ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Economics and Business Economics

Date final version: 25/10/2020

Skewness preference in the Dutch stock market

Student
Simone Bradley
488607

Supervisor
Barahona R.
Second assessor
Zhu, H.

Abstract

This paper evaluates the hypothesis that stocks with high idiosyncratic skewness have lower expected returns. The data set analysed consists of companies trading on the Amsterdam stock exchange from March 1983 up until April 2020. Panel regressions are used to model investor perceptions of expected idiosyncratic skewness. Besides lagged idiosyncratic skewness momentum and firm specific characteristics are also incorporated in the model. When sorting stocks into three portfolios each month according to their expected idiosyncratic skewness, the portfolio with the highest expected idiosyncratic skewness has an average monthly return 0.82% lower than the portfolio with the lowest skewness, at a 5% significance level. Adjusting for risk, the CAPM alpha is also lower for portfolios with high expected idiosyncratic skewness. In addition, this paper evaluates portfolio returns over time and does not find significant evidence linking the magnitude of these differences to certain time periods.

Table of contents

1. Introduction and problem formulation

It may be appealing to invest a stock with a potential for very high returns, even if the chance is very low, rather than a stock with a larger chance for small returns. However, the common preference for these lottery-like stocks comes at a cost. Several studies have found that idiosyncratic skewness (IS) is negatively associated with stock returns (Ayadi, Cao, Lazrak, & Wang, 2019; Barberis & Huang, 2008; Boyer, Mitton, Vorkink, 2009, Brunnermeier, Gollier, & Parker 2007; Mitton & Vorkink, 2007). In addition, this expensive taste in stocks has shown to limit investor diversification (Mitton & Vorkink, 2007) and decrease the efficiency of markets (Blau & Whitby, 2018). IS refers to the degree of asymmetrical probabilities surrounding the mean in a firm specific return distribution. A positively (or right) skewed return distribution contains frequent small losses and a few extreme gains. Negatively (or left) skewed distributions have frequent small gains and a few extreme losses.

The objective of this paper is to further research these findings, to increase the understanding of investor preferences. The majority of the referenced studies have been done using data from the American stock market. This paper will contribute to this by examining the Dutch stock market, using data from the AEX for a time period from March 1983 to April 2020. In addition to the differing market and extended time frame, this paper will contribute by evaluating the magnitude of the possible lower stock returns for positively skewed stocks over time. An evaluation will be made to examine if stocks with high IS are overpriced in certain time periods. The central aim of this paper is to evaluate what characteristics are important predictors for IS in the Dutch stock market, and then examine if and to what extent high idiosyncratic skewness leads to lower average stock returns. These returns for different levels of skewness will also be evaluated over time. The following research question is formulated:

'To what extent is the expected upside potential (positive skewness) of a stocks return distribution overpriced in the Dutch stock market and how does this vary over time?'

The rest of this paper is organized as follows. First an overview of theories covering the efficient pricing of stocks, the cumulative prospect theory and skewness preference is presented. This is followed by a review of existing literature and its findings to substantiate the empirical investigation of this paper. The data set will be summarized and motivated. After this the hypotheses to support the main research question is formulated and explained and followed by the methodology of this paper. Empirical findings are shown, and this paper will ends with a summary, conclusion and discussion.

2. Theoretical framework

In this section a foundation for this paper will be made by providing an overview of existing theory regarding stock returns and skewness preference.

The efficient market hypothesis (EMH) assumes that all publicly available information relevant to the value of an asset is directly reflected in the price. At any point in time a stock should equal its present discounted value of all future earnings and dividends. This is the cornerstone of neoclassical finance and directly implies that "beating the market" should not be possible as prices only react to new information. For the EMH to hold, investor rationality is required such that any inefficient pricing is corrected with arbitrage. The well-known CAPM model builds on the EMH assumption and finds systematic market risk to be relevant in explaining security returns. However, empirical tests continuously fail in confirming this risk-return relationship and anomalies in stock pricing persist.

Economists and psychologists have gained insight in risk attitude using extensive experimental evidence. They find, when making decisions under risk, people move away from the classic expected utility theory. The most prominent non expected utility model is the cumulative prospect theory (CPT). The CPT is a modified version of Kahneman and Tversky's (1979) original prospect theory, and used observations to develop a descriptive model for decisions made under risk. They find that people examine possible outcomes relative to a reference point with losses described as a subtraction from the reference point and gains an addition. Another well-known finding is that people are loss averse, leading to differing risk attitudes around gains and losses. Losses have a bigger impact then gains, generally a loss is neutralized with approximately twice the gain. When weighting probabilities for possible outcomes in unknown situations, with the weights in line with the expected probability of the outcomes, people aren't likely to assign these weights in a well-constructed objective manner. Instead research shows that people overweight the tails of the distribution to which it applies to. The probability of unlikely events is over estimated whereas common events is underestimated. Overweighting unlikely events ensures that lottery and insurance companies stay profitable. People that buy lottery tickets show a willingness to pay more than the expected value, for a small chance of winning. When evaluating risks, if people over weight the small chance of a negative event happening they pay too much for insurance. The probability weighting part of the CPT is the most relevant for this paper as it can model the common preference for a positively skewed or lottery like wealth distribution.

The skewness in a distribution measures the asymmetrical probabilities surrounding the mean. A positively (or right) skewed return distribution contains frequent small losses and few extreme gains. This can be compared to the distributions of lottery tickets, as most people end up with a small loss and sometimes one (or a few) people experience an extreme gain. Negatively (or left) skewed distributions have frequent small gains and few extreme losses. When translated to the stock market, overweighting extreme outcomes, could result in the common preference for positively skewed stocks. Securities with upside potential could therefore be overpriced resulting in negative abnormal returns on average.

3. Literature review

Previous research and literature covering the skewness preference of investors provides an important starting point and substantiation for the empirical investigation in this paper. Therefore this section reviews the most relevant findings of foregoing literature. Numerous studies that have researched possible asset pricing implications of idiosyncratic skewness. They are reviewed and discussed below.

Individual investors exhibit speculative behaviour on the stock market. Theoretical models attempt to examine the impact of these types of investors. Mitton and Vorkink (2007) develop a model of heterogeneous preference for skewness, assuming some investors have a preference for positive skewness while others optimize the Sharpe ratio of their portfolios. What Mitton and Vorkink (2007) describe as ‘lotto investors’ are investors that choose to hold undiversified portfolios enabling them to have a higher exposure to positive skewness. They document that investors with these undiversified portfolios have returns that are significantly more positively skewed than those of diversified portfolios. They show that mean-variance efficiency is, in this case, sacrificed for a higher exposure to skewness. They find that IS has an impact on prices in equilibrium. This relationship between undiversified portfolio’s and skewness is not a coincidence as undiversified investors choose stocks with a higher IS. This results in negative alphas relative to the market portfolio for stocks with high IS. Barberis and Huang (2008) use the cumulative prospect theory of Tversky and Kahneman

(1992), where they particularly focus on the probability weighting component, to create a model for investors with preference for positively skewed assets. The reason they use the cumulative prospect theory is because of its strong ability to capture risks attitudes in experimental settings. Their main goal is to evaluate if such a model can help improve understanding of investor behaviour present in financial markets. In line with Mitton and Vorkink (2007) they show that these preferences lead to the overpricing of positively skewed securities, which therefore generate a negative excess return on average.

As the forecasting of skewness is based on predicting small probability events it is a challenging task. Harvey and Siddique (1999) find that lagged IS is a weak predictor of current IS. Thus, a more appropriate way to predict skewness must be found. Chen, Hong and Stein (2001) introduce an approach that use frim characteristics and idiosyncratic volatility (IV) and find that these are also predictors of skewness.

The most the most relevant to this paper is the analysis of Boyer et al. (2009). They test the above theories by using a cross-sectional model that estimates expected IS. In their model they incorporate firm characteristics like size, industry, turn over and momentum in addition to lagged skewness and IV. They find that IS and returns are negatively correlated. They find that the Fama and French (1993) alpha of low-expected skewness quintiles exceeds that of the high expected quintile with 1% per month.

A more recent paper, Ayadi, et al. (2019) is also able to corroborate that expected IS is a significantly and constantly priced component of stocks and find a negative relationship between skewness and expected returns. Furthermore, their results suggest that stocks decline in price in the month following extreme positive returns, leading to negative returns. Blau et al. (2018) examine the association between skewness and the efficiency of stock prices. They reveal that positively skewed stocks are more prone to inefficiency, because lottery-like preferences lead to overvaluation. They find a positive relationship between price delay and skewness. This price delay measures the delay of market wide information reflected in stock prices. Boyer et al. (2009) additionally find that low average returns of stocks with high IV (Guo et al., 2006) can possibly be (partly) explained by the positive relation they find between IV and IS. The reverse could also be true leaving the reason for the low average return to be possibly caused by one, both or a combination of overlapping explanatory power that they could attribute to either.

4. Data

The purpose of this section is to provide a clear description of the data and underpin the choice of the stock market researched in this paper. Prior research focusses mainly on companies listed on the American stock market. This paper complements this by using companies currently traded on the Amsterdam stock market, the AEX. Using the AEX gives a new perspective as it is a market where skewness preference of stocks has not been extensively researched. In addition prior researchers focussing on the U.S. stock market, results in lots of empirical studies being performed on the same data sets. This can cause a bias commonly referred to as ‘data snooping’. Although this will never be fully eliminated, examining different markets and therefore different data can be a valuable addition to existing research.

To enable effective application of the methodology the data is filtered so it can comply too the following restrictions.

i. Daily returns of the companies are available from March 1983 through to April 2020.
ii. All companies are still listed on the AEX in April 2020.
iii. The monthly turnover, market capitalization and the standard industrial classification (SIC) code are available for the companies from March 1988 through April 2020.

Turnover gives a representation of how easily a stock can be traded i.e. the liquidity of the stock. Market capitalization refers to the market value of a company’s stock times the number of stocks outstanding and allows investors to evaluate the relative size of a company. The SIC code is a system that classifies stocks into industries by using a four digit code. The methodology of this paper involves panel regressions, thus the main reason for implementing these restrictions is to create a strongly balanced panel data set. This leads to a data set consisting of 23 companies categorized into seven different industries from April 1983 through April 2020. An overview of the companies, their average market capitalization over the specified time period and their industry is represented in table 1. In addition to the data mentioned above, daily AEX returns are used as a proxy for the market return and the Dutch three month, end of period, interbank rate is used as a proxy for the risk free rate. All data is obtained from Bloomberg.

Table 1
Overview of companies and their characteristics present in the data set

	Company	Industry	Average Market Value (in million €)
1	Royal Dutch Shell	Manufacturing	81202.842
2	Philips Elten. Koninklijke	Manufacturing	20909.939
3	Koninklijke Ahold Delhaize	Retail Trade	11299.358
4	Akzo Nobel	Manufacturing	9716.507
5	Heineken Holding	Manufacturing	7706.238
6	Wolters Kluwer	Services	5756.834
7	Aegon	Finance, Insurance and Real Estate	14529.486
8	Hal Trust	Finance, Insurance and Real Estate	3738.740
9	Boskalis Westminster	Construction	1591.659
10	Corbion	Manufacturing	1297.426
11	Hunter Douglas	Manufacturing	1194.372
12	SBM Offshore	Mining	1589.733
13	THK Group	Manufacturing	548.916
14	BAM Group Kon.	Construction	581.647
15	Amsterdam Commodities	Wholesale Trade	152.944
16	Bever Holding	Finance, Insurance and Real Estate	38.987
17	DGB Group	Manufacturing	53.809
18	Kendrion	Manufacturing	158.546
19	Nedap	Manufacturing	132.471
20	Novisource	Service	51.294
21	Porcelyne fles	Manufacturing	6.215
22	Stern Group	Retail Trade	65.258
23	Wereldhave	Finance, Insurance and Real Estate	1097.282

The table gives an overview of companies in the data set that ranges from March 1983 through April 2020. The first column lists the companies, the second column contains the industry of each company and the third column contains the average market capitalization of each company over time in millions of euros. The average market value is determined over the time period of March 1988 through April 2020. The stocks are traded on the AEX.

5. Hypotheses and Methodology

This section begins with explaining and motivating the three hypotheses that are used to support the main research objective. This is followed by an extensive outline of this papers methodology that shows which methods are used to test the hypotheses. Each part includes an explanation as to why certain steps are taken the way they are and thereby shows the careful thought process that has led to this methodology.

The methodology of this paper is summarized as follows. First historical measures of IV and IS are defined. Then panel regressions are performed to test the significance of the correlations of variables to IS. By incorporating the significant variables, a model of expected skewness is constructed. Monthly portfolios with stocks sorted on their skewness are then formed and evaluated whether the highest and lowest skewed portfolio have different returns.

Finally, the same method is applied to different time periods to examine possible differences over time.

5.1 Hypotheses

The aim of this paper is to evaluate what characteristics are important predictors for IS, and then examine if and to what extent high IS leads to lower average stock returns. These possibly differing returns for different levels of skewness will also be evaluated over time.

Boyer et al. (2009) and Chen et al. (2001) find that during periods of heavy trading volume, stocks are more likely to be negatively skewed. Therefore, it is expected that turnover has a negative correlation with IS. Boyer et al. (2009), Chen et al. (2001) and Viva et al. (2017) also find that lagged momentum is negatively correlated with IS. Thus, the expectation is that low past returns could lead to higher IS. In addition, Boyer et al. (2009) and Viva et al. (2017) find that lagged IV is also positively correlated with skewness. Prior empirical research also finds a strong and negative relation between firm size and IS (Boyer et al., 2009; Chen et al., 2001; Hugues, 2020; Viva et al. 2017). Boyer et al. (2009) also shows that certain industries are more or less prone to skewed returns.

Considering the above research the first hypothesis is stated as:

Lagged idiosyncratic skewness, lagged idiosyncratic volatility, lagged turnover, lagged momentum, lagged size and industry are significant predictors for idiosyncratic skewness.

This hypothesis will be tested by performing a sample wide panel regression with the companies as cross-sectional variable. The IS will be the dependant variable and lagged IS, lagged IS, lagged turnover, lagged momentum and lagged size will be the independent variables. In addition, industry dummies will be added.

To continue, numerous prior research finds that returns for portfolios of stocks consisting of higher expected IS have lower average returns (Barberis & Huang, 2008; Brunnermeier et al., 2007; Mitton & Vorkink, 2007; Boyer et al., 2009; Ayadi et al., 2019). The theory that can explain this is the CPT (Kahneman & Tversky, 1992). Investors overweight the tails of the distribution, leading to a preference for positively skewed stocks. This preference could cause these types of assets to be overpriced and negatively skewed stocks to be undervalued leading to a difference in average return between the two. The second hypothesis tests this and is:

There is a significant positive difference in average returns between portfolios consisting of stocks with the highest expected skewness and the lowest expected skewness.

This hypothesis will be tested by using the significantly correlated variables for IS to construct expected future IS. The stocks will then be sorted into portfolios ranging evenly from low to high expected IS and a paired t-test will be used to evaluate the difference between the average monthly returns of the portfolios. It has not been extensively researched whether these possibly different returns could be more or less pronounced in certain time periods. It could perhaps be expected that lottery type assets do even worse in certain time periods or market conditions, indicating that there are times when it is even more expensive to hold positively skewed stocks. In order to also include a time aspect to the skewness pricing the third and last hypothesis is:

The magnitude of the difference in average return between the portfolios consisting of stocks with the highest expected skewness and the lowest expected skewness is significantly different in certain time periods.

This will be tested by a paired t-test to evaluate the difference between the average monthly returns of the portfolios, but this time it will be done for specified periods of time instead of over the whole sample.

5.2 Historical idiosyncratic volatility and skewness

In this paper an investment horizon is defined as the time period over which investors hope to experience an extreme positive outcome. The choice of investment horizon is made to evaluate a time period that captures the typical investors focus on the long-run upside potential of stocks. Therefore an investment horizon of 60 months is used.

The historical IS and IV of the stocks will be defined as follows. Let T be the investment horizon in months, denote S(t) as the set of trading days from the first day of the month t-T+1 through the end of the month t and let N(t) be the number of trading days in this set. The CAPM model (shaper linter 1973) is used to determine the regression residual $\varepsilon_{i,d}$ for the i^{th} firm on day d.

$$R_{i,d} - R_{f,d} = \alpha_{i,d} + \beta_{i,d}\left(R_{m,d} - R_{f,d}\right) + \varepsilon_{i,d} \tag{1}$$

As a proxy for the risk-free rate, the three month end of period interbank rate of the bank of the Netherlands is used. Daily AEX returns are used as a proxy for the Market return. For each of the companies in the sample, the daily CAPM residuals are obtained. Then using $\varepsilon_{i,d}$ the historical estimates of monthly IV and IS respectively for firm *i* using data for the days in S (t) are defined as:

$$IV_{i,t} = \left(\frac{1}{N(t)} \sum_{d \in S(t)} \varepsilon_{i,d}^2\right)^{1/2} \tag{2}$$

$$IS_{i,t} = \frac{1}{N(t)} \frac{\sum_{d \in S(t)} \varepsilon_{i,d}^3}{IV_{i,t}^3} \tag{3}$$

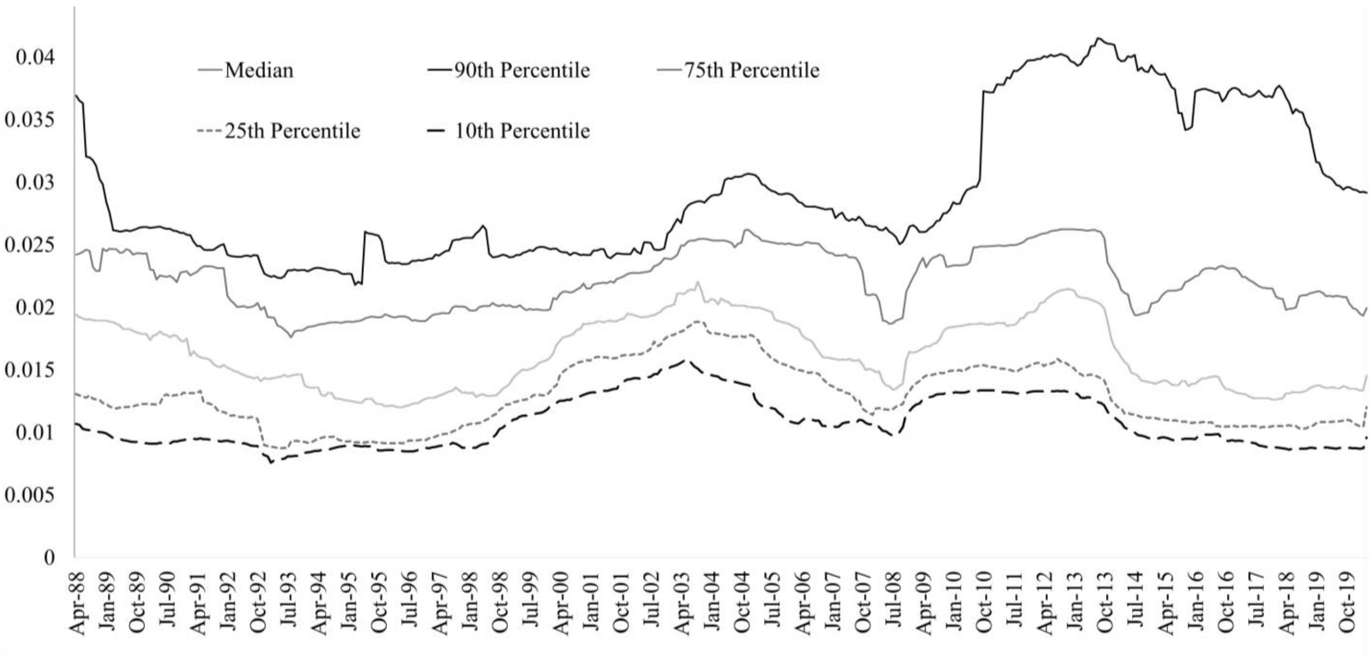

Figure 1
Cross-sectional distribution of firm level volatility
This Figure plots the 10th, 25th, 50th, 75th and 90th percentiles of the monthly cross-sectional distribution of idiosyncratic volatility for 23 companies trading on the Dutch stock market (AEX) from March 1988 through April 2020. When calculating idiosyncratic volatility a time horizon of 60 months is used and estimated as $IV_{i,t} = (\frac{1}{N(t)} \sum_{d \in S(t)} \varepsilon_{i,d}^2)^{1/2}$

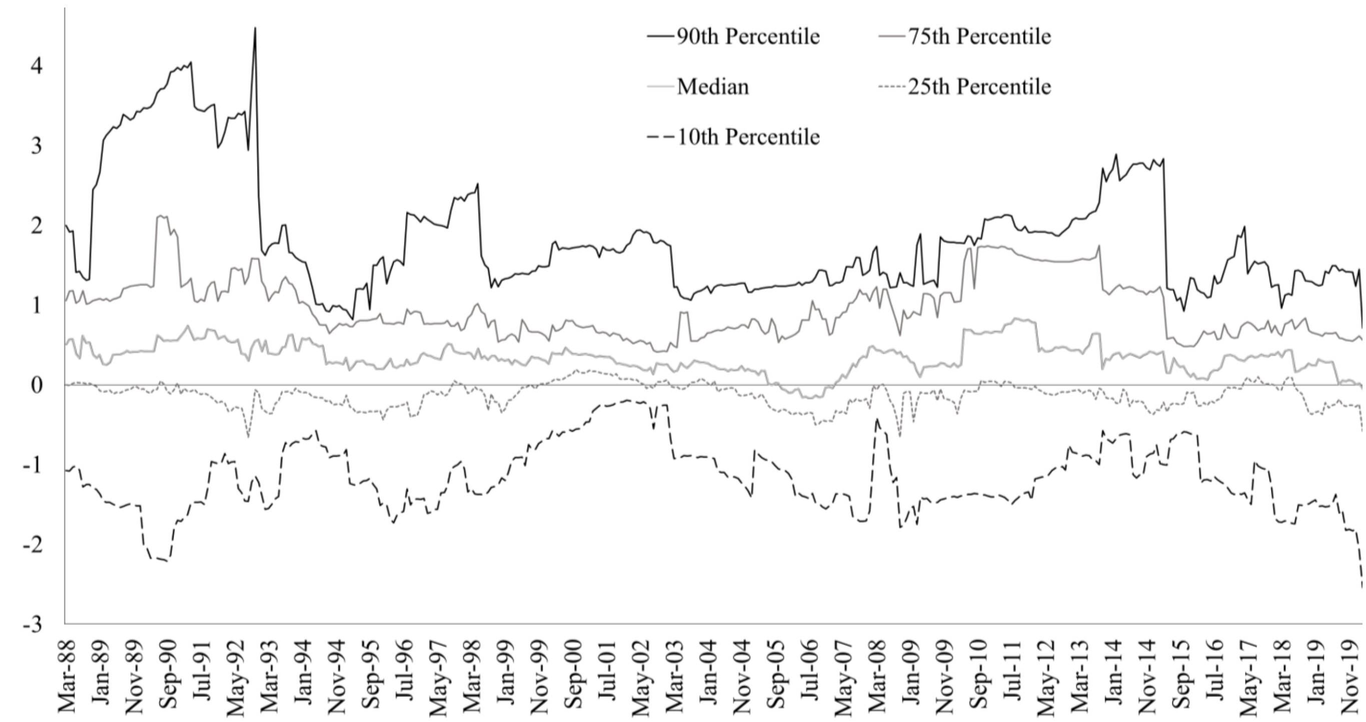

Figure 2
Cross-sectional distribution of firm level skewness
This Figure plots the 10th, 25th, 50th, 75th and 90th percentiles of the monthly cross-sectional distribution of idiosyncratic firm skewness for 23 companies trading on the Dutch stock market (AEX) from March 1988 through April 2020. When calculating idiosyncratic skewness a time horizon of 60 months is used and is estimated as $IS_{i,t} = \frac{1}{N(t)} \frac{\sum_{d \in S(t)} \varepsilon_{i,d}^3}{IV_{i,t}^3}$

To clarify on the time horizon, for the estimation of $IS_{i,t}$ five years of prior data is used. In figure 1 and 2, IV and IS respectively are plotted over time. This is the cross-sectional distribution from April 1988 through to April 2020 using a time horizon of 60 months to calculate IS and IV. The figures show the 10th, 25th, 50th, 75th, and 90th percentile each month. In figure 2 can be seen that the 90th percentile has the greatest variation followed by the 10th percentile.

5.3 *Panel regressions*

A model is developed of expected skewness using a similar method as Boyer et al (2009). The model will incorporate trading volume, past returns and firm characteristics like industry, size and book to market ratios. In this paper sample wide panel regressions are used. This is motivated by Chen, Hong and Stein (2001) and Boyer et al (2009). Boyer et al (2009) use cross-sectional regressions estimated separately each month. They do this in order to allow the relations between skewness and the firm specific variables to vary across time. However, they mention that in unreported results they make estimations based on panel regressions and find similar results, therefore panel regressions are used in this paper.

Chen et al. (2001) also use panel regressions. The first company characteristic that will be touched upon is Momentum. The variable $MOM_{i,t-T}$ is defined as the cumulative return for firm i over the month's t-T-12 to t-T-1. This is based on Chen et al (2001) and Boyer et al. (2009) that find that momentum is negatively correlated with estimated skewness. Next $TURN_{i,t-T}$ is the firm's average daily turnover in the month t-T. This is motivated by Chen et al. (2003) who predict that during periods of increased trading volume negative skewness is most prominent. $SIZE_{i,t-T}$ is included to control for firm size, which is the market capitalisation of the firm each month. Lastly industry dummies are added leaving out manufacturing as the reference category to prevent perfect multicollinearity. A time horizon of T = 60 months is used. The panel regressions are estimated for the months t from March 1988 through April 2015.

$$\begin{aligned} IS_{i,t} = {} & \beta_0 + \beta_1 IS_{i,t-T} + \beta_2 IV_{i,t-T} + \beta_3 MOM_{i,t-T} + \beta_4 SIZE_{i,t-T} \\ & + \beta_5\, TURN_{i,t-T} + \beta_{6,} SER_i + \beta_{7,} MIN_i + \beta_{8,} CON_i \\ & + \beta_9 WHOLE_i + \beta_{10} RETAIL_i + \beta_{11} FIN_i + \varepsilon_{i,t} \end{aligned} \tag{4}$$

In table 2 there is an overview of the panel regression coefficients. First a model with all the coefficients mentioned above is evaluated. Turnover is an insignificant predictor of IS and is thus left out in the next model. Lagged IV also turns out to be an insignificant predictor of IS and is thus also left out in the second model, shown on the right side of the table. The industry dummies vary in significance, presumably due to the different frequencies of the industries prevalent in the dataset and are thus kept in the model. The final model used is:

$$\begin{aligned} IS_{i,t} = {} & \beta_0 + \beta_1 IS_{i,t-T} + \beta_2 MOM_{i,t-T} + \beta_3 SIZE_{i,t-T} + \beta_{4,} SER_i \\ & + \beta_{5,} MIN_i + \beta_{6,} CON_i + \beta_7 WHOLE_i + \beta_8 RETAIL_i + \beta_9 FIN_i + \varepsilon_{i,t} \end{aligned} \tag{5}$$

Table 2
Skewness predictive panel regressions

IS	Coefficient	P-value	Coefficient	P-value
ISi,t-T	-0.1379	0.000	-0.149	0.000
IVi,t-T	-1.464	0.295		
MOMi,t-T	0.757	0.000	0.7483	0.000
SIZEi,t-T	-0.0000124	0.000	-0.0000103	0.000
TURNi,t-T	0.0000158	0.196		
SERi,t-T	1.773	0.000	1.741	0.000
MINi,t-T	-0.4028	0.458	-0.4058	0.541
CONi,t-T	-0.8332	0.037	-0.8397	0.086
WHOLEi,t-T	0.7349	0.176	0.7305	0.271
RETi,t-T	-0.3582	0.370	-0.3362	0.491
FINi,t-T	0.4153	0.172	0.4368	0.239
Constant	0.5540	0.001	0.542	0.005

This table contains the coefficients and their P-values from two panel regressions of the form

$$\mathrm{IS}_{i,t} = \beta_0 + \beta_1 \mathrm{IS}_{i,t-T} + \beta_2 \mathrm{IV}_{i,t-T} + \beta_3 \mathrm{MOM}_{i,t-T} + \beta_4 \mathrm{SIZE}_{i,t-T} + \beta_5\, \mathrm{TURN}_{i,t-T} + \beta_{6,} \mathrm{SER}_i + \beta_{7,} \mathrm{MIN}_i + \beta_{8,} \mathrm{CON}_i + \beta_9 \mathrm{WHOLE}_i + \beta_{10} \mathrm{RETAIL}_i + \beta_{11} \mathrm{FIN}_i + \varepsilon_{i,t}$$

Where $IS_{i,t-T}$ and $IV_{i,t-T}$ are idiosyncratic skewness and volatility respectively and are estimated with a time horizon of 60 months. $MOM_{i,t-T}$ is defined as the cumulative return for firm *i* over the month's *t-T-12* to *t-T-1*. $SIZE_{i,t-T}$ is the market capitalisation of the firm each month. $TURN_{i,t-T}$ is the firm's average daily turnover in the month *t-T*. Where $IS_{i,t-T}$ and $IV_{i,t-T}$ are idiosyncratic skewness and volatility respectively. The coefficients of $\mathrm{IV}_{i,t-T}$ and $\mathrm{TURN}_{i,t-T}$ are not significant in the first panel regression and are therefore left out in the second. A time horizon of T = 60 months is used. The panel regressions are estimated for the months *t* from March 1988 through April 2015 and includes stocks that trade on the AEX.

The panel regression yields estimated historical relationships between IS and past company characteristics. Using these coefficients an estimate of expected skewness over time is obtained for each company. The first month for which an estimate of expected idiosyncratic skewness exists is March 1993. This is because skewness is predicted a horizon of five years in advance and the first $IS_{i,t}$ is available in March 1988 due to its calculation described above. Note that considering equation four that also incorporates lagged idiosyncratic skewness it could be presumed return data five years prior to March 1993 is also needed. An advantage however of using sample wide panel regressions is that coefficients can be used for the whole time period. Expected IS is obtained in the following way.

$$E_t(IS_{i,t+T}) = \beta_{0,} + \beta_1 IS_{i,t} + \beta_2 MOM_{i,t} + \beta_3 SIZE_{i,t} + \beta_4 SER_{i,t} + \beta_5 MIN_{i,t} + \beta_{6,} CON_{i,t} + \beta_7 WHOLE_{i,t} + \beta_8 RETAIL_{i,t} + \beta_9 FIN_{i,t} \qquad (6)$$

5.4 Portfolio formation

To be able to investigate the properties of stocks with high skewness in comparison with low skewness, portfolios are formed. After the calculation of expected skewness each month the companies are sorted into portfolios accordingly. Portfolio one consists of the lowest expected skewness and portfolio five consists of the highest expected skewness. Initially five portfolios are formed each month in the following manner. At each month t, a range of expected skewness is determined by subtracting the lowest value from the highest expected skewness value. This range is then divided into five even groups, and each value is placed into the representative group. This method of portfolio formation differs from previous literature that ensure portfolios have similar amounts of stocks, as stocks are sorted in quintiles (Boyer et al., 2009). However, the skewness-seeking investors that Barberis and Huang (2008), Brunnermeier et al. (2007), and Mitton and Vorkink (2007) model are not so much interested by the portfolio's idiosyncratic skewness. They are more interested in a single stocks high return potential. Therefore, it is decided to construct the portfolios in the above described manner to ensure that exactly those extreme estimations of expected idiosyncratic skewness do not lose their power. When dividing stocks evenly or using quintiles, the most extreme skewness estimations will be softened by less extreme observations. There could be months in which a few stocks differ largely in skewness expectations compared to the other stocks, leading to a few high skewness expectations and lots of lower skewness expectations. Additionally, the finding that investors have the tendency to stay undiversified in order to capture upside potential (Mitton & Vorkink, 2007) it seems forcing less skewed stocks into portfolios that are meant to represent the highest possible expected idiosyncratic skewness each month will distort these preferences and interests. As a natural consequence of this method not all portfolios contain the same amount of stocks, and specifically portfolio four has months in which there are no expected skewness values that fit in the range. Portfolio four therefore has less time period observations in the further methodology of this paper.

Table 3
Overview of portfolios

		Average across portfolio	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4	Portfolio 5
		Idiosyncratic skewness	1.453	0.918074	0.1588	0.1352	0.2878
		Idiosyncratic volatility	0.02463	0.02038	0.01883	0.01847	0.02656
		Expected Idiosyncratic skewness	-0.4674	0.3024	0.7460	1.216	2.3611
		Market Value (in Million €)	19337	7053	3878	1506	2716
		Turnover	2763	1972	1319	542	688
		Momentum (%)	5.438	3.406	15.74	21.94	23.16
	Company	Industry	frequency	frequency	frequency	frequency	frequency
1	Royal Dutch Shell	Manufacturing	193	113	18	0	0
2	Philips Elten. Koninklijke	Manufacturing	10	178	131	5	0
3	Koninklijke Ahold Delhaize	Retail Trade	24	209	87	4	0
4	Akzo Nobel	Manufacturing	6	179	136	3	0
5	Heineken Holding	Manufacturing	5	113	205	1	0
6	Wolters Kluwer	Services	0	1	5	12	306
7	Aegon	Finance, Insurance and Real Estate	1	60	220	43	0
8	Hal Trust	Finance, Insurance and Real Estate	8	11	201	104	0
9	Boskalis Westminster	Construction	277	46	1	0	0
10	Corbion	Manufacturing	6	113	194	11	0
11	Hunter Douglas	Manufacturing	4	150	160	10	0
12	SMB Offshore	Mining	8	87	224	5	0
13	THK Group	Manufacturing	6	106	202	10	0
14	BAM Group Kon.	Construction	303	21	0	0	0
15	Amsterdam Commodities	Wholesale Trade	0	4	156	162	2
16	Bever Holding	Finance, Insurance and Real Estate	48	62	137	68	9
17	DGB Group	Manufacturing	33	119	162	10	0
18	Kendrion	Manufacturing	9	153	162	0	0
19	Nedap	Manufacturing	4	134	185	1	0
20	Novisource	Service	27	21	12	23	241
21	Porcelyne fles	Manufacturing	33	191	96	4	0
22	Stern Group	Retail Trade	69	229	26	0	0
23	Wereldhave	Finance, Insurance and Real Estate	0	6	206	112	0

The table gives an overview portfolio characteristics. In this table companies are sorted into five portfolios according to their expected idiosyncratic skewness. Portfolio five has the highest expected idiosyncratic skewness and portfolio one the lowest. The table lists the companies and their corresponding industries and how often each company is placed in each portfolio. The average idiosyncratic skewness, volatility, expected idiosyncratic skewness, market value, turnover and momentum of each portfolio over a time period of May 1993 through April 2020 is also displayed in the top right of the table. The stocks are traded on the AEX.

In table 3 there is an overview of the portfolios and their features. The top part of the table summarizes the average IS, IV, expected IS, market value, turnover and momentum in each portfolio. The bottom half of the table displays the frequency of each company in each portfolio and the industry. Service seems to be a dominant industry in the portfolio with the highest expected skewness, whereas construction is more often present in the lowest expected skewness portfolio. Due to the relatively small sample size methods of sorting the stocks in three and two portfolios will also be investigated. Average returns and risk adjusted returns will be derived for each of the portfolios. It could be possible that certain portfolios differ in risk and therefore might have different returns thus, to be able to control for these possible differences in risk, for each portfolio the time series average of the CAPM alpha is derived. This makes it possible to investigate the differences in excess returns across the portfolios adjusted for risk.

$$R_{p,t} - R_{f,t} = \alpha_{p,t} + \beta_{p,t}(R_{m,t} - R_{f,t}) + \varepsilon_{p,t} \tag{7}$$

5.5 Time variation

Next an analysis is made around the difference in the average returns over time. It will be evaluated if there are certain time periods that generate more pronounced differences in average returns of the portfolios. This section examines differences in returns over time with stocks sorted into three portfolios. Figure four lays out the visual differences in portfolios over time. The line plotted represents the difference in return between portfolio one and three. Figure three plots the monthly AEX return. The figure indicates a few time periods that seem to have a larger dispersion in portfolio returns. These time periods could be assigned to less financially stable times. The two most prominent being in the period of 1999 through 2003 and the period 2007 through 2010. These could possibly be attributed to the dot.com bubble and the real estate bubble and their respective aftermaths.

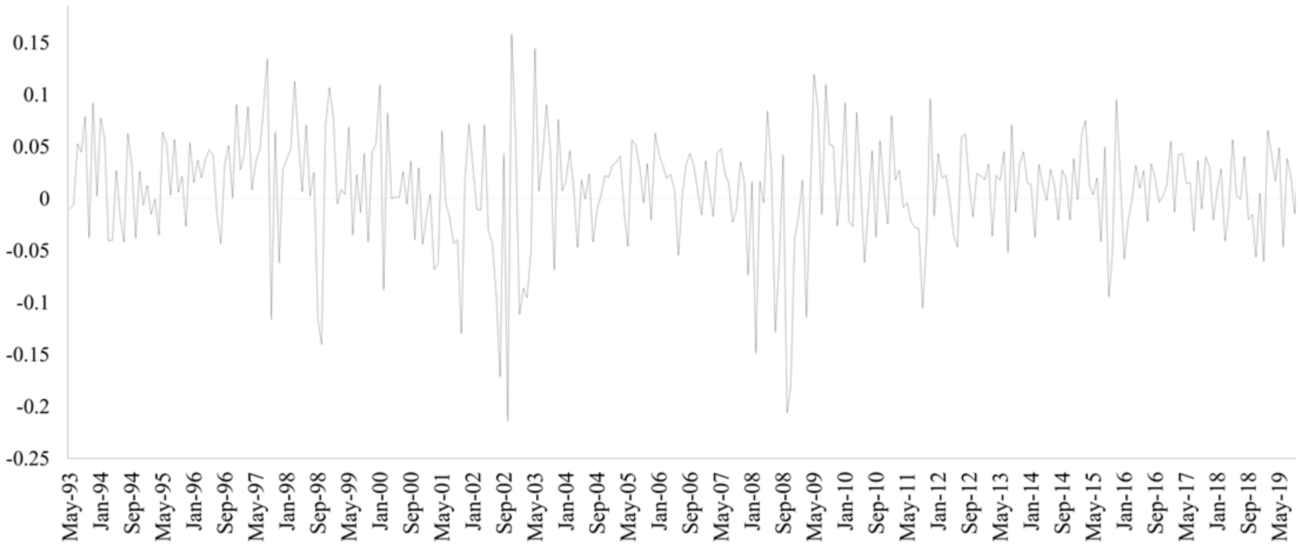

Figure 3
AEX return over time
This Figure plots the monthly AEX returns on the Dutch stock market from May 1993 through April 2020.

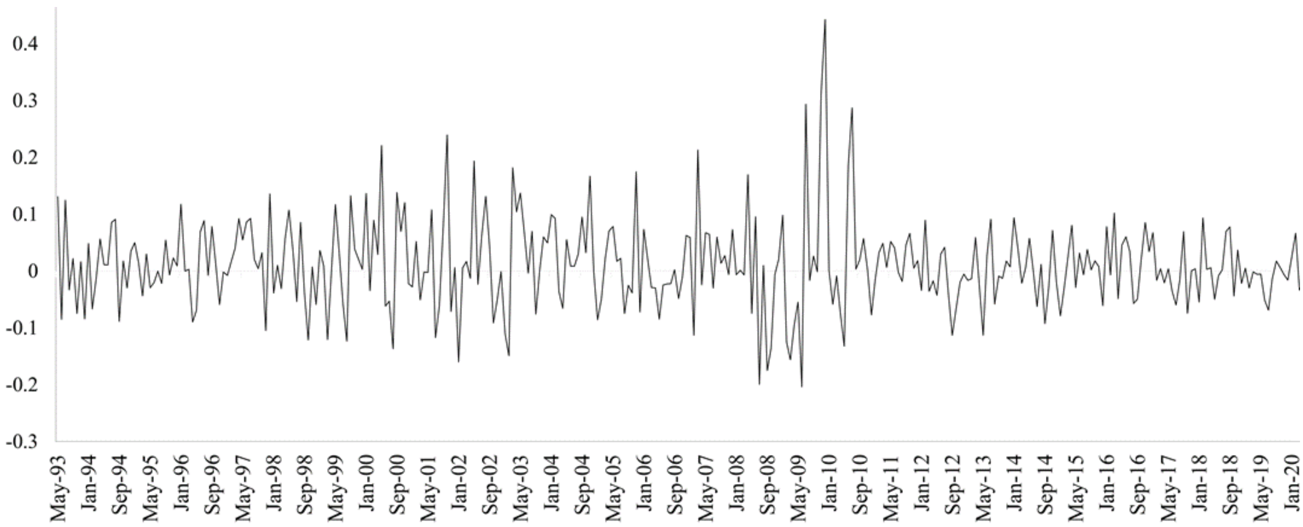

Figure 4
Difference in return of portfolio three and one
This figure plots the difference in returns between portfolio three and one over time from May 1993 through April 2020. These portfolios are made by sorting 23 companies trading on the Dutch stock market (AEX) according to their expected idiosyncratic skewness. Portfolio one is the portfolio with the lowest expected skewness and portfolio three has the highest. The difference in returns each month is defined as the return of portfolio one minus the return of portfolio three.

Figure three serves as a reference to the market returns, displaying the more volatile times. The following six time periods are defined. The first time periods go from 1993 to 1998, 1999 to 2003, 2004 to 2006, 2007-2010 and lastly 2011 to 2020. In order to quantify these possible differences a paired t-test is done to evaluate the difference between the average monthly returns of the portfolios, it will be done for specified periods of time mentioned above.

6. Empirical findings

6.1 cross-sectional differences

The first hypothesis tested if lagged IS, lagged IS, lagged turnover, lagged momentum, lagged size and industry are significant predictors for IS. As shown in table 2, lagged IV and lagged turnover are insignificant predictors.

Table 4
Monthly portfolio returns by estimated skewness

Portfolio	Obsertvations	Mean (%)	Standard deviation	CAPM alpha (%)	P-value
1(low)	324	1.284	13.34	1.163	0.121
2	324	1.034	5.580	0.9149	0.004
3	324	0.7305	4.908	0.5977	0.030
4	257	0.7904	7.067	0.6121	0.172
5(high)	324	0.3172	8.470	0.2926	0.539
1(low)	324	1.118	6.064	1.114	0.001
2	324	0.6581	4.597	0.6028	0.020
3(high)	234	0.2991	8.168	0.3204	0.485
1(low)	324	1.004	4.762	0.8742	0.001
2(high)	324	0.5286	5.022	0.4211	0.136

This table gives an overview of the average (mean) returns of portfolios. The portfolios are created by sorting stocks into five, three or two portfolios according to their expected idiosyncratic skewness. Portfolio, five, three and two have the highest expected idiosyncratic skewness and portfolio one the lowest. Additionally the table shows the CAPM alpha for each portfolio and the corresponding p-values. The CAPM alpha is estimated by the following regression. $R_{p,t} - R_{f,t} = \alpha_{p,t} + \beta_{p,t}(R_{m,t} - R_{f,t}) + \varepsilon_{p,t}$ The portfolio returns are calculated over a time period of May 1993 through April 2020. The stocks are traded on the AEX.

Table 4 contains an overview of the mean returns of each portfolio. The top part of the table shows the differences when dividing the stocks into five portfolios and the middle and bottom part show the stocks divided into three and two portfolios respectively. T-tests are performed on each of the three differences between the average returns in the highest and lowest expected IS portfolios and are summarized in table 5. When using five portfolios the difference in mean returns from the highest and lowest IS portfolio is insignificant with a p-value of 0.12. However, when using three and respectively two portfolios mean returns are substantially larger for the lowest skewness portfolio (0.82% and 0.48%). This means that the second hypothesis that tested if there is a significant positive difference in average returns between portfolios consisting of stocks with the highest expected skewness and the lowest expected skewness cannot be rejected. Adjusted for risk, results also carefully seem to indicate lower returns for stocks with higher expected IS.

Table 5
Statistical significance

Number of portfolios	Difference in average monthly return(%)	Pr(T>t)
5	0.9664	0.1205
3	0.8191	0.0440
2	0.4760	0.0177

This table gives an overview of the difference in average (mean) returns of portfolios. The portfolios are created by sorting stocks into five, three or two portfolios according to their expected idiosyncratic skewness. The table also reports the p-value of the difference that is obtained by preforming a paired t-test. This difference is defined as the difference between the average monthly return of portfolio with the lowest expected skewness and the highest expected skewness. The portfolio returns are calculated over a time period of May 1993 through April 2020. The stocks are traded on the AEX.

6.2 differences across time

The following five time periods are defined. The time periods go from 1993 to 1998, 1999 to 2003, 2004 to 2006, 2007 to 2010 and lastly 2011 to 2020. Table 6 summarizes the average return for each portfolio. It also shows the difference between portfolios one and three each time period, and the significance of this difference that is once again tested with a paired t test.

The magnitude of the difference in average return between the portfolios consisting of stocks with the highest expected skewness and the lowest expected skewness is not significantly different in certain time periods.

Table 6
Return differences at specified time periods

Time period	Portfolio	Average monthly return (%)	Standard deviation	Difference (%)	Pr(T>t)
1993-1998	1(low)	1.990	5.078	1.207	0.0515
	2	2.050	4.468		
	3(high)	0.7825	6.574		
1999-2003	1(low)	0.6580	5.321	1.416	0.1515
	2	-0.1373	5.010		
	3(high)	-0.7576	10.46		
2004-2006	1(low)	2.512	4.419	-0.0006238	0.5141
	2	1.417	3.129		
	3(high)	2.574	10.18		
2007-2010	1(low)	1.631	10.53	1.503	0.2247
	2	-0.08659	5.683		
	3(high)	0.1280	11.02		
2011-2020	1(low)	0.4000	4.640	0.2557	0.2805
	2	0.5392	4.175		
	3(high)	0.1443	4.897		

This table gives an overview of the difference in average (mean) returns of portfolios in five different time periods. The time periods go from 1993 to 1998, 1999 to 2003, 2004 to 2006, 2007 to 2010 and lastly 2011 to 2020. The portfolios are created by sorting stocks into three portfolios according to their expected idiosyncratic skewness. This difference is defined as the difference between the average monthly return of portfolio with the lowest expected skewness and the highest expected skewness. Portfolio three has the highest expected idiosyncratic skewness and portfolio one the lowest. Paired t-tests are performed each time period to evaluate the significance of the difference. The portfolio returns are calculated over a time period of May 1993 through April 2020. The stocks are traded on the AEX.

7. Conclusion, limitations and suggestions for further research

This section provides a summary of the paper and reviews and concludes the empirical findings. Then the limitations will be laid out and discussed in detail. Finally, a few suggestions for further research will be made.

Numerous theories have described that IS could be a priced feature of stocks. Investors exist that have skewness preference, meaning that they overvalue stocks with (high) upside potential. This paper attempts to empirically test this in the Dutch stock market for a time period between 1993 and 2020. A model is estimated of predicted IS using a time horizon of 60 months to describe preference in long-run upside potential. The results of the panel regression indicate that important predictors of IS are lagged IS, momentum, size and industry. Interestingly and in contrast to previous literature that finds a positive relationship (Boyer et al., 2009), lagged IV does not seem to be significant in explaining IS in this data set. Another contrast in that lagged IS has a negative correlation with IS, whereas previous literature finds a positive relationship. This predicted IS is then used to explain the cross-sectional returns. Stocks are sorted into portfolios according to their expected IS. When dividing the stocks into three portfolios the most interesting results are found. The portfolio with the lowest expected IS outperforms the portfolio with the highest expected IS on average by 0.82% per month with a significance level of 5%. Adjusting for risk also results into higher CAPM alphas for stocks with lower expected skewness. This suggests that there could also be investors trading on the Dutch stock market that overvalue high skewness.

Boyer et al. (2009) additionally find that low average returns of stocks with high IV can possibly be (partly) explained by the positive relation they find between IV and IS. The reverse could also be true leaving the reason for the low average return to be possibly caused by one, both or a combination of overlapping explanatory power that they could attribute to either. In all cases they do find that both keep their explanatory power when controlling for one another. Interestingly, this paper finds no significant correlation between IS and IV when preforming the panel regression. Suggesting that the lower returns in this case are solely due to the higher IS. This is consistent with their theoretical view that IS should have the upper hand in explaining these low returns given the established theoretical explanations (Barberis & Huang, 2007; Brunnermeier & Parker, 2005; Mitton & Vorkink, 2007).

Although the methodology of this paper is thoughtfully constructed, there are limitations. When selecting the data, strong restrictions have been placed leading to the following problems. First, the companies selected have been trading on the Amsterdam stock exchange for about 27 years, creating a survivorship bias. A survivorship bias exists when the data restrictions have caused certain entities in a population to be overlooked because they did not make it past a selection process. In this case stocks have not been included that have gone bankrupt, left the stock market or have been taken over in the last 27 years. This results in a data set with only strong and well established companies. The next limitation regarding the data set is the cross-sectional variation. This paper only examines 23 companies leading to limited cross-sectional variation, especially regarding the industries. There is only one mining company and two companies in service and construction. The two service companies were frequently placed into the highest IS portfolio and the two construction companies were frequently placed into the lowest IS portfolio. This variation between the two industries could be interesting to further research using a much larger data set, however with so little companies per industry no conclusions could be made around this. There is also limited cross-sectional variation in the market capitalization of the companies. Other possible problems arise when forming portfolios. When using a larger data set the chance that each portfolio has more observations is bigger and the difference in average returns in the highest and lowest expected IS when dividing the stocks into five portfolios might be significant.

To improve on these limitations a larger data must be chosen and stocks that have not survived for this extensive period of time need to be included into the sample. For even more cross-sectional variation research could create a data set that incorporates more than just one country. Cross-sectional variation in countries might also be a predictor for expected skewness and it would be interesting to evaluate whether skewness preference differs over cultures. Regarding the time variation a larger time period could have been used as this would lead to more periods. Even though the difference overtime is not very significant, the results do carefully point in the direction of a larger dispersion of returns over portfolios ranging from high to low idiosyncratic skewness in crisis times. Further research could give more insight into this possible relationship. Further research could also look into what industries are affected most by these times and different time horizons could also be used to evaluate this.

8. Literature

Ayadi, M., Cao, X., Lazrak, S., & Wang, Y. (2019). Do idiosyncratic skewness and kurtosis really matter? *North American Journal of Economics and Finance* 50: 101008.

Bali, T., N. Cakici, & R. Whitelaw. (2011). Maxing Out: Stocks as Lotteries and the Cross-Section of Stock Returns. *Journal of Financial Economics* 78: 243–276.

Blau, B. & Whitby, R. (2018) Skewness, short interest and the efficiency of stock prices, *Applied Economics*, 50:20, 2229-2242.

Boyer B., Mitton, T., & Vorkink, k. (2009). Expected idiosyncratic skewness. *The review of financial studies* 23: 170-202.

Brunnermeier, M., Gollier, C. & Parker, J. (2007). Optimal Beliefs, Asset Prices and the Preference for Skewed Returns. *American Economic Review Papers and Proceedings* 97:159–65.

Chen, J., H. Hong, & J. Stein. (2001). Forecasting Crashes: Trading Volume, Past Returns, and Conditional Skewness in Stock Prices. *Journal of Financial Economics* 61:345–81.

Fama, E. F., & K. R. French. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33:3–56.

Fama, E. F., & J. MacBeth. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy* 71:607–36.

Guo, H. & Savickas, R. (2006). Idiosyncratic Volatility, Stock Market Volatility, and Expected Stock Returns, *Journal of Business & Economic Statistics*, 24:1, 43-56.

Harvey, C. R., & A. Siddique. (1999). Autoregressive Conditional Skewness. *Journal of Financial and Quantitative Analysis* 34:465–88.

Harvey, C. R., & A. Siddique. (2000). Conditional Skewness in Asset Pricing Tests. *Journal of Finance* 55:1263– 95.

Hong, H., & J. Stein. (2003). Differences of Opinion, Short-Sales Constraints and Market Crashes. *Review of Financial Studies* 16:487–525.

Hugues, L. (2020). Measuring skewness premia. *Journal of Financial Economics* 135: 399-424.

Kahneman, D. & A. Tversky. (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263-91.

Kahneman, D. (2011). Thinking, fast and slow. (1st Ed.) Farrar, Straus and Giroux, New York.

Mitton, T., & K. Vorkink. (2007). Equilibrium Underdiversification and the Preference for Skewness. *Review of Financial Studies* 20:1255–88.

Sharpe, William F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*. 19 (3): 425–442.

Thaler, R. & Sunstein C. (2008). Nudge. (1st Ed.) Business Contact.

Viva, L., Kasanen, E. & Trigeorgis, L. (2017). Real Options, Idiosyncratic Skewness, and Diversification. *Journal of financial and quantitative analysis* 52: 215-241.