UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Downside risk: is downside risk priced in the U.S. stock market?

EFI390

Bachelor's Project 15 hp
Authors: Arsalan Hakimi & Raouf Bahsoun
Tutor: Marcin Zamojski
Spring Term 2020

Abstract

This paper aims to add further research to the field of downside risk, and downside risk measures' influence on the average returns in the U.S. stock market. The study also examines and compares how well the Fama-French three-factor model, Carhart four-factor model, Fama-French five-factor Model, q-four factor model, and q-five factor model explain these average returns. This was done by constructing zero-cost portfolios, split into two weight classes of stocks in the portfolios. The study shows relatively strong results for a major group of the downside risk measures. The measures of the major group show significance and good explanatory power; this could lay ground for further research and use of downside risk measures in financial contexts. Regarding the minor group of the downside risk measures, the result gives ambiguous implications about the way the asset pricing models can explain those residual mean returns. Therefore, the minor group could not establish what asset pricing model is preferred over other models.

Keywords: Excess kurtosis, skewness, Value-at-Risk, Expected shortfall, semi deviation, downside beta, Sortino ratio, Fama-French three-factor model, Fama-French five-factor model, Carhart four-factor model, q-four factor model, q-five factor model, asset pricing, U.S. stock market.

Acknowledgement

Most importantly, we would like to thank our humble and helpful supervisor, Dr. Marcin Zamojski, for the continuous support during the writing process. We would also like to thank the students and opponents for their underpinning feedback and their challenging questions.

Contents

1. Introduction

In the context of finance, risk is one of the most important factors to consider. In everyday life, risk is associated with the avoidance of losing wealth. However, traditional risk measures treat bad outcomes and good outcomes equally. In this paper we focus on downside risk. We put this into practical test to emphasize and stress the cause of narrowing the overall view of what risk is, how it is perceived and how investors cope with it.

In this research paper, we look at listed U.S. companies between January 1960 and December 2019. Several downside risk measures are chosen and examined for their relevance in stock pricing. More specifically, an examination of chosen downside risk measures is performed to see if the risk measures in question are being priced in the expected stock returns. If they are not being priced, they could be interpreted as potential additional risk factors. This is done by studying the alphas generated by regressions of zero-cost portfolios for each risk measure.

Lately, larger attention is being paid to downside risk. According to Farago and Tédongap (2018), the definition of downside risk could be identified as the risk that investors associate with "bad" outcomes such as negative returns or, more generally, returns below their expectations. Investors do not associate risk with large positive returns, returns above their expectations, or upside swings in general. Ang, Chen, and Xing (2006) examines if there is an existing premium for holding stocks with high downside risk. A cross-sectional regression analysis of the stock returns is performed and a downside premium of approximately 6% per annum was presented. The premium does, however, not reflect a compensation for regular market beta; neither was it explained by coskewness, liquidity risk, size or value and momentum characteristics.

In this paper, we use seven downside risk measures: excess kurtosis, skewness, Value-at-Risk, Expected shortfall, semi deviation, downside beta and the Sortino ratio. Excess kurtosis and skewness of an underlying distribution are used to measure the level in which downside risk is priced in the stocks of the U.S. stock markets. Excess kurtosis captures the size (e.g., fat versus thin, and long versus small) of the tails of the underlying distribution in the U.S. stock returns. Skewness captures the direction in which the underlying distribution is skewed. Value-at-Risk and Expected shortfall attempt to quantify the total risk in a portfolio in form of monetary loss (Hull, 2015). Furthermore, the Sortino ratio is similar to the Sharpe ratio, with the key distinction that it is based on semi deviation. Semi deviation is a variation measure just like standard deviation but solely captures the downside direction of a stock fluctuation. Lastly, downside beta is similar to the traditional market beta, although it solely captures the downside direction of a stock fluctuation.

The downside risk measures are used to construct factor-mimicking portfolios. We use five asset pricing models to test whether these portfolios correspond to anomalies or potential risk factors. We adapt the testing procedure from Hou et. al. (2015, 2018). The first two of the models are the Fama and French (1993) three-factor model and the Fama and French (2015) five-factor model. The former model is based on three factors affecting expected stock return: market beta, size, and book-to-market ratio. The latter model distinguishes from the former with two additional factors: investment and profitability. Henceforth, other models used in this paper are Carhart (1997) four-factor model, as well as q-four and q-five factor models of Hou et. al. (2015, 2018). The Carhart four-factor model is similar to the Fama-French three-factor model but deviates with its additional factor that captures the effect of momentum. The q-factor models have minor differences compared to the previous models, which is the set of factors included.

The main finding of the paper is that for equally-weighted portfolios, none of the asset pricing models are able to fully explain the variation in the factor-mimicking portfolios (alpha is significant). The sign of the intercept is consistent with the interpretation of the factor-mimicking portfolios as risk factors, i.e., we find that taking a long position in low downside risk stocks and a short position in high downside risk stocks generates a positive abnormal return. Finally, the intercepts are highly significant with t-statistics ranging from 0.92 to 12.07.

Furthermore, we find that for four out of seven risk measures (Value-at-Risk, Expected shortfall, semi deviation and standard deviation), some of the variation can be explained with the average *Adjusted* R^2 of 0.5024. For the skewness and kurtosis, the asset pricing models provide worse fit, with the average *Adjusted* R^2 of 0.1314. The highest *Adjusted* R^2 is found for the downside beta portfolios, with an average *Adjusted* R^2 of 0.6850.

For the equivalent risk measures of the value-weighted zero-cost portfolios, an opposite pattern (compared to equally-weighted zero-cost portfolios) are identified. In absolute terms, the intercepts increase for every newly added factor of asset pricing models. Some of the variation of these risk measures can be explained with the average *Adjusted* R^2 of 0.5769. For the skewness and kurtosis, these asset pricing models also provide worse fit, with the average *Adjusted* R^2 of 0.1997. In comparison to the equally-weighted zero-cost portfolios, downside beta does not distinguish itself as the risk measure with best providing *Adjusted* R^2. The results are different suggesting that downside risk matters more for smaller companies than for large companies.

The remainder of the paper is structured as follows: in Section 3, downside risk measures, asset pricing models, and the methods used in this thesis are presented. Section 4 describes how raw data has been retrieved and processed. The empirical results are exhibited and discussed in Section 5. In connection to that, Section 6 concludes the thesis by summing up and uplifting a few aspects of the thesis.

2. Literature review

In this section, previous studies of downside risk and downside risk measures are presented. Studies about asset pricing models, such as the Fama-French three-factor and Fama-French five-factor models, Carhart four-factor model and q-factor models are also presented.

2.1 Downside risk measures

Ang, Chen, and Xing (2006) claim that stocks with high covariation with the market, with respect to downside risk have higher returns. For downside risk, the cross section of returns shows a premium. This is consistent with a market where participants place more weight in losses and less weight in gains. Therefore, premiums are required by investors for holding assets with higher downside risk; thus, assets with higher downside risk have higher returns. Ang, Chen, and Xing (2006) also conclude that past downside beta provides good predictions of future covariation with downward market movements, but for stocks with high volatilities, past downside beta predicts poorly of future downside risk.

Downside risk in asset prices is further discussed by Farago and Tédongap (2018). Apart from market returns and market volatility, they find other *disappointment-related factors*: a downstate factor, a market downside factor and a volatility downside factor. Downside risk can also be related to a rise in market volatility. The provided empirical test strengthens the argument that the factors are priced in the cross-section of various asset classes.

Investors possess different opinions of risk in modern portfolio history (Estrada 2006). Investors do not place equal weights in upsides and downsides, which is why the downside risk framework is quickly gaining acceptance among academics. Semi deviation captures the downside dispersion that investors try to avoid. The measure gives a good estimate of risk when the distribution is skewed, and the benchmark return is other than the mean. Estrada (2006) further describes that according to the downside beta, downward market swings are risky and upward market swings are not necessarily as risky. Some of the risk measures are very popular and have received a lot of attention in the literature (VaR, ES), while others like the Sortino ratio are not.

Acharya et al. (2016) provide simple ways to measure banks' contributions to systemic risk and propose ways to limit this. Value-at-Risk and Expected shortfall are initially introduced as measures to limit systemic risks. Their theory further suggests that the regulation of systemic risk should depend on the systemic expected shortfall (SES) of each firm. The components of SES could have been used to predict the 2007-2008 financial crisis (Acharya et al. 2016). They conclude that financial institutions have incentives to take risks that could affect everyone, except for the case when the external costs of systemic risk are internalized.

2.2 Factor models

Not all factors that are proposed in the literature are genuine risk factors. Some of them are simply anomalies. These anomalies can be explained by sophisticated asset pricing models or due to multiple hypothesis testing fallacy (Campbell, Yan and Heqing, 2016). Recently, Hou et al. (2015, 2018) look at 158 anomalies and find that most of them can be explained by a five-factor model. The testing procedures in these papers are GRS-tests and significance of alphas. We largely follow the testing procedure from these papers.

Fama and French (1993) study whether common risk factors in stocks and bonds are captured in the cross-section of average returns. Their empirical results show that firms with small market capitalizations and high book-to-market outperform the market. The three-factor model is introduced and contains a market factor (the CAPM factor), a size factor and a book-to-market factor. In a later research, Fama and French (2015) obtain empirical results which suggest that five factors provide a good explanation to the variation in bond and stock returns, as well as cross-section of average returns. In addition to size and book-to-market, patterns in average returns related to operating profitability and investment are found.

Novy-Marx (2013) states that firms with higher profitability generate relatively higher average returns, regardless of bigger size and low book value. He further mentions that valuation and profitability strategies are composed to obtain productivity cheaply and leads to larger abnormal returns. Value strategies and profitability strategies are negatively correlated, which is why they work well when put together, which in turn reduces the volatility. This goes hand in hand with the reasoning of the value premium that is argued by Fama and French (1993). Titman, Wei, and Xie (2004) find that a second factor would be in place: investments.

Carhart (1997) describes that selling the bottom-decile mutual funds and buying top-decile funds in 1996 yields 8%. Momentum explains 4.6% of this spread, 0.7% is explained by expense ratios and differences in transaction costs explain 1%. Carhart (1997) ends his conclusion by stating three rules-of-thumb for wealth-maximizing mutual fund investors. The first rule is to avoid funds with continuous bad performance. The second rule is that funds with high returns in the previous year will yield higher-than-average expected returns the following year but will decline in the following years. Last rule-of-thumb is that the investment costs of expense ratios, load fees and transaction costs all have negative impacts on performance.

Over the past 25 years, the Fama-French model has failed to account for many asset pricing anomalies. Hou, Xue, Zhang (2015) perform an examination of approximately 80 anomalies and yielded two major findings. For the high-minus-low deciles created by value-weighted returns and NYSE breakpoints, insignificant mean returns are found for almost 40 of the anomalies. Their evidence suggests that a q-factor model with four factors (the market factor, a size factor, an investment factor and a profitability factor) outperforms the Fama-French and Carhart models.

Hou et al. (2018) further extend the q-four factor model by adding an expected growth factor and forming the q-five factor model. Although the Fama-French five-factor model is the best value-versus-growth model, it does not show any explanatory power for momentum, according to Hou et al. (2018). They eventually claim that the best performing model is the q-five factor model.

3. Methodology

In this section, downside risk measures, asset pricing models and our testing procedures are presented. Subsections 3.1 to 3.7 discuss the downside risk measures. Subsections 3.8 to 3.12 discuss the asset pricing models that are used when running our regressions. Subsections 3.13 and 3.14 present the methods used to test the downside risk measures.

3.1 Skewness

Skewness quantifies the extent of asymmetry of a given distribution. Skewness is defined as:

$$s = \frac{\sqrt{n(n-1)}}{n-2} \frac{\frac{1}{n}\Sigma_i (x_i - \underline{x})^3}{(\frac{1}{n}\Sigma_i (x_i - \underline{x})^2)^{3/2}}$$

where n is the sample size, x_i is a random variable and $\underline{x}$ is the mean of the random variables. The normal distribution is symmetric and has a skewness of zero. We distinguish between *positive* and *negative* skewness that are also referred as *right-skewed* and *left-skewed*, respectively. The direction reference of the skewness comes from the fact that the left or right tail is longer. For a positively skewed distribution, the mode is smaller than the median, which in turn is smaller than the mean, and vice versa for negatively skewed distributions.

3.2 Excess kurtosis

Excess kurtosis is a statistical measure that describes a probability, or return distribution, that has a kurtosis coefficient larger than the coefficient associated with a normal distribution, which is approximately 3. The degree of excess kurtosis is calculated with the following formula (Pearson, 1905):

$$\eta = \beta_2 - 3 = \frac{\mu_4 - 3\mu_2^2}{\mu_2^2}$$

where β_2 is the characteristic coordinate of the second moment; μ_2 and μ_4 are the means of the second and fourth central moments.

There are three regimes of excess kurtosis (Pearson, 1905). The first regime is a *mesokurtic* distribution. The form of this distribution is quite similar to the normal distribution. Comparing tops, the mesokurtic distribution is almost equally flat-topped. Recall the excess kurtosis equation, the degree of excess kurtosis would be $\eta = 0$. The second regime is a *leptokurtic* distribution. This kind of distribution exhibits a greater degree of excess kurtosis than a mesokurtic distribution. It is distinguished by its fatter tails. For this type of distribution, the given degree of excess kurtosis is $\eta > 0$. Top of the normal curve of a leptokurtic distribution is less flat-topped. The third regime is a *platykurtic* distribution, which is characterized by its greater flat-toppedness. Consequently, the given degree of excess kurtosis is $\eta < 0$. The tails are much thinner, which clearly implies that there are fewer outliers than for a leptokurtic or mesokurtic distribution. So, there are fewer extreme outliers that deviates from the mean compared to leptokurtic distribution. It is important to emphasize that no distribution could be legitimately described as normal unless both the excess kurtosis and the skewness are zero. In the real world, this kind of distribution could rarely be found.

3.3 Value-at-Risk

Value-at-Risk, henceforth VaR, was first presented by JPMorgan (1990), and published in their own system, RiskMetrics (1994). VaR is used to make a statement of how much money that could be lost within a certain time period, given a certain confidence level. In other words, VaR for a portfolio can be calculated from a probability distribution of losses during a time horizon (days, weeks, months or years) and the portfolio value (Hull, 2015). This relationship is described by following equation:

$$Pr(X \leq VaR) = \alpha,$$

where X denotes returns of the portfolio and α is a predetermined percentile point on a probability distribution.[1] VaR is, however, flawed in one perspective. It does not possess all risk properties that are needed to maintain relevance within a risk measure. VaR meets the monotonicity, translation invariance and homogeneity properties, but not subadditivity.[2]

---

[1] There are three methods for measuring VaR: parametric method, Monte Carlo simulation and historical method. The parametric method assumes that the returns of each risk factor impacting the values of the assets in a portfolio, will follow a normal distribution. The equation can then be simplified to one that links the returns of those risk factors to the values of the assets. The Monte Carlo simulation is based on defining a stochastic model for each risk factor that can impact the value of the portfolio. The historical method is based on the rankings of the profits and losses of the current portfolio, where for each instrument and portfolio changes in value, historical returns are applied on each risk factor for these profits and losses.

[2] Monotonicity: If a portfolio presents worse results than other portfolios, ceteris paribus, its risk measure should be greater. Translation invariance: If *K* amount of cash is added to the portfolio, its risk measure should decrease by *K*. Homogeneity: Changing the size of a portfolio by a factor λ, ceteris paribus, their risk measure should be multiplied by λ. Subadditivity: After two portfolios have been merged, the risk measure should be lower than before they were merged.

3.4 Expected shortfall

Expected shortfall, henceforth ES, calculates the expected loss above a given certain percentile on a loss distribution, during a certain time period. ES possesses all the properties, including the subadditivity property. ES of a portfolio is the mean on days when the loss exceeds the predetermined VaR point (Hull, 2015). It is defined by the following equation:

$$\int_{-\infty}^{\alpha} x\, f(x)$$

where α is a predetermined percentile point on a probability distribution and x denotes the returns of the portfolio. Using ES instead of VaR limits the probability of taking unacceptable risks. This is based on the grounds that VaR only reflects the loss on one specific percentile point, whereas ES reflects the average loss on that percentile point and above. ES captures the advantages of diversification is therefore a more comprehensive measure (Hull, 2015).

3.5 Semi deviation

Semi deviation is a measure of dispersion, which captures an investment's mean return but differs from standard deviation in the sense that semi deviation only takes negative returns into account. The formula is presented in the following equation:

$$\sum_{B} = \sqrt{(1/T) \times \sum_{t=1}^{T} \quad \{Min(R_t - B, 0)\}^2}$$

This measures the volatility below the chosen benchmark return, B, for T observations, during the time period t and R_t denotes the returns for an asset (Estrada 2006). Some investors prefer using a limit that reflects their own risk tolerance more accurately. This is called a benchmark return. A benchmark return could be any return that an investor chooses to use; risk-free rate, the mean return of a set of returns (S&P 500 or Russell 3000) or a zero-return could be used among many others.

3.6 Downside beta

Downside beta measures the probability of an asset falling below the least accepted return, the benchmark return. The downside beta was popularized by Ang, Chen, and Xing (2006). In this paper, we use the formulation from Estrada (2006):

$$\beta_B{}^D = \frac{\sum_{t=1}^{T}\{Min(R_t - B, 0) \times Min(R_{MKT} - B_{MKT}, 0)\}}{\sum_{t=1}^{T}\{Min(R_{MKT} - B_{MKT}, 0)\}^2},$$

where t indicates time and T indicates the number of observations, R_{MKT} and R_t denote the average market excess return and the excess return of an asset; B_{MKT} and B_t denote the benchmark market return and the asset benchmark return. If the downside beta for an asset is 1.5, it explains that, on average, when the market falls by 1% below the benchmark return, the asset will fall 1.5% below the benchmark return. The bigger the downside beta, the more sensitive is the return of the stock to the movements of the market.

3.7 Sortino ratio

The *Sortino ratio* follows the same structure as the Sharpe ratio but deviates through its risk-adjusted factor - the semi deviation. The excess return of the Sortino ratio can be calculated with risk-free rate as well as mean return of the same portfolio, zero, interbank interest rate, or some other benchmark return (like the benchmark return included in the formulas for semi deviation and downside beta). The formula is explained by the following equation:

$$T_P = \frac{E(R_P) - B}{\Sigma_{B_p}}$$

$E(R_P)$ denotes the expected return of a specific portfolio, B denotes the benchmark return chosen for the specific portfolio, and Σ_{B_P} denotes the semideviation of the portfolio with respect to the benchmark return B (Estrada, 2006). The values that the Sortino ratio is expected to attain are similar to those of the Sharpe ratio. Negative values of the Sortino ratio are undesired. As the value of the Sortino ratio increases, the performance per unit of downside risk increases. Therefore, the bigger value of Sortino ratio implies the better performance of the portfolio.

3.8 Fama-French three-factor model

Fama-French three-factor model is a multi-factor asset pricing model. Based on previous research, Fama and French (1993) draw the conclusion that small firms in terms of market capitalization and firms with high book-to-market ratio seem to outperform the market. Therefore, the size and valuation factors are added to the formula of Capital Asset Pricing Model (CAPM). This results in the Fama-French three-factor model:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{i,MKT}MKT_t + \beta_{i,SMB}SMB_t + \beta_{i,HML}HML_t + \varepsilon_{i,t},$$

where α denotes the unexplained contribution to the mean return that this specific multi-factor asset pricing model generates. Further, ε denotes the error factor which intended to be zero if there is no error in the data set. *MKT* stands for the excess return on a market proxy. Fama and French (2015) state that the size factor, *SMB*, is the difference between the mean of the three smallest portfolio returns and the mean of the three biggest portfolio returns. The factor is based on independent sorts of stocks into two size groups. The size breakpoint is the NYSE median market capitalization. Historically, on average, stocks with small market capitalization have been outperforming those with large market capitalization. This is translated into a higher $\beta_{i,SMB}$ due to the higher risk premium that stocks with small market capitalizations have. Fama and French (2015) also state that the value factor, *HML*, is the difference between the mean of the two highest book-to-market ratio portfolio returns and the mean of the two lowest book-to-market ratio portfolio returns. The factor is based on independent sorts of stocks, divided into three book-to-market ratio groups. The book-to-market ratio breakpoints are the 30th and 70th percentiles of book-to-market ratios for NYSE stocks. Usually, high and low book-to-market ratios are described as *value* and *growth stocks*, respectively. Value stock is defined as a big, mature company with low growth expectations but instead generates good profitability. Growth stock is the opposite - it is defined as a small company with high growth expectations and low profitability (because of its size; the company is in the first period of the company's life cycle). A value stock tends to have its beta estimated as a positive *HML* factor while a growth stock tends to have its beta estimated as a negative *HML* factor.

3.9 Carhart four-factor model

Carhart (1997) concludes that the Fama-French three-factor model was not able to explain the cross-sectional variation in momentum-sorted portfolio returns, although the Fama-French three-factor model improved average CAPM pricing errors. Consequently, Carhart extends the Fama-French three-factor model by adding a momentum factor. The *Carhart four-factor model* uses the following formula:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{i,MKT}MKT_i + \beta_{i,SMB}SMB_i + \beta_{i,HML}HML_t + \beta_{i,MOM}MOM_t + \varepsilon_{i,t}$$

MOM denotes the difference between a portfolio of past winners and a portfolio of past losers for the trailing twelve months. Carhart (1997) points out that the Carhart four-factor model explains more than half the spread[3] in return on the one-year-return portfolios. Further, the Carhart four-factor model explains a smaller fraction of the spread in return on the two- to four-year portfolios, and none of the spread in the five-year portfolios.

3.10 Fama-French five-factor model

The *Fama-French five-factor model* is an extension of the original Fama-French three-factor model. Fama and French (2015) add two factors that capture the effect of investments and profitability on the examined average stock returns. This was an effect of several researchers finding evidence of Fama-French three-factor model being incomplete. The incompleteness derives from the alpha factor not being zero, which implies that there are other factors affecting the average returns of stocks. Novy-Marx (2013) presents the finding of profitable firms generating significantly higher average returns, despite their relatively big size and low book value (i.e., low book-to-market ratio). Titman, Wei, and Xie (2004) find the grounds for the second additional factor: investments. The finding suggests that firms that substantially increase capital expenditures would achieve less positive or negative stock returns. Fama and French (2015) conclude that with these findings, they can form the Fama-French five-factor model:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{i,MKT} MKT_t + \beta_{i,SMB} SMB_t + \beta_{i,HML} HML_t + \beta_{i,RMW} RMW_t + \beta_{i,CMA} CMA_t + \varepsilon_{i,t}$$

The construction of *RMW* and *CMA* are similar to the construction of *HML*, except that the second sort is either on operating profitability or investment. These factors can be interpreted as the means of profitability and investment factors for small and big stocks (Fama and French 2015). *RMW* factor can be based on operating profit, based on *Return on Equity*. The *CMA* factor is the change in total assets from the previous fiscal year to the current fiscal year, divided by previous fiscal year's total assets. Anomaly variables from the three-factor model could induce problems since they do not properly capture the variables in question, which is why it is relevant to include investment and profitability factors. Fama and French (2015) conclude that patterns in mean returns are acknowledged when put in relation to size, book-to-market equities, profitability and investment.

[3] Buying last year's top-decile mutual funds and selling last year's bottom-decile funds yields a return of 8 percent per year.

3.11 q-four-factor model

The *q-four-factor model* is a multi-factor asset pricing model with a distinguishing set of mentioned factors: market, size, investment, and profitability. The formula of *q-four-factor model* is written in the following manner:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{i,MKT}MKT_t + \beta_{i,SMB}SMB_t + \beta_{i,RMW}RMW_t + \beta_{i,CMA}CMA_t + \varepsilon_{i,t}$$

In their paper, Hou, Xue and Zhang (2015) empirically test that the q-four factor model summarizes the cross section of average stock returns. The test was done by examining nearly 80 anomalies, to reveal that almost one-half of the anomalies are insignificant. Besides this, the performance was at least comparable to, and in many cases better than that of the Fama-French three and Carhart four-factor models in explaining the average excess return of stocks.[4]

3.12 q-five-factor model

Hou et. al. (2018) add a new factor to the previous q-four-factor model: an expected growth factor *EG*. The newly added factor shows strong explanatory power in the cross-sectional regressions. The expected growth factor outperforms other recently proposed asset pricing models such as the Fama-French six-factor model. The new model is called *q-five factor model* and is presented in the following formula:

$$\begin{aligned} R_{i,t} - R_{f,t} = \alpha_i &+ \beta_{i,MKT}MKT_t + \beta_{i,SMB}SMB_t \\ &+ \beta_{i,RMW}RMW_t + \beta_{i,CMA}CMA_t + \beta_{i,EG}EG_t + \varepsilon_{i,t} \end{aligned}$$

The model has shown to be the overall best performing model.

3.13 Regression analysis (time-series)

For every risk measure, regressions are performed for portfolios with equally weighted returns and value weighted returns. Using both methods reflect two different investment methods, giving a broader understanding of the downside risk measures. These portfolios are meant to be factor-mimicking portfolios. The first- and the tenth quantiles are calculated for the portfolio returns; a long position is taken for the tenth quantile and a short position for the first quantile, or vice versa, depending on which downside risk measure is being examined. Implication of this is to state that a downside factor is considered by creating these zero-cost portfolios that

[4] Based on some exemptions

mimic the measures. Furthermore, the portfolios are zero-cost portfolios, not one-sided portfolios, which is why excess returns ($R_{i,t} - R_{f,t}$) are not needed for these regressions.

A total of $14 \times 5 \times 2$ time-series regressions are constructed, 70 regressions per weight class. *Alpha*, *Alpha t-statistic* and *Adjusted R^2* are presented for our regressions. These regressions are made with HAC standard errors with 10 lags.[5] The error structure is assumed to be heteroskedastic and possibly serially correlated. Regressions with robust standard errors are also made, but regressions with HAC standard errors are mainly used due to their higher robustness (see appendix for the regressions with robust standard errors).

The risk measures are constructed as zero-cost portfolios and are split into equally-weighted and value-weighted portfolios. Alpha is the mean of the part of the portfolio that is not being explained in the asset pricing models. Alpha t-statistic is used to see how significant the alpha is (to set the significance level). A significant alpha suggests that the asset pricing models can be extended with our downside risk factor. *Adjusted R^2* is presented because it penalizes when adding new factors to the asset pricing models.

3.14 Construction of the portfolios

The outputs are every investable stock from the beginning of the period until the end of the period examined. January 3rd, 1967 to June 28th, 2019. Rebalancing dates are set every year on the 30th of June. This is done to reflect new companies that entered the market, also companies that were delisted. Table 2 is presented to reflect the plausible ranges for the different risk measures. The table shows that the Sortino ratio and semi deviation are the same order of magnitude since they are related; the same parallel is drawn for VaR and ES.

[5] This is an arbitrary number and could be optimized, but for the interest of time it was not optimized.

Table 2: Summary statistics for every risk measure, where semi deviation and Sortino ratio without notations are constructed with respect to the portfolio mean returns.

	Observations	Mean	Std	Min	25%	50%	75%	Max
Skewness	237811	0.0465	1.3490	-50.0710	-0.1878	0.1346	0.4687	36.0639
Kurtosis	237811	15.3177	43.3003	-0.4680	4.4577	7.6505	14.1492	3508.5234
VaR 90.0%	237819	-3.2880	1.7958	-40.5465	-4.2560	-2.9322	-2.0176	0
VaR 95.0%	237819	-4.7970	2.5473	-69.3147	-6.1076	-4.2560	-2.9509	0
VaR 99.0%	237818	-8.7111	4.7247	-91.6355	-11.0780	-7.6919	-5.3110	0
ES 90.0%	237811	-5.6142	3.0287	-58.6023	-7.2437	-5.0173	-3.4451	0
ES 95.0%	237811	-7.3808	3.9182	-78.7904	-9.4480	-6.5491	-4.5201	0
ES 99.0%	237811	-12.1613	6.6844	-130.7754	-15.5369	-10.6305	-7.3156	0
Semi deviation	237811	2.2559	1.1744	0.2082	1.3702	1.9806	2.8661	20.7655
Semi deviation risk-free	237811	2.2662	1.1980	0.0342	1.3811	2.0134	2.9087	20.7348
Standard deviation	237811	3.2759	1.6983	0	2.0242	2.9270	4.1885	29.7643
Downside beta	237819	0.6981	0.5463	0	0.2865	0.5780	0.9657	11.6356
Sortino	237811	-0.0151	0.0309	-0.3550	-0.0318	-0.0145	-2.87E-05	0.2655
Sortino risk-free	237811	-0.0052	0.0324	-0.7737	-0.0231	-0.0052	0.0106	0.6958

4. Data

4.1 Data sample

We use data obtained from CRSP (Center for Research in Security Prices) through Wharton Research Data Services (*wrds*). The initial data sample consists of daily stock returns of all American-listed companies between January 1960 and December 2019, with a total of 18231 stocks. The data sample consists of ten different query variables.[6] Some duplicates in stocks are identified. The duplicates are removed, which is further explained in the forthcoming section. We include cumulative factors for prices and shares outstanding to adjust for changes in numbers of shares outstanding (for instance, stock splits and share buybacks), acquisitions and sales of stocks. This is also to calculate adjusted prices and adjusted shares outstanding. This makes it possible to further calculate the correct market capitalization for every tradable date.

The Fama-French factors and the momentum factor (Carhart four-factor model) used in the report are obtained from Kenneth French's website (2019). The q-factors are obtained from the global-q website (2020).

4.2 Data processing description

The raw data that we download from CRSP needs to be processed to account for missing values (blank; not presented) and multiple share classes. We remove mutual funds, trusts, ETFs (e.g., iShares), etc. For the filtration process and further processing of data, programming language Python is used. Missing observations are filtered out and companies with several share classes are identified. Among those companies, we identify and filter out share classes with low volumes traded. We assume that share classes that traded bigger volumes 75% of the time, are the relevant ones for our analysis. Given two share classes, we picked the share class with the longer history. Share classes with short time series are removed, as well as shorter price series; series with fewer bid/ask quotes are also removed. A comparison is also made to pick the share classes with the bigger market capitalizations. The effect of distinguishing share classes in this manner was to make sure that the dataset was narrowed down to only including tradeable share classes.

[6] The following query variables downloaded for each tradable company were PERMNO (*Permanent ID-number of a Security*), PERMCO (*Permanent ID-number of a Company*), DT (*Tradable Dates*), COMNAM (*Company Name*), SHRCLS (*Share Class*), PRIMEXCH (*Primary Exchange*), SHROUT (*Number of Shares Outstanding*), PRC (*Closing Price or Bid/Ask Average*), CFACSHR (*Cumulative Factor to Adjust Shares Outstanding*) and CFACPR (*Cumulative Factor to Adjust Price*).

5. Empirical results

5.1 Descriptive statistics

This subchapter presents a summary of descriptive statistics for our risk measure regressions, using asset pricing models. Table 3 reflect numbers for a set of risk measures from both weight classes, although equally-weighted portfolios are the main focus in the results.

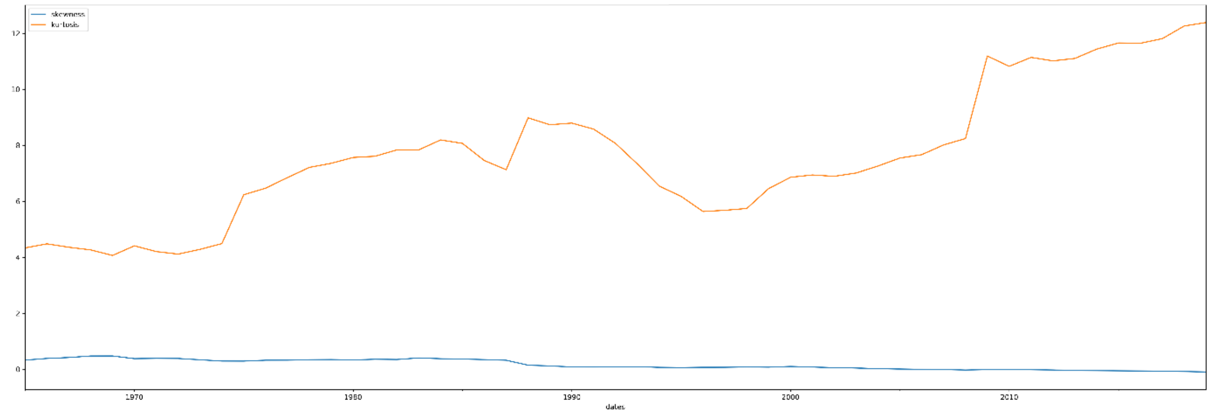
Table 3: Descriptive statistics for a set of risk measures' zero-cost portfolios on every date; EW denotes equally-weighted portfolios and VW denotes value-weighted portfolios

	Observations	Mean	Std	Min	25%	50%	75%	Max
ES 95% EW	13576	0.0752	0.9075	-9.4748	-0.4210	0.0390	0.5284	10.2654
ES 95% VW	13576	-0.0881	1.2037	-11.8416	-0.7120	-0.1129	0.5021	17.2985
VaR 99% EW	13576	0.0727	0.8877	-8.8909	-0.4108	0.0381	0.5177	10.1668
VaR 99% VW	13576	-0.0838	1.1682	-11.2647	-0.6916	-0.1040	0.5004	16.5246
Sortino risk-free EW	13576	0.0401	0.6825	-7.4641	-0.2931	0.0385	0.3754	6.9159
Sortino risk-free VW	13576	0.0008	0.8453	-5.7156	-0.4408	0.0086	0.4409	7.9846
Semi deviation EW	13576	-0.0754	0.8604	-9.6672	-0.5036	-0.0407	0.3849	8.7624
Semi deviation VW	13576	0.0938	1.1230	-15.3593	-0.4687	0.1124	0.6786	12.0278
Downside beta EW	13576	0.0027	0.8764	-9.7094	-0.4108	-0.0255	0.4359	10.3266
Downside beta VW	13576	0.0052	0.9294	-7.7154	-0.4670	0.0021	0.4714	8.8541
Kurtosis EW	13576	-0.0132	0.4517	-3.2214	-0.2698	-0.0051	0.2518	6.3558
Kurtosis VW	13576	0.0059	0.7072	-10.3702	-0.3402	-0.0001	0.3523	8.0680
Skewness EW	13576	-0.0134	0.3828	-3.9035	-0.2197	-0.0142	0.1998	3.5137
Skewness VW	13576	0.0036	0.6175	-8.2017	-0.3161	0.0123	0.3330	8.0386

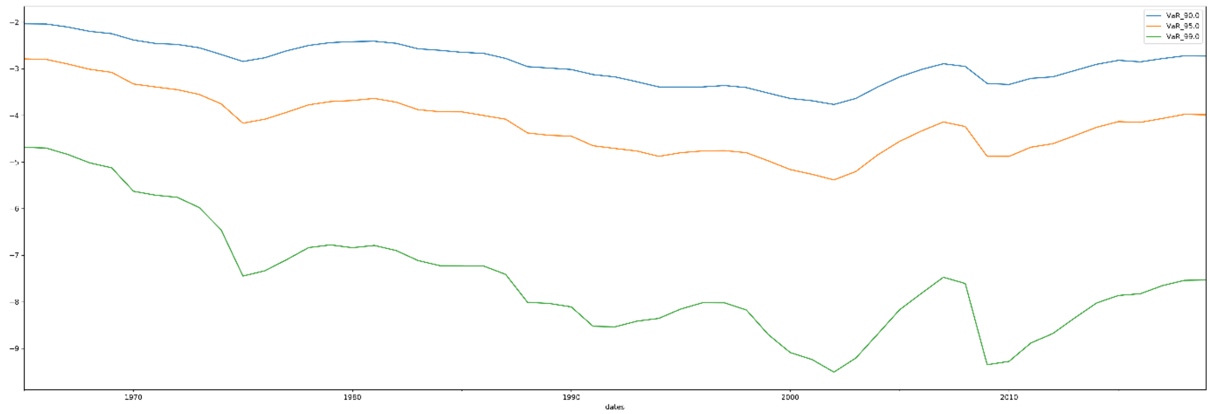
In Figure 1, we can see how the median returns of the portfolios change for every risk measure over time, except for the Sortino ratio. In the first panel, kurtosis increases from 4 to 12, while skewness slightly decreases from 0.2 to 0. In the second panel, VaR is measured; the range goes from -2 to -9. All VaR measures slightly decreases over time, but the smallest one is VaR 99%. In Panel 3, ES follows the same trend as VaR, but the range goes from -3.5 to -13. Panel 4 shows how semi deviation and standard deviation increases over time, where the biggest increase is reflected in semi deviation with respect to the risk-free rate. Sortino ratio and Sortino ratio with respect to the risk-free rate has not changed over time, and has steadily been at 0, which is reflected in Panel 5. Downside beta has, however, decreased from 1.5 to 0.5 over the time frame given.

Figure 1: Median over time for every risk measure

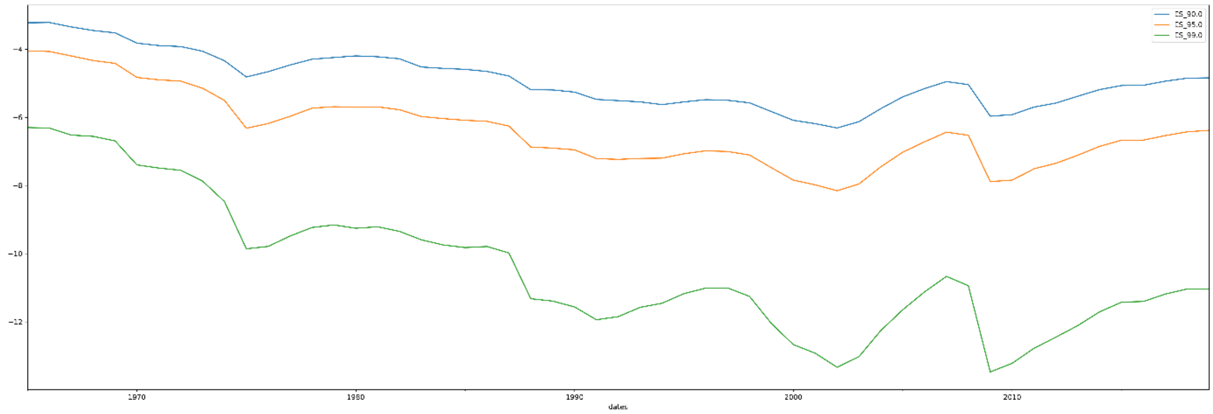
Panel 1: Skewness in blue and kurtosis in orange



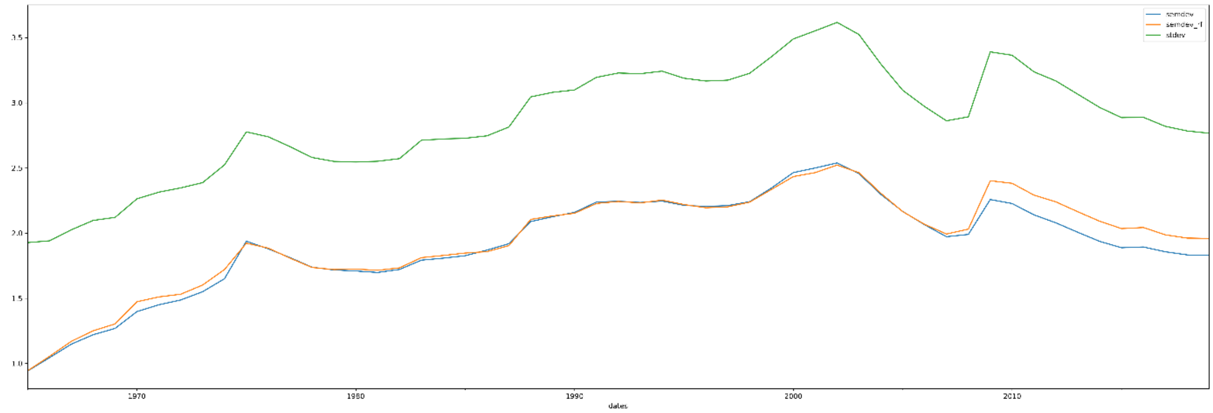
Panel 2: VaR 90% in blue, 95% in orange and 99% in green



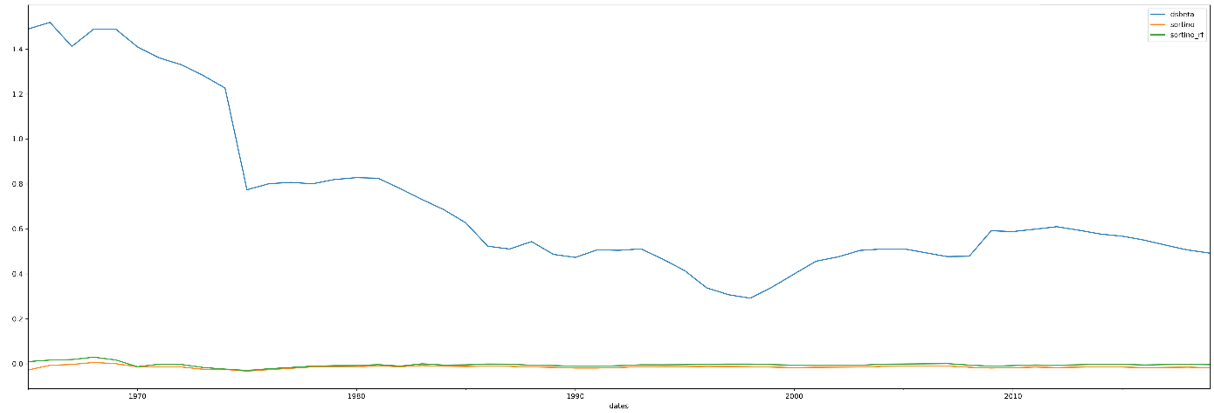
Panel 3: ES 90% in blue, 95% in orange and 99% in green



Panel 4: Semi deviation in blue, semi deviation risk-free in orange and standard deviation in green



Panel 5: Downside beta in blue, Sortino ratio in orange and Sortino ratio risk-free in green



5.2 Main findings

We present the time-series regressions in Table 4. First and foremost, we find that nearly all the alphas are significant and negative (all except for the Sortino ratio and Sortino ratio with respect to the risk-free rate). This suggests that companies with higher exposure to downside risk earn lower returns and vice versa after controlling for other risk factors. This finding is consistent with downside risk being traded at a discount. The economic interpretation of the findings is that investors are either oblivious to downside risk, their risk preferences are skewed towards higher variance portfolios, or they are not investing in efficient portfolios. Ang, Chen, and Xing (2006) arrive at a different conclusion, that the cross section of stock returns reflects a downside risk premium. They also conclude that stocks with high downside betas have higher average returns, which they further describe as consistent with a market where participants place more weight in losses and less weight in gains. Their conclusions can be partly explained by different sample groups in the papers. Ang, Chen, and Xing (2006) base their study on a concentrated sample group, i.e., NYSE, while our study contains all U.S. stock exchanges. It is likely that they intend to avoid the illiquidity effect of including small firms.

The zero-cost portfolios are then potential risk factors, where none of the asset pricing models are fully able to explain the variation in our portfolios. The alphas are also very significant in the sense that they have high t-statistics, which suggests that this is not due to multiple hypothesis testing or fortuitous numbers in the results. Considering the high *Adjusted* R^2, the asset pricing models are not completely ignored. The factors in the models are indeed correlated with other factors. However, this is not an issue since the factors from asset pricing models themselves are correlated (for example SMB and MKT in the Fama-French three-factor model).

For many models (VaR 90%, VaR 95%, VaR 99%, ES 90%, ES 95%, ES 99%, semi deviation, semi deviation with respect to the risk-free rate and standard deviation), we find significant alphas. Alpha of the q-five factor model is closest to zero, followed by the alphas of q-four, Fama-French five, Carhart four and Fama-French three-factor models. Although, the most significant alpha is the one of Fama-French three-factor model (biggest absolute t-statistic). As we keep on extending models from the Fama-French three-factor model to Fama-French five and q-factor models, we can explain additional parts of the variation, which translates to added explanatory power (higher *Adjusted* R^2). Despite that, there is a small deviation: *Adjusted* R^2

of the q-four and q-five factor models are smaller than those of the Carhart four and Fama-French five-factor models, respectively (see Table 4).

Table 4: Equally weighted returns using HAC; numbers in brackets denominate *Alpha t-statistics*

	***Alpha* FF3 [%]**	***Adj.* R^2 FF3**	***Alpha* FF5 [%]**	***Adj.* R^2 FF5**	***Alpha* Carhart [%]**	***Adj.* R^2 Carhart**	***Alpha* q-four [%]**	***Adj.* R^2 q-four**	***Alpha* q-five [%]**	***Adj.* R^2 q-five**
ES at 90.0%	-0.0948	0.5034	-0.0853	0.5363	-0.0920	0.5061	-0.0813	0.4977	-0.0703	0.5098
	(-12.07)		*(-11.63)*		*(-11.59)*		*(-10.57)*		*(-9.18)*	
ES at 95.0%	-0.0932	0.4956	-0.0831	0.5311	-0.0902	0.4988	-0.0789	0.4923	-0.0682	0.5041
	(-11.69)		*(-11.12)*		*(-11.18)*		*(-10.04)*		*(-8.73)*	
ES at 99.0%	-0.0893	0.4984	-0.0786	0.5389	-0.0857	0.5031	-0.0741	0.5001	-0.0645	0.5101
	(-11.31)		*(-10.66)*		*(-10.72)*		*(-9.60)*		*(-8.39)*	
VaR at 90.0%	-0.0929	0.5167	-0.0836	0.5509	-0.0902	0.5190	-0.0796	0.5100	-0.0687	0.5215
	(-11.84)		*(-11.46)*		*(-11.37)*		*(-10.38)*		*(-8.95)*	
VaR at 95.0%	-0.0919	0.5066	-0.0825	0.5392	-0.0889	0.5096	-0.0781	0.5020	-0.0672	0.5139
	(-11.68)		*(-11.20)*		*(-11.16)*		*(-10.07)*		*(-8.70)*	
VaR at 99.0%	-0.0904	0.4843	-0.0806	0.5187	-0.0872	0.4879	-0.0764	0.4832	-0.0657	0.4955
	(-11.46)		*(-10.89)*		*(-10.93)*		*(-9.82)*		*(-8.53)*	
Downside beta	-0.0198	0.6878	-0.0208	0.6898	-0.0183	0.6887	-0.0176	0.6775	-0.0119	0.6812
	(-3.91)		*(-4.18)*		*(-3.59)*		*(-3.49)*		*(-2.31)*	
Kurtosis	-0.0091	0.1212	-0.0042	0.1507	-0.0081	0.1225	-0.0048	0.1460	-0.0087	0.1522
	(-2.17)		*(-1.03)*		*(-1.87)*		*(-1.16)*		*(-2.06)*	
Skewness	-0.0097	0.1219	-0.0058	0.1473	-0.0114	0.1274	-0.0079	0.1066	-0.0035	0.1179
	(-2.60)		*(-1.63)*		*(-2.95)*		*(-2.04)*		*(-0.92)*	
Semi deviation	-0.0877	0.4279	-0.0760	0.4791	-0.0850	0.4309	-0.0730	0.4162	-0.0634	0.4266
	(-10.81)		*(-10.09)*		*(-10.29)*		*(-9.06)*		*(-7.92)*	
Semi deviation risk-free	-0.0950	0.5031	-0.0844	0.5418	-0.0919	0.5063	-0.0803	0.5006	-0.0700	0.5110
	(-11.68)		*(-11.13)*		*(-11.16)*		*(-10.06)*		*(-8.80)*	
Sortino ratio	-0.0268	0.4807	0.0307	0.5191	0.0277	0.4811	0.0237	0.4926	0.0240	0.4926
	(4.21)		*(4.95)*		*(4.20)*		*(3.78)*		*(3.78)*	
Sortino ratio risk-free	0.0357	0.4006	0.0367	0.4440	0.0367	0.4012	0.0308	0.4386	0.0304	0.4386
	(5.29)		*(5.66)*		*(5.25)*		*(4.74)*		*(4.65)*	
Standard deviation	-0.0951	0.5081	-0.0842	0.5474	-0.0923	0.5106	-0.0806	0.5014	-0.0698	0.5126
	(-11.64)		*(-11.08)*		*(-11.13)*		*(-9.97)*		*(-8.67)*	

VaR and ES are significant at 99% but are not significant when we want to account for multiple hypothesis testing. This is something that comes out after many of the factors are false positive.

Many models exhibit relatively strong significance in explaining the mean return of the zero-cost portfolios. The q-five factor model stands out as the best explaining model with its smallest absolute alpha. The increase in *Adjusted R^2* through Fama-French three, Carhart four, and Fama-French five-factor models seem logical, where every newly added factor increases the credibility of the models and in explaining the alpha. Therefore, it is reasonable to take a cautious look on why *Adjusted R^2* of the q-four and q-five factor models are smaller than those of the Carhart four and Fama-French five-factor models. It appears that the set of factors from the q-factor models give rise to minor differences in the *Adjusted R^2*, compared to the Carhart four and Fama-French five-factor models. That is due to the idea that the Carhart four and Fama-French five-factor models go better together with the downside risk measures that we form our zero-cost portfolios from. In Table 4, the remaining measures give grounds for the second group of the split.[7] This group consists of deviating characteristics.

5.3 Other findings

5.3.1 Downside beta (EW)

The alpha presented by the downside beta using the Fama-French five-factor model deviates from the previous findings from Section 5.2, by having the biggest absolute value. T-statistics among the zero-cost portfolios are small (in absolute values) when compared to Section 5.2, which raises thoughts for underlying reasons. The reason could be that the Fama-French five-factor model is better at pricing the assets. Lower t-statistics would probably not survive the multiple hypothesis testing.

5.3.2 Skewness (EW)

For skewness, alpha of the q-five factor model is closest to zero. The significance is reflected in the t-statistics, which show small absolute values, where the t-statistic from the Carhart four-factor model shows that its alpha is statistically significant. The variation in the returns of skewness-formed zero-cost portfolios are, to a certain extent, explained by their small absolute *Adjusted R^2*.

5.3.3 Kurtosis (EW)

In Table 4, kurtosis-formed zero-cost portfolios constructed by the Fama-French five-factor model show the smallest, absolute values. While this is current, the most significant alpha is

[7] The second group consists of downside beta, kurtosis, skewness, Sortino ratio and Sortino ratio with respect to the risk-free rate.

the one of the Fama-French three-factor model (t-statistics in Table 4). Further, significantly lower *Adjusted* R^2 are reflected in the returns of the kurtosis-formed portfolios, compared to those from Section 5.2. This gives rise to uncertainty regarding the intercepts. The conclusion that could be drawn is that these values show that there is low explanatory power in these alpha factors. That can be reflected in the small absolute value of the t-statistic from the zero-cost portfolio formed by kurtosis.

Kurtosis and skewness fit well into a theoretical framework, but when put into a practical perspective, they seem to lose explanation power regarding the unexplained variables of asset pricing models. The alphas do not have statistical explanatory power when attempting to see how much of the mean return that is explained by priced risk measures like skewness and kurtosis.

5.3.4 Sortino ratio (EW)

The benchmark return used in this measure is the mean return of the portfolio. The q-four factor model performs well when explaining some of the variation in its returns, although alpha is not zero. When it comes to the significance of t-statistics, they are smaller than those in Section 5.2, in absolute values. The alpha of the most significant t-statistic is the one from the Fama-French five-factor model. These findings give rise to a contradicting linkage when compared to Section 5.2.

5.3.5 Sortino ratio with respect to the risk-free rate (EW)

The zero-cost portfolios formed from the Sortino ratio with respect to the risk-free rate also present deviating findings in the results. The q-five factor model manages to explain some of the variation in the returns, which is reflected in its alpha, experiencing the smallest absolute value. Like the Sortino ratio with respect to the portfolio mean, the different details stress a linkage that are not similar to the linkages found in Section 5.2. The contradicting linkage that Sortino ratio and Sortino ratio with respect to the risk-free rate seem to support, may reflect the distinguishing characteristic that Sortino ratio possesses in coherence with U.S. stocks.

5.3.6 Miscellaneous comments

There are some details in Table 4 that could further be stressed. As we increase the confidence level of ES and VaR, alphas come closer to zero; the higher the confidence level, the more the accuracy increases. The Sortino ratio with respect to the risk-free rate seems to be the only measure to experience positive alphas.

5.4 Equally-weighted vs. value-weighted

There are some differences between the equally-weighted zero-cost portfolios and the value-weighted zero-cost portfolios. In Table 5, alphas of the risk measures discussed in Section 5.2, follow an opposite trend, i.e., as we add more factors to the regressions, alphas become larger in absolute value. The alphas are backed up by different t-statistics which are not similar to the t-statistics previously mentioned (compare t-statistics in Table 4 and 5). A similar trend can be reflected for the zero-cost portfolios formed from downside beta (compare downside beta in Table 4 and Table 5). Further comments of Table 5 are that the kurtosis and skewness have similar characteristics compared to the findings in Section 5.3. They are surrounded by very small *Adjusted* R^2 and irregular t-statistics like small absolute values (compared to those in Section 5.3). The low *Adjusted* R^2 show that the factors in the models have low correlation with other factors. Similar trends can be reflected for the zero-cost portfolios formed from the Sortino ratio (with respect to the mean return) and Sortino ratio with respect to the risk-free rate (see Table 5).

The big absolute values of the alphas of risk measures equivalent to those in Section 5.2 can be indications of value-weighted stock portfolios not cooperating well with asset pricing models. It shows disturbance in the returns when adding new irrelevant factors that do not co-work with the volume factor. Perhaps a model designed with a volume factor should be tested to see if the alphas show better significance. The few similarities with the outliers can reflect value-weighted stock portfolios working slightly better than the ones in Section 5.2. The value-weighted zero-cost portfolios are considered not to be convenient for our results. It seems that downside risk is priced mostly for smaller companies.

Table 5: Value weighted returns using HAC; numbers in brackets denominate *Alpha t-statistics*

	***Alpha* FF3 [%]**	***Adj.* R^2 FF3**	***Alpha* FF5 [%]**	***Adj.* R^2 FF5**	***Alpha* Carhart [%]**	***Adj.* R^2 Carhart**	***Alpha* q-four [%]**	***Adj.* R^2 q-four**	***Alpha* q-five [%]**	***Adj.* R^2 q-five**
ES at 90.0%	-0.0707	0.5242	-0.0978	0.6374	-0.0743	0.5266	-0.0988	0.5460	-0.1130	0.5571
	(-8.55)		*(-13.85)*		*(-8.76)*		*(-12.15)*		*(-13.84)*	
ES at 95.0%	-0.0686	0.5389	-0.0960	0.6638	-0.0736	0.5440	-0.0995	0.5851	-0.1132	0.5964
	(-8.57)		*(-14.31)*		*(-9.04)*		*(-13.08)*		*(-14.95)*	
ES at 99.0%	-0.0652	0.5000	-0.0962	0.6522	-0.0708	0.5063	-0.0968	0.5494	-0.1097	0.5591
	(-8.17)		*(-13.94)*		*(-8.61)*		*(-12.31)*		*(-13.77)*	
VaR at 90.0%	-0.0654	0.5255	-0.0924	0.6359	-0.0695	0.5284	-0.0920	0.5219	-0.1031	0.5281
	(-7.56)		*(-12.37)*		*(-7.85)*		*(-10.55)*		*(-11.82)*	
VaR at 95.0%	-0.0710	0.5528	-0.0993	0.6799	-0.0768	0.5591	-0.1017	0.5838	-0.1138	0.5920
	(-8.60)		*(-14.39)*		*(-9.07)*		*(-12.66)*		*(-14.20)*	
VaR at 99.0%	-0.0640	0.5358	-0.0904	0.6586	-0.0683	0.5396	-0.0926	0.5794	-0.1072	0.5929
	(-8.24)		*(-13.73)*		*(-8.52)*		*(-12.43)*		*(-14.46)*	
Downside beta	-0.0078	0.4198	-0.0125	0.4271	-0.0081	0.4198	-0.0152	0.4103	-0.0100	0.4131
	(-1.23)		*(-2.00)*		*(-1.25)*		*(-2.36)*		*(-1.51)*	
Kurtosis	0.0117	0.2045	-0.0011	0.2789	0.0075	0.2144	-0.0002	0.2633	-0.0022	0.2640
	(2.16)		*(-0.21)*		*(1.37)*		*(-0.04)*		*(-0.41)*	
Skewness	-0.0033	0.1491	0.0010	0.1629	-0.0032	0.1490	0.0045	0.1537	0.0083	0.1569
	(-0.67)		*(0.19)*		*(-0.63)*		*(0.88)*		*(1.63)*	
Semi deviation	0.0785	0.5630	0.1038	0.6890	0.0816	0.5652	0.1054	0.5876	0.1153	0.5943
	(10.48)		*(17.04)*		*(10.55)*		*(14.58)*		*(15.98)*	
Semi deviation risk-free	0.0738	0.5431	0.1030	0.6800	0.0778	0.5463	0.1044	0.5768	0.1177	0.5868
	(9.17)		*(15.28)*		*(9.44)*		*(13.40)*		*(14.99)*	
Sortino ratio	0.0155	0.5289	0.0256	0.6047	0.0137	0.5300	0.0109	0.5178	0.0083	0.5186
	(2.65)		*(4.65)*		*(2.32)*		*(1.84)*		*(1.36)*	
Sortino ratio risk-free	0.0069	0.4995	0.0101	0.5547	0.0041	0.5026	-0.0040	0.5181	-0.0071	0.5192
	(1.13)		*(1.78)*		*(0.67)*		*(-0.70)*		*(-1.22)*	
Standard deviation	0.0683	0.5446	0.0989	0.6873	0.0719	0.5469	0.0992	0.5704	0.1130	0.5806
	(8.13)		*(14.45)*		*(8.32)*		*(12.11)*		*(13.79)*	

6. Conclusion and discussion

This paper aims to emphasize further knowledge and research about downside risk as an academic subject but also to uplift its use in practical, financial contexts. The data output is used to construct zero-cost portfolios. We run 140 time-series regressions where factors from Fama-French three, Fama-French five, Carhart four, q-four and q-five factor models are used. We find that the alphas for our regressions are significant, and the results using data from the q-factor models possess the most explanatory power.

The reason behind the differences in equally-weighted portfolios and value-weighted portfolios could be subject of further studies. The results of the paper open up for further improvements. Other theories can be considered to develop the paper. For instance, a relatable subject that can be linked to this paper is the *Extreme Value Theory*. This branch of tail risk would provide a theory for a closer look into extreme probability outcomes that could occur. Our findings could also be linked to market bubble literature. Another improvement could be to run cross-sectional regressions, *Fama-MacBeth regressions*. These cross-sectional regressions estimate the betas and risk premiums for any risk factors that are expected to determine asset returns. Multiple assets across time can be regressed. This has beneficial implications to analyze the value of the correlation created in the cross-sections of the multiple assets and to see how much of the asset returns are priced by the downside risk measures.

Lastly, another improvement that could be done regarding this paper is to run a *Gibbons-Ross-Shanken* (GRS) test. The GRS test is used to study the alpha intercepts from the regressions of our asset pricing models, leading us into a comparison of the models' explanatory power. In this case, the test enables to test the hypothesis that all alphas would be jointly equal to zero. Reason for this statistical test is to diversify the paper further and to give a broader view of downside risk measures and to convince the reader about the weight of using downside risk measures in practice. The mentioned improvements are interesting to follow and see the outcomes of, but we leave that work for future researches.

References

Acharya, Viral; Pedersen, Lasse; Philippon, Thomas; Richardson, Matthew (2016). "Measuring Systemic Risk". In: *The Review of Financial Studies*, Vol. 30, No. 1, pp. 2-47. DOI: https://academic-oup-com.ezproxy.ub.gu.se/rfs/article/30/1/2/2682977

Ang, Andrew; Chen, Joseph; Xing, Yuhang (2006). "Downside Risk". In: *The Review of Financial Studies*, Vol. 19, No. 4, pp. 1191-1239. DOI: https://www-jstor-org.ezproxy.ub.gu.se/stable/4123472?seq=1#metadata_info_tab_contents.

Campbell, Harvey; Liu, Yan; Heqing, Zhu (2016). "... and the Cross-Section of Expected Returns". In: *The Review of Financial Studies*, Vol. 29, No. 1, pp. 5-68. DOI: https://academic.oup.com/rfs/article/29/1/5/1843824

Carhart, Mark M. (1997). "On Persistence in Mutual Fund Performance". In: *The Journal of Finance*, Vol. 52, No. 1, pp. 57-82. DOI: https://www-jstor-org.ezproxy.ub.gu.se/stable/2329556?seq=1#metadata_info_tab_contents.

Doane, David P.; Seward, Lori E. (2011). "Measuring Skewness: A Forgotten Statistic?". In: *Journal of Statistics Education*, Vol. 19, No. 2, pp. 1-18. DOI: https://www.tandfonline.com/doi/abs/10.1080/10691898.2011.11889611.

Estrada, Javier (2006). "Downside Risk in Practice". In: *Journal of Applied Corporate Finance - A Morgan Stanley Publication*, Vol. 18, No. 1, pp. 117-125. DOI: http://sfxeu11.hosted.exlibrisgroup.com/sfxgub?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc.

Farago, Adam; Tédongap, Roméo (2018). "Downside risks and the cross-section of asset returns". In: *Journal of Financial Economics*, Vol. 129, No. 1, pp. 69-86. DOI: https://www-sciencedirect-com.ezproxy.ub.gu.se/science/article/pii/S0304405X18300813.

Fama, Eugene F.; French, Kenneth R. (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of Financial Economics*, Vol. 33, No. 1, pp. 3-56. DOI: https://www-sciencedirect-com.ezproxy.ub.gu.se/science/article/pii/0304405X93900235.

Fama, Eugene F.; French, Kenneth R. (2015). “A five-factor asset pricing model”. In: *Journal of Financial Economics*, Vol. 116, No. 1, pp. 1-22. DOI: https://www-sciencedirect-com.ezproxy.ub.gu.se/science/article/pii/S0304405X14002323.

Hou, Kewei; Xue, Chen; Zhang, Lu (2015). “Digesting Anomalies: An Investment Approach”. In: *The Review of Financial Studies*, Vol. 28, No.3, pp. 650-705. DOI: https://www-jstor-org.ezproxy.ub.gu.se/stable/24465724?seq=1#metadata_info_tab_contents.

Hou, Kewei; Mo, Haitao; Xue, Chen; Zhang, Lu (2018). “q^5”. In: *National Bureau of Economic Research*, NBER Working Paper 24709. DOI: https://www-nber-org.ezproxy.ub.gu.se/papers/w24709.

Hull, John C. (2015). “Risk Management and Financial Institutions”. In: *John Wiley & Sons, Inc.*, Fourth Edition. DOI: https://ebookcentral-proquest-com.ezproxy.ub.gu.se/lib/gu/detail.action?docID=1895807.

Novy-Marx, Robert (2013). “The other side of value: The gross profitability premium”. In: *Journal of Financial Economics*, Vol. 108.1, No. 1, pp. 1-28. DOI: https://www-sciencedirect-com.ezproxy.ub.gu.se/science/article/pii/S0304405X13000044.

Pearson, Karl (1905). “Skew variation, a rejoinder.”. In: *Biometrika*, Vol. 4, No.1, pp. 169-212.

Sharpe, William F. (1966). "Mutual Fund Performance". In: *Journal of Business*, Vol. 39, No. 1, pp. 119–138. DOI: https://www-jstor-org.ezproxy.ub.gu.se/stable/2351741#metadata_info_tab_contents.

Titman, Sheridan; Wei, K. C. John; Xie, Feixue (2004). “Capital Investments and Stock Returns”. In: *Journal of Financial and Quantitative Analysis*, Vol. 39, No. 4, pp. 677-700. DOI: https://www-jstor-org.ezproxy.ub.gu.se/stable/30031881?seq=1#metadata_info_tab_contents.

Appendix

Table A1: Equally-weighted returns using robust errors; numbers in brackets denominate *Alpha t-statistics*

	Alpha FF3 (%)	R^2 **FF3**	**Alpha FF5 (%)**	R^2 **FF5**	**Alpha Carhart (%)**	R^2 **Carhart**	**Alpha q-four (%)**	R^2 **q-four**	**Alpha q-five (%)**	R^2 **q-five**
ES at 90.0%	0.0948	0.5035	0.0853	0.5364	0.092	0.5063	0.0813	0.4979	0.0703	0.5100
	(16.71)		*(15.58)*		*(15.96)*		*(14.12)*		*(12.22)*	
ES at 95.0%	0.0932	0.4957	0.0831	0.5313	0.0902	0.4989	0.0789	0.4924	0.0682	0.5043
	(16.56)		*(15.34)*		*(15.78)*		*(13.85)*		*(11.97)*	
ES at 99.0%	0.0893	0.4985	0.0786	0.539	0.0857	0.5032	0.0741	0.5003	0.0645	0.5103
	(16.24)		*(14.92)*		*(15.38)*		*(13.40)*		*(11.63)*	
VaR at 90.0%	0.0929	0.5168	0.0836	0.551	0.0902	0.5192	0.0796	0.5102	0.0687	0.5217
	(16.17)		*(15.11)*		*(15.46)*		*(13.65)*		*(11.76)*	
VaR at 95.0%	0.0919	0.5068	0.0825	0.5394	0.0889	0.5098	0.0781	0.5021	0.0672	0.5141
	(16.22)		*(15.09)*		*(15.47)*		*(13.60)*		*(11.70)*	
VaR at 99.0%	0.0904	0.4844	0.0806	0.5189	0.0872	0.4881	0.0764	0.4833	0.0657	0.4957
	(16.23)		*(15.02)*		*(15.43)*		*(13.58)*		*(11.70)*	
DS beta	-0.0198	0.6879	-0.0208	0.69	-0.0183	0.6888	-0.0176	0.6776	-0.0119	0.6814
	(-4.67)		*(-4.93)*		*(-4.26)*		*(-4.07)*		*(-2.72)*	
Kurtosis	-0.0091	0.1214	-0.0042	0.1511	-0.0081	0.1228	-0.0048	0.1462	-0.0087	0.1525
	(-2.44)		*(-1.14)*		*(-2.14)*		*(-1.30)*		*(-2.32)*	
Skew.	-0.0097	0.1221	-0.0058	0.1477	-0.0114	0.1277	-0.0079	0.1069	-0.0035	0.1182
	(-3.12)		*(-1.87)*		*(-3.59)*		*(-2.46)*		*(-1.09)*	
Semidev.	-0.0877	0.428	-0.076	0.4793	-0.0849	0.4311	-0.073	0.4163	-0.0634	0.4268
	(-15.44)		*(-14.04)*		*(-14.69)*		*(-12.58)*		*(-10.90)*	
Semidev. risk-free	-0.095	0.5032	-0.0844	0.5419	-0.0919	0.5065	-0.0803	0.5008	-0.07	0.5112
	(-16.56)		*(-15.34)*		*(-15.78)*		*(-13.84)*		*(-12.05)*	
Sortino	0.0268	0.4809	0.0307	0.5192	0.0277	0.4813	0.0237	0.4928	0.024	0.4928
	(6.05)		*(7.14)*		*(6.10)*		*(5.31)*		*(5.29)*	
Sortino risk-free	0.0357	0.4007	0.0367	0.4442	0.0367	0.4013	0.0308	0.4388	0.0304	0.4388
	(7.71)		*(8.20)*		*(7.74)*		*(6.73)*		*(6.56)*	
Standard deviation	-0.0951	0.5082	-0.0842	0.5476	-0.0923	0.5108	-0.0806	0.5016	-0.0698	0.5128
	(-16.44)		*(-15.20)*		*(-15.70)*		*(-13.70)*		*(-11.86)*	

Table A2: Value-weighted returns using robust errors; numbers in brackets denominate *Alpha t-statistics*

	Alpha FF3 (%)	R^2 **FF3**	**Alpha FF5 (%)**	R^2 **FF5**	**Alpha Carhart (%)**	R^2 **Carhart**	**Alpha q-four (%)**	R^2 **q-four**	**Alpha q-five (%)**	R^2 **q-five**
ES at 90.0%	-0.0707	0.5243	-0.0978	0.6376	-0.0743	0.5267	-0.0988	0.5461	-0.1131	0.5573
	(-9.43)		*(-14.80)*		*(-9.79)*		*(-13.30)*		*(-15.32)*	
ES at 95.0%	-0.0686	0.5390	-0.0960	0.6639	-0.0736	0.5442	-0.0994	0.5853	-0.1132	0.5966
	(-9.65)		*(-15.67)*		*(-10.26)*		*(-14.55)*		*(-16.66)*	
ES at 99.0%	-0.0652	0.5001	-0.0962	0.6524	-0.0708	0.50655	-0.0968	0.5495	-0.1097	0.5593
	(-8.74)		*(-15.25(*		*(-9.36)*		*(-13.40)*		*(-15.19)*	
VaR at 90.0%	-0.0654	0.5256	-0.0924	0.6360	-0.0695	0.5285	-0.0920	0.5221	-0.1031	0.5283
	(-8.37)		*(-13.33)*		*(-8.81)*		*(-11.49)*		*(-12.87)*	
VaR at 95.0%	-0.0710	0.5529	-0.0993	0.6800	-0.0768	0.5592	-0.1017	0.5839	-0.1138	0.5922
	(-9.89)		*(-16.16)*		*(-10.59)*		*(-14.42)*		*(-16.20)*	
VaR at 99.0%	-0.06404	0.5359	-0.0904	0.6587	-0.0683	0.5398	-0.0926	0.5796	-0.1072	0.5931
	(-9.25)		*(-15.15)*		*(-9.74)*		*(-13.91)*		*(-16.24)*	
DS beta	-0.0078	0.4199	-0.0124	0.4274	-0.0081	0.4200	-0.0152	0.4104	-0.0100	0.4134
	(-1.30)		*(-2.10)*		*(-1.33)*		*(-2.50)*		*(-1.63)*	
Kurtosis	0.01167	0.2047	-0.0011	0.2792	0.0075	0.2146	-0.0002	0.2636	-0.0022	0.2643
	(2.08)		*(-0.20)*		*(1.33)*		*(-0.04)*		*(-0.40)*	
Skew.	-0.0033	0.1493	0.0009	0.1532	-0.0031	0.1493	0.0045	0.1539	0.0083	0.1572
	(-0.66)		*(0.19)*		*(-0.63)*		*(0.89)*		*(1.64)*	
Semidev.	0.0785	0.5631	0.1038	0.6891	0.0816	0.5654	0.1054	0.5877	0.1153	0.5944
	(12.19)		*(18.97)*		*(12.49)*		*(16.63)*		*(18.15)*	
Semidev. risk-free	0.0738	0.5432	0.1030	0.6801	0.0778	0.5464	0.1044	0.5769	0.1177	0.5870
	(10.22)		*(16.90)*		*(10.67)*		*(14.76)*		*(16.71)*	
Sortino	0.0155	0.5290	0.0256	0.6048	0.0137	0.5302	0.0109	0.5179	0.0083	0.5188
	(2.97)		*(5.31)*		*(2.57)*		*(2.04)*		*(1.51)*	
Sortino risk-free	0.0069	0.4996	0.0101	0.5549	0.0041	0.5027	-0.0040	0.5182	-0.0071	0.5194
	(1.31)		*(2.03)*		*(0.77)*		*(-0.78)*		*(-1.36)*	
Standard dev.	0.0683	0.5447	0.0987	0.6874	0.0719	0.5471	0.0992	0.5706	0.1129	0.5808
	(9.20)		*(15.84)*		*(9.52)*		*(13.45)*		*(15.33)*	

A3: Data processing and zero-cost portfolio code

Downside risk

Import packages

In [1]:
```
import pandas as pd
pd.set_option('display.max_columns', 50)
import numpy as np
import pickle
import gc
from tqdm import tqdm
from scipy import stats
from collections import defaultdict
from matplotlib import pyplot as plt
```

Read file

In [60]:
```
dataset=pd.read_csv(
    '31a2de501f1a06e2.csv.gz',
    compression='gzip',
    parse_dates=['date'],
    dtype={
        'PERMNO': 'Int32',
        'SICCD': 'str',
        'COMNAM': 'str',
        'SHRCLS': 'str',
        'PRIMEXCH': 'str',
        'PERMCO': 'Int32',
        'PRC': np.float64,
        'SHROUT': np.float64,
        'CFACPR': np.float64,
        'CFASHR': np.float64
    },
    #Don't load columns that we do not use
    usecols=[
        'PERMNO',
        'date',
        'COMNAM',
        'PERMCO',
        'PRC',
        'SHROUT',
        'CFACPR',
        'CFACSHR'
    ],
)
```

Remove ETFs and mutual funds

In [3]:
```
def logical_ors(x):
    if len(x)>2:
        return np.logical_or(x[0], logical_ors(x[1:]))
    else:
        return np.logical_or(x[0], x[1])

fund_ix=[dataset['COMNAM'].apply(lambda y: x in str(y)) for x in [
    'TRUST',
    'E T F',
    'EXCH TRD',
    ' FUNDS',
    'EXCHANGE TRADED',
    'ETF',
    'ISHARES',
]]
fund_ix=logical_ors(fund_ix)
fund_ix=dataset.PERMCO.isin(set(dataset.loc[fund_ix,'PERMCO']))
dataset=dataset[~fund_ix]
```

In [4]:
```
dataset=dataset[~dataset['PRC'].apply(np.isnan)]
```

In [5]:
```
dataset["SHROUT_ADJ"]=dataset["SHROUT"]*dataset["CFACSHR"]
dataset["PRC_ADJ"]=dataset["PRC"].apply(np.fabs)/dataset["CFACPR"]
dataset['BidAsk']=dataset['PRC']<0
dataset['MktCap']=dataset["SHROUT_ADJ"]*dataset["PRC_ADJ"]
```

In [6]:
```
for c in ['SHRCLS', 'SICCD', 'TICKER', 'PRIMEXCH', 'PRC', 'SHROUT', 'CFACPR', 'CFACSHR']:
    try:
        del dataset[c]
    except:
        continue
gc.collect()
```

Out[6]: 48

Multiple share classes

In [8]:
```python
#Slower
# multiple_share_classes={}
# for pm in tqdm(set(dataset['PERMCO'])):
#     x=set(dataset.loc[dataset.PERMNO==pm,'PERMNO'])
#     if len(x)>1:
#         multiple_share_classes[pm]=list(x)

#Faster
multiple_share_classes=dataset[['PERMCO','PERMNO']].groupby('PERMCO').agg(set)
ix=multiple_share_classes['PERMNO'].apply(len)>1
multiple_share_classes=multiple_share_classes.loc[ix, 'PERMNO'].to_dict()
```

In [9]:
```python
all_permnos_to_remove=set()
for pm in multiple_share_classes.keys():
    ix=dataset.PERMCO==pm
    dft=dataset.loc[ix,['date','PRC_ADJ','MktCap','PERMNO','BidAsk']]
    permnos=set(dft.PERMNO)
    dft=pd.pivot_table(dft, index=['date'], values=['PRC_ADJ', 'MktCap','BidAsk'], columns=['PERMNO'])

    #What percentage of prices are midpoints
    #Remove series with more bid/ask quotes
    criterion_bidask=dft['BidAsk'].sum()/dft['PRC_ADJ'].count()
    threshold=criterion_bidask.min()*1.25
    pms_left=criterion_bidask[criterion_bidask<threshold].index.values.tolist()

    if len(pms_left)>1:
        #Remove shorter price series
        dftc=dft['PRC_ADJ'].copy()
        dftc=dftc[pms_left]
        criterion_length=dftc.count()
        threshold=criterion_length.max()*0.75
        pms_left=criterion_length[criterion_length>threshold].index.values.tolist()
        if len(pms_left)>1:
            #Remove smaller market caps
            dftc=dft['MktCap'].copy()
            dftc=dftc[pms_left]
            row_max=dft['MktCap'].max(axis=1)
            for c in dftc.columns:
                dftc[c]=dftc[c]==row_max
            criterion_mktcap=dftc.sum()
            threshold=criterion_mktcap.max()*0.75
            pms_left=criterion_mktcap[criterion_mktcap>threshold].index.values.tolist()
            if len(pms_left)>1:
                #Assume these are increasing with time and pick first
                pms_left=[np.min(pms_left)]
    permnos_to_remove=permnos-set(pms_left)
    all_permnos_to_remove = all_permnos_to_remove | permnos_to_remove
dataset=dataset[~dataset.PERMNO.isin(all_permnos_to_remove)]
```

In [10]:
```python
#Slower
# multiple_share_classes={}
# for pm in tqdm(set(dataset['PERMCO'])):
#     x=set(dataset.loc[dataset.PERMNO==pm,'PERMNO'])
#     if len(x)>1:
#         multiple_share_classes[pm]=list(x)

#Faster
multiple_share_classes=dataset[['PERMCO','PERMNO']].groupby('PERMCO').agg(set)
ix=multiple_share_classes['PERMNO'].apply(len)>1
multiple_share_classes=multiple_share_classes.loc[ix, 'PERMNO'].to_dict()
```

In [11]:
```python
for k in multiple_share_classes.keys():
    if len(multiple_share_classes[k])>1:
        ix=dataset.PERMCO==k
        print(list(set(dataset.loc[ix,'COMNAM'])))
        print("="*30)
```

In [12]:
```python
for pm in multiple_share_classes.keys():
    ix=dataset.PERMCO==pm
    dft=dataset.loc[ix,['date','PRC_ADJ','MktCap','PERMNO','BidAsk']]
    dft=pd.pivot_table(dft, index=['date'], values=['PRC_ADJ', 'MktCap','BidAsk'], columns=['PERMNO'])
    f=plt.figure(figsize=(20,5))
    ax=plt.subplot(1,3,1)
    dft['PRC_ADJ'].plot(ax=ax, title=dataset.loc[ix,'COMNAM'].values[-1])
    ax=plt.subplot(1,3,2)
    dft['MktCap'].plot(ax=ax, title=str(pm))
    ax=plt.subplot(1,3,3)
    dft['BidAsk'].applymap(lambda x: 1 if x==True else 0).plot(ax=ax, title=str(pm), lw=0.1)
    plt.show()
    plt.close()
```

In [13]:
```python
with open('dataset.pickle','wb') as fo:
    pickle.dump(dataset, fo, -1)
```

First pickle

In [2]:
```
with open('dataset.pickle','rb') as fo:
    dataset=pickle.load(fo)
```

One rebalancing date sandbox to implement sorts

In [5]:
```
rebalancing_dates=[pd.to_datetime("{}0630".format(i)) for i in range(dataset.date.min().year+5, dataset.date.max().
```

In [19]:
```
i_rebalancing_date=len(rebalancing_dates)-1
rebalancing_date=rebalancing_dates[i_rebalancing_date]
dataset_rebalancing=dataset[dataset.date<=rebalancing_date].copy()
dataset_rebalancing=pd.pivot_table(dataset_rebalancing, index='date', columns=['PERMCO'], values=['MktCap','PRC_ADJ
#What can I invest in?
investable_permcos=set(dataset_rebalancing.resample('M').last().iloc[-2:-1,:].T.dropna().T['PRC_ADJ'].columns.value

#Remove short time series
dft=dataset_rebalancing['PRC_ADJ'][investable_permcos].count()
investable_permcos = investable_permcos & set(dft[dft>265*2].index.values.tolist())

#Remove very small caps
dft=dataset_rebalancing['MktCap'].resample('M').last().iloc[-2:-1,:].T.dropna()
threshold=dft.quantile(0.05).values[0]
investable_permcos = investable_permcos & set(dft[dft.iloc[:,0]>threshold].T.columns.values.tolist())

dataset_rebalancing_mktcap=dataset_rebalancing['MktCap'][investable_permcos]
dataset_rebalancing_adjprc=dataset_rebalancing['PRC_ADJ'][investable_permcos]
dataset_rebalancing_returns=100*(dataset_rebalancing_adjprc.apply(np.log)-dataset_rebalancing_adjprc.apply(np.log).

with open('dataset_rebalancing_last_date.pickle','wb') as fo:
    pickle.dump({
        'dataset_rebalancing_mktcap': dataset_rebalancing_mktcap,
        'dataset_rebalancing_adjprc': dataset_rebalancing_adjprc,
        'dataset_rebalancing_returns': dataset_rebalancing_returns,
    }, fo, -1)

```

Second pickle

In [3]:
```
with open('dataset_rebalancing_last_date.pickle','rb') as fo:
    dataset_rebalancing_last_date=pickle.load(fo)
dataset_rebalancing_mktcap=dataset_rebalancing_last_date['dataset_rebalancing_mktcap']
dataset_rebalancing_adjprc=dataset_rebalancing_last_date['dataset_rebalancing_adjprc']
dataset_rebalancing_returns=dataset_rebalancing_last_date['dataset_rebalancing_returns']
```

Computing all returns on portfolio sorts

In [4]:
```
def compute_portfolio_returns(i_rebalancing_date, Q1_stocks, Q10_stocks, all_portfolio_results, name):
    Q1_returns = dataset_rebalancing_returns.loc[rebalancing_dates[i_rebalancing_date-1]:, Q1_stocks].iloc[1:,:]
    Q1_returns_EW=Q1_returns.mean(axis=1)
    weights=dataset_rebalancing_mktcap.loc[rebalancing_dates[i_rebalancing_date-1]:, Q1_stocks].copy().iloc[1:, :]
    totals=dataset_rebalancing_mktcap.loc[rebalancing_dates[i_rebalancing_date-1]:, Q1_stocks].sum(1)
    for c in weights.columns:
        weights.loc[:,c]/=totals
    Q1_returns_VW=(Q1_returns*weights).sum(1)

    Q10_returns = dataset_rebalancing_returns.loc[rebalancing_dates[i_rebalancing_date-1]:, Q10_stocks].iloc[1:,:]
    Q10_returns_EW=Q10_returns.mean(axis=1)
    weights=dataset_rebalancing_mktcap.loc[rebalancing_dates[i_rebalancing_date-1]:, Q10_stocks].copy().iloc[1:, :]
    totals=dataset_rebalancing_mktcap.loc[rebalancing_dates[i_rebalancing_date-1]:, Q10_stocks].sum(1)
    for c in weights.columns:
        weights.loc[:,c]/=totals
    Q10_returns_VW=(Q10_returns*weights).sum(1)

    all_portfolio_results['Q1_{}_EW'.format(name)].append(Q1_returns_EW)
    all_portfolio_results['Q1_{}_VW'.format(name)].append(Q1_returns_VW)
    all_portfolio_results['Q10_{}_EW'.format(name)].append(Q10_returns_EW)
    all_portfolio_results['Q10_{}_VW'.format(name)].append(Q10_returns_VW)
```

In [5]:
```
def get_Qx_stocks(formula,formula_name,q_l,q_h,portfolios):
    Q1_threshold=formula.quantile(q_l)
    Q10_threshold=formula.quantile(q_h)
    Q1_stocks=formula[formula<Q1_threshold].index.values.tolist()
    Q10_stocks=formula[formula>Q10_threshold].index.values.tolist()
    portfolios[formula_name]={'Q1':Q1_stocks,'Q10':Q10_stocks}
```

In [6]:
```
def get_ES(dataset_rebalancing_returns,q):
    threshold=dataset_rebalancing_returns.quantile(q)
    ix=dataset_rebalancing_returns>threshold
    dft=dataset_rebalancing_returns.copy()
    dft[ix]=np.nan
    return dft.sum()/dft.count()
```

In [7]:

```python
def get_semdev():
    m=(dataset_merged[['Mkt-RF','RF']].sum(1))
    dft=dataset_rebalancing_returns.add(-m,axis=0)
    ix=dataset_rebalancing_returns>0
    dft[ix]=0.0
    return np.sqrt(dft.pow(2).mean())
def get_semdev_rf():
    m=(dataset_merged[['RF']].sum(1))
    dft=dataset_rebalancing_returns.add(-m,axis=0)
    ix=dataset_rebalancing_returns>0
    dft[ix]=0.0
    return np.sqrt(dft.pow(2).mean())
```

In [8]:

```python
dataset_2 = pd.read_csv(
    'F-F_Research_Data_5_Factors_2x3_daily.CSV',
    parse_dates=['date'],
        dtype={
        'Mkt-RF': np.float64,
        'SMB': np.float64,
        'HML' : np.float64,
        'RMW' : np.float64,
        'CMA' : np.float64,
        'RF' : np.float64
    },
    usecols=[
        'date',
        'Mkt-RF',
        'RF'
    ],
)
```

In [9]:

```python
def get_dsbeta():
    port_m=dataset_merged[dataset_merged.keys()[:-2]].mean()
    port_dft=dataset_merged[dataset_merged.keys()[:-2]].copy()
    port_dft-=port_m
    port_ix=port_dft>0
    port_dft[port_ix]=0.0

    mkt_m=(dataset_merged['Mkt-RF'] + dataset_merged['RF']).mean()
    mkt_dft=(dataset_merged['Mkt-RF'] + dataset_merged['RF']).copy()
    mkt_ix=mkt_dft>0
    mkt_dft[mkt_ix]=0.0

    return port_dft.multiply(mkt_dft, axis=0).sum()/mkt_dft.pow(2).sum()
```

In [10]:

```python
def get_sortino():
    m=(dataset_merged[['Mkt-RF','RF']].sum(1)).mean()
    port_m=dataset_merged[dataset_merged.keys()[:-2]].mean()
    semdev_m=get_semdev()
    return (port_m-m).div(semdev_m)

def get_sortino_rf():
    m=(dataset_merged[['RF']].sum(1)).mean()
    port_m=dataset_merged[dataset_merged.keys()[:-2]].mean()
    semdev_m=get_semdev_rf()
    return (port_m-m).div(semdev_m)
```

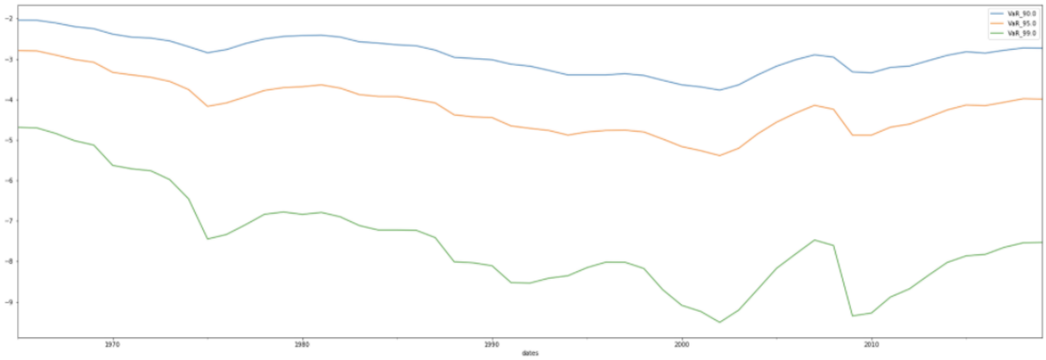
In [ ]:

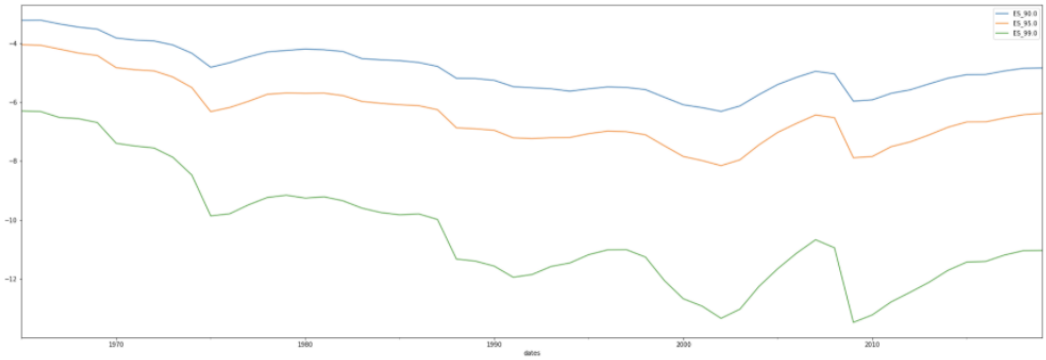
```
all_portfolio_results = defaultdict(list)
portfolios={}
risk_measures=defaultdict(list)
q_l,q_h=0.1,0.9
rebalancing_dates=[pd.to_datetime("{}0630".format(i)) for i in range(dataset.date.min().year+5, dataset.date.max().
for i_rebalancing_date in tqdm(range(len(rebalancing_dates))):
    rebalancing_date=rebalancing_dates[i_rebalancing_date]
    dataset_rebalancing=dataset[dataset.date<=rebalancing_date].copy()
    dataset_rebalancing=pd.pivot_table(dataset_rebalancing, index='date', columns=['PERMCO'], values=['MktCap','PRC


    if i_rebalancing_date!=0:
        dataset_rebalancing_mktcap=dataset_rebalancing['MktCap'][investable_permcos]
        dataset_rebalancing_adjprc=dataset_rebalancing['PRC_ADJ'][investable_permcos]
        dataset_rebalancing_returns=100*(dataset_rebalancing_adjprc.apply(np.log)-dataset_rebalancing_adjprc.apply(

        for ss in portfolios.keys():
            compute_portfolio_returns(i_rebalancing_date, portfolios[ss]['Q1'], portfolios[ss]['Q10'], all_portfoli

    #What can I invest in?
    investable_permcos=set(dataset_rebalancing.resample('M').last().iloc[-2:-1,:].T.dropna().T['PRC_ADJ'].columns.v

    #Remove short time series
    dft=dataset_rebalancing['PRC_ADJ'][investable_permcos].count()
    investable_permcos = investable_permcos & set(dft[dft>265*2].index.values.tolist())

    #Remove very small caps
    dft=dataset_rebalancing['MktCap'].resample('M').last().iloc[-2:-1,:].T.dropna()
    threshold=dft.quantile(0.05).values[0]
    investable_permcos = investable_permcos & set(dft[dft.iloc[:,0]>threshold].T.columns.values.tolist())

    dataset_rebalancing_mktcap=dataset_rebalancing['MktCap'][investable_permcos]
    dataset_rebalancing_adjprc=dataset_rebalancing['PRC_ADJ'][investable_permcos]
    dataset_rebalancing_returns=100*(dataset_rebalancing_adjprc.apply(np.log)-dataset_rebalancing_adjprc.apply(np.l

    dataset_merged = pd.merge(dataset_rebalancing_returns,dataset_2, left_index=True, right_on='date', how='inner')
    dataset_merged = dataset_merged.set_index('date')

    #Measures of downside risk and stock selection for next period
    #Skewness
    skew=dataset_rebalancing_returns.skew()
    risk_measures['skewness'].append(skew)
    get_Qx_stocks(skew,'skewness',q_l,q_h,portfolios)

    #Kurtosis
    kurt=dataset_rebalancing_returns.kurtosis()
    risk_measures['kurtosis'].append(kurt)
    get_Qx_stocks(kurt,'kurtosis',q_l,q_h,portfolios)

    #VaR
    for qVaR in [0.1,0.05,0.01]:
        VaR = dataset_rebalancing_returns.quantile(qVaR)
        risk_measures['VaR_{}'.format(100*(1-qVaR))].append(VaR)
        get_Qx_stocks(VaR,'VaR_{}'.format(100*(1-qVaR)),q_l,q_h,portfolios)

    #ES
    for qES in [0.1,0.05,0.01]:
        ES = get_ES(dataset_rebalancing_returns,qES)
        risk_measures['ES_{}'.format(100*(1-qES))].append(ES)
        get_Qx_stocks(ES,'ES_{}'.format(100*(1-qES)),q_l,q_h,portfolios)

    #Semi deviation (benchmarket=market)
    semdev = get_semdev()
    risk_measures['semdev'].append(semdev)
    get_Qx_stocks(semdev, 'semdev',q_l,q_h,portfolios)

    #Semi deviation (benchmarket=risk-free)
    semdev_rf = get_semdev_rf()
    risk_measures['semdev_rf'].append(semdev_rf)
    get_Qx_stocks(semdev_rf,'semdev_rf',q_l,q_h,portfolios)

    #Standard deviation
    stdev = dataset_rebalancing_returns.std()
    risk_measures['stdev'].append(stdev)
    get_Qx_stocks(stdev,'stdev',q_l,q_h,portfolios)

    #Downside beta
    dsbeta = get_dsbeta()
    risk_measures['dsbeta'].append(dsbeta)
    get_Qx_stocks(dsbeta,'dsbeta',q_l,q_h,portfolios)

    #Sortino ratio (benchmark=market)
    sortino = get_sortino()
    risk_measures['sortino'].append(sortino)
    get_Qx_stocks(sortino,'sortino',q_l,q_h,portfolios)

    #Sortino ratio (benchmark=risk-free)
    sortino_rf = get_sortino_rf()
    risk_measures['sortino_rf'].append(sortino_rf)
    get_Qx_stocks(sortino_rf,'sortino_rf',q_l,q_h,portfolios)

```

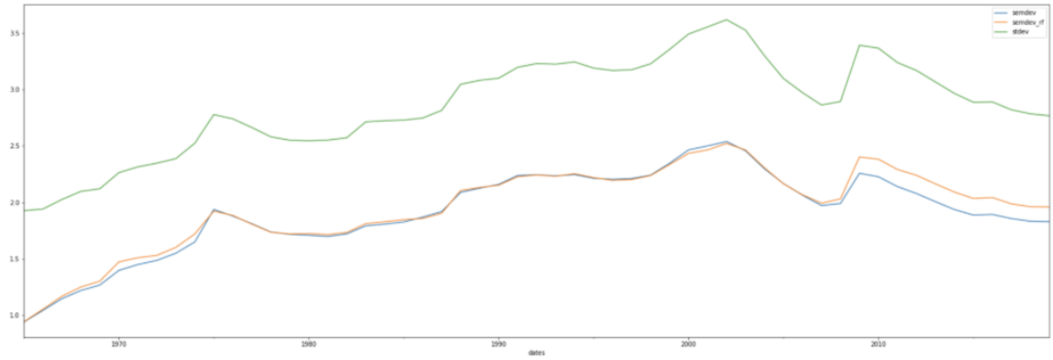
Summary statistics for every risk measure on rebalancing dates ¶

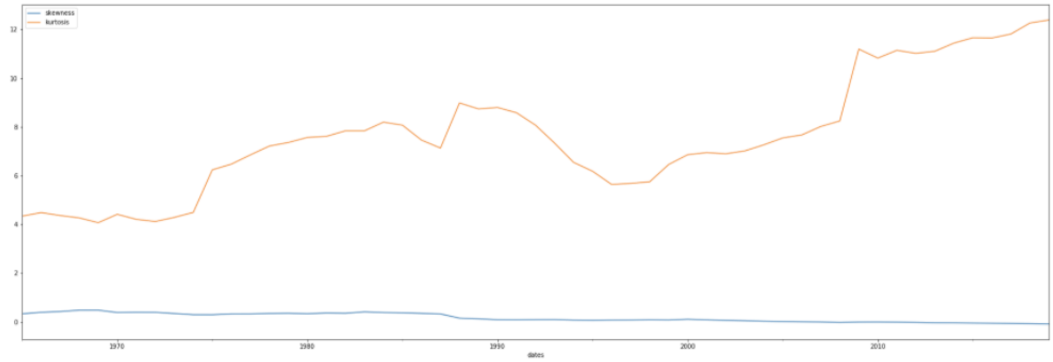
In [57]:

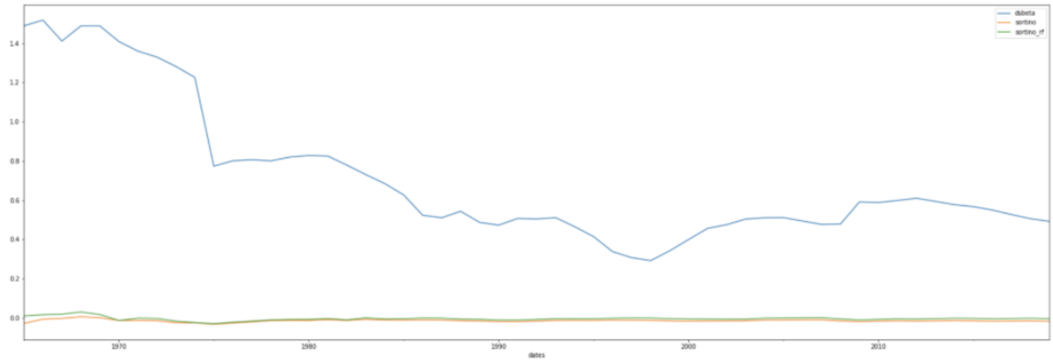
```
f=plt.figure(figsize=(30,10))
ax=plt.subplot(111)
for k in ['VaR_90.0', 'VaR_95.0', 'VaR_99.0']:
    pd.DataFrame({
        k: list(map(lambda x: x.median(), risk_measures[k])),
        'dates':rebalancing_dates}).plot(x='dates', ax=ax)
plt.savefig('figure1.png', dpi=300)

f=plt.figure(figsize=(30,10))
ax=plt.subplot(111)
for k in ['ES_90.0', 'ES_95.0', 'ES_99.0']:
    pd.DataFrame({
        k: list(map(lambda x: x.median(), risk_measures[k])),
        'dates':rebalancing_dates}).plot(x='dates', ax=ax)
plt.savefig('figure2.png', dpi=300)

f=plt.figure(figsize=(30,10))
ax=plt.subplot(111)
for k in ['semdev', 'semdev_rf', 'stdev']:
    pd.DataFrame({
        k: list(map(lambda x: x.median(), risk_measures[k])),
        'dates':rebalancing_dates}).plot(x='dates', ax=ax)
plt.savefig('figure3.png', dpi=300)

f=plt.figure(figsize=(30,10))
ax=plt.subplot(111)
for k in ['skewness', 'kurtosis']:
    pd.DataFrame({
        k: list(map(lambda x: x.median(), risk_measures[k])),
        'dates':rebalancing_dates}).plot(x='dates', ax=ax)
plt.savefig('figure4.png', dpi=300)

f=plt.figure(figsize=(30,10))
ax=plt.subplot(111)
for k in ['dsbeta', 'sortino', 'sortino_rf']:
    pd.DataFrame({
        k: list(map(lambda x: x.median(), risk_measures[k])),
        'dates':rebalancing_dates}).plot(x='dates', ax=ax)
plt.savefig('figure5.png', dpi=300)
```











In [53]:

```
#Data description
t=pd.DataFrame({k:pd.concat(risk_measures[k]).replace([np.inf, -np.inf], np.nan).dropna().describe() for k in risk_
t.to_clipboard()
t
```

Out[53]:

	count	mean	std	min	25%	50%	75%	max
skewness	237811.0	0.046528	1.348970	-50.070978	-0.187772	0.134628	0.468742	36.063946
kurtosis	237811.0	15.317666	43.300339	-0.467973	4.457748	7.650460	14.149167	3508.523432
VaR_90.0	237819.0	-3.288001	1.795807	-40.546511	-4.255961	-2.932160	-2.017604	0.000000
VaR_95.0	237819.0	-4.796869	2.547318	-69.314718	-6.107556	-4.255961	-2.950882	0.000000
VaR_99.0	237818.0	-8.711159	4.724710	-91.635473	-11.079988	-7.691904	-5.310983	0.000000
ES_90.0	237811.0	-5.614236	3.028679	-58.602310	-7.243731	-5.017260	-3.445141	0.000000
ES_95.0	237811.0	-7.380823	3.918238	-78.790421	-9.448027	-6.549080	-4.520060	0.000000
ES_99.0	237811.0	-12.161343	6.684433	-130.775437	-15.536917	-10.630483	-7.315597	0.000000
semdev	237811.0	2.255855	1.174426	0.208207	1.370247	1.980587	2.866073	20.765519
semdev_rf	237811.0	2.266215	1.197897	0.034211	1.381145	2.013351	2.908684	20.734774
stdev	237811.0	3.275948	1.698266	0.000000	2.024195	2.927029	4.188505	29.764342
dsbeta	237819.0	0.698136	0.546320	0.000000	0.286549	0.577993	0.965655	11.635589
sortino	237811.0	-0.015098	0.030869	-0.354966	-0.031794	-0.014511	-0.000029	0.265521
sortino_rf	237811.0	-0.005244	0.032363	-0.773680	-0.023052	-0.005228	0.010582	0.695814

Export to Stata

In [ ]:

```python
keys=list(all_portfolio_results.keys())
k=keys[0]
df_final_all=pd.DataFrame(pd.concat(all_portfolio_results[k]), columns=[k])
for k in keys[1:]:
    df_final_all=pd.merge(
        df_final_all, pd.DataFrame(pd.concat(all_portfolio_results[k]), columns=[k]),
        left_index=True, right_index=True
    )
#df_final_all
```

In [ ]:

```python
keys=sorted(list(set(map(lambda x: '_'.join(x.split('_')[1:-1]), df_final_all.columns))))
k=keys[0]
df_final=pd.DataFrame(df_final_all['Q10_{}_EW'.format(k)]-df_final_all['Q1_{}_EW'.format(k)], columns=['Q10-Q1_{}_E
for k in keys[1:]:
    df_final=pd.merge(
        df_final, pd.DataFrame(df_final_all['Q10_{}_EW'.format(k)]-df_final_all['Q1_{}_EW'.format(k)], columns=['Q1
        left_index=True, right_index=True
    )
for k in keys:
    df_final=pd.merge(
        df_final, pd.DataFrame(df_final_all['Q10_{}_VW'.format(k)]-df_final_all['Q1_{}_VW'.format(k)], columns=['Q1
        left_index=True, right_index=True
    )

df_final=pd.merge(
    df_final,
    pd.read_csv(
        'F-F_Research_Data_5_Factors_2x3_daily.CSV',
        parse_dates=['date'],
            dtype={
                'Mkt-RF': np.float64,
                'SMB': np.float64,
                'HML' : np.float64,
                'RMW' : np.float64,
                'CMA' : np.float64,
                'RF' : np.float64
        },
    ),
    how='outer', left_index=True, right_on='date'
).set_index('date')

df_final=pd.merge(
    df_final,
    pd.read_csv(
        'F-F_Momentum_Factor_daily.CSV',
        parse_dates=['date'],
            dtype={
                'Mom': np.float64
        },
    ),
    how='outer', left_index=True, right_on='date'
).set_index('date')

df_final=pd.merge(
    df_final,
    pd.read_csv(
        'q5_factors_daily_2019.CSV',
        parse_dates=['date'],
            dtype={
                'R_F': np.float64,
                'R_MKT': np.float64,
                'R_ME': np.float64,
                'R_IA': np.float64,
                'R_ROE': np.float64,
                'R_EG': np.float64
        },
    ),
    how='outer', left_index=True, right_on='date'
).set_index('date')

df_final.dropna().to_csv('df_final.csv')
```

In [56]:

```python
#Visualize zero-cost portfolio summary statistics
df_final_t = df_final.describe().T
df_final_t.to_clipboard()
df_final_t
```