliefs, which is proxied by a high IVOL, stocks have the tendency to be overvalued and this also makes a future price reversal highly likely.

Johnson (2004) developed a model that equity valuation appears to be valued as an option on the assets of levered firms. Businesses with higher measured idiosyncratic volatility tend to have higher equity value and a lower expected equity return over the fixed terminal value of the firm.

The low-volatility anomaly has been very impactful on the theory and practice of money management. Investors have in practice been exploiting the low volatility anomaly through mutual and exchange traded funds. Ramos & Hans (2016) reports that in the past years investors have allocated more than 10 billion dollars to these low volatility funds, which are strongly growing, however this is still marginal when compared to the mutual fund industry size of 20 trillion dollars. Most existing strategies only account for the risk factor, with little to no attention paid to the other characteristics known to influence portfolio results. One can think of analyzing time-varying nature of low-risk strategies and quantifying the nature of the relationship based on other characteristics besides risk. Examples of this include size, value momentum, and other macroeconomic variables.

Which way risk is defined (Volatility or Beta), or what size of stocks is analyzed, low risk stocks consistently outperform high risk stocks over the period. The low volatility anomaly appears present when using either residual volatility or total volatility. Remarkably an investor aggressively pursuing high volatility loss appears to have almost a full loss speaking in real terms over the past four decades as reported by Baker et al (2009), this of course accounts for the devaluation of currency through inflation. This shows that there are important behavioral biases to be discovered in market data, as the rational models fail to explain the observed market returns.

By researching a pile of broad international developed markets, Ang et al (2009) discovered that stocks with recent high idiosyncratic volatility had lower average future returns. 23 markets were investigated, across the sample the average difference between top and bottom quintile portfolios was -1,31% per month. This effect was individually significant for every G7 country. This leads to a conclusion that this phenomenon is not just a country specific effect but globally present. Haugen (2012) confirms the existence of this effect in every country specific equity market that was tested.

Several potential causes were found for the idiosyncratic volatility anomaly. The volatility deciles appear to be largely driven by news about the firms expected future earnings. This presents a relevant perspective by which to assess the inverse relation between idiosyncratic volatility and stock returns.

Behavioral aspects

Another perspective is investor's preference for high volatility stocks being derived from their biases. The preference for lotteries, the overconfidence bias and also the representativeness bias are important to emphasize here. Kahneman and Tversky received the Nobel Prize for their work in these psychological phenomena. In turn they all lead to increased volatile stock demand that is not rational, but has underlying behavioral causes.

Behavioral modeling of security prices seeks to combine two approaches. Firstly, some market participants are behaving irrationally one way or the other. Putting this into context, the low risk anomaly could be caused by the preference for lotteries and other established biases such as the ones mentioned in the previous paragraph. Investors proportionally buy up more high volatility stocks reducing their expected returns, inexplicable by stock fundamental analysis. Secondly there are limits to the ability of arbitrage to get rid of price inefficiencies. Institutional investors do not always offset the price impact of irrational market buying. This can be playing its part in the low volatility anomaly, the limits to arbitrage are likely underestimated in rational models as is addressed by Barberis and Thaler (2002).

Many institutional arbitrageurs have fixed benchmark mandates which are weighted by capitalization, this by nature discourages the investment in low volatility stocks. The implications drawn from the research of Brennan (1993) using the model of agency and assets prices reveals that the traditional fixed benchmark mandates constrained on leverage cause smart money to avoid the great risk-return ratio of the stock portfolios with low volatility. The institutions are thus not a stabilizing force in the markets, they are more so a destabilizing factor contributing to the existence of the anomaly.

Further (behavioral) explanations for the low risk anomaly are given by Karceski (2002). Irrationality comes from a different source in this paper. He determined that mutual fund investors chase the returns over time and across funds, this could be due to the stickiness of their flows of funds and also extrapolation bias. These biases make money managers more likely to care about the outperformance of the bull markets than the bear markets, hence increasing the demand for high beta stocks, even with their worse expected returns. This conclusion is complementary to the other behavioral biases that are found in other research. The focus for this paper will be on the idiosyncratic low volatility anomaly.

The section finalizing the behavioral aspects takes inspiration from the views of Chen et al (2009) considering idiosyncratic volatility and possible explanations. The idiosyncratic volatility anomaly is a popular topic of debate among finance researchers recently. A substantial amount of research reports an array of possible explanations for the presence of the anomaly. Examples of this include, but are not limited to Kapadia (2006) and Mitton, Boyer, Vorkink (2009) reporting that the anomaly is consistent with investor skewness preferences. Barinov (2008) uses a real-option model for the explanation of low expected returns of stocks with high idiosyncratic volatility. Huang et al (2009) show that idiosyncratic volatility no longer is a predictor of future stock returns after controlling for past stock returns. This applies in context of a cross-sectional regression setting.

Avramov, Cederburg, and Hore (2009) produced a rational asset pricing model that linked low expected return with high idiosyncratic volatility. The reason for this is that these stocks have low exposure to the modelled long-run growth factor. Boehme et al (2009) argues that the combination of short-sale constraints and heterogeneity of investor opinion may cause the negative volatility return relation. The argument is that in an information-segmented economy as from Merton (1987), the short sale constraint effect on stocks hypothesized by Miller (1977) can produce the inverse relationship between the idiosyncratic volatility and future stock return. Jiang et al (2009) associate the anomaly to selective corporate disclosure, in which management tends to report on goods news more extensively then on bad news. Evidence is disclosed within the paper that high IVOL stocks tend to have lower future returns.

Ang, Hodrick, Xing, and Zhang (2006) conducted research by which they found an anomaly that stock with higher idiosyncratic return volatility have lower future returns on average. Specifically the bottom quarter of idiosyncratic volatility outperform stocks in the top 25% by 1.06% per month. The results stay robust when controlled for firm size, value and stock momentum, volume, liquidity and dispersion of analyst forecasts. Extra evidence shows that this anomaly holds for the international stock market and is not explained by trading frictions, asymmetric investor information or skewness and other higher moments of returns as reported by Ang et al (2006).

Balance account anomalies

The accruals anomaly entails that firms that have consistent deviations of earnings from cash flows are punished for this by the market, they are perceived as firms presenting greater uncertainty and risk. Bandyopadhyay et al (2010) confirm the robustness of this anomaly across the NYSE/NASDAQ/AMEX-listed firms on the merged CRSP and Compustat database for the period of 1976 to 2008. One of the possible causes for this is the common appearance of compromising the earnings of companies. Firms with high accruals are expected to have low future stock returns and possibly higher idiosyncratic volatility.

R&D and Capex intensity significantly impact company performance and hence stock returns. According to Tubbs (2007) it appears companies can gain competitive advantage by increasing R&D in times of economic disparity. A substantial study looking into R&D expenses between 1974 and 2001 finds significant evidence for increased long term abnormal returns for companies that have higher R&D and capex expenses. Abnormal stock returns of between 30 and 56 basis points were found. It is hypothesized investors underreact to changes in R&D expenditure.

Sales growth is found to be of negative effect on stock returns due to excess extrapolation of analysts expected future sales growth. This result is confirmed by Lakonishok et el (1994) for US data. Lau et al (2002) confirm a similar result for the stock markets of Malaysia and Singapore. Sales growth also appears to interact strangely with the January effect. Sales growth appears to be a significant control variable only for non-January months, according to Lau et al (2002).

Equal weighted returns vs value weighted returns and benchmarking

When looking into the idiosyncratic volatility anomaly, its causes, and the future stock returns many papers calculate both value-weighted portfolio returns instead of equal-weighted portfolio returns. The study by Ang et al (2006) mainly reports VW results. The difference between value-weighted portfolio returns and equal weighted portfolio returns is generally interpreted as the effect of individual stock market size on the data. In the case of value-weighted portfolios the returns are not considered to be of the size effect. A study conducted by Bali and Cakici (2008) found existing IVOL anomalies only in the value-weighted portfolios.

Chen et al (2009) argue that this is not necessarily a valid interpretation of the data. Non-common stocks and penny stocks have significantly smaller market caps in general. Considering this these will have much lower weight in value-weighted portfolios in comparison to equal-weighted portfolios. From this one can conclude that the value returns come mainly from non-penny common stocks. They find that the negative return to IVOL relationship is significant among non-penny common stocks, all common stocks and the whole CRSP stock universe.

Market cycles

Bull and bear stock markets are economic phenomena that attract a lot of attention. The mean return shift appears consistently for extended periods of time. Gonzalez et al (2004) finds persistent differences between bull and bear market cycles. Volume is seen increasing while stock prices increase. It is documented that volatility is much higher when stock returns are negative. Volatility is lower when the stock market is consistently rising. The stock market appears to only have become increasingly more volatile over time. The good states become even better and the bad states become even worse. It is also found that bear markets appear to last a shorter 15 months on average compared to 21 months for a bull market according to the characteristics reported in Gonzalez et al (2004).

The January effect

The January effect is a well-known anomaly in stocks. Ritter (1988) finds that this effect is contained largely to small cap stocks. D'mello, Ferris and Hwang (2003) took note of abnormal selling pressure prior to the year-end for stocks that experienced large losses, their trade size also decreased. This phenomenon occurred inversely for capital gain stocks in January. They suggest this is due to individual tax-loss selling, implying the origin to be with private investors and not institutional ones. However, this effect also appears to be present in countries where the tax system works differently and is not dependent on January according to Haug and Hirschey (2006).

The January effect appears absent in large-cap stocks, this supports the thesis that the January effect is largely a small cap stock occurrence. Lokonishok and Smidt argue that the January effect is not a real tradeable anomaly because of the large bid-ask spreads and thin volume making it hard to trade profitably on this effect. Sullivan Timmerman, and White argue that the January effect is a form of the so called data-snooping hypothesis. This entails that the January effect is caused by investors having a preference for taking buying decisions in the beginning of the New Year.

In relation to the idiosyncratic volatility anomaly Jiang et al (2009) finds that the anomaly is not present in January. Their research using volatility deciles shows that stocks have higher returns and alphas in higher volatility deciles for January. The relationships changes significantly when isolating January, in fact it seems to flip.

Hypotheses

The results from this study are expected to confirm the conclusion from the mainstream literature on the anomaly, this means that the anomaly will be present and of negative sign throughout the data sample. The idea is to stick with the findings of previous literature as a benchmark measure so an outperformance for the top decile and underperformance for the bottom quantile of stocks sorted by the idiosyncratic volatility for a cross section of stocks.

The anomaly is not robust to the January effect, the relationship between idiosyncratic volatility and return is likely not of negative sign. The relationship will flip for January based on the fact that previous analysis from Chen et al (2009) has these findings. This paper has a different time horizon but I expect the conclusion to be the same.

Sector analysis is likely to produce results that sectors possessing large co-movement in their stock will see less significant idiosyncratic volatility anomaly. This because the measure is based on individual stock volatility and the analysis does not account for sector specific correlations among stocks. It would make sense that the sectors with more earnings uncertainty and higher stock volatility will show a higher idiosyncratic volatility anomaly. The sectors that appear to have the highest stock volatility are energy, commodities and financials. Their volatility exceeds that of technology stocks. However, in the analysis idiosyncratic volatility is used, some sectors may experience more systemic risk than others and therefore it is difficult to hypothesize.

I expect the over performance of low volatility stocks to be more pronounced in bear markets then in bull markets. This is likely due to the fact that the volatility in bear markets tends to be higher than the volatility in bull markets. A number of authors have found evidence for this, this includes Gomez Biscarri and Perez de Gracia (2004), Guidolin and Timmermann (2005), Tu (2006), Maheu and McCurdy (2000), Edwards et al (2003), Joneset al (2004), Gonzalez et al (2005) and Nishina et al (2006) amongst others. The expectation is that in January the IVOL anomaly has the opposite sign or is at the least weaker than for other months, as is found by Chen et al (2009).

The size and value effect in the computation of the idiosyncratic volatility anomaly are likely to have a slight influence on the analysis and robustness, but nothing major is expected, the same is expected for the use of an equal-weighted or alternate index. Generally the difference between equal and value weighted returns is looked upon as a size effect. All methods are expected to produce an idiosyncratic volatility anomaly that is highly significant when tested over the top minus bottom decile. The influence of several balance accounts on the idiosyncratic volatility anomalies will be mapped and tested. This analysis is run with a different investment horizon through the use of annual data.

Sales growth is expected to have a negative impact on future earnings; high growth tends to produce future company earnings shocks, which leads to likely a larger IVOL as identified by Jiang et al (2009). Lagged R&D expenditure as a ratio of market capitalization and as a ratio of total assets is expected to be of positive sign when dependent on the volatility estimates. This effect is likely strongly significant as found by Jiang et al (2009).

According to Sloan (1996) the accruals anomaly entails that stock with higher accruals tend to have lower future stock returns. Higher accrual companies usually have more uncertain earnings and therefore a higher idiosyncratic volatility (anomaly).

Methodology & Data

In this section the method of finding and testing negative relationship between the idiosyncratic volatility anomaly, also referred to as IVOL, and returns will be explained. The potential causes for the idiosyncratic anomaly that are tackled in this paper include research and development expenditure, capital expenditure, past sales growth and the accounting accruals effect. Circumstances analyzed include the January effect and the impact of the market being in a bull or bear phase of the magnitude of the idiosyncratic volatility anomaly. This section is concluded with sector specific IVOL analysis.

The daily stock data from the CRSP will be the focus of research regarding the IVOL anomaly. The total returns will be used, including dividends. The returns will be value weighted to exclude interpretation of the anomaly as a size effect. The data that is used is stock returns, prices, shares outstanding. A large dataset of around 65.000.000 observations is used, in order to have predictive power for all parts of the analysis, the stock data is recorded at daily frequency. The data sample contains around 28000 stocks from the CRSP institute. These are all stocks that are or have been included in the CRSP stock universe for the analyzed timeframe. From Compustat all the available data from WRDS is used. The focus will be on CRSP stocks from a variety of sectors. Every month the portfolio will be rebalanced. The IVOL will also be estimated on a monthly basis using the three-factor model from Fama and French (1993) and without using the model. Some time-lags/leads will be included as regressors in the analysis to account for nonsynchronous trading. All variables used are initially Winsorized. The Fama and French factors daily, monthly and annual data can be sourced from Ken French's website. Balance account data can be obtained from Compustat. From the Compustat data sample companies that have negative capital expenditure, revenue, total assets or negative total capital will be excluded from the sample. Firms that have less than four years of observations will be filtered out of the sample.

For accurate representation of stock volatility a stock has to have data for at least 15 days or 3 trading weeks in any particular month considered. Stocks trading at prices below 0 are removed from the sample. The sample period of the stocks will be the past 50 years (from Q4 1969 up until and including Q4 2019) IVOL estimation regressions have to be performed on a monthly basis for every stock. Following Fama and French firms with negative book values are excluded.

Deciles will be formed based on the IVOL analysis. The data will later be split up into sectors in order to analyze the behavior of the anomaly over different sectors. The selected portfolios should contain at minimum 20 stocks at any time. The data will also be split on five decade long time intervals, and comparing the behavior of the anomaly during times of recession and times of economic boom, the January effect in the data is analyzed.

In the latter part the approach is explained for analyzing other anomalies and their relationship to the IVOL. The effect of sales growth, capital expenditure and R&D expense and accruals are included. At last the behavior of the IVOL in different stock sectors will be studied.

The IVOL estimation will be done according to the following formula, specifically the IVOL is the standard deviation of the residuals from the regression below. This approach is very similar to the one used by Ang et al. (2006) This is to capture only the idiosyncratic risk measure filtering out the systemic risk. The time lags are included to control for the effects related to nonsynchronous trading. Following the approach from Chen et al. (2009) The model will also be estimated without time lags as regressors in order to check for empirical consistency of the measure.

$$r_t = \alpha + \sum_{i=0}^{3} \gamma_i r_{m,t-i} + \epsilon_t \tag{1}$$

Here $r_t$is the daily stock return, $r_{m,t-i}$ is the daily CRSP value-weighted index return.

A second estimation will be done too using the three factor model by Fama and French.

$$r_t = \alpha + \beta_1 HML_t + \beta_2 SMB_{\ t} + \sum_{i=0}^{3} \gamma_i r_{m,t-i} + \epsilon_t \tag{2}$$

Here $r_t$is the daily stock return, $r_{m,t-i}$ is the daily CRSP value-weighted index return. HML_t and $SMB_{\ t}$ are the daily Fama-French book-to-market and size factors.

For comparison purposes the IVOL will also be estimated based on equal-weighted index returns, in conjunction stocks will be benchmarked against the SP500 value weighted returns benchmark instead of using the CRSP benchmark solely. The difference in IVOL estimation will be compared.

As a robustness check a model without any lagged variables is also contemplated as well as model that has three leads and lags for each variable to determine the consistency amongst results. There will also be controlled for past monthly returns and for past annual returns to see if this explains away the IVOL predictive power for the sample using Fama-Macbeth regression.

There will be an alternative index measure used in conjunction with the CRSP value-weighted index return. The CRSP equal-weighted index return will be used to further clarify the impact of stock sizes on the analysis. Stocks will be held to the S&P500 index benchmark to draw a comparison.

From equation one three lags were implemented, the standard deviation of the residuals coming from the regressions estimates can be compared to see how the different lags impact the estimation process of the IVOL measure.

For accurate representation of stock volatility a stock has to have data for at least 15 days or 3 trading weeks in any particular month considered. The sample period of the stocks will be the past 50 years(from Q1 1970 up and until Q4 2019) regressions have to be performed on a monthly basis for every stock.

Deciles will be formed based on the IVOL analysis. The data will later be split up into sectors in order to analyze the behavior of the anomaly over different sectors. The selected portfolios should contain 25 stocks at any one time. The data will also be split on decades, and also comparing the magnitude of the anomaly during times of recession and times of economic boom. As a risk return measure the Sharpe ratio will be computed for both idiosyncratic and total risk measures for every month and stock.

Stock returns are investigated for a holding period of a month after the measured of the IVOL according to the regression. The stocks will be ranked at the end of each period respectively, to form the equal weighted deciles which are then held until the next period passes. The IVOL estimation month can be referred to as the IVOL estimation period similar to Ang et al. (2006a) In all cases only information before the end of the month is used.

When calculating holding period returns Shumway(1997) is followed by treating delisting returns and replace missing delisting returns by 0 if not performance related and otherwise -30%. The time series means of the monthly holding periods are to be calculated as well as their t-statistics. The time series t-statistics is computed using the Newey-West (1987) heteroscedasticity and autocorrelation consistent covariance estimator. This procedure makes use of 12 months or 1 annum of lag(s).

All portfolio characteristics are linked to the determined cross-section of stock returns, therefore further controlling for effects is necessary. This is done by following the Carhart (1997) four factor model.

$$r_t - r_{f,t} = \alpha + \beta_1(r_{m,t} - r_{f,t}) + \beta_2 HML_t + \beta_3 SMB_t + \beta_4 UMD_t + \epsilon_t \qquad (3)$$

In this equation r_t is the monthly portfolio return, $r_{f,t}$ is the monthly T-bill yield, and $r_{m,t}$ is the monthly return of the CRSP value-weighted index. HML_t, SMB_t and UMD_t are the Fama-French monthly size, book-to-market, and momentum factors.

If the IVOL anomaly is indeed present the regressions will show a significantly higher idiosyncratic risk to return ratio in the data. So the more the IVOL measure increases the less the return relatively increases. Then the difference in relative returns between high and low volatility portfolios can be calculated in order to test the significance of this difference across the different quintiles of idiosyncratic volatility. The stocks will be sorted in terms of IVOL magnitude into equal weighted deciles, this will allow analysis and testing of the risk average risk return ratio between different groups. The alpha should be considerably higher for lower deciles of idiosyncratic risk then for the higher deciles.

The presence of the anomaly will be tested using the Newey-West as explained before approach on the difference between the highest and the lowest idiosyncratic volatility equal weighted decile IVOL statistics over time. The Newey-West procedure accommodates heteroscedasticity and autocorrelation. The same procedure will be done on the individual portfolio returns and the alpha generated by the monthly Carhart four-factor alphas from regression (3). This way we can investigate to what degree the volatility in anomaly present within the data and how this relates to stock returns. The t-statistics have to be sufficient for the results to hold at 5% significance at the minimum but 1% is preferred, so that the coefficient is significantly different from 0, to reject the null hypothesis that there is no volatility anomaly present. Which corresponds to a t-statistic of 1,96 for either sign.

For each idiosyncratic volatility decile the collective alpha will be tested with the Newey-West procedure. If there is no significant excess return the alphas will be either insignificantly different from zero or negative. This entails that there is no significant excess return over the market when accounting for all the factors from the equations above, and vice versa.

From the second regression results a conclusion can be drawn about the effect of the Fama-Macbeth 3-factor model. The influence and significance of the market size growth as well as the book-to-market growth will show how these factors impact the IVOL anomaly estimation and stock returns.

From the third regression results a conclusion can be drawn about the effect of the Carhart 4-factor model. The influence and significance of the market size growth, book-to-market growth and momentum factor, this will show how these factors impact the IVOL anomaly estimation and stock returns.

The January effect

The effect for January will be analyzed through two sample selection on the different IVOL-return relations, either being in January or not in January. Doran, Jiang and Peterson(2008) document an interesting sign flip of the relationship in January. The IVOL-return is significantly positive during January while it is significantly negative in other months. The analysis previously described is then run for each category of observations. The following regression is used.

To test the January effect the procedure explained in the main section before will be used. In short this entails conducting a high minus low decile Newey-West procedure test. Followed by a Fama-Macbeth regression as a robustness check. The Fama-macbeth regression is done as follows. It is run once without dummy and once with as depicted here.

$$r_t - r_{f,t} = \alpha + \beta_1 IVOL_t + \beta_2 HML_t + \beta_3 SMB_{\ t} + \beta_4 UMD_{\ t} + D_1 \alpha \qquad (4)$$

$$r_t - r_{f,t} = \alpha + \beta_1 IVOL_t + \beta_2 HML_t + \beta_3 SMB_{\ t} + \beta_4 UMD_{\ t} + \beta_5 D_1 \qquad (5)$$

Here $r_t$is the annual sector stock return, $IVOL_t$is the annually compounded idiosyncratic stock volatility as came out of the monthly regression (2). HML_t and $SMB_{\ t}$ are the annual Fama-French book-to-market and size factors. $UMD_{\ t}$ encompasses the annual Fama-French momentum factor. $RET_{\ t}$ is the stock return during month t. RET0 is the stock return in a year up until month t. The dummy is 1 if the month is January and otherwise zero. Both regressions are run separately due to multicollinearity issues between the dummy and it’s constant.

Further robustness checks

Market cycles

All bear periods are collectively analyzed as a single dummy variable in order to compare the behavior of the IVOL anomaly in different market environments. In the sample period of 1970-2020 all years where the CRSP value weighted index produces a negative return will be classified under the bear market dummy. The analysis previously described is then run for each category of observations. After this the constants of both the bear and bull cycles can be tested to see if there is a significant difference.

To test the market cycle effect the procedure explained in the main section before will be used. In short this entails conducting a high minus low decile Newey-West procedure test. Followed by a Fama-Macbeth regression as a robustness check (see equation 4). Logically the dummy is now 1 if there is a bear market as defined above or otherwise zero.

Equal weighted vs value weighted index returns and alternate benchmark

The difference between IVOL generated based on value weighted index returns will be compared to the IVOL generated by equal weighted index returns. In order to see if this makes notable difference in the magnitude of the IVOL and the magnitude of the size effect in large cap stocks. The IVOL generated based on the respective stock index benchmark (SP500 vs CRSP) is compared to see the effect on magnitude of using a different benchmark.

Annual robustness checks

The annual data will be broken into deciles and statistics are reported for each individual decile, in order to judge the distribution of each of the variables.

The following regression is used as a starting point for the annual data. These regressions are similar to the ones used to determine the alphas in equation three.

$$r_t - r_{f,t} = \alpha + \beta_1(r_{m,t} - r_{f,t}) + \beta_2 HML_t + \beta_3 SMB_t + \beta_4 UMD_t + \epsilon_t \qquad (3)$$

In this equation r_t is the annual portfolio return, $r_{f,t}$ is the annual T-bill yield, and $r_{m,t}$ is the annual return of the CRSP value-weighted index. HML_t, SMB_t and UMD_t are the Fama-French annual size, book-to-market, and momentum factors. Each balance account variable explained below is then added independently and in combination on a trial and error basis to attempt to uncover interesting findings. The most relevant findings are reported in the results section.

Sales growth

Firms with high past growth; generally also having a high idiosyncratic volatility anomaly have an increased risk of future negative income shocks. Investors are known to suffer from extrapolation bias and this may explain why stocks with high idiosyncratic volatility have negative future earnings shocks Lakonishok et al (1994) concludes.

The influence of sales growth will be probed using a similar approach to Lakonishok et al (1994) the firms past growth is measured using the previous year its annual sales growth rate. In order to see whether firms with higher growth do indeed have future income shocks. The investor bias is left for what it is in this paper.

In this and the following sections all analysis will be done from year-to-year. The entire sample period is used for research which is 1970-2020. The IVOL will be compounded over the twelve months for the portfolio holding period, which is from the beginning of the year to the end of the year. The Fama French factors annual data is used in this part, and the next.

The investment anomaly

When firms engage in large capital ventures or expenditures, their business fundamentals are often dramatically impacted by this undertaking. Mainly the uncertainty about future cash flows is increased. Therefore it seems logical there is a link between idiosyncratic stock return volatility and firms capital expenditure. It is sensible to use the R&D expenditure and CAPEX of firms to attempt to explain the empirically documented relationships among IVOL, future returns etc.

The intensity of R&D expenditure is measured by the ratio of R&D expenditure in the fiscal year that ends in calendar year t – 1 to the market capitalization of the firm at the end of June of year t. Data used to compute R&D and CAPEX measures are from Compustat. R&D includes both intangible and tangible investments. To track the effects of only tangible investments the CAPEX effect is measured using the capital expenditure in the fiscal year that ends in calendar year t – 1, scaled by total assets at the beginning of the fiscal year.

The accruals anomaly

In 1996 Sloan has shown that stocks that have high accruals have poor future stock returns. The explanation provided by Sloan is poor investor reaction to information. Most investors do not recognize the difference between the accruals and cash flow component of earnings in their stock valuation process. The accruals anomaly may actually be a component of the idiosyncratic volatility anomaly, which will be tested through the following approach.

Then same approach is used that Sloan(1996) did. The accruals for individual firms are measured using annual balance sheet and income statement data from Compustat. This data is used in the following formula to determine the extent of the accruals anomaly.

$$Accruals = \frac{[(\Delta CA - \Delta CASH) - (\Delta \mathrm{CL} - \Delta \mathrm{STD} - \Delta \mathrm{TP}) - DEP]}{ATA} \tag{4}$$

In this function ΔCA is the change in current assets from the previous fiscal year. $\Delta CASH$ is the change in cash and short term investments, $\Delta \mathrm{CL}$ is the change in current liablities, $\Delta \mathrm{STD}$ is the change in the debt component of the current liabilities, $\Delta \mathrm{TP}$ is the change in income taxes payable. DEP is the depreciation and amortization expense. ATA is the average of the total assets in the beginning and end of the fiscal year. So in conjunction with the investment anomaly the accruals data is also measured for the fiscal year ending in t – 1.

Sector analysis

The prevalence of IVOL through different sectors will be analyzed by estimating a Fama-Macbeth regression for different sectors which can be identified through the North American industry classification code from Compustat. It would seem logical that stocks in sectors that have a higher volatility would have a higher IVOL volatility anomaly. NASDAQ stocks are grouped separately as well to see what the different IVOL magnitude and significance is for tech stocks. The following regression is run for every sector that is included with abundant observations in the data, the specific sectors are defined in the tables of results.

$$r_t = \alpha + \beta_1 IVOL_t + \beta_2 HML_t + \beta_3 SMB_t + \beta_4 UMD_t \qquad (5)$$

Here $r_t$is the annual sector stock return, $IVOL_t$is the annually compounded idiosyncratic stock volatility as came out of the monthly regression (2). HML_t and SMB_t are the annual Fama-French book-to-market and size factors. UMD_t encompasses the annual Fama-French momentum factor.

Results

The first section looks into the idiosyncratic anomaly estimations. Here the distinction is made between estimates accounting for non-synchronous trading only, contemporaneously accounting for Fama-French size and book-to-market effect. Estimates are done using both equal-weighted and value-weighted returns, and using CRSP index as a benchmark as well as SP500 index.

After the main statistics, the decile Newey-West procedure tests of the anomaly are presented, in conjunction to dissection of the anomaly that is present. The aspects of the dissection include looking at market cycles and the January effect for monthly data. For annual data I look at Sales growth and investment, accruals and analyzing the anomaly over different sectors.

Monthly analysis

Decade	IVOL	IVOL P10	IVOL P90	N
1970-79	2,20	1,97	2,54	3472
1980-89	2,35	2,05	2,74	5680
1990-99	3,10	2,56	3,72	9742
2000-09	2,27	1,94	2,66	4208
2010-19	1,47	1,29	1,72	4852

Decade	IVOLFM	IVOLFM P10	IVOLFM P90
1970-79	2,00	1,79	2,31
1980-89	2,13	1,86	2,50
1990-99	2,82	2,33	3,39
2001-09	2,06	1,75	2,42
2010-19	1,33	1,16	1,56

Table 1 reports summary statistics of idiosyncratic volatility during various decades that are included in the sample period. This includes the 90th percentile referred to as P90 and the 10th percentile referred to as P10, the mean is reported in conjunction with the number of stocks that is included in any decade indicated as N. The idiosyncratic volatility(IVOL/IVOLFM) is computed by calculating the standard deviation of residual estimates from regressing daily individual stock returns onto contemporaneous daily CRSP value weighted market returns and three lags thereof(T-1 to T-3). IVOLFM strips away size and value effect according to the Fama-French factors. In every decade the 90th percentile, 10th percentile and mean are computed. Each stock has traded at least 15 days in any particular month to qualify to be represented in the data. Stocks that end their delisting due to performance related issues have their final daily stock return replaced by -0,3. Results are reported as a percentage.

Over time the monthly idiosyncratic volatility is seen fluctuating substantially peaking at an average of 3,10% for the 90s decade, and bottoming at 1,47% in the 10s decade. The relative range of both percentiles is seen increasing with the volatility. The distribution of the IVOL data appears to be right skewed. The number of stocks varies greatly over the decades at its peak 9742 stocks are included in the sample, in the first decade of analysis around 3472 stocks are included. The Fama-French size and book-to-market effect leads to lower idiosyncratic volatility estimates, yielding a more conservative range of estimates in idiosyncratic volatility.

IVOL decile	IVOL	IVOL P10	IVOL P90	IVOLFM	IVOLFM P10	IVOLFM P90	N
1	0,39	0,14	0,60	0,34	0,10	0,53	3783
2	0,79	0,67	0,91	0,71	0,59	0,82	1809
3	1,08	0,97	1,19	0,97	0,87	1,07	1865
4	1,36	1,25	1,48	1,23	1,12	1,34	1957
5	1,68	1,54	1,81	1,51	1,39	1,64	2161
6	2,04	1.88	2,20	1,84	1,70	1,99	2419
7	2,48	2,28	2,68	2,25	2,07	2,43	2559
8	3,06	2,79	3,34	2,78	2,54	3,04	2767
9	3,95	3,51	4,46	3,60	3,20	4,07	2597
10	6,42	4,82	8,75	5,86	4,39	8,01	2033
Full sample	2,32	1,99	2,74	2,11	1,80	2,49	27954

Table 2 reports summary statistics of idiosyncratic volatility broken up into deciles for the sample period. This includes the 90th percentile referred to as P90 and the 10th percentile referred to as P10, the mean is reported. The idiosyncratic volatility(IVOLFM) is computed by calculating the standard deviation of residual estimates from regressing daily individual stock returns onto contemporaneous daily CRSP value weighted market returns and *three lags thereof(T-1 to T-3), furthermore the Fama-French size and book-to-market daily values are used . In every decade the 90th percentile, 10th percentile and mean are computed. Each stock has traded at least 15 days in any particular month to qualify to be represented in the data. Stocks that end their delisting due to performance related issues have their final daily stock return replaced by -0,3. Results are reported as a percentage. N is the average amount of stocks per portfolio used in both measures.*

Regarding the distribution of the idiosyncratic volatility in decile format, the average idiosyncratic volatility ranges between 0,39% and 6,4% at the respective bottom and top deciles. For the IVOLFM this is 0,33% to 5,8% penny stocks are found at the top of the idiosyncratic volatility distribution which explains why the range is wide. The percentiles of the top decile are wider then that of the others, probably due to presence of penny-stocks in this part of the distribution. The statistics confirm the previous notion of the effect of the Fama and French factors. With average idiosyncratic volatility being 0,21% lower when accounting for them.

IVOL decile	IVOL	PRET	Alpha	Idiosyncratic Sharpe	Sharpe ratio
1	0,39	0,41	-0,70(-46,95)	-8,57E+13	0,05
2	0,79	0,64	-0,47(-34,98)	0,82	0,12
3	1,08	0,55	-0,47(-26,63)	0,52	0,10
4	1,36	0,47	-0,47(-19,85)	0,35	0,08
5	1,68	0,40	-0,56(-17,26)	0,24	0,06
6	2,04	0,36	-0,55(-12,98)	0,18	0,05
7	2,48	0,22	-0,62(-12,11)	0,09	0,03
8	3,06	0,11	-0,69(-11,02)	0,04	0,01
9	3,95	-0,28	-1,07(-11,84)	-0,07	-0,04
10	6,42	-1,62	-2,19(-8,18)	-0,23	-0,15
H-L	6,02	-2,02	-1,50	-1,05	-0,20
T-stat	123,42	-90,23	-47,79	-74,56	-105,73
Full sample	2,32	0,13	-0,78	-8,57E+12	0,03

Table 3 reports average idiosyncratic volatility(IVOL), average monthly Carhart(1997) four factor alphas(Alpha), and average monthly portfolio returns(PRET) for equal-weighted decile portfolios, which were sorted on the idiosyncratic volatility. In each month the stocks are sorted into deciles based on the IVOL and form equal-weighted portfolios. The idiosyncratic volatility is computed by calculating the standard deviation of residual estimates from regressing daily individual stock returns onto contemporaneous daily CRSP value weighted market returns and three lags thereof(T-1 to T-3). The portfolios are all held for one month, and are formed using the entire sample of large cap common stock data. The average number of stocks in each portfolio is reported as N. Alphas and returns are in percentage points, and Newey-West t-statistics are reported in parenthesis. All portfolios are required to have at least 20 stocks in any particular month otherwise the returns in that month are excluded from further computation. Results are reported as a percentage.

The idiosyncratic volatility decile test provides a very significant difference in idiosyncratic volatility between the top and bottom decile, the substantial absolute difference was noticeable in the data already. The test confirms the significance of this difference. The difference in volatility is 6,02%.

In terms of returns the difference is significant up to the 1% level as well. This is evidence that the return is substantially higher in the bottom decile while the volatility is significantly lower. The Sharpe ratio of the second decile is 0,2 which is substantially higher than the -0,2 that is calculated for the top decile. This rejects the notion put forth by the CAPM and confirms the anomaly is present within the data. The idiosyncratic risk does not scale proportionally with returns for this dataset. The difference in return is -2,02%. The risk return trade-off seems to be somewhat intact for the top two deciles and after this the absolute return is already seen decreasing. The top decile shows a negative stock return of -1,62% monthly.

In terms of Alpha the alpha is significantly lower for the top decile then for the bottom decile. There appears to be no positive alpha for the lower deciles which is unusual. This is potentially due to the momentum effect absorbing the Alpha. The alpha peaks at the bottom of the decile distribution. There is a sizeable difference between the Alpha H-L decile. The Sharpe ratio peaks at the second decile indicating the best risk return relationship in this portfolio. The idiosyncratic Sharpe peaks around the second decile as well.

IVOLFM decile	IVOLFM	PRET	Alpha	Idiosyncratic Sharpe	Sharpe ratio
1	0,34	0,31	-0,79(-45,52)	-2,51E+13	0,03
2	0,71	0,62	-0,47(-34,31)	0,90	0,12
3	0,97	0,54	-0,48(-26,59)	0,56	0,10
4	1,23	0,46	-0,49(-20,49)	0,38	0,08
5	1,51	0,37	-0,56(-17,15)	0,24	0,06
6	1,84	0,35	-0,56(-13,04)	0,19	0,05
7	2,25	0,28	-0,58(-10,66)	0,12	0,04
8	2,78	0,13	-0,69(-11,09)	0,05	0,01
9	3,60	-0,14	-0,91(-9,28)	-0,04	-0,02
10	5,86	-1,22	-1,91(-6,32)	-0,19	-0,12
H-L	5,53	-1,53	-1,12	-1,09	-0,15
T-stat	114,23	-67,13	-34,56	-78,76	-67,87
Full sample	2,11	0,17	0,75	-2,50E+12	0,04

Table 4 reports average idiosyncratic volatility(IVOLFM), average monthly Carhart(1997) four factor alphas(Alpha), and average monthly returns(PRET) for equal-weighted decile portfolios, which were sorted on the idiosyncratic volatility. In each month the stocks are sorted into deciles based on the IVOL and form equal-weighted portfolios. The idiosyncratic volatility is computed by calculating the standard deviation of residual estimates from regressing daily individual stock returns onto contemporaneous daily CRSP equal weighted market returns and three lags thereof(T-1 to T-3). The portfolios are all held for one month, and are formed using the entire sample of large cap common stock data. The average number of stocks in each portfolio is reported as N. Alphas and returns are in percentage points, and Newey-West t-statistics are reported in parenthesis. All portfolios are required to have at least 20 stocks in any particular month otherwise the returns in that month are excluded from further computation. Results are reported as a percentage. The H-L test uses the second decile in the idiosyncratic Sharpe.

The idiosyncratic volatility decile test provides a very significant difference in idiosyncratic volatility between the top and bottom decile, the substantial absolute difference was noticeable in the data already. The test confirms the significance of this difference. The difference in volatility is 5,53%. The confirmed magnitude of the anomaly is lower when accounting for Fama and French factors.

In terms of returns the difference is significant up to the 1% level as well. This is evidence that the return is substantially higher in the bottom decile while the volatility is significantly lower. This rejects the notion put forth by the CAPM and confirms the anomaly is present within the data. The difference in return is -1,96%. The risk return trade-off seems to be somewhat intact for the top two deciles and after this the absolute return is already seen decreasing. The top decile shows a negative stock return of -1,53% monthly. The difference in portfolio returns is of lesser magnitude when accounting for Fama and French factors.

In terms of Alpha the alpha is significantly lower for the top decile then for the bottom decile. There appears to be no positive alpha for the lower deciles which is unusual. In the literature it is quite common to find positive Alphas when looking into the low volatility anomaly, such as found by Chen et al (2009). This is potentially due to the momentum effect absorbing the Alpha. Carhart (1997) finds that when using a three factor model that includes size, book-to-market momentum the Alpha becomes slightly negative. He finds that excess returns are primarily driven by the one year momentum effect. This analysis used a different kind of data, namely mutual funds return data.

The alpha peaks at the bottom of the decile distribution. There is a sizeable difference between the Alpha H-L decile. The alpha appears slightly more significant than when using IVOL. The Sharpe ratio peaks at the second decile indicating the best risk return relationship in this portfolio.

	(1)	(2)	(3)
VARIABLES	CRSP whole	Bear market	Bull market
IVOL	-1,074***	-0,916***	-1,135***
	(-85,59)	(-42,25)	(-76,25)
HML	-0,00685***	-0,00754***	-0,00659***
	(-113,3)	(-62,34)	(-94,93)
SMB	0,00192***	0,00188***	0,00194***
	(39,59)	(19,16)	(34,86)
RETT	0,0326***	-0,00183	0,0476***
	(23,05)	(-0,74)	(28,02)
RET0	6,28e-06***	1,30e-05***	4,66e-06**
	(3,120)	(2,63)	(2,379)
Constant	-0,439***	-0,450***	-0,435***
	(-1,182)	(0,000689)	(-1,039)
Observations	3.224.636	861.099	2.363.537

t-statistics in parentheses
*** p<0,01, ** p<0,05, * p<0,1

Table 5 reports the results of the following Fama-MacBeth regressions. The dependent variable is the monthly individual stock returns during the month after the contemporaneous month, the month when idiosyncratic volatility is measured. The explanatory variables include idiosyncratic volatility (IVOL), fama French market size factor (SMB), fama French book-to-market factor (HML), stock returns during the 11 months prior to the contemporaneous month. RETT is the stock return during month t. The cross-sectional regressions are performed in each month. The time-series averages of regression coefficients are reported, their corresponding t-statistics, as well as the adjusted Rsquares. RET0 is included as an explanatory variable which is the stock return in year up until month t. The Newey-west t-statistics, reported in parenthesis. The sample uses data from 1970 up to 2020.

The idiosyncratic volatility has a decreasing effect on the next month's return of the entire CRSP stock universe. This effect appears less pronounced in bear markets then in bull markets. This is likely due to the systemic risk factor increasing in market downturns, price correlations are known to increase in bear markets.

The book-to-market factor has a negative effect on future returns and slightly more severely for the bear market data. The size factor has a positive effect on future returns this appears not to be affected a lot by the market cycles.

The return of the previous month appears to impact the return significantly positively, however the significance of this variable disappears entirely when the market is in a bear phase. The returns of the previous year have a small but significant effect on the returns of next month, the magnitude of this effect appears more marginal for bear phases.

VARIABLES	(1) CRSP whole	(2) Bear market	(3) Bull market	(5) January effect dummy	(6) January constant
IVOL	-1,086***	-0,915***	-1,139***	-1,088***	-1,088***
	(-85,59)	(-42,10)	(-101,3)	(-85,69)	(-85,69)
HML	-0,00685***	-0,00754***	-0,00659***	-0,00685***	-0,00685***
	(-113,2)	(-62,33)	(-98,75)	(-113,2)	(-113,2)
SMB	0,00192***	0,00188***	0,00194***	0,00192***	0,00192***
	(39,57)	(19,16)	(31,10)	(39,57)	(39,57)
January				0,00578***	
				(10,02)	
January constant					-0,0133***
					(-10,02)
Constant	-0,438***	-0,450***	-0,434***	-0,439***	-0,439***
	(-1.178)	(-654,1)	(-1.365)	(-1.174)	(-1.174)
Observations	3.224.636	861.099	2.363.537	3.224.636	3.224.636

t-statistics in parentheses
*** p<0,01, ** p<0,05, * p<0,1

Table 6 reports the results of the following Fama-MacBeth regressions. The dependent variable is the monthly individual stock returns during the month after the contemporaneous month, the month when idiosyncratic *volatility is measured. The explanatory variables include idiosyncratic volatility (IVOL), fama French market size factor (SMB), fama French book-to-market factor (HML), stock returns during the 11 months prior to the contemporaneous month. Controlling for previous returns is left out here. The fourth model includes a January dummy. The fifth model includes a January only regression constant. The Newey-west t-statistics, reported in parenthesis. The sample uses data from 1970 up to 2020.*

For completeness and clarity these regressions are included without controlling for previous return, in conjunction with a full sample January dummy. The effect of controlling for the previous time periods returns appears very limited. Januarys appear to generate a 0,578% extra monthly return across the full sample. The alpha for January is marginally different but is significantly different then the full sample constant.

Non-January

IVOLFM decile	IVOLFM	PRET	Alpha	Idiosyncratic Sharpe	Sharpe ratio
1	0,34	0,25	-0,82	-5,10E+11	0,07
2	0,71	0,59	-0,49	0,83	0,12
3	0,97	0,50	-0,50	0,52	0,09
4	1,23	0,40	-0,51	0,34	0,07
5	1,51	0,28	-0,62	0,20	0,05
6	1,84	0,21	-0,68	0,14	0,04
7	2,25	0,06	-0,77	0,05	0,01
8	2,78	-0,23	-1,00	-0,05	-0,03
9	3,60	-0,73	-1,47	-0,16	-0,09
10	5,86	-2,77	-3,38	-0,39	-0,24
H-L	5,52	-3,02	-2,56	-0,39	-0,31
T-stat	112,56	-106,57	-65,79	-10,35	-108,26
Full sample	2,10	-0,14	-1,03	-5,15E+10	0,01

January

IVOLFM decile	IVOLFM	PRET	Alpha	Idiosyncratic Sharpe	Sharpe ratio
1	0,34	1,08	-0,46	3,57	0,26
2	0,71	1,00	-0,24	1,34	0,20
3	0,97	0,97	-0,22	0,92	0,18
4	1,23	1,16	-0,20	0,87	0,21
5	1,51	1,35	0,00	0,81	0,22
6	1,84	1,84	0,73	0,91	0,29
7	2,25	2,61	1,51	1,07	0,38
8	2,78	3,93	2,60	1,29	0,53
9	3,60	5,94	4,86	1,51	0,72
10	5,94	13,72	12,27	2,00	1,19
H-L	5,60	12,64	12,73	-1,57	0,94
T-stat	125,6	169,13	112,67	-82,56	234,72
Full sample	2,22	3,58	2,34	1,41	0,44

Table 7 reports average idiosyncratic volatility(IVOL), average monthly Carhart(1997) four factor alphas(Alpha), and average monthly returns(PRET) for equal-weighted decile portfolios, which were sorted on the idiosyncratic volatility. In each month the stocks are sorted into deciles based on the IVOL and form equal-weighted portfolios. The idiosyncratic volatility is computed by calculating the standard deviation of residual estimates from regressing daily individual stock returns onto contemporaneous daily CRSP value weighted market returns and three lags thereof(T-1 to T-3). The portfolios are all held for one month, and are formed using the entire sample of large cap common stock data. The average number of stocks in each portfolio is reported as N. Alphas and returns are in percentage points, and Newey-west t-statistics are reported in parenthesis. The results are computed for January and compared to non-January months for the sample period of 1970 to 2020. Results are reported as a percentage. The H-L test uses the second decile in the idiosyncratic Sharpe.

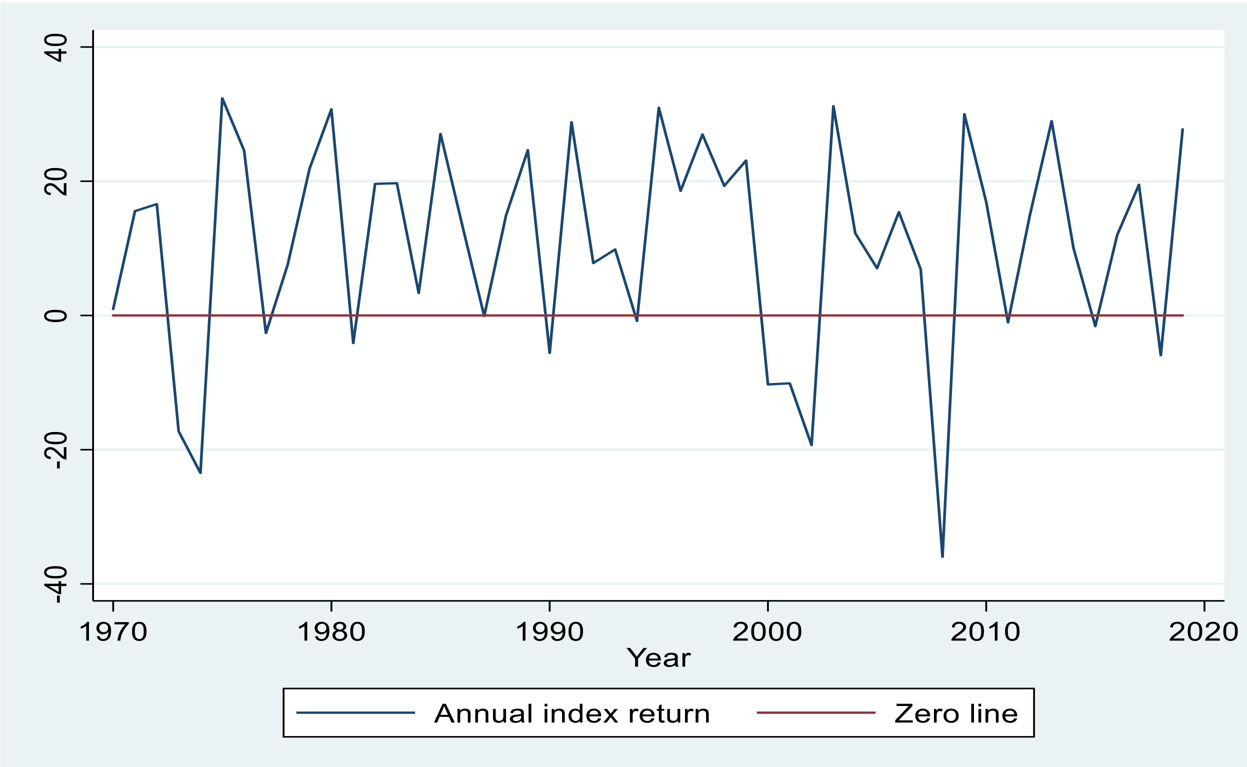
January is contrary to other months in its result. The H-L test of returns flips. It becomes highly significantly positive after exclusively analyzing January. The relationship between idiosyncratic risk and return is in agreement with the notion the CAPM puts forth. The returns now increase as the idiosyncratic volatility of returns rises, across all deciles. However one could now suggest picking stocks with a higher idiosyncratic volatility depending on the risk aversity of investors, this did not make sense based on the full sample.

The Alphas behave very differently in January's the top deciles now seem to be strongly performing in this month. The Sharpe ratio in the top decile is a lot higher than in the bottom decile here, this is contrary to the findings in the general sample.

For the idiosyncratic Sharp the idiosyncratic sharp becomes very high for the first decile. This is due to the idiosyncratic risk being very low in this area, while January comparatively gives good stock returns overall. The major source of risk for the bottom decile of stocks is therefore non-idiosyncratic. It could be that the nature of this risk is sector specific risk or that it is general market risk. That is beyond the scope of this paper.

Robustness checks

First the influence of market cycles on the robustness of the anomaly is assessed. The graph below shows the CRSP index returns over the years and the respective bear/bull market periods.



Graph 1 indicates which years have been marked as bear market. When the annual return falls below zero the year is marked as a bear year. The index used to determine this is the CRSP value-weighted index. The Annual index return is indicated as a percentage.

Bear market

IVOLFM decile	IVOLFM	PRET	Alpha	Idiosyncratic Sharpe	Sharpe ratio
1	0,33	-0,52	-0,26	-1,90E+12	-0,10
2	0,71	-0,30	0,17	-0,26	-0,06
3	0,97	-0,53	0,01	-0,43	-0,10
4	1,23	-0,76	-0,03	-0,51	-0,13
5	1,51	-1,02	-0,25	-0,58	-0,17
6	1,84	-1,24	-0,41	-0,58	-0,19
7	2,25	-1,44	-0,35	-0,55	-0,21
8	2,78	-1,99	-1,01	-0,63	-0,26
9	3,61	-2,96	-1,85	-0,73	-0,35
10	5,99	-4,66	-3,23	-0,72	-0,42
H-L	5,66	-4,13	-2,97	-0,46	-0,32
T-stat	145,78	-116,53	-57,54	-54,34	-64,55
Full sample	2,33	-1,71	-0,87	-1,77E+11	-0,21

Bull market

IVOLFM decile	IVOLFM	PRET	Alpha	Idiosyncratic Sharpe	Sharpe ratio
1	0,34	0,59	-0,97	2,34	0,14
2	0,71	0,90	-0,66	1,21	0,18
3	0,97	0,87	-0,63	0,86	0,16
4	1,23	0,86	-0,63	0,66	0,15
5	1,51	0,83	-0,67	0,53	0,14
6	1,84	0,91	-0,61	0,47	0,14
7	2,24	0,90	-0,66	0,39	0,13
8	2,78	0,94	-0,57	0,33	0,13
9	3,60	1,05	-0,51	0,29	0,13
10	5,79	0,69	-1,18	0,15	0,06
H-L	5,45	0,11	-0,21	-1,07	-0,08
T-stat	109,76	7,8	-23,79	-65,09	-20,78
Full sample	2,03	0,86	-0,70	0,74	0,14

Table 8 reports average idiosyncratic volatility(IVOLFM), average monthly Carhart(1997) four factor alphas(Alpha), and average monthly returns(PRET) for equal-weighted decile portfolios, which were sorted on the idiosyncratic volatility. In each month the stocks are sorted into deciles based on the IVOL and form equal-weighted portfolios. The idiosyncratic volatility is computed by calculating the standard deviation of residual estimates from regressing daily individual stock returns onto contemporaneous daily CRSP value weighted market returns and three lags thereof(T-1 to T-3). The portfolios are all held for one month, and are formed using the entire sample of large cap common stock data. The average number of stocks in each portfolio is reported as N. Alphas and returns are in *percentage points, and Newey-west t-statistics are reported in parenthesis. The results are computed for bear market and non-bear market years for the sample period of 1970 to 2020. The bear market dummy is defined as a particular month in the previous year having a return that is a negative percentage. Results are reported as a percentage. The H-L test uses the second decile in the idiosyncratic Sharpe.*

The bear market displays slightly higher idiosyncratic volatility a further indication of the idea that the systemic risk factor increases in market downturns. The returns are obviously all negative, the distribution does appear slightly different where the second IVOL decile now has the best returns. The anomaly strongly persists in bear markets.

The alphas are a lot more negative for the bear market years indicating stronger underperformance of high idiosyncratic volatility stocks when the general market is down. The differences in alpha become a lot less noticeable when the bear market years are filtered out of the sample, the same goes for the Sharpe ratio. Despite their being relatively fewer bear market years it is evident that they have a strong influence on the results of this study.

In bull markets the anomaly appears weaker than in the entire sample. The idiosyncratic volatility is lower than in the full sample. The returns are now positive for all deciles leading to a significantly positive H-L IVOL decile test. Proportionally to risk the return for the top decile is still lower.

For the idiosyncratic Sharp the idiosyncratic sharp becomes very high for the first decile. This is due to the idiosyncratic risk being very low in this area, while bull markets comparatively give better stock returns overall. The major source of risk for the bottom decile of stocks is therefore non-idiosyncratic. It could be that the nature of this risk is sector specific risk or that it is general market risk. That is beyond the scope of this paper.

In order to assess the robustness of the current methodology of calculating idiosyncratic risks several alternate approaches have been explored.

The methodology for computing the IVOLE variable is very similar to the IVOL variable. Instead of the value-weighted CRSP market index returns, equal-weighted index returns are used. For further explanation refer to the methodology section.

Decade	IVOLE	IVOLE P10	IVOLE P90
1970-79	2,19	1,97	2,53
1980-89	2,35	2,06	2,74
1990-99	3,09	2,56	3,71
2000-09	2,25	1,93	2,65
2010-19	1,47	1,29	1,70

IVOLE decile	IVOLE	IVOLE P10	IVOLE P90	IVOLEFM	IVOLE P10	IVOLE P90
1	0,41	0,16	0,61	0,35	0,12	0,53
2	0,80	0,68	0,92	0,71	0,60	0,82
3	1,08	0,97	1,19	0,97	0,87	0,11
4	1,36	1,25	1,48	1,23	1,13	1,34
5	1,67	1,54	1,81	1,52	1,40	1,64
6	2,03	1,88	2,19	1,85	1,70	1,99
7	2,46	2,27	2,66	2,24	2,07	2,43
8	3,03	2,78	3,32	2,78	2,53	3,04
9	3,93	3,49	4,43	3,60	3,19	4,06
10	6,38	4,80	8,71	5,85	4,39	8,00
Full sample	2,32	1,98	2,73	2,11	1,80	2,49

Table 9 reports summary statistics of idiosyncratic volatility during various decades that are included in the sample period. This includes the 90th percentile referred to as P90 and the 10th percentile referred to as P10, the mean is reported. The idiosyncratic volatility(IVOLE/IVOLEFM) is computed by calculating the standard deviation of residual estimates from regressing daily individual stock returns onto contemporaneous daily CRSP equal weighted market returns and three lags thereof(T-1 to T-3). In every decade the 90th percentile, 10th percentile and mean are computed. Each stock has traded at least 15 days in any particular month to qualify to be represented in the data. Stocks that end their delisting due to performance related issues have their final daily stock return replaced by -0,3. Results are reported as a percentage.

Estimating idiosyncratic volatility based on equal weighted CRSP index return has little impact on the numbers in the distribution. The entire distribution becomes marginally narrower in case of using the equal weighted return, both when looking at the distribution of individual deciles and in general. The estimates have a marginally lower magnitude of idiosyncratic volatility.

The methodology for computing the IVOL2/IVOL2FM variable is very similar to the other IVOL2FM/IVOLFM variables. Instead of the value-weighted CRSP market index returns, the value-weighted SP500 index returns are used. For further explanation refer to the methodology section.

Decade	IVOL2FM	IVOL2FM P10	IVOL2FM P90
1970-79	2,00	1,79	2,31
1980-89	2,13	1,87	2,50
1990-99	2,82	2,33	3,39
2000-09	2,06	1,76	2,42
2010-19	1,33	1,16	1,56

IVOL2 Decile	IVOL2	IVOL2 P10	IVOL2 P90	IVOL2FM	IVOL2FM P10	IVOL2FM P90
1	0,40	0,15	0,61	0,34	0,12	0,54
2	0,80	0,67	0,92	0,71	0,59	0,82
3	1,09	0,97	1,20	0,97	0,87	1,07
4	1,37	1,25	1,49	1,23	1,13	1,34
5	1,68	1,55	1,82	1,52	1,40	1,64
6	2,05	1,89	2,21	1,85	1,71	2,00
7	2,49	2,30	2,70	2,25	2,07	2,44
8	3,07	2,81	3,36	2,78	2,54	3,05
9	3,96	3,53	4,53	3,60	3,20	4,13
10	6,43	4,80	8,66	5,86	4,36	7,90
Full sample	2,33	1,99	2,75	2,11	1,80	2,49

Table 10 reports summary statistics of idiosyncratic volatility during various decades that are included in the sample period. This includes the 90th percentile referred to as P90 and the 10th percentile referred to as P10, the mean is reported. Secondly idiosyncratic volatility (IVOL2/IVOL2FM) is broken up into deciles. The idiosyncratic volatility is computed by calculating the standard deviation of residual estimates from regressing daily individual stock returns onto contemporaneous daily SP500 value weighted market returns and three lags thereof(T-1 to T-3). In every decade the 90th percentile, 10th percentile and mean are computed. Each stock has traded at least 15 days in any particular month to qualify to be represented in the data. Stocks that end their delisting due to performance related issues have their final daily stock return replaced by -0,3. Results are reported as a percentage.

When estimating IVOL based on the SP500 index return this leads to near identical estimates. Similar to equal-weighted estimates, the distribution appears marginally narrower when looking at individual deciles and in general.

Due to the limited insight that generating the alternate IVOL measures provides, the further analysis is only conducted using the idiosyncratic volatility accounting for non-synchronous trading, and the idiosyncratic volatility accounting for non-synchronous trading and Fama-French factors. Both computed using the CRSP value-weighted index. The differences are not nearly substantial enough to assume significant impact on the analysis.

Time robustness checks and annual variables

First it is examined how the variables are related to the idiosyncratic return volatility. For each fiscal year the stocks are sorted in decile portfolios based on IVOL. The results are reported below.

IVOLFM Decile	IVOLFM	Accruals	Capital exp. ratio	Capital exp. ratio(assets)	R&D exp. ratio	R&D exp. ratio(assets)	Sales growth	Cash and equivalents
1	1,16	3,02	4,50	5,81	3,41	1,756	1,17	39,10
2	3,3	2,44	5,03	4,98	3,08	1,98	1,15	34,57
3	4,45	2,38	5,56	5,66	3,46	2,63	1,15	31,57
4	5,49	15,62	6,04	6,44	4,30	3,43	1,17	23,10
5	6,60	2,01	6,28	7,15	6,07	4,33	1,21	21,46
6	7,89	1,91	6,25	8,39	8,36	5,80	1,85	14,12
7	9,46	1,53	5,93	8,28	12,05	8,37	1,73	11,47
8	11,45	2,05	5,74	9,40	15,44	12,15	1,39	8,15
9	14,42	1,59	5,74	11,37	16,31	14,84	1,44	6,47
10	21,47	1,59	5,49	14,19	17,07	20,99	1,51	3,42
Full sample	7,625	3,93	5,79	8,21	9,87	8,27	1,33	17,27126

Table 11 reports the average sales growth (SG), analyst forecasts of long-term earnings growth (LTG), capital expenditure intensity (CAPEX), R&D intensity (R&D), Cash and equivalents (CH), accruals (ACC) of deciles formed on idiosyncratic volatility. The idiosyncratic volatility is computed by calculating the standard deviation of residual estimates from regressing daily individual stock returns onto contemporaneous daily CRSP value weighted market returns and three lags thereof(T-1 to T-3). The idiosyncratic volatility is compounded to annual timeframe.

The accruals appear negatively related to idiosyncratic volatility. Capital expenditure seems to have a U-shaped relationship to IVOL when based on market capitalization but positively related when based on total assets. Research and development expenditure appears to be positively related to idiosyncratic volatility, and even more so when based on total assets. Sales growth appears positively related to idiosyncratic volatility. Cash and cash equivalents appear negatively related to IVOL.

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5
IVOLFM	-13,04***	-10,31***	-5,760***	-9,319***	-6,441***
	(-93,60)	(-81,73)	(-17,45)	(-63,43)	(-39,47)
SMB		0,0344***	0,0154***	0,0597***	0,0322***
		(43,32)	(10,69)	(82,01)	(37,59)
HML		-0,0450***	-0,0466***	-0,0410***	-0,0463***
		(-94,35)	(-48,30)	(-85,15)	(-86,43)
Mom		-0,0463***	-0,590***	-0,0430***	-0,0498***
		(-129,7)	(-80,05)	(-121,2)	(-120.7)
Accruals			-2,00e-05		
			(-1,267)		
Cash equivalents				0,000133***	
				(5,056)	
Capital expenditures					-5,995***
					(-36,20)
Constant	-1,817***	-1,727***	-2,568***	-1,705***	-1,834***
	(-121,5)	(-130,2)	(-76,41)	(-95,71)	(-89,09)
Observations	136.600	136.600	37.875	98.124	103.980

	Model 6	Model 7	Model 8	Model 9
IVOLFM	-8,006***	-13,43***	-12,31***	-6,426***
	(-35,56)	(-70,14)	(-84,12)	(-38,46)
SMB	0,0368***	0,0360***	0,0330***	0,0327***
	(32,89)	(46,82)	(56,11)	(37,73)
HML	-0,0470***	-0,0285***	-0,0265***	-0,0478***
	(-68,38)	(-47,42)	(-55,74)	(-87,53)
Mom	-0,0490***	-0,0351***	-0,0356***	-0,0516***
	(-91,07)	(-82,29)	(-106,8)	(-121,3)
R&D expenditure	1,433***			
	(16,52)			
Capital expenditures(assets)		0.476***		
		(10,60)		
R&D expenditure(assets)			-0,164***	
			(-5,081)	
Sales growth				0,000552***
				(12,00)
Constant	-2,044***	-0,549***	-0,648***	-2,211***
	(-77,57)	(-26,65)	(-40,89)	(-110,2)
Observations	61.596	40.507	65.815	102.344

	Model 10	Model 11	Model 12
IVOLFM	-7,984***	-13,33***	-9,422***
	(-36,23)	(-69,38)	(-32,10)
SMB	0,0359***	0,0361***	0,0595***
	(32,51)	(46,75)	(46,90)
HML	-0,0439***	-0,0283***	-0,0374***
	(-64,54)	(-46,86)	(-43,52)
Mom	-0,0462***	-0,0347***	-0,0452***
	(-89,26)	(-81,69)	(-74,47)
Capital expenditures	-9,612***		
	(-34,52)		
R&D expenditure	1,191***		
	(15,71)		
Capital expenditures		-0,537***	
		(-7,475)	
Capital expenditures(assets)		0,510***	
		(11,37)	
Accruals			0,00429***
			(7,532)
Cash equivalents			0,000135***
			(3,788)
Constant	-1,568***	-0,530***	-1,938***
	(-57,60)	(-25,55)	(-63,90)
Observations	61.086	40.270	33.419

t-statistics in parentheses
*** p<0,01, ** p<0,05, * p<0,1

Table 12 reports the capital expenditure intensity (CAPEX), R&D intensity (R&D), Cash and equivalents(CH), accruals (ACC), average sales growth (SG) and of decile portfolios formed on idiosyncratic volatility. Numbers in parentheses below the variable names are the t-statistics for each variable. The dependent variable is the excess stock return. The Newey-West T-statistics are computed with an annum lag. The idiosyncratic volatility is computed by calculating the standard deviation of residual estimates from regressing daily individual stock returns onto contemporaneous daily CRSP value weighted market returns and three lags thereof(T-1 to T-3). The idiosyncratic volatility is compounded to annual timeframe.

The various anomalies appear to absorb some the idiosyncratic volatility effect. The accruals effect would only be significant for 20% significance so this effect appears unclear and small, however when controlling for cash and cash equivalents the anomaly becomes highly significant and positive. Cash and cash equivalents on its own is of significant positive impact on return.

The capital expenditures appear to have a negative impact on stock return, however the sign flips when the capital expenditures is taken as a ratio of total assets. R&D expenditure gives a positive impact on stock return but when taken as a ratio of total assets the sign of the relationship flips. When regressed together, capital expenditure becomes of stronger negative influence on stock return. R&D expenditure influence becomes positive. Sales growth appears to have a mildly positive effect on stock returns.

The idiosyncratic volatility anomaly is robust to controlling for the other anomalies. It is therefore not a manifestation of other known market anomalies. Most variables to appear to absorb a part of the coefficient of the idiosyncratic volatility, but this is not enough to come close to touching the significance of the idiosyncratic volatility anomaly.

VARIABLES	(1) Agriculture etc.	(2) Mining	(3) Utilities	(4) Manufacturing	(5) Wholesale trade
IVOL	-2,937***	-1,898***	-1,920***	-1,692***	-1,799***
	(-7,865)	(-10,61)	(-8,17)	(-25,52)	(-6,884)
SMB	0,00742*	0,0115***	0,00321***	0,0111***	0,0107***
	(1,929)	(13,14)	(6,000)	(30,74)	(8,144)
HML	0,000271	0,00249***	0,00241***	-0,00301***	0,000834
	(0,185)	(3,957)	(5,917)	(-11,33)	(0,973)
Mom	-0,00243*	-0,00550***	-0,00153***	-0,00382***	-0,00525***
	(-1,666)	(-8,065)	(-5,375)	(-17,08)	(-4,829)
Constant	0,309***	0,264***	0,216***	0,277***	0,285***
	(7,939)	(16,64)	(19,41)	(50,33)	(16,46)
Observations	359	5.911	3.996	49.799	3.760

	(6) Retail trade	(7) Transportation and warehousing	(8) Information	(9) Finance and insurance	(10) Scientific and technical services
IVOL	-2,010***	-2,376***	-1,359***	-1,100***	-1,841***
	(-8,040)	(-9,637)	(-10,16)	(-10,18)	(-10,95)
SMB	0,00779***	0,00886***	0,0147***	0,00192***	0,0126***
	(6,690)	(10,07)	(15,57)	(6,736)	(10,37)
HML	-0,000296	0,00104	-0,00971***	-0,000131	-0,00550***
	(-0,453)	(1,591)	(-14,55)	(-0,717)	(-5,462)
Mom	-0,00570***	-0,00320***	-0,00322***	-0,00378***	-0,00324***
	(-6,823)	(-4,612)	(-6,447)	(-26,05)	(-4,350)
Constant	0,329***	0,288***	0,266***	0,111***	0,311***
	(18,03)	(17,58)	(23,58)	(34,00)	(20,93)
Observations	4.577	3.615	11.263	28.088	4.578

	(11) Management of enterprises	(12) Administrative and support	(13) Educational services	(14) Health Care and social assistance
IVOL	-1,841***	-2,091***	-2,541***	-1,844***
	(-10,95)	(-10,17)	(-3,992)	(-5,581)
SMB	0,0126***	0,00857***	0,0808**	0,0132***
	(10,37)	(7,420)	(2,112)	(5,476)
HML	-0,00550***	-0,00107	0,000915	-0,000490
	(-5.462)	(-1,193)	(0,427)	(-0,249)
Mom	-0,00324***	-0,00304***	-0,00186	-0,00176*
	(-4,350)	(-4,228)	(-1,503)	(-1,709)
Constant	0,311***	0,294***	0,338***	0,309***
	(20,93)	(16,34)	(6,247)	(10,08)
Observations	4.578	2.431	433	1.897

t-statistics in parentheses
*** p<0,01, ** p<0,05, * p<0,1

*Table 13 reports average idiosyncratic volatility (IVOL), across different sectors indicated by their North American industry classification. The idiosyncratic volatility is computed by calculating the standard deviation of residual estimates from regressing daily individual stock returns onto contemporaneous daily CRSP value weighted market returns and three lags thereof(T-1 to T-3). The idiosyncratic volatility is compounded to annual timeframe. The dependent variable is stock return. The average number of stocks in each portfolio is reported as N. Returns are in percentage points, and t-statistics are reported in parenthesis. All portfolios are required to have at least 20 stocks in any particular month otherwise the returns in that month are excluded from further computation. Numbers in parentheses below the variable names are the sample period for each variable. The Newey-West t-statistics are computed with an annum lag. *Note that less insightful sector analysis is put in the extra analysis section of the appendix*

The agricultural sector stock returns appear to be the most affected by the idiosyncratic volatility. Transportation and warehousing and the retail trade sector are highly affected by the idiosyncratic volatility. Sectors stock returns that are relatively little impacted by stock returns appear to be the finance and insurance and the information sector. What can be a general explanation for these sector fluctuations in idiosyncratic volatility effect is the prevalence of systemic risk throughout different sectors. A sector that has bigger portion of systemic risk is thus less affected by the idiosyncratic stock risk.

Conclusion

Stocks with high idiosyncratic risk tend to have low future returns, as was found first in Ang, Hodrick, Xing and Zhang(2006). This has been named the idiosyncratic volatility anomaly in the financial research world. Multiple studies have since attempted to uncover explanations for the prevalence of the anomaly. The purpose of this study was to evaluate the robustness of the idiosyncratic volatility anomaly, in particular with respect to the January effect and market cycles. This paper looked into the effect of using different systemic risk factors on the idiosyncratic volatility variable as well as the effect of a diversity of other anomalies on the idiosyncratic volatility anomaly. The paper finds that the idiosyncratic risk anomaly is very robust as well as identifying several potential driving forces and constituents.

The systemic risk factor in the form of index market movements chosen is does not appear very important, as long as it represents the economic entity that stocks are analyzed from. The resulting estimates are roughly the same for using the CRSP value or equal weighted index as well as using the S&P500 value weighted index.

The idiosyncratic volatility anomaly that is found is of lesser magnitude when there is accounted for size and book-to-market anomaly effects, but the anomaly is still highly significant afterward. The H-L decile yields a difference in idiosyncratic volatility of 5,52% and difference in returns of -1,5% per month. The alpha tends to be negative around -1,5% per month. This confirms the hypothesis that the anomaly is present within the data.

The IVOL anomaly appears to be present strongly throughout both bull and bear markets, though the return effect is weaker in bull markets. The H-L decile for stock returns in bull markets is 0,11% per month, absolute relative return is non negative here. Autocorrelation appears to be of lesser importance in bear markets. In January the IVOL relationship flips around and the risk return relationship appears as expected in CAPM theory, with a slight deviation in lower deciles. Here the H-L decile idiosyncratic volatility comes forth as 2,22% in January the difference in return is 12,6% in a month. In January the alpha relationship flips and becomes significantly positive. This confirms the hypothesis that the anomaly is not present in the data when only considering January.

Some balance account variables seem in worse shape with higher IVOL stocks. Accruals and cash/equivalents are samples of this. Accruals, capital expenditures and sales growth appear to absorb a good part of the idiosyncratic anomaly. The variation in these variables appears to explain away the idiosyncratic volatility anomaly at least in part, other variables appear to have some impact but to a lesser degree.

Bibliography

Ang, H. X. (2006). The cross-section of volatility and expected returns. *The journal of finance*.

Baker, B. a. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial analysts journal*.

Baker, H. a. (1991). The efficient market inefficiency of capitalization–weighted stock portfolios. *The Journal of portfolio management*.

Brennan. (1971). Capital market equilibrium with divergent borrowing and lending rates. *Journal of financial and quantitative analysis*.

Carhart. (1997). On persistence in mutual fund performance. *The Journal of finance*.

Chen, J. X. (2009). Dissecting the idiosyncratic volatility anomaly. *SSRN*.

Constantinides. (1974). Optimal stock trading with personal taxes: Implications for prices and the abnormal January returns. *Journal of financial economics*.

Haugen, B. a. (2012). Low risk stocks outperform within all observable markets of the world. *SSRN*.

Heins, H. a. (1975). Risk and the rate of return on financial assets: Some old wine in new bottles. *Journal of Financial and Quantitative Analysis*.

Huang. (2009). The cross section of cashflow volatility and expected stock returns. *Journal of Empirical finance*.

Huang, B. a. (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review*.

Jensen, B. a. (1972). The capital asset pricing model: Some empirical tests. *SSRN*.

Johnson, Y. a. (2004). Bond market volatility vs. stock market volatility: the Swiss experience. *Financial markets and portfolio management*.

Lintner. (1965). Security prices, risk, and maximal gains from diversification. *The journal of finance*.

Ma, J. a. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*.

Miller. (1977). Idiosyncratic risk. *Journal of financial economics*.

Mossin. (1966). Equilibrium in a capital asset market. *Journal of the econometric society*.

Paradox. (2020, June 15). *Paradox investing*. Retrieved from Paradox: https://www.paradoxinvesting.com/

Sharpe. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*.

Thorley, D. S. (2006). Minimum-variance portfolios in the US equity market. *The Journal of portfolio management*.

Vliet, B. a. (2007). The volatility effect. *The Journal of portfolio management*.

Appendix

Extra analysis

	(1)	(2)	(3)	(4)
VARIABLES	Monthly 1	Monthly 2	Monthly 3	Annual 1
IVOL	-1.055***	-1.076***	-1.057***	-10.68***
	(-83.49)	(-85.65)	(-83.55)	(-85.41)
HML	-0.00685***	-0.00685***	-0.00685***	0.0343***
	(-113.3)	(-113.3)	(-113.3)	(43.78)
SMB	0.00192***	0.00192***	0.00192***	-0.0465***
	(39.64)	(39.58)	(39.64)	(-98.83)
RETT	0.0303***	0.0319***	0.0296***	
	(21.43)	(22.51)	(20.88)	
RET0	6.30e-06***	6.28e-06***	6.30e-06***	
	(3.098)	(3.120)	(3.098)	
Bearmarket	-0.00937***		-0.00939***	0.649***
	(-19.78)		(-19.81)	(44.82)
January		0.00457***	0.00466***	
		(7.902)	(8.046)	
				(-131.9)
Constant	-0.436***	-0.439***	-0.437***	-1.860***
	(-1,140)	(-1,178)	(-1,136)	(-136.8)
Observations	3,224,636	3,224,636	3,224,636	136,600

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14 reports the results of the following Fama-MacBeth regressions. The dependent variable is the monthly individual stock returns during the month after the contemporaneous month, the month when idiosyncratic volatility is measured. The explanatory variables include idiosyncratic volatility (IVOL), fama French market size factor (SMB), fama French book-to-market factor (HML), stock returns during the 11 months prior to the contemporaneous month. RETT is the stock return during month t. Bear market is the bear market dummy, equals 1 if there is a bear market. January is the January dummy which equals 1 if the month is a January. The cross-sectional regressions are performed in each month. The time-series averages of regression coefficients are reported, their corresponding t-statistics, as well as the adjusted Rsquares. RET0 is included as an explanatory variable which is the stock return in year up until month t. The Newey-west t-statistics, reported in parenthesis. The sample uses data from 1970 up to 2020. Model 4 is an annual data robustness check.

The current bear market definition appears to give higher future returns when the investment horizon is a year, instead of a month. Other than that this check does not add much.

VARIABLES	(1) Monthly 1	(2) Monthly 2	(3) Monthly 3
IVOL	-1.064***	-1.088***	-1.066***
	(-83.37)	(-85.69)	(-83.44)
HML	-0.00685***	-0.00685***	-0.00685***
	(-113.3)	(-113.2)	(-113.3)
SMB	0.00192***	0.00192***	0.00192***
	(39.63)	(39.57)	(39.63)
January		0.00578***	0.00476***
		(10.02)	(7.121)
Bearmarket	-0.0101***		-0.0104***
	(-21.28)		(-21.10)
January*Bearmarket			0.00380***
			(2.747)
Constant	-0.436***	-0.439***	-0.436***
	(-1,135)	(-1,174)	(-1,128)
Observations	3,224,636	3,224,636	3,224,636

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 15 reports the results of the following Fama-MacBeth regressions. The dependent variable is the monthly individual stock returns during the month after the contemporaneous month, the month when idiosyncratic volatility is measured. The explanatory variables include idiosyncratic volatility (IVOL), fama French market size factor (SMB), fama French book-to-market factor (HML), stock returns during the 11 months prior to the contemporaneous month. January is a dummy for the month January, Bearmarket is a dummy for the bear market as defined in the methodology. The interaction between the January and Bearmarket dummy is included in the third model. The Newey-west t-statistics, reported in parenthesis. The sample uses data from 1970 up to 2020.

The main insight derived here is that the January effect appears to be robust during bear markets.

Stata code

```
//Start working with CRSP dataset

import delimited "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\CRSP stock data\Full CSRP stock data\CSRP stock data part 1.CSV"


//Removing double observations

duplicates report permno date

duplicates list permno date

duplicates tag permno date, gen(isdup)
```

```
drop if isdup

drop isdup


//formatting different dates

format %tdmon_YYYY,_MMDD date

ssc install numdate

numdate daily daydate = date, pattern(YMD)

extrdate month month = daydate

extrdate year fyear = daydate

gen monthdate = ym(fyear, month)

format monthdate %tm

drop month


//Initial data cleaning

bysort monthdate permno: gen month_obs = _N

drop if month_obs < 15

drop if prc <= 0

drop if prc == .

drop month_obs

destring ret, replace force

drop if ret == .

winsor2 vwretd , suffix(W) cuts(1 99) by(daydate)

winsor2 ewretd, suffix(W) cuts(1 99) by(daydate)

winsor2 ret, suffix(W) cuts(1 99) by(daydate)

winsor2 sprtrn, suffix(W) cuts(1 99) by(daydate)
```

```
drop ret vwretd ewretd sprtrn


//Setting time series and handling delisting
tsset permno daydate
gen daterun = .
gen permnorun = .
by permno: replace daterun = cond(L.daterun == ., 1, L.daterun + 1)
bysort permno: gen permno_l = permno[_n-1]
gen lastobs = cond(missing(permno_l), 1, 0)
gen lastdate = lastobs*daydate
replace lastobs = 0 if lastdate == 3654
count if lastobs>0
gen delisting = lastobs[_n-1]
gen srn = 0
replace srn = 1 if retW < 0
gen delistingp = srn*delisting
replace retW = 0 if delisting
replace retW = -0.3 if delistingp
drop delisting delistingp srn lastobs lastdate daterun permnorun permo_l


//Time leads and lags for regressions
//SP500 benchmark value weighted
by permno: gen sprtrn_LG1 = sprtrnW[_n-1]
by permno: gen sprtrn_LG2 = sprtrnW[_n-2]
by permno: gen sprtrn_LG3 = sprtrnW[_n-3]
```

```
by permno: gen sprtrn_L1 = sprtrnW[_n+1]

by permno: gen sprtrn_L2 = sprtrnW[_n+2]

by permno: gen sprtrn_L3 = sprtrnW[_n+3]

//CSRP benchmark equal weighted

by permno: gen ewretd_LG1 = ewretdW[_n-1]

by permno: gen ewretd_LG2 = ewretdW[_n-2]

by permno: gen ewretd_LG3 = ewretdW[_n-3]

by permno: gen ewretd_L1 = ewretdW[_n+1]

by permno: gen ewretd_L2 = ewretdW[_n+2]

by permno: gen ewretd_L3 = ewretdW[_n+3]

//CSRP benchmark value weighted

by permno: gen vwretd_LG1 = vwretdW[_n-1]

by permno: gen vwretd_LG2 = vwretdW[_n-2]

by permno: gen vwretd_LG3 = vwretdW[_n-3]

by permno: gen vwretd_L1 = vwretdW[_n+1]

by permno: gen vwretd_L2 = vwretdW[_n+2]

by permno: gen vwretd_L3 = vwretdW[_n+3]


//This analysis is ran four times for every quartile of PERMNO data until collapse

//Regression(1) computation of the idiosyncratic volatility and checks

ssc install asreg

egen monthreg = group(permno monthdate)

egen yearreg = group(permno fyear)

bys monthreg: asreg retW vwretdW, fit

bysort monthdate permno: egen ivol_NL = sd(_residuals)
```

drop _Nobs _R2 _adjR2 _b_vwretdW _b_cons _fitted _residuals

bys monthreg: asreg retW vwretdW vwretd_LG1 vwretd_LG2 vwretd_LG3, fit

bysort monthdate permno: egen ivol = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_vwretdW _b_vwretd_LG1 _b_vwretd_LG2 _b_vwretd_LG3 _b_cons _fitted _residuals

bys yearreg: asreg retW vwretdW vwretd_LG1 vwretd_LG2 vwretd_LG3, fit

bys fyear permno: egen ivoly = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_vwretdW _b_vwretd_LG1 _b_vwretd_LG2 _b_vwretd_LG3 _b_cons _fitted _residuals

bys monthreg: asreg retW vwretdW vwretd_LG1 vwretd_LG2 vwretd_LG3 vwretd_L1 vwretd_L2 vwretd_L3, fit

bysort monthdate permno: egen ivol_L = sd(_residuals)

//Regression(1) for SP500 index return

drop _Nobs _R2 _adjR2 _b_vwretdW _b_vwretd_LG1 _b_vwretd_LG2 _b_vwretd_LG3 _b_vwretd_L1 _b_vwretd_L2 _b_vwretd_L3 _b_cons _fitted _residuals _fitted _residuals

bys monthreg: asreg retW sprtrnW, fit

bysort monthdate permno: egen ivol2_NL = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_sprtrnW _b_cons _fitted _residuals

bys monthreg: asreg retW sprtrnW sprtrn_LG1 sprtrn_LG2 sprtrn_LG3, fit

bysort monthdate permno: egen ivol2 = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_sprtrnW _b_sprtrn_LG1 _b_sprtrn_LG2 _b_sprtrn_LG3 _b_cons _fitted _residuals

bys yearreg: asreg retW sprtrnW sprtrn_LG1 sprtrn_LG2 sprtrn_LG3, fit

bysort fyear permno: egen ivol2y = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_sprtrnW _b_sprtrn_LG1 _b_sprtrn_LG2 _b_sprtrn_LG3 _b_cons _fitted _residuals

bys monthreg: asreg retW sprtrnW sprtrn_LG1 sprtrn_LG2 sprtrn_LG3 sprtrn_L1 sprtrn_L2 sprtrn_L3, fit

bysort monthdate permno: egen ivol2_L = sd(_residuals) drop _Nobs _R2 _adjR2 _b_sprtrnW _b_sprtrn_LG1 _b_sprtrn_LG2 _b_sprtrn_LG3 _b_sprtrn_L1 _b_sprtrn_L2 _b_sprtrn_L3 _b_cons _fitted _residuals

//Run regression (1) for equal weighted returns

bys monthreg: asreg retW ewretdW, fit

bysort monthdate permno: egen ivole_NL = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_ewretdW _b_cons _fitted _residuals

bys monthreg: asreg retW ewretdW ewretd_LG1 ewretd_LG2 ewretd_LG3, fit

bysort monthdate permno: egen ivole = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_ewretdW _b_ewretd_LG1 _b_ewretd_LG2 _b_ewretd_LG3 _b_cons _fitted _residuals

bys monthreg: asreg retW ewretdW ewretd_LG1 ewretd_LG2 ewretd_LG3 ewretd_L1 ewretd_L2 ewretd_L3, fit

bysort monthdate permno: egen ivole_L = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_ewretdW _b_ewretd_LG1 _b_ewretd_LG2 _b_ewretd_LG3 _b_ewretd_L1 _b_ewretd_L2 _b_ewretd_L3 _b_cons _fitted _residuals _fitted _residuals

//Merge fama-french factors

import delimited "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Thesis combined\Fama french factors combined model.csv", delimiter(";")

save "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Fama French\fama french combined model.dta"

use "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Thesis second run analysis\Thesis combined data and analysis.dta", clear

merge m:m date using "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Fama French\fama french combined model.dta"

drop if _merge==1

drop if _merge==2

drop _merge

save "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Thesis combined\Stock data and fama french factors.dta"

//Cleaning of merged factors

winsor2 mktrf, suffix(W) cuts(1 99) by(daydate)

winsor2 smb, suffix(W) cuts(1 99) by(daydate)

winsor2 hml , suffix(W) cuts(1 99) by(daydate)

winsor2 rf, suffix(W) cuts(1 99) by(daydate)

winsor2 mom, suffix(W) cuts(1 99) by(daydate)

winsor2 rmw, suffix(W) cuts(1 99) by(daydate)

winsor2 cma, suffix(W) cuts(1 99) by(daydate)

drop mktrf smb hml rf mom rmw cma

//Regression(2) extra estimation of IVOL with Fama-French factors and checks

bys monthreg: asreg retW smbW hmlW vwretdW, fit

bysort monthdate permno: egen ivolfm_NL = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_smbW _b_hmlW _b_vwretdW _b_cons _fitted _residuals

bys monthreg: asreg retW smbW hmlW vwretdW vwretd_LG1 vwretd_LG2 vwretd_LG3, fit

bysort monthdate permno: egen ivolfm = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_smbW _b_hmlW _b_vwretdW _b_vwretd_LG1 _b_vwretd_LG2 _b_vwretd_LG3 _b_cons _fitted _residuals

bys yearreg: asreg retW smbW hmlW vwretdW vwretd_LG1 vwretd_LG2 vwretd_LG3, fit

bysort fyear permno: egen ivolfmy = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_smbW _b_hmlW _b_vwretdW _b_vwretd_LG1 _b_vwretd_LG2 _b_vwretd_LG3 _b_cons _fitted _residuals

bys monthreg: asreg retW smbW hmlW vwretdW vwretd_LG1 vwretd_LG2 vwretd_LG3 vwretd_L1 vwretd_L2 vwretd_L3,fit bysort monthdate permno: egen ivolfm_L = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_smbW _b_hmlW _b_vwretdW _b_vwretd_LG1 _b_vwretd_LG2 _b_vwretd_LG3 _b_vwretd_L1 _b_vwretd_L2 _b_vwretd_L3 _b_cons _fitted _residuals

//Run regression(2) for SP index return

bys monthreg: asreg retW smbW hmlW sprtrnW, fit

bysort monthdate permno: egen ivol2fm_NL = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_smbW _b_hmlW _b_sprtrnW _b_cons _fitted _residuals

bys monthreg: asreg retW smbW hmlW sprtrnW sprtrn_LG1 sprtrn_LG2 sprtrn_LG3, fit

bysort monthdate permno: egen ivol2fm = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_smbW _b_hmlW _b_sprtrnW _b_sprtrn_LG1 _b_sprtrn_LG2 _b_sprtrn_LG3 _b_cons _fitted _residuals

bys yearreg: asreg retW smbW hmlW sprtrnW sprtrn_LG1 sprtrn_LG2 sprtrn_LG3, fit

bysort fyear permno: egen ivol2fmy = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_smbW _b_hmlW _b_sprtrnW _b_sprtrn_LG1 _b_sprtrn_LG2 _b_sprtrn_LG3 _b_cons _fitted _residuals

bys monthreg: asreg retW smbW hmlW sprtrnW sprtrn_LG1 sprtrn_LG2 sprtrn_LG3 sprtrn_L1 sprtrn_L2 sprtrn_L3, fit

bysort monthdate permno: egen ivol2fm_L = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_smbW _b_hmlW _b_sprtrnW _b_sprtrn_LG1 _b_sprtrn_LG2 _b_sprtrn_LG3 _b_sprtrn_L1 _b_sprtrn_L2 _b_sprtrn_L3 _b_cons _fitted _residuals

//Run regression(2) for equal weigthed returns

bys monthreg: asreg retW smbW hmlW ewretdW, fit

bysort monthdate permno: egen ivolefm_NL = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_smbW _b_hmlW _b_ewretdW _b_cons _fitted _residuals

bys monthreg: asreg retW smbW hmlW ewretdW ewretd_LG1 ewretd_LG2 ewretd_LG3, fit

```
bysort monthdate permno: egen ivolefm = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_smbW _b_hmlW _b_ewretdW _b_ewretd_LG1 _b_ewretd_LG2
_b_ewretd_LG3 _b_cons _fitted _residuals

bys monthreg: asreg retW smbW hmlW ewretdW ewretd_LG1 ewretd_LG2 ewretd_LG3 ewretd_L1
ewretd_L2 ewretd_L3,fit

bysort monthdate permno: egen ivolefm_L = sd(_residuals)

drop _Nobs _R2 _adjR2 _b_smbW _b_hmlW _b_ewretdW _b_ewretd_LG1 _b_ewretd_LG2
_b_ewretd_LG3 _b_ewretd_L1 _b_ewretd_L2 _b_ewretd_L3 _b_cons _fitted _residuals


//Variable monthly conversion

ascol vwretdW, tom returns(simple) keep(all)

drop month_id

ascol sprtrnW, tom returns(simple) keep(all)

drop month_id

ascol ewretdW, tom returns(simple) keep(all)

drop month_id

ascol momW, tom returns(simple) keep(all)

drop month_id

ascol retW, tom returns(simple) keep(all)

drop month_id


//Variable annual conversion

ascol vwretdW, toy returns(simple) keep(all)

drop year_id

ascol retW, toy returns(simple) keep(all)

drop year_id

ascol sprtrnW, toy returns(simple) keep(all)
```

```
drop year_id
gen year_ivol = ivol*4
gen year_ivolfm = ivolfm*4
gen year_ivol2 = ivol2*4
gen year_ivol2fm = ivol2fm*4

//Bear market, penny stock and january dummy
gen bearmarket = 0
replace bearmarket = 1 if year_vwretdW < 0
gen monthdate_txt = string(monthdate, "%tm")
gen january = 0
gen monthdate_sub = substr(monthdate_txt, 5,6)
replace january = 1 if monthdate_sub == "m1"
drop monthdate_txt
gen penny = 0
replace penny = 1 if prc <= 5

//volumechanges
bysort permno: gen volchanges = vol/vol[_n-1]
ascol volchanges, tom returns(simple) keep(all)
drop month_id

//Preparing merger after collapse
drop date
gen date = monthdate
```

//Labeled all variables for easy recognition(no code provided)

//Save before collapsing data and collapsing for monthly and annual analysis

save "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Thesis second run analysis\Thesis combined data and analysis ran.dta"

collapse month_vwretdW month_sprtrnW month_retW month_ewretdW ivol ivol2 ivolfm ivol2fm ivole ivolefm bearmarket january penny year_vwretdW year_retW month_volchanges, by(permno monthdate date)

save "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Thesis second run analysis\Thesis combined data and analysis ran(monthly collapse).dta"

use "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Thesis second run analysis\Thesis combined data and analysis ran.dta"

collapse year_vwretdW year_retW year_sprtrnW year_ivol year_ivolfm year_ivol2 year_ivol2fm ivoly ivolfmy ivol2y ivol2fmy, by(permno ticker fyear)

save "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Thesis second run analysis\Thesis combined data and analysis ran(annual collapse).dta"

//Start working with COMPUSTAT dataset

//Import compustat data

import delimited "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Compustat variables data\Compustat variables data.csv", delimiter(comma)

//Removing double observations

duplicates report gvkey datadate

duplicates list gvkey datadate

duplicates tag gvkey datadate, gen(isdup)

drop if isdup

```
drop isdup


//Generate capital expenditures as a ratio
bysort gvkey: gen capxratio = capx/at
bysort gvkey: gen capxratiom = capx/mkvalt
bysort gvkey: gen rdexpratio = xrd/at
bysort gvkey: gen rdexpratiom = xrd/mkvalt


//Generate accruals components and salesgrowth(CHECH another measure for cash changes)
bysort gvkey: gen cashchanges = ch/ch[_n-1]
bysort gvkey: gen cashchangessi = che/che[_n-1]
bysort gvkey: gen clchanges = lct/lct[_n-1]
bysort gvkey: gen cachanges = act/act[_n-1]
bysort gvkey: gen debtincl = dlc/dlc[_n-1]
bysort gvkey: gen salesgrowth = revt/revt[_n-1]
bysort gvkey: gen tpchanges = txp/txp[_n-1]
bysort gvkey: gen salesturnovergrowth = sale/sale[_n-1]
bysort gvkey: gen averageassets = at/at[_n-1]
bysort gvkey: gen dpaexpense = dp/dp[_n-1]


//Data cleaning
drop if capx < 0
drop if revt < 0
drop if at < 0
drop if lse < 0
```

```
bysort gvkey : drop if _N < 4


//Winsorizing usable variables

winsor2 capxratio, suffix(W) cuts(1 99) by(datadate)

winsor2 capxratiom, suffix(W) cuts(1 99) by(datadate)

winsor2 rdexpratio, suffix(W) cuts(1 99) by(datadate)

winsor2 rdexpratiom, suffix(W) cuts(1 99) by(datadate)

winsor2 chech, suffix(W) cuts(1 99) by(datadate)

winsor2 dpaexpense, suffix(W) cuts(1 99) by(datadate)

winsor2 debtincl, suffix(W) cuts(1 99) by(datadate)

winsor2 cashchanges, suffix(W) cuts(1 99) by(datadate)

winsor2 cashchangessi, suffix(W) cuts(1 99) by(datadate)

winsor2 clchanges, suffix(W) cuts(1 99) by(datadate)

winsor2 cachanges, suffix(W) cuts(1 99) by(datadate)

winsor2 salesgrowth, suffix(W) cuts(1 99) by(datadate)

winsor2 salesturnovergrowth, suffix(W) cuts(1 99) by(datadate)

winsor2 tpchanges, suffix(W) cuts(1 99) by(datadate)

winsor2 averageassets, suffix(W) cuts(1 99) by(datadate)


//Generate accruals variable

bysort gvkey: gen accrualscom = (cachangesW - cashchangesW) - (clchangesW - debtinclW -
tpchangesW) - dpaexpenseW

bysort gvkey: gen accruals = accrualscom/averageassets

drop accrualscom


//Labeled all variables for easy recognition(no code provided)
```

```
save "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Compustat variables data\Compustat variables processing.dta"


//Merge monthly fama french factors

use "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Thesis combined\Stock data and fama french factors(monthly collapse).dta"

format date %tmCYN ym(2019, 5)

merge m:m date using "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Fama French\fama french momentum monthly"

drop if _merge==1

drop if _merge==2

drop _merge

merge m:m date using "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Fama French\fama french monthly model"

drop if _merge==1

drop if _merge==2

drop _merge


//Winsorizing new merged variables and new time series definition

winsor2 MktRF, suffix(W) cuts(1 99) by(Date)

winsor2 SMB, suffix(W) cuts(1 99) by(Date)

winsor2 HML, suffix(W) cuts(1 99) by(Date)

winsor2 RF, suffix(W) cuts(1 99) by(Date)

winsor2 Mom, suffix(W) cuts(1 99) by(Date)

tsset permno monthdate
```

```
//Decade dummies
extrdate year fyear = monthdate
gen decade = 0
replace decade = 5 if fyear < 2021
replace decade = 4 if fyear < 2011
replace decade = 3 if fyear < 2001
replace decade = 2 if fyear < 1991
replace decade = 1 if fyear < 1981

//Create IVOL deciles for analysis
xtile ivol_10m = ivol, nq(10)
xtile ivolfm_10 = ivolfm, nq(10)
bysort ivol_10: egen ivol_p10 = pctile(ivol), p(10)
bysort ivolfm_10: egen ivolfm_p10 = pctile(ivolfm), p(10)
bysort ivol_10: egen ivol_p90 = pctile(ivol), p(90)
bysort ivolfm_10: egen ivolfm_p90 = pctile(ivolfm), p(90)
xtile ivol2_10 = ivol2, nq(10)
xtile ivol2fm_10 = ivol2fm, nq(10)
bysort ivol2_10: egen ivol2_p10 = pctile(ivol2), p(10)
bysort ivol2fm_10: egen ivol2fm_p10 = pctile(ivol2fm), p(10)
bysort ivol2_10: egen ivol2_p90 = pctile(ivol2), p(90)
bysort ivol2fm_10: egen ivol2fm_p90 = pctile(ivol2fm), p(90)
xtile ivole_10 = ivole, nq(10)
xtile ivolefm_10 = ivolefm, nq(10)
bysort ivole_10: egen ivole_p10 = pctile(ivole), p(10)
```

```
bysort ivolefm_10: egen ivolefm_p10 = pctile(ivolefm), p(10)

bysort ivole_10: egen ivole_p90 = pctile(ivole), p(90)

bysort ivolefm_10: egen ivolefm_p90 = pctile(ivolefm), p(90)


//Number of stocks per ivol decile

sort permno monthdate

gen permno_l = permno[_n-1]

gen nstock = 0

replace nstock = 1 if permno_l != permno

bysort ivol_10: count if nstock == 1

bysort ivolfm_10: count if nstock == 1

bysort decade: count if nstock == 1

count if nstock == 1


//Report different IVOL statistics

tabstat ivol ivol_p10 ivol_p90 , by(ivol_10) s(mean)

tabstat ivolfm ivolfm_p10 ivolfm_p90 , by(ivolfm_10) s(mean)

tabstat ivol2 ivol2_p10 ivol2_p90 , by(ivol_10) s(mean)

tabstat ivol2fm ivol2fm_p10 ivol2fm_p90 , by(ivolfm_10) s(mean)

tabstat ivole ivole_p10 ivole_p90 , by(ivole_10) s(mean)

tabstat ivolefm ivolefm_p10 ivolefm_p90 , by(ivolefm_10) s(mean)

tabstat ivol ivol_p10 ivol_p90 , by(decade) s(mean)

tabstat ivolfm ivolfm_p10 ivolfm_p90 , by(decade) s(mean)

tabstat ivol2 ivol2_p10 ivol2_p90 , by(decade) s(mean)

tabstat ivol2fm ivol2fm_p10 ivol2fm_p90 , by(decade) s(mean)
```

```
tabstat ivole ivole_p10 ivole_p90 , by(decade) s(mean)

tabstat ivolefm ivolefm_p10 ivolefm_p90 , by(decade) s(mean)


//Calculate returns minus t bill yield

gen month_vwretrf = month_vwretdW - RFW

gen month_retrf = month_retW - RFW

gen month_sprtrnrf = month_sprtrnW - RFW


//Regression(3) run monthly Carhart four factor model generating alphas and checks

bys monthdate ivol_10: asreg month_retrf month_vwretrf HMLW SMBW MomW, fit

bysort monthdate ivol_10: gen CarhartA =(_b_cons)

drop _Nobs _R2 _adjR2 _b_SMBW _b_HMLW _b_MomW _b_month_vwretrf _b_cons _fitted _residuals

bys monthdate ivolfm_10: asreg month_retrf month_vwretrf HMLW SMBW MomW, fit

bysort monthdate ivolfm_10: gen CarhartfmA =(_b_cons)

drop _Nobs _R2 _adjR2 _b_SMBW _b_HMLW _b_MomW _b_month_vwretrf _b_cons _fitted _residuals

bys ivol_10: reg month_retrf month_vwretrf HMLW SMBW MomW

bys ivolfm_10: reg month_retrf month_vwretrf HMLW SMBW MomW


//generate further statistics(high minus low quintile etc.)

//Returns

bys ivol_10: egen portfolioST = mean(month_retW)

bys ivolfm_10: egen portfoliofmST = mean(month_retW)

bys monthdate: egen STq10 = mean(cond(ivol_10 == 10, month_retW, .) )

bys monthdate: egen STq1 = mean(cond(ivol_10 == 1, month_retW, .) )

bys monthdate: egen STfmq10 = mean(cond(ivolfm_10 == 10, month_retW, .) )
```

```
bys monthdate: egen STfmq1 = mean(cond(ivolfm_10 == 1, month_retW, .) )

gen STqdiff = STq10 - STq1

gen STfmqdiff = STfmq10 - STfmq1

//IVOL

bys ivol_10: egen portfolioivol = mean(ivol)

bys ivolfm_10: egen portfolioivolfm = mean(ivolfm)

bys monthdate: egen ivolq10 = mean(cond(ivol_10 == 10, ivol, .) )

bys monthdate: egen ivolq1 = mean(cond(ivol_10 == 1, ivol, .) )

bys monthdate: egen ivolfmq10 = mean(cond(ivolfm_10 == 10, ivolfm, .) )

bys monthdate: egen ivolfmq1 = mean(cond(ivolfm_10 == 1, ivolfm, .) )

gen ivolqdiff = ivolq10 - ivolq1

gen ivolfmqdiff = ivolfmq10 - ivolfmq1

//Alphas

bys ivol_10: egen portfolioA = mean(CarhartA)

bys ivolfm_10: egen portfoliofmA = mean(CarhartfmA)

bys monthdate: egen alphaq10 = mean(cond(ivol_10 == 10, CarhartA, .) )

bys monthdate: egen alphaq1 = mean(cond(ivol_10 == 1, CarhartA, .) )

bys monthdate: egen alphafmq10 = mean(cond(ivolfm_10 == 10, CarhartfmA, .) )

bys monthdate: egen alphafmq1 = mean(cond(ivolfm_10 == 1, CarhartfmA, .) )

gen alphaqdiff = alphaq10 - alphaq1

gen alphafmqdiff = alphafmq10 - alphafmq1


//Generate sharpe ratio for every IVOL decile

bysort permno: egen volt = sd(month_vwretdW)

gen vol = volt + ivol
```

```
gen volfm = volt + ivolfm

gen sharpivol = month_retW/ivol

gen sharpivolfm = month_retW/ivolfm

gen sharpvol = month_retW/vol

gen sharpvolfm = month_retW/volfm

//high minus low quintile

bys ivol_10: egen sharpportfolio = mean(sharpvol)

bys ivolfm_10: egen sharpportfoliofm = mean(sharpvolfm)

bys monthdate: egen SSTq10 = mean(cond(ivol_10 == 10, sharpvol, .) )

bys monthdate: egen SSTq1 = mean(cond(ivol_10 == 1, sharpvol, .) )

bys monthdate: egen SSTfmq10 = mean(cond(ivolfm_10 == 10, sharpvolfm, .) )

bys monthdate: egen SSTfmq1 = mean(cond(ivolfm_10 == 1, sharpvolfm, .) )

gen SSTqdiff = SSTq10 - SSTq1

gen SSTfmqdiff = SSTfmq10 - SSTfmq1

bys ivol_10: egen sharpiportfolio = mean(sharpivol)

bys ivolfm_10: egen sharpiportfoliofm = mean(sharpivolfm)

bys monthdate: egen SISTq10 = mean(cond(ivol_10 == 10, sharpivol, .) )

bys monthdate: egen SISTq1 = mean(cond(ivol_10 == 1, sharpivol, .) )

bys monthdate: egen SISTfmq10 = mean(cond(ivolfm_10 == 10, sharpivolfm, .) )

bys monthdate: egen SISTfmq1 = mean(cond(ivolfm_10 == 1, sharpivolfm, .) )

gen SISTqdiff = SISTq10 - SISTq1

gen SISTfmqdiff = SISTfmq10 - SISTfmq1


//Ivol deciles high low Newey regression test

sort permno monthdate
```

```
newey STqdiff, lag(12) force

newey STfmqdiff, lag(12) force

newey ivolqdiff, lag(12) force

newey ivolfmqdiff, lag(12) force

newey alphaqdiff, lag(12) force

newey alphafmqdiff, lag(12) force

newey SSTqdiff, lag(12) force

newey SSTfmqdiff, lag(12) force

newey SISTqdiff, lag(12) force

newey SISTfmqdiff, lag(12) force


//tabstat statistics

tabstat ivol portfolioST month_retW CarhartA sharpivol sharpvol, by(ivol_10) s(mean)

tabstat ivolfm portfoliofmST month_retW CarhartfmA sharpivolfm sharpvolfm, by(ivolfm_10) s(mean)


//January and market cycle effect statistics

bysort january: tabstat ivolfm month_retW CarhartfmA sharpivol sharpvol, by(ivolfm_10) s(mean)

bysort bearmarket: tabstat ivolfm month_retW CarhartfmA sharpivol sharpvol, by(ivolfm_10) s(mean)


//Bearmarket and januarys plot

bysort fyear: egen avgyear_vwretdW = mean(year_vwretdW)

scatter avgyear_vwretdW fyear

bysort fyear: egen avgyear_retW = mean(year_retW)

bysort ivolfm_10: egen iavgyear_retW = mean(year_retW)

scatter avgyear_retW fyear
```

```
scatter iavgyear_retW fyear


//generate further statistics(high minus low quintile etc.)
//Returns
bys january: egen jportfolioret = mean(month_retW)
bys bearmarket: egen bportfolioret = mean(month_retW)
bys january: egen jretq10 = mean(cond(ivolfm_10 == 10, month_retW, .) )
bys january: egen jretq1 = mean(cond(ivolfm_10 == 1, month_retW, .) )
bys bearmarket: egen bretq10 = mean(cond(ivolfm_10 == 10, month_retW, .) )
bys bearmarket: egen bretq1 = mean(cond(ivolfm_10 == 1, month_retW, .) )
gen Jdiff = jretq10 - jretq1
gen Bdiff = bretq10 - bretq1
//IVOL
bys january: egen jportfolioivol = mean(ivol)
bys bearmarket: egen bportfolioivolfm = mean(ivolfm)
bys january: egen jivolfmq10 = mean(cond(ivolfm_10 == 10, ivolfm, .) )
bys january: egen jivolfmq1 = mean(cond(ivolfm_10 == 1, ivolfm, .) )
bys bearmarket: egen bivolfmq10 = mean(cond(ivolfm_10 == 10, ivolfm, .) )
bys bearmarket: egen bivolfmq1 = mean(cond(ivolfm_10 == 1, ivolfm, .) )
gen Jivolqdiff = jivolfmq10 - jivolfmq1
gen Bivolfmqdiff = bivolfmq10 - bivolfmq1
//Alphas
bys january: egen jportfolioA = mean(CarhartfmA)
bys bearmarket: egen bportfoliofmA = mean(CarhartfmA)
bys january: egen jalphaq10 = mean(cond(ivolfm_10 == 10, CarhartfmA, .) )
```

```
bys january: egen jalphaq1 = mean(cond(ivolfm_10 == 1, CarhartfmA, .) )

bys bearmarket: egen balphafmq10 = mean(cond(ivolfm_10 == 10, CarhartfmA, .) )

bys bearmarket: egen balphafmq1 = mean(cond(ivolfm_10 == 1, CarhartfmA, .) )

gen Jalphaqdiff = jalphaq10 - jalphaq1

gen Balphafmqdiff = balphafmq10 - balphafmq1


//Ivol deciles high low Newey regression test

sort permno monthdate

newey Jdiff, lag(12) force

newey Bdiff, lag(12) force

newey Jivolqdiff, lag(12) force

newey Bivolfmqdiff, lag(12) force

newey Jalphaqdiff, lag(12) force

newey Balphafmqdiff, lag(12) force

//Fama-macbeth regressions

bysort permno: gen month_retWL = month_retW[_n+1]

bysort permno: gen HMLWL = HMLW[_n+1]

bysort permno: gen SMBWL = SMBW[_n+1]

asreg month_retrfl ivol HMLWL SMBWL

newey month_retrfl ivol HMLWL SMBWL month_retW year_retW, lag(12) force

bysort bearmarket: newey month_retrfl ivol HMLWL SMBWL month_retW year_retW, lag(12) force

newey month_retrfl ivol HMLWL SMBWL, lag(12) force

drop _Nobs _R2 _adjR2 _b_ivol _b_ivol _b_HMLWL _b_SMBWL _b_cons

newey month_retrfl ivol HMLWL SMBWL january, lag(12) force

bysort bearmarket: newey month_retrfl ivol HMLWL SMBWL, lag(12) force
```

```
//Extra regressions

newey month_retrfl ivol HMLWL SMBWL month_retW year_retW bearmarket, lag(12) force

newey month_retrfl ivol HMLWL SMBWL month_retW year_retW january, lag(12) force

newey month_retrfl ivol HMLWL SMBWL month_retW year_retW january bearmarket, lag(12) force

newey month_retrfl ivol HMLWL SMBWL bearmarket, lag(12) force

newey month_retrfl ivol HMLWL SMBWL january, lag(12) force

newey month_retrfl ivol HMLWL SMBWL january bearmarket, lag(12) force//Merge annual fama fench factors

//constant test for january

save "C:\Users\miche\Documents\Documents - Copy\Studie Fe\Thesis low volatility anomaly\Thesis data\Thesis second run analysis\Thesis combined data and analysis ran(monthly collapse).dta", replace

drop if january = 0

asreg month_retrfl ivol HMLWL SMBWL

gen constantj = _b_cons

drop _Nobs _R2 _adjR2 _b_ivol _b_ivol _b_HMLWL _b_SMBWL _b_cons

use save "C:\Users\miche\Documents\Documents - Copy\Studie Fe\Thesis low volatility anomaly\Thesis data\Thesis second run analysis\Thesis combined data and analysis ran(monthly collapse).dta", clear

gen constantj = -0.43443564 if january != 0

newey month_retrfl ivol HMLWL SMBWL constantj, lag(12) force

use "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Thesis combined\Stock data and fama french factors(annual collapse).dta"

gen date = fyear

format date %ty

rename ticker tic

merge m:m date using "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Fama French\fama french momentum annually"
```

```
drop if _merge==1

drop if _merge==2

drop _merge

merge m:m date using "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Fama French\fama french annual model"

drop if _merge==1

drop if _merge==2

drop _merge


//Preparing merger of COMPUSTAT data

duplicates report tic fyear

duplicates list tic fyear

duplicates tag tic fyear, gen(isdup)

drop if isdup

drop isdup

sort tic fyear


//Start working with COMPUSTAT dataset


//Import compustat data

import delimited "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Compustat variables data\Compustat variables data.csv", delimiter(comma)


//Removing double observations

duplicates report gvkey datadate

duplicates list gvkey datadate
```

```
duplicates tag gvkey datadate, gen(isdup)
drop if isdup
drop isdup

//Generate capital expenditures as a ratio
gen capxratio = capx/at
gen capxratiom = capx/mkvalt
gen rdexpratio = xrd/at
gen rdexpratiom = xrd/mkvalt

//Generate accruals components and salesgrowth(CHECH another measure for cash changes)
gen cashchanges = ch/ch[_n-1]
gen cashchangessi = che/che[_n-1]
gen clchanges = lct/lct[_n-1]
gen cachanges = act/act[_n-1]
gen debtincl = dlc/dlc[_n-1]
gen salesgrowth = revt/revt[_n-1]
gen tpchanges = txp/txp[_n-1]
gen salesturnovergrowth = sale/sale[_n-1]
gen averageassets = at/at[_n-1]
gen dpaexpense = dp/dp[_n-1]

//Data cleaning
drop if capx < 0
drop if revt < 0
```

```
drop if at < 0

drop if lse < 0

bysort gvkey : drop if _N < 4


//Winsorizing usable variables

winsor2 capxratio, suffix(W) cuts(1 99) by(datadate)

winsor2 capxratiom, suffix(W) cuts(1 99) by(datadate)

winsor2 rdexpratio, suffix(W) cuts(1 99) by(datadate)

winsor2 rdexpratiom, suffix(W) cuts(1 99) by(datadate)

winsor2 chech, suffix(W) cuts(1 99) by(datadate)

winsor2 dpaexpense, suffix(W) cuts(1 99) by(datadate)

winsor2 debtincl, suffix(W) cuts(1 99) by(datadate)

winsor2 cashchanges, suffix(W) cuts(1 99) by(datadate)

winsor2 cashchangessi, suffix(W) cuts(1 99) by(datadate)

winsor2 clchanges, suffix(W) cuts(1 99) by(datadate)

winsor2 cachanges, suffix(W) cuts(1 99) by(datadate)

winsor2 salesgrowth, suffix(W) cuts(1 99) by(datadate)

winsor2 salesturnovergrowth, suffix(W) cuts(1 99) by(datadate)

winsor2 tpchanges, suffix(W) cuts(1 99) by(datadate)

winsor2 averageassets, suffix(W) cuts(1 99) by(datadate)


//Generate accruals variable

gen accrualscom = (cachangesW - cashchangesW) - (clchangesW - debtinclW - tpchangesW) -
dpaexpenseW

gen accruals = accrualscom/averageassets

drop accrualscom
```

```
//Labeled all variables for easy recognition(no code provided)

save "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Compustat variables data\Compustat variables processing.dta"

//Merging annual data
duplicates report tic fyear
duplicates list tic fyear
duplicates tag tic fyear, gen(isdup)
drop if isdup
drop isdup
sort tic fyear
merge m:m tic fyear using "C:\Users\miche\Downloads\Thesis low volatility anomaly\Thesis data\Thesis combined\Stock data and fama french factors(annual collapse)"

//Getting rid of duplicates of gvkey and fyear and introduce different time series
duplicates report gvkey fyear
duplicates list gvkey fyear
duplicates tag gvkey fyear, gen(isdup)
drop if isdup
drop isdup
tsset gvkey fyear

//Winsorizing new merged variables
winsor2 mktrf, suffix(W) cuts(1 99) by(fyear)
```

```
winsor2 smb, suffix(W) cuts(1 99) by(fyear)

winsor2 hml, suffix(W) cuts(1 99) by(fyear)

winsor2 rf, suffix(W) cuts(1 99) by(fyear)

winsor2 mom, suffix(W) cuts(1 99) by(fyear)

drop mktrf smb hml rf mom


//Create IVOL deciles for analysis

xtile year_ivol_10 = year_ivol, nq(10)

xtile year_ivolfm_10 = year_ivolfm, nq(10)

bysort year_ivol_10: egen year_ivol_p10 = pctile(year_ivol), p(10)

bysort year_ivolfm_10: egen year_ivolfm_p10 = pctile(year_ivolfm), p(10)

bysort year_ivol_10: egen year_ivol_p90 = pctile(year_ivol), p(90)

bysort year_ivolfm_10: egen year_ivolfm_p90 = pctile(year_ivolfm), p(90)

xtile year_ivol2_10 = year_ivol2, nq(10)

xtile year_ivol2fm_10 = year_ivol2fm, nq(10)

bysort year_ivol2_10: egen year_ivol2_p10 = pctile(year_ivol2), p(10)

bysort year_ivol2fm_10: egen year_ivol2fmy_p10 = pctile(year_ivol2fm), p(10)

bysort year_ivol2_10: egen year_ivol2_p90 = pctile(year_ivol2), p(90)

bysort year_ivol2fm_10: egen year_ivol2fm_p90 = pctile(year_ivol2fm), p(90)


//Calculate returns minus t bill yield

gen year_vwretrf = year_vwretdW - rfW

gen year_retrf = year_retW - rfW

gen year_sprtrnrf = year_sprtrnW - rfW
```

```
//Report IVOL statistics and explanations by decile

tabstat year_ivolfm year_ivolfm_p10 year_ivolfm_p90 , by(year_ivolfm_10) s(mean)

tabstat year_ivolfm accruals capxratioW capxratiomW rdexpratioW rdexpratiomW salesgrowthW chechW, by(year_ivolfm_10) s(mean)


//Generate general fama-macbeth regressions for idiosyncratic volatility explanations(Newey-west procedure)

sort gvkey fyear

newey year_retrf year_ivolfm, lag(1) force

newey year_retrf year_ivolfm smbW hmlW momW, lag(1) force

newey year_retrf year_ivolfm smbW hmlW momW  accruals, lag(1) force

newey year_retrf year_ivolfm smbW hmlW momW  chechW, lag(1) force

newey year_retrf year_ivolfm smbW hmlW momW capxratioW, lag(1) force

newey year_retrf year_ivolfm smbW hmlW momW capxratiomW, lag(1) force

newey year_retrf year_ivolfm smbW hmlW momW rdexpratioW, lag(1) force

newey year_retrf year_ivolfm smbW hmlW momW rdexpratiomW, lag(1) force

newey year_retrf year_ivolfm smbW hmlW momW salesgrowthW, lag(1) force

//Regress Capex, R&D and accruals, chech simultaneosly

newey year_retrf year_ivolfm smbW hmlW momW capxratioW rdexpratioW, lag(1) force

newey year_retrf year_ivolfm smbW hmlW momW capxratiomW rdexpratiomW, lag(1) force

newey year_retrf year_ivolfm smbW hmlW momW  accruals chechW, lag(1) force


//Sector analysis

gen industry = 0

replace industry = 1 if substr(string(naics), 1, 2) == "11"

replace industry = 2 if substr(string(naics), 1, 2) == "21"
```

```
replace industry = 3 if substr(string(naics), 1, 2) == "22"

replace industry = 4 if substr(string(naics), 1, 2) == "31"

replace industry = 4 if substr(string(naics), 1, 2) == "32"

replace industry = 4 if substr(string(naics), 1, 2) == "33"

replace industry = 5 if substr(string(naics), 1, 2) == "42"

replace industry = 6 if substr(string(naics), 1, 2) == "44"

replace industry = 6 if substr(string(naics), 1, 2) == "45"

replace industry = 7 if substr(string(naics), 1, 2) == "48"

replace industry = 7 if substr(string(naics), 1, 2) == "49"

replace industry = 8 if substr(string(naics), 1, 2) == "51"

replace industry = 9 if substr(string(naics), 1, 2) == "52"

replace industry = 10 if substr(string(naics), 1, 2) == "54"

replace industry = 11 if substr(string(naics), 1, 2) == "55"

replace industry = 12 if substr(string(naics), 1, 2) == "56"

replace industry = 13 if substr(string(naics), 1, 2) == "61"

replace industry = 14 if substr(string(naics), 1, 2) == "62"


bysort industry fyear: newey year_retW year_ivol smbW hmlW momW, lag(1) force


//outreg2 using sumcor.doc, replace ctitle(Model 1) tstat label

//outreg2 using sumcor.doc, append ctitle(Model 1) tstat label


//estpost tabstat ivol ivol_p10 ivol_p90 , by(decade) s(mean)

//esttab . using tabstatdecade.rtf, main(mean) label replace
```