A Study of Implied Volatility Spreads

– and –

Their Relation to Stock Returns

A study examining the effect from implied volatility spread of the option market on the returns of the stock market with portfolio formation, factor estimation, risk premias and ex-post evaluation.

Marie Dyveke Olafsen
201808138

Supervisor: Jonas Nygaard Eriksen
Associate Professor, Aarhus University

Aarhus University, Department of Economics and Business Economics
M.Sc. Finance Master's Thesis

Hand-in date: June 1st 2023
The thesis may be published

Abstract

This projects presents the analysis of the implied volatility spread and its effect upon the stock market. Implied volatility spreads are, in efficient markets, only present between american options and has historically a negative average. This implicates that calls are cheap and puts are expensive, when correcting the option's intrinsic value and signals a belief of the market participants to the future distribution of the stock price.

This I deem a relevant proxy for risk in the stock market. The approach for the analysis of the dependence will be portfolio sorting with an estimated long-short factor. Data is accessed through OptionMetrics, CRSP, and Open Asset Pricing, with the sample spanning from 1996 to 2021 of available observations at a daily granularity of stocks and options traded at American exchanges. I have estimated two relevant signals for the portfolio formation, with the first being the value of the implied volatility spread and the second signal being the recent change in the implied volatility spread. The choices made in the portfolio formation and later analysis is evaluated explicitly, to ensure that the results are robust against data-snooping.

I set out to investigate the relation between the implied volatility spread, as it is a proxy of the mispricing of options after correcting for the option's intrinsic value, and the stock returns. The option market is in general very interesting, as it provides an expectation of the future distribution of the stock price, given probabilities inferred from the prices. I find that the returns are significantly higher for the tail portfolios, and even more so when sorting into several portfolios, the factor is not proxied by other external factors, but it does not demand a significant risk premium when included in the traditional tests. An ex-post economic analysis show a limited inclusion of the portfolios in the optimal portfolio.

Most of my analysis has been based on an entire sample view, but as the option markets have matured significantly over the sample period, it would be interesting to split the analysis into subperiods. These subperiods could be determined by liquidity factors of both markets, or in a chronological order. The signals deployed could be supplemented by variations to look at only the negative values of the signal, or to investigate the absolute value. This would proxy mispricings in general, and not the directional mispricings of my analysis.

A common critique point of these kinds of projects is the applicability for an investor in the markets. None of the results include transaction costs or borrowing constraints, which would severely affect the possible earnings incurred for an investor wishing to be exposed to the risk factors. Furthermore, the mean-variance framework of the economic analysis should be challenged, as the implied volatility spread proxies for a loss-aversion not seen in rational investors.

Contents

1 Introduction

The American Stock market has been the victim of extensive scrutiny from researchers, investors and analysts among others. We have looked at anything affecting the returns and dived deep into all kinds of models or assumptions for the analysis to explain and predict the common variation of the stocks.

Starting from a no-arbitrage assumption, Ross (1976) introduced a factor model, which should comprise all the factors affecting the returns. These factors should signify systematic risk in the markets, as any investor could diversify their portfolios to avoid idiosyncratic risk. These factors were not specified in the original article, and ever since, researchers have tried to identify the factors which demand a risk premium, and those that do not. The factors demanding a risk premium are systematic risk factors, and can be either traded, or non-traded. Traded factors are portfolios formed on the factor, while non-traded factors could be macro-economic factors. The factors that do not affect the stock market or does not demand a risk premium will be denoted quasi factors. Not all of these factors are observable, and most of them seems to be correlated or proxy for the same risk, which provides problems when doing traditional regressions.

The arbitrage theory did not specify how many and of which origin, factors were. So researchers have formed three relevant groups of factors. The first group is the macroeconomic factors, which are not traded, very hard to measure, infrequently updated and in general prone to error in the estimation. The second group consists of statistical factors, found through principal components analysis which maps the space of variation in the returns. And the last group is consisting of the firm-specific characteristics. These could be accounting measures, technical analysis of earlier price movements, or related signals from industries or other financial markets.

I have chosen to dive deeper into the option markets, and how they are related to the stock market. The main point of using option data, is the inherent forward looking property of the option markets, as they are contracts for differences dependent on the future states of the stocks.

This inherent forward looking property of the option market is shown in a theoretical setting. We can estimate the risk-neutral distribution of the stock price given a lot of different options on the stock according to formula (2) in Breeden and Litzenberger (1978) (and elaborated in appendix A of Isakas and Melenberg (2020)). This distribution is under the $\mathbb{Q}$ measure but can be approximated to the $\mathbb{P}$ measure and used for forecasting stock prices. The distribution of the expected future value of the stock price is found by evaluting the twice differentiated call price with respect to the strike at a variety of strikes at each maturity. Figlewski and Webb (1993) deploys this data and estimate the market's risk neutral densities of the stock prices, and they find a predictive power for the stock prices.

This approach has been used extensively for european options on indexes, as their options are far more liquid and indicative of the market's expectations, than the american options on individual stocks. Using the option markets to get an idea of the future stock prices are, however, relevant, even with the lower liquidity. Cremers and Weinbaum (2010) introduces the use of the implied volatility spread, as a proxy for the mispricings of the puts and calls with the same maturity, strike and underlying stock. This implied volatility signifies the relative price of a put and a call after being corrected for the intrinsic value. A positive value of the implied volatility spread indicates a cheap put, or an expensive call. These values are further indicating the market's belief about the future of the stock. If the market increases the demand for call options, that would increase the price of the option, further increasing the implied volatility spread, and indicate that the market expects the stock price to increase in the future.

I deem it relevant to investigate this signal further, as I form two signals from the implied volatility spread. The first being the value, to proxy for the market beliefs of future movements of the stock price, and second being the recent change in the implied volatility spread. The last signal is not chosen to investigate how fast the information from the option market's affect the stock markets, but instead to proxy for the change in beliefs.

To analyse this, I choose portfolio formation to reduce the idiosyncratic noise, and form long-short factors based on the portfolios. The choice of portfolios relies on the absence of a priori assumptions. This approach also follows the methods in Cremers and Weinbaum (2010). They find the signal to have predictive power, and more amplified when only looking at the stocks where the liquidity of the options was high.

Other related factors encompass the skewness factor of Harvey and Siddique (2000b). The skewness of stock prices are priced in the markets, which indicates that the investors are pricing the risk

Same with Amin et al. (2004)

The project is structured as follows. The following section will set the stage, and explain the relevant theoretical frameworks built upon in this project. The third section will outline the data, how it was obtained and filtered for the analysis. The fourth section builds upon the theoretical section and explains how the analyses will be conducted, and argues for the choices made in the process. The analyses of the data are done in section five. The sixth section discusses further research and limitations regarding the choices made in this project, and section 7 concludes.

2 Setting the Stage

The setting of this study is elaborated in this section. I will start out by explaining the implied volatility spread, how it is derived and why it is a relevant signal. Then I will go on to discuss the ideas underlying the factor models and how they came to be through the arbitrage pricing theory. Next the rational investor's preference will be discussed to support the analysis of optimal portfolio formation, and subsequently a few concepts from prospect theory will be outlined to give some nuance to discussion. Lastly I will include a small note on the statistical properties of the data, in regards to problems arising from heteroscedasticity, autocorrelation and stationarity.

2.1 Option Pricing

The interest of this entire project is centered around the implied volatility spread of american options on stocks across the American stock market within the last 28 years.

The optionmarket is inherently forward looking, as options are contracts for future differences between the strike and stock price. In general, the price of a call option with time to maturity, T, and strike, K, is defined as follows:

$$C_t(K,T) = \mathbb{E}_t^{\mathbb{Q}}\left[(S_T - K)^+\right] \tag{2.1}$$

where the right hand side of the equation just shows the expectation under the $\mathbb{Q}$-measure[1] to the positive difference between the stock price, S_T, at maturity and the strike while assuming the risk free rate, r, to be zero (and constant). As an option is a contract for difference, where the buyer chooses if they want to exercise the contract for the difference at prespecified moments before maturity, they provide an alternative to taking an outright position in the underlying for an investor with different views (or information) than the market. In addition to exposing the investor towards this outright stock price, the non-linearity of the option pay-off also allows the investor to take exposure in the volatility of the stock.

The optionmarkets are characterised by trading a lot of different european options upon indexes and the like, while the options on individual stocks are american. These two kinds of options share many similarities, and I will therefore start by outlining the theory for pricing an european option and then dive into the differences and how this affects the prices of american options. The focus within these classes of options will be both puts and calls. I will refrain from including any other kinds of options, as their pricing and payoff structure is different, and therefore might affect the results in an unintended way.

Given the historic context of the American stock market increasing in value over the last decades, it seems natural that an investor will expect the future value of a stock to be higher than the current value. Therefore, one would conclude that calls should be priced higher than puts, if the strike is equal to the current traded price. However, some investors use the puts to lock in their profits, as they secure a future salesprice at the

[1] The $\mathbb{Q}$ denotes risk-neutral pricing and will be elaborated upon further in the following subsections.

strike level of the put. And according to prospect theory[2], this loss aversion is rational and very present in all financial markets.

With a put and a call with the same strike, maturity and underlying, it is natural to conclude that the prices of these two products should be related. This is defined as the put-call parity, and it holds only for european options:

$$C_t(K,T) + Ke^{-r(T-t)} = P_t(K,T) + S_t \tag{2.2}$$

where the price of a call in addition with the present value of the strike (discounted at the risk-free rate) is equal to the sum of the put price and the current stock price. This relation holds as a position in a call option and lending at the risk free rate with a payoff of K at time T, against a position in one put option and the share, results in the exact same payoff at time T. This should be priced the same according to the law of one price. This also means that one can replicate the payoff from a call option by investing in the opposite site of the equation, leading to arbitrage opportunities if the option is mispriced according to the put-call parity. Any mispricings in the 'real' world would therefore be due to transaction costs, restrictions on borrowing or shorting constraints on the underlying stock as arbitrageurs would ensure the markets stay efficient.

2.1.1 Implied Volatility

The implied volatility stems from the general framework for European option prices of Black and Scholes (1973). Assuming lognormal distributed returns, constant volatility of the stock and a constant relevant interest rate, and of course no transaction costs or difference between borrowing and lending rates, they present the iconic pricing formula for an european Call option:

$$C_t(K,T) = SN(d_1) - Ke^{-r(T-t)}N(d_2) \tag{2.3}$$

where

$$d_1 = \frac{\ln(S_0/K) + \left(r - \delta + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}$$

and

$$d_2 = \frac{\ln(S_0/K) + \left(r - \delta - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}$$

with S as the stock price today, $N(j)$ being the cumulative probability function evaluated in j with a standard normal distribution, K is the strike of the option, r being the risk free rate and $T-t$ being the time to maturity. In the latter two equations, the constant term of the volatility of the underlying, σ, is present. The corresponding price for a european put option is:

$$P_t(K,T) = Ke^{-r(T-t)}N(-d_2) - SN(-d_1) \tag{2.4}$$

[2] See subsection **??**.

These functions were derived by Merton (1973) with a simple assumption of investors preferring more to less. Combining the put-call parity with the price function of the european call option, I can infer the corresponding price of a put, which means that the implied volatility of a call option will be equal to that of a put option, if their strike, maturity, and underlying stock is the same. Thus any implied volatility spread different from zero will indicate market inefficiencies and/or arbitrage opportunities. This result holds regardles of the Black & Scholes framework, as any implied volatility spread different from zero on european options represents an arbitrage opportunity in frictionless markets.

When calculating the model-based price of an european call option, I would need to make an estimate of the volatility on the underlying stock's price. In other words, the volatility of the stock return is deterministic in regards to the option price. Therefore, given a set of option prices, I can derive the implied volatility, assuming that Equations 2.3 and 2.4 holds.

As the implied volatility is deterministic, I can plug in the volatility of a stock return into Equation 2.3 and get the theoretical implied price of a call option.

2.1.2 Implied Volatility Spread

For the american options, the pricing equation is slightly modified, as an american option always will be priced higher than their european counterpart. This is due to the early excercise premium, given that american options can be excercised at any time before maturity. The inequality of the american call option is:

$$C_t^a(K,T) \geq S_t N\left(d_1\right) - Ke^{-r(T-t)} N\left(d_2\right) \tag{2.5}$$

with the options being the same value only when they are far out of the money. As the call price now is an inequality and not an equality, the put-call-parity is now of course not exact either.

To model the value of an option, the volatility of the stock price is portrayed in a tree-like structure. Essentially, the stock return follows a Brownian motion in the Black and Scholes (1973) framework, and this process can be approximated with a binomial process using different estimations of the up- and down-movements and the risk-neutral probability of an up-movement. This binomial process is drawn to resemble a tree and shows the movement of the stock price and the corresponding value of the option in each time-increment assuming that the stock price will either move up by u or down by d. These values are approximated by several different approaches, but most commonly is the Cox-Ross-Rubinstein (CRR) formula (Cox et al., 1979):

$$\begin{cases} u = e^{\sigma\sqrt{\Delta t}} \\ d = e^{-\sigma\sqrt{\Delta t}} \end{cases} \tag{2.6}$$

which then can be used to calculate the risk-neutral probabilities of an up-movement of the underlying, q:

$$q = \frac{e^{r\Delta t} - d}{u - d} \tag{2.7}$$

As the inherent discrete time nature of binomial trees, when the time increments goes towards zero, the price of the option will approach the Black-Scholes model implied price.

It should be noted here, that the choice of the CRR model is an industry standard, and I assume that a different choice of for example the Second Order CRR or the Jarrow-Rudd model (Jarrow and Rudd, 1982), will not change the results drastically, as they only show a small modification to CRR. Furthermore, I am looking at the difference in implied volatility and not the absolute value, which means that only a model with asymmetric assignment of the volatility to calls versus puts would give a significant different estimation of the implied volatility spread. In addition to the choice of model, I could also use trinomial trees instead, which would resemble a possibility of the stock to either move up, down or be constant in the next period. This would make the analysis more nuanced, but in my view not provide enough new insights to outweight the increased computational burden. Another feasible way to valuing these american options would have been to use a simulation with a simple least squares approch as presented by **??**.

Thus I can approximate the price of an option given a constant, known value of the volatility and a constant risk-free rate. This price will then be risk-neutral, and means that according to a risk-neutral view on the financial markets, an option should have that price to reflect the market's view on the future value of the option.

Using the binomial trees, it is possible to infer a price for an american option, by assuming that at any future state, the investor will exercise the option if the future expected value is less than the (at the time) current value. According to OptionMetrics (2015) the implied volatility is calculated using the binomial tree approach of Cox et al. (1979) and letting the time increments be sufficiently small to proxy continous time. For the calculation of a call option's value at time i the following function is used:

$$C_i = \max \left\{ \begin{array}{c} \left(qC_{i+1}^{up} + (1-q)\,C_{i+1}^{down}\right)/R \\ S_i - K \end{array} \right\} \tag{2.8}$$

with q being the risk-neutral probability of an up-movement in the stock price and R is the relevant continous risk-free rate less any relevant dividends. Using this formula iteratively from maturity and back to the present, this will yield a fair price for the option. However, when I have only the price and not the volatility, I can use this setup to find the implied volatility. Thus I iterate through different values for σ until the estimated price of the option converges to the midpoint between best bid and offer of the option at market close every day.

Completely analogously to the european option case, if I plug in the volatility of a stock return recursively into Equation 2.8, it would result in the model-implied call price.

Traditionally when approximating the implied volalility, the researcher uses the price of the option, which incorporates all current knowledge of future dividends or stocksplits, and has the following form:

$$f_t = S_t e^{R(T-t)}. \tag{2.9}$$

As the put-call parity does not hold for american options, it is possible to encounter a non-zero implied volatility spread between a pair of call and put options. This spread is indicative of a relative expensive call or put. Following the notation in Cremers and Weinbaum (2010) the implied volatility spread is found through:

$$IV_{i,t}^{call} - IV_{i,t}^{put} = VS_{i,t} \tag{2.10}$$

where t, i and j specifies the time, stock and combination of strike and maturity. The equaiton indicates that if the spread is positive (negative) then the call is relatively expensive (cheap) or the put is cheap (expensive). I would expect the spread to be close to zero and have a fairly small standard deviation, as I assume the markets are efficient and pricing the products close to their risk-neutral values.

2.1.3 Risk-Neutral Pricing

To provide a short note on the risk-neutral pricing, I will elaborate on the difference between the $\mathbb{Q}$ and $\mathbb{P}$ measure, and why it is relevant to keep their definitions in mind when doing an empirical asset pricing exercise on the basis of options and their risk-neutral inferred volatilities.

As the $\mathbb{Q}$ measure is only signaling an adapted set of probabilities of future states which allows for risk-neutral pricing of any asset, it is related to the true probabilities under the $\mathbb{P}$ measure. The relation is defined in Girsanov's theorem (Girsanov, 1960).

On a general level, the measure determines which probabilities are assigned to uncertain events. Under the $\mathbb{P}$ measure (also known as the physical/historical measure), the probabilities are the ones pertaining to the real world. This is the measure under which we observe asset dynamics.

The $\mathbb{Q}$-measure is another particular measure that is famously known from the fundamental theorem of asset pricing (Delbaen and Schachermayer, 1994). It corresponds to the measure that would prevail in a world where all investors were risk-neutral, which would drive risk-premia to zero. In particular, under this measure, all asset price processes are driftless (i.e., martingales) when deflated by the value of the bank account. Another way to say this is that all assets earn the risk-free rate under the Q measure. The FTAP states that all assets can be priced as a discounted expectation under the Q measure, as in (2.1).

As is apparent from the Girsanov theorem, $\mathbb{P}$ and $\mathbb{Q}$ measures are related through a risk premium, which is closely related to stochastic discount factors (SDF's). On an intuitive level, one may view the $\mathbb{Q}$-measure as arising from a combination of the physical probabilities and a correction that is related to the risk-aversion of investors (and hence the SDF).

Thus I am observing mispricings under the $\mathbb{Q}$-measure in the option markets, and basing my analysis of the stock returns under a $\mathbb{P}$-measure on the magnitude of these mispricings. This results in an analysis based on the $\mathbb{P}$-measure which identifies risk premia in the stock market.

2.2 Arbitrage Pricing Theory

Arbitrage pricing theory was introduced by Ross (1976) through the simple assumption of no arbitrage. He argues that investors should diversify their portfolios, so their exposure only is towards systemic risk factors and not idiosyncratic risk. And that all returns are primarily affected by a selected amount of factors. These returns are, of course, only a relative pricing mechanism, as Ross only states that the factors help determine the relative price levels of the assets.

These factors driving the returns should be fewer than the amount of assets available for investment. Ross derives the approximate relation between a diversified investment in a variety of assets and an unknown set of factors. The derivation is based on the statement that a zero-beta net-zero investment in a set of assets also should provide zero return, so as to not provide arbitrage opportunities. A net-zero investment in a diversified portfolio with no exposure to any factors, should yield no expected payoff.

Thus Ross builds the arguments as to all returns being approximately driven by a linear dependence on the factors. The factor model can as an extension of the APT model of Ross be written as:

$$r_{t,i} = \mathbb{E}_{t-1}\left(r_{t,i}\right) + \sum_{k=1}^{K} \beta_{k,i} \cdot f_{k,t-1} + \sum_{j=1}^{J} \beta_{j,i} \cdot g_{j,t-1} + \sum_{h=1}^{H} \beta_{h,i} \cdot p_{h,t} + \epsilon_{t,i} \tag{2.11}$$

where the returns of an asset in period t is determined through its expected value, its exposure to the priced factors, f, the quasi factors g, which are not priced, and lastly the characteristics of the asset, p, which is only observable within the return period. Any of the relevant factors might be traded.

According to the Efficient Market Hypothesis Fama (1970), the investor should already incorporate any public knowledge into the pricing of the security, and the investors should demand compensation for systemic risk. This means that any idiosycratic risk factors do not demand a risk premium. The market should therefore have homogenous expectations regarding the exposures and risk factors.

2.3 Rational Investors' Preferences

A rational investor is characterised by preferring more to less. In the mean-variance model of Markowitz (1952), the returns approximately follows a gaussian distribution, which leads to the natural conclusion that investors will make their investment decisions solely based on the expected return and the variance of the return.

This consideration of only mean and variance means that we can implement the portfolio optimization proposed by Markowitz (1952) and Markowitz (1959). Here the covariance matrix of the returns, the expected value and the variance of the returns are assumed to remain somewhat stable. And the investor should then combine the portfolios on the efficient frontier, and identify the relevant split between the risk free asset and the optimal portfolio.

In this setup, I am not assuming risk aversion, but only that the only source of risk relevant to the investor is the variance. This is because I do not need the additional assumptions of the utility framework introduced by Von Neumann and Morgenstern (2007), because I am limiting the focus to the optimal portfolio formation.

Inclusion of the utility framework would result in a choice regarding the utility function. This would model the rational investors preference under an expected utility view, and not account for several behavioural biases, which I have deemed relevant in the choice of implied volatility spreads. Namely the loss aversion which is evident as a backbone of the prospect theory introduced by Kahneman and Tversky (2013).

Prospect Theory To make a brief comment on prospect theory. The choice of including the implied volatility spreads data, is partly based on the historical average being negative. I interpret this negative value to be a signal of the put options to frequently have a relatively high price, even after correcting for the intrinsic value of the option. A relatively high price of the put indicates that investors are bidding up the price of the right to sell in the future. Thus they are willing to pay a higher price to ensure that their profits are locked in. This indicates loss aversion, as they are not behaving rationally. If all investors were rational, then the implied volaatility spread should be fairly small and close to zero, as it only would proxy the extra information of the delayed market close of the optionmarkets, and the tendency of informed investors to trade in the option markets before moving to the stock market.

2.4 A Small Note on Statistical Properties

The statistical properties of the returns are relevant for the interpretation and validity of the results. For the regressions deployed, the returns should follow a gaussian distribution. OLS is assumed to be the efficient estimator, but the returns show some problems related to autocorrelation and heteroscedasticity.

The autocorrelation of asset returns is a common problem, and as such I have deployed Newey-West standard errors to correct for this potential issue. This also takes care of the problems with heteroscedasticity.

Some of the models are deployed in the literature using both OLS and in a GMM framework, but to keep the simplicity of the methods, I have refrained from using GMM. The benefits from including GMM would be the robust standard errors and removing the errors-in-variables problem arising from Fama-Macbeth regressions, but this will be corrected through the use of robust standard errors and a correction by Shanken for the errors-in-variables.

Correcting for these problems allows me to interpret and conduct inference upon the results.

The stationarity of the signal is shown in the appendix, along with the plots of the autocorrelation function of the different portfolios.

3 Data

Data is downloaded from OptionMetrics, CRSP, Kenneth French's Library and Open Asset Pricing. The extensive data has been filtered to only include the stocks registered on American exchanges. And they are only included in the portfolio formation if a signal is observed in the relevant period, which is the preceding week. All data have been joined together using the SECIDs and PERMNOs.

Wharton's Research Database has option data available through OptionMetrics on a daily basis from the first of January 1996 until December 31st 2021 (as of 1st of May 2023). Center for Research in Security Prices has data available through the same accesspoint with daily holding period returns for all stocks registered in the US. To compliment the portfolios formed based on the implied volatility spread, I have included portfolios formed on size and value by Kenneth French[3] as well as some of the factors calculated both with a daily frequency and a weekly frequency. To compliment the factors already retrieved from French, I also include selected factors available from Open Asset Pricing (Chen and Zimmermann).

The risk free rate deployed to get excess returns is the 1-month t-bill rate supplied on a daily basis in Kenneth French's Library.

The returns of each stock can be calculated in different ways. The first approach is to use the supplied opening price and closing price of each stock at each day, and the second approach is to use the holding period return supplied in the dataset as well. I have chosen the second option, as this takes any stocksplits or dividend payouts into account. This can clearly be seen in Figure A1, where six representative sample years are plotted with all daily returns as calculated through the two different approaches. Note that the axes are different, and if these two methods were equal, the observations would lie on a perfect line of $x = y$ (which they clearly do not).

The conditions for excluding some of the data regarding the implied volatilities are introduced in Cremers and Weinbaum (2010). The conditions are elaborated below. Note, however, that as also discussed in Shang (2016), I will do the analysis for a dataset with the conditions imposed, but also include the results from the analysis without the conditions imposed in the appendix, to show the effect of the conditions and the robustness of the results.

The signal is the spread between the implied volatility of call and put options with the same maturity and strike. These values are observed at market close at a daily frequence. I will take an average across maturities and strikes everyday for each stock to have a simple signal. This average can both be a simple average across every available datapoint or a weighted average using the open interest reported at market close for each pair. Cremers and Weinbaum (2010) argue that the latter approach incorporates the liquidity aspect of the options. I assume that the more liquidity an option has, the more fair is the price, and the more reflective of the market's opinion is it. Thus the latter approach for weighting the implied volatility spread ensures that the signal incorporates the market's view.

[3] The data is available at his webpage: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

Conditions The conditions for including the options in the signal formation is described below. The first condition is in regards to the implied volatility of each option. These conditions are not related to the definition of the following unconditional and conditional analysis.

Condition 3.1: Restrictions on Implied Volatility Level

$$0 \leq IV_{j,t}^{i,call} \leq 1.5 \text{ and } 0 \leq IV_{j,t}^{i,put} \leq 1.5$$

The implied volatility should be within these limits for both puts and calls for the spread to be included in the signal, which naturally limits the distribution of the implied volatility spread. The historical volatility of the stock markets represented by the VIX Index reached a max of 0.7 in 2017, therefore a limit of 1.5 on individual options is sensible.

The second condition relates to the time to maturity of the option pairs, measured in days.

Condition 3.2: Restrictions on Time To Maturity

$$7 \leq TTM_{days} \leq 365$$

Options with time to maturity within these limits are the most liquid and by excluding the imminent maturing options, the signal will only contain data on options with maturity prevailing the return period.

For the last condition, the forward price of the stock is estimated at the point of daily closing prices for the individual option. Optionmetrics describes the calculation as follows: "The forward security price is calculated based on the last closing security price, plus the interest, less projected dividends" - OptionMetrics (2015) and the standard form of the equation is given in Equation 2.9.

Condition 3.3: Restrictions on Moneyness

$$0.7 \leq \frac{F_{0,i}}{K} \leq 1.3$$

The moneyness (ratio between forward price of the underlying stock and the strike price of the option) should be between those limits, which ensures that the options included are somewhat close to being at-the-money and decreases the noise from illiquid deep-in-the-money or deep-out-of-the-money options.

The average of these is taken per day, and the observed signal for return predictability is the latest day within the last 7 days before the return period begins. If there is no observed spread in this period, the stock is not considered for portfolio formation.

Signals Deployed The two main signals formed on the data are the actual average value of the implied volatility spread and the recent change in the implied volatility spread for each individual stock.

The subsample of relevant options and stocks will be formed to fulfill the above-mentioned conditions. Furthermore, as each stock has multiple pairs of put-calls and therefore also multiple implied volatility spreads observed each day across maturities and strikes, I have two subsample regarding either a weighted mean of these implied volatility spreads using the sum of open interest, and a subsample with just a simple mean of all these values. The difference in the distribution of these means across the time period seems rather stationary. In Figure A2 I show the median, 95% and 5% percentile of each of these subsamples across the entire sample period. Thus I note that the signal seems very stationary, and that prominent outliers are present in all four subsample with (C) having the most volatile median and (A) having more volatile percentiles.

It is interesting to investigate these plots, as they show very little difference between them, and as such, it seems an arbitrary choice to filter out the option pairs, unless the economic arguments are convincing. Furthermore, a weighting according to the open interest is also argued for by Cremers and Weinbaum (2010) and likewise deployed in Shang (2016). Some of the most prominent results will therefore also be reported for each of these subsamples, to investigate the effect from making these choices.

To calculate the second signal, I take the difference of the earliest reported average implied volatility spread and the latest reported value within a subsample of the last seven days before portfolio formation. Thus a change is reported and used for sorting portfolios. I believe this signal to be more informative and relevant in a combination with the value of the implied volatility spread and not a suitable candidate signal for univariate portfolio sorting. The reasoning is based on the object of this project to find a signal to proxy for the risk associated with mispricing of options, especially related to the too-high priced put options on stocks which might proxy some loss aversion mechanics on the stock market, and as such, I am not interested in how this mispricing changes week-on-week only, but it is relevant when combined with the value of the implied volatility spread.

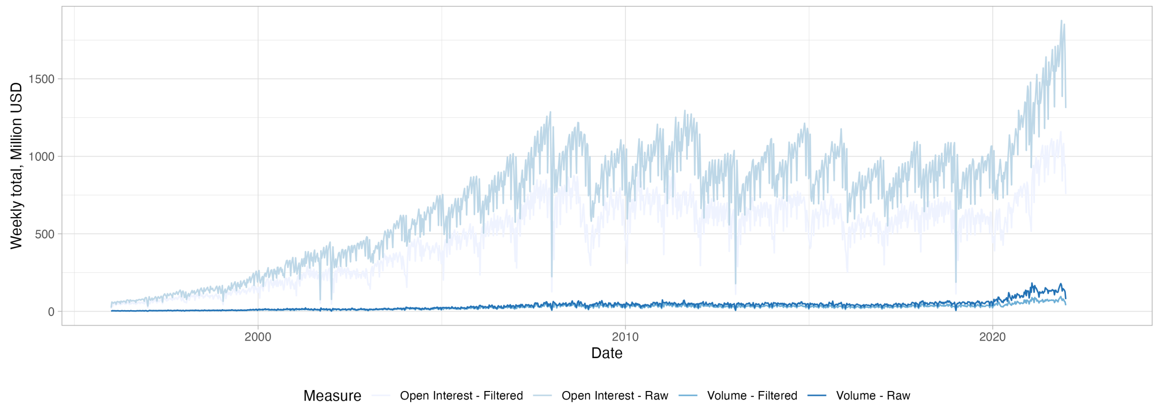
Metadata Different measures of metadata are plotted in Figure 3.0.1. These plots show the evolution of some metrics across the sample period. In general most of the data has increased over the sample period, and a picture of effect from the IT-bubble in the early 2000's, the financial crisis in 2008 and the COVID-19 shock and subsequent upward trend following.

COMMENT MORE HERE

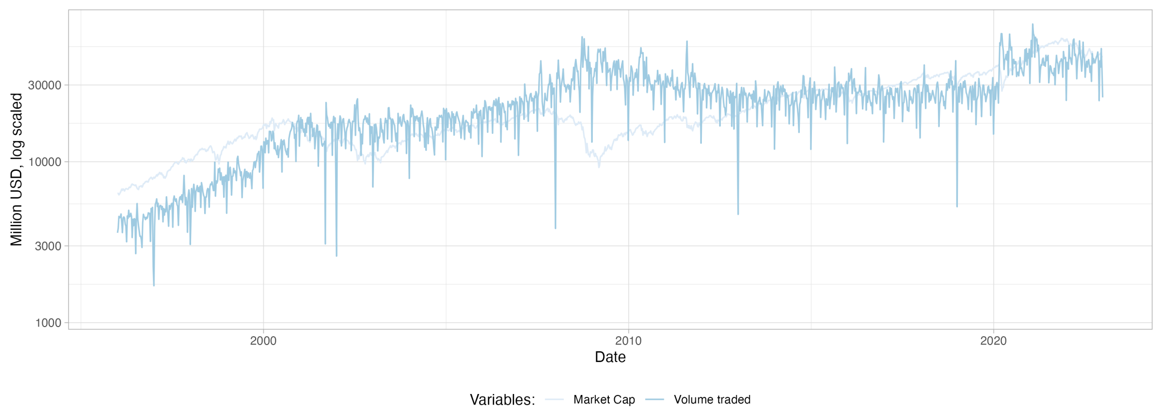
Other Use of Available Data If time, scope and focus had allowed it, I would also have loved to deploy the ratio of open interest in the stock market versus the open interest in the option market. This could have been a relevant proxy for the market liquidity and how well the market prices resembled the market's beliefs. This could have been done through just a simple continuous variable which could factor in through interaction terms with the dummies of the different portfolios, and it could also be incorporated through a split into different 'levels', with the sample splitted into five subsamples resembling the five different formalized levels of the ratio. In addition to this, it could be relevant to factor in a de-trending of this ratio, to not make the earliest years with very few options offered constitute the lowest group.

Figure 3.0.1: Volume traded, Number of Firms and Median bid-ask spread for CRSP and OptionMetrics

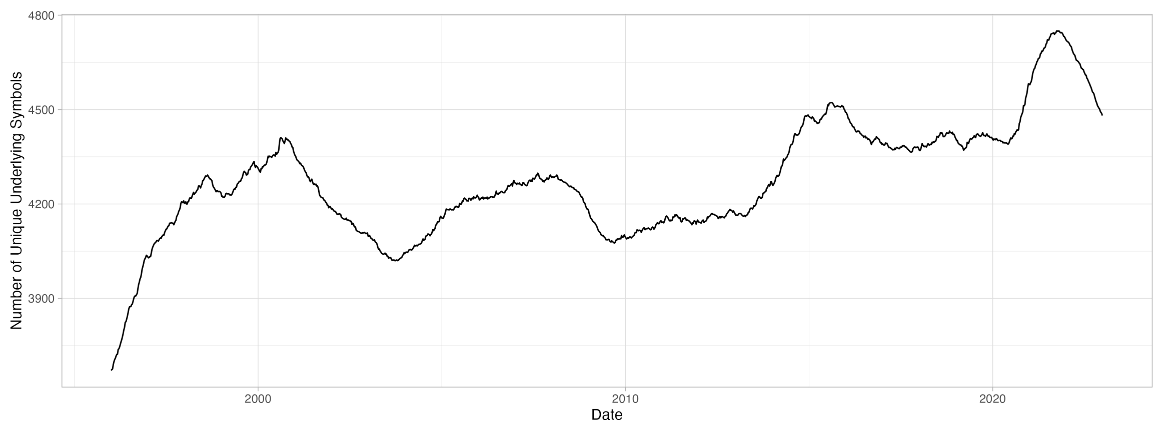
(a) OptionMetrics : Open Interest and Volume across two subsamples



(b) CRSP : Volume Traded and total Market Value



(c) CRSP : Number of Firms



Note: Figure shows different measures of metadata for both CRSP data and OptionMetrics data. The two subsampled in subplot (A) consist of a subsample with the conditions mentioned above imposed, and another subsample without these restrictions imposed. Note plot (B) has a log scaled Y axis.

4 Empirical Implementation

The empirical implementation of the signal and it's effect on the stock market will be elaborated upon in this section. It will start by outlining the distribution of the signal and choices made regarding both the estimation of the signal and the formation of the portfolios, whereafter I will outline the factors, how they are made, and what other considerations were relevant for the analysis. Afterwards, I will go through the different frameworks for analyzing the relation between the stock market and the signal, and an investigation into the risk premias of the factors. The section will conclude with an overview of the economic evaluation of the factors.

4.1 Estimating the Signal

For the estimation of the signal, several choices have to be made. The first and most important aspect, in my view, is the choice of which period to observe the signal in. In Fama and French (1992) only information available at portfolio formation is used, while Cremers and Weinbaum (2010) include portfolios formed on signals observed at market-close Wednesday and holding period return measured from market-close on the same day until the following week's Wednesday. On the base of wanting to include the overnigth returns.

For every single option in the database, the implied volatility is found at end of each trading day. As mentioned above, the put-call-parity no longer holds for american options, and therefore there might be a difference in the implied volatility of a put and call with the same strike, maturity, and underlying stock. A high implied volatility means the options is worth more, or priced higher, and vice versa.

I wanted to evaluate the relative pricings of options, and their magnitude, so the signal is then based on the implied volatility spread. This implied volatility spread essentially controls for the intrinsic value of the options and makes it possible to compare the 'mispricing' across maturities, strikes, time, and stocks. Following the notation in Cremers and Weinbaum (2010) and the definition of the implied volatility spread in Equation 2.10, the signal is then estimated across time and stocks:

$$VS_{i,t} = \sum_{j=1}^{N_{i,t}} w_{j,t}^{i} \left(IV_{j,t}^{i,call} - IV_{j,t}^{i,put} \right) \tag{4.1}$$

where t, i and j specifies the time, stock and combination of strike and maturity. $w_{j,t}^{i}$ is the weight for the specific maturity, strike and time. This weight can be either a simple average or based on the sum of open interest in the pair of options.

4.2 Forming Portfolios

The analysis of the relation between stock returns and the signal defined above is conducted using portfolios. Thus I will split the different firms into portfolios with the use of the signal.

The choice of forming portfolios compared to analyzing specific stocks over a period of time, is made based on the flexibility of the setup. The approach requires no a priori assumptions about the effect, and also allows for the discovery of more than just a linear effect. This approach also has some diasadvantages, as the flexibility gives rise to a bigger risk of datasnooping.

The choice of using portfolios for the analysis is further made based on the extensive literature within asset pricing where a variety of signals have been evaluated through portfolio formation. Furthermore, the choice is based on an approach looking at the effect of the signal on the entire American stock market, and not on specific stocks, therefore, an aggregation and combination of these stocks provides a clearer picture of the effect from the signal and reduces the noise and idiosyncratic factors from other individual characteristics.

The portfolios are formed using breakpoints in distribution of the signal. Thus I start by choosing the amount of portfolios I would like. The traditional choice would be to follow Fama and French (1992) and make three portfolios with breakpoints at the 30 percentile and 70 percentile, to make portfolios consisting of the lowest 30% of the signal, then the middle 40% and the last portfolio of the highest 30% from the sample. In Cremers and Weinbaum (2010) they use five equal sized portfolios, and I will compare with their example, but instead primarily use the results with 10 equal-sized portfolios. An overview of the different versions of the portfolios is shown in the Appendix in Table B3.

The use of equal-sized portfolios stems from an ambition to evaluate the effect of the entire distribution of the signal and not just evaluate the tails as done in the traditional approach. Furthermore, my main analysis will be based on 5 equal sized portfolios to ensure a sufficient amount of stocks are in each. This means that I have a smaller sample of portfolio returns to evaluate, which I thought a reasonable tradeoff given the inclusion of the results from the analysis with different amount of portfolios in the appendix.

The statistical properties of portfolio sorting have been further analyzed by Cattaneo et al. (2020), and they find that contrary to what has been the trend in the literature of only focusing on 3-10 portfolios, a higher number of portfolios will give more robust and asymptotically valid results. To supplement the main focus of a bivariate portfolio sort with 10 times 10 portfolios, I will also evaluate a 15 times 15, even though it is not very common in the literature, the above-mentioned article highlights the robustness of such more extreme formations.

Furthermore, the signal identified above might not be the sole signal that we are interested in. A combination of the value of the implied volatility spread and the recent change in the implied volatility spread might also provide some insights.

Therefore, I will conduct an analysis focusing only on the univariate portfolio sorting based on the value, and following that a bivariate portfolio sorting with both a dependent and independent sorting with the implied volatility spread as the first sorting factor and the recent change in the implied volatility spread as the second sorting factor.

The *univariate* portfolio sorting has the following steps:

Step 1: Computing the breakpoints Identifying the signal of stock, i, at time, t, as $S_{i,t}$, I will start from the cross-sectional distribution of the signal, forming the number of portfolios I need, denoted n_ρ, I find for all the $n_\rho - 1$ breakpoints the k'th breakpoint as follows:

$$\mathcal{B}_{k,t} = \text{Percentile}_{\rho_k}\left(S_{i,t}\right) \tag{4.2}$$

The sample used for the breakpoints determination is the entire available dataset. This is done to ensure equal sized portfolios across the analysis. It might however affect the liquidity aspect of the conclusions, as inclusion of smaller and less liquid stocks might skew the picture. This is due to the fact that smaller and less liquid stocks often have higher transactions costs and tighter borrowing constraints, and therefore not priced precisely as the markets expect. An alternative would be to calculate the breakpoints for the portfolios from a sample of S&P 500 stocks and their corresponding implied volatility spreads. I deem however, that only considering the largest 500 stocks in the US for breakpoints determination might skew the picture when including more than 2000 stocks per period, and therefore I use value-weighted returns to correct for any effects from illiquid stocks (and by weighting the implied volatility by the open interest). Furthermore, I would assume that only the larger and more traded stocks have options written on them, and of these a smaller fraction of the options are actually traded every week.

Step 2: Forming the Portfolios The portfolios are then formed using the breakpoint, $\mathcal{B}_{k,t}$, found in the previous step. A stock, i, is then included in portfolio, k, if it's signal is within the breakpoints identified:

$$P_{k,t} = \{i | \mathcal{B}_{k-1,t} \leq S_{i,t} < \mathcal{B}_{k,t}\} \tag{4.3}$$

When forming the last portfolio (the portfolio consisting of stocks with the highest value of the signal), I will add the last few stocks with a signal value equal to the maximum signal value of the sample. In contrast some researchers include less than or equal sign on both side of the signal in the formation of their portfolios, which means that some stocks might be included in two portfolios. To counteract this I choose the above approach insted, which results in my last portfolio having slightly more stocks included, but no stocks will be in two portfolios at once.

Step 3: Calculating the Returns The returns will be calculated based on a value-weighted investment. Thus they are reported in simple value-weighted form with the market value, $MV_{i,t-1}$ observed as the shares outstanding timed the opening price at portfolio formation, and calculated as follows:

$$r_{k,t} = \frac{\sum_{i=1}^{N_{k,t}} MV_{i,t-1} \cdot r_{i,t}}{\sum_{i=1}^{N_{k,t}} MV_{i,t-1}} \tag{4.4}$$

I choose a value-weighted approach compared to equal-weighted, as I want to make sure that small and illiquid stocks, which are difficult to trade, have a small impact on the results.

For a *bivariate* portfolio sorting, the steps look a little different. First and foremost, this will make it possible to control for two different sorting signals compared to just one in the univariate portfolio sorting. The inclusion of bivariate sorting to take a second signal into account leaves more choices to be made. Of course, there is the amount of portfolios and the percentiles of the second signal, but another important choice regarding the sorting of the stocks is also relevant, namely whether to do independent or dependent sorting. Which means that when we do the breakpoints calculation of the second signal, we should either find them for the entire sample or from the grouping from the first signal.

In my case, I will look at the recent change in implied volatility spread as the second signal. Given the combination of the value of the same dataseries being the first sorting signal, it makes sense to do a dependent sorting. If an independent sorting were to be made, and given that the signal is relatively stable over the period as seen in Figure A2, there would be very few stocks with a low value of the signal (thus sorted in the first portfolio in the first sorting) and also having a big positive change in the signal value (sorted into the last portfolio in the second sorting) and vice versa. An independent sorting would therefore result in an almost empty portfolio of stocks with high (low) value and big negative (positive) change in the signal, due to the correlation between the two signals. A dependent sorting is chosen, and the following steps from above is adapted.

Adapted Step 1: Breakpoints for Dependent Bivariate Sorting To find the breakpoints dependent on the first sorting signal, I use the following approach. Note that as mentioned above, the last portfolios are including the stocks with maximum value of the signal. The first sorting breakpoints for portfolio k are defined as above in Equation 4.2, and the second sorting signal's breakpoints for portfolio j is defined below:

$$\mathcal{B}^2_{k,j,t} = \text{Percentile}_{\rho_j}\left(S^2_{i,t} | \mathcal{B}^1_{k,t} \leq S^1_{i,t} < \mathcal{B}^1_{k,t}\right) \tag{4.5}$$

Which leaves me with $n_{p_1} - 1$ breakpoints for the first portfolio sorting and $n_{p_1} \cdot (n_{p_2} - 1)$ breakpoints for the second portfolio sorting.

Adapted Step 2: Forming Portfolios with Dependent Bivariate Sorting Instead of allocating stocks to different portfolios according to only the first signal, I use both set of breakpoints simultaneously.

$$P_{k,j,t} = \left\{i | \mathcal{B}^1_{k-1,j,t} \leq S_{i,t} < \mathcal{B}^1_{k,j,t}\right\} \cap \left\{i | \mathcal{B}^2_{k,j-1,t} \leq S_{i,t} < \mathcal{B}^2_{k,j,t}\right\} \tag{4.6}$$

This results in $n_{p_1} \cdot n_{p_2}$ portfolios with approximately an equal amount of stocks included in each, with only the last portfolios including a slightly larger amounts of stocks.

Adapted Step 3: Calculating Returns with Dependent Bivariate Sorting The third step is only slightly changed to account for the second dimension of portfolios:

$$r_{k,j,t} = \frac{\sum_{i=1}^{N_{k,j,t}} MV_{i,t-1} \cdot r_{i,t}}{\sum_{i=1}^{N_{k,j,t}} MV_{i,t-1}} \tag{4.7}$$

This concluding the adapted steps for the bivariate sorting. Any structure of analysis from here will look largely the same.

A further sorting of the stocks into a third formation is not feasible, even though it might make sense to want to account for more signals. This is due to how portfolio sorting scales with the amount of sorting signals. And as such, most of the literature focus on either univariate or bivariate sorting. In an attempt to account for more variables, in particular the liquidity of the option market, we can reduce the initial amount of stocks to only include these that have a ratio of open interest in options to open interest in stocks to be larger than a certain (time-variant) threshold, and compare these results to the full sample. Essentially a third portfolio formation with a sole interest in the subsample with the highest ratio.

The portfolio horizon will be limited to 1 week, and no forecasting of a future implied volatility spread will be made to enable a longer horizon of portfolio sorting.

With the portfolio returns being computed, the factors should be estimated. Both the portfolios and the factors will be analyzed through an unconditional analysis and a conditional analysis.

4.3 Forming the Factors

A factor can be formed as a long-short zero-net investment in the portfolios, or it could be formed based on an average across different intersecting portfolios as in the Fama and French (2015). The choice of forming the factors as a position in only the tail portfolios means that the more portfolios considered, the more extreme might the factor be and as such, it would resemble only the outmost exposure to the signal and not a general position.

To inspect the factors, I will compare the factor competition[4] results of three different univariate portfolio sorts based on the level of the implied volatility spread. The results are shown in Table 4.3.1.

The results of Table 4.3.1 shows that the factors have the same level of significance, but the intercept varies as it increases in the amounts of portfolios. This signifies a clear linear dependence between a high value of the signal and a high return in the subsequent week, as a factor formed on less stocks in the far out tails have a higher intercept, and thus a higher average.

The descriptive statistics of the portfolios in general, will, of course, be elaborated upon in the empricial section. For now, I will move on to describe the relevant metrics for an unconditional analysis of the portfolios.

4.4 Unconditional Analysis

I will define the unconditional analysis as to be the analysis of the distribution of the returns and a few selected statistical tests upon only the portfolios without involving any other factors or models.

The returns of any portfolio cannot be comprised by only the average return. In the finance literature in general, the risk associated with any portfolio is mainly described by its standard deviation. This is due to

[4]See subsection 4.5.2.

TABLE 4.3.1: Factor Competition Intercept based on 3 different Portfolio Formations

Dep. Variable	Intercept, 3	Intercept, 5	Intercept, 10
IMPVOL	0.294 (***)	0.392 (***)	0.552 (***)
MKT_RF	0.254 (***)	0.251 (***)	0.26 (***)
SMB	0.054	0.063	0.042
HML	-0.077	-0.06	-0.057
RMW	0.014	0.028	0.05
CMA	0.04	0.028	0.041
BetaLiquidityPS	0.02	0.094	0.009
betaVIX	0.097	0.029	0.077
CoskewACX	0.294 (***)	0.333 (***)	0.278 (***)
Coskewness	0.15 (*)	0.095	0.12
OptionVolume1	0.464 (***)	0.455 (***)	0.475 (***)
OptionVolume2	-0.075	-0.058	-0.076
skew1	0.461 (***)	0.435 (***)	0.458 (***)

Note: the table shows the intercept and it's significance across 3 different portfolio sorts. The first portfolio sort has 3 portfolios with a 30% - 40% - 30% split, the second has 5 equal-sized portfolios and the last has 10 equal-sized portfolios. The regression span the entire sample period and consists of weekly returns. The significance of the coefficients are coded according to the p-value: $0 < (***) < 0.001 < (**) < 0.01 < (*) < 0.05$. The scenarioIDs are 12, 1 and 13 respectively.

the returns often being assumed normally distributed. As will be discussed later, different measures of risk might be relevant, and as such, I will report both the skewness and kurtosis for each portfolio. These metrics are all reported to give an idea about the distribution of the returns. All metrics are being calculated across the entire sample period.

Using this assumption of gaussian distributed returns, a relevant metric of the performance of portfolios in the mean-variance space is the Sharpe Ratio:

$$SR = \frac{R_p - R_f}{\sigma_p} \tag{4.8}$$

which is too calculated across the entire sample period. In addition to this, a common known metric is the Jensen's Alpha, which is essentially just the intercept of the CAPM. This metric will not be reported among the general descriptive statistics, but the intercept will be tested among a general time-series regression against an extended factor model which includes the FF5 factors.

The portfolios formed upon the signals gives rise to a factor formation. Traditionally a long-short version of factors has been the default, where they resemble a zero-net-investment in the tail portfolios. Thus a factor denoted IMPVOL will signify a long investment in portfolio N and a short investment in portfolio 1, and in the case of bivariate portfolio sorting; a long investment in portfolio N_N and a short investment in portfolio 1_1. In the bivariate portfolio sorting case, I will include a second factor of a long position in portfolio N_1 and a short position in portfolio 1_N. In Fama and French (2015) they choose a slightly different formation of the portfolios based on a zero-net-investment long-short position of average across different portfolio sortings. This setup was deemed a slightly too complex in comparison with the goal of highlighting the effect of the signal upon the stock returns. All denotion of the portfolios have the portfolio numbering of the first sorting as the first number and the portfolio numbering of the second sorting.

While I am interested in evaluating the factors, I will also investigate the patterns of the returns in the portfolios. In particular, I deem it interesting to see if the returns are monotically increasing or decreasing in the portfolio order. A statistical test will be introduced in the following subsection before I will dive deeper into the statistical test of the factor, and its relation to other factors from the literature.

4.4.1 Monotonicity Test

The pattern of the returns have traditionally been inspected through visual analysis of the mean return of each portfolio. To do this in a more structured way, Wolak (1987, 1989) introduce a test for monotinicity in returns with a null hypothesis of a clear pattern, and Fama (1984) proposes a similar test using Bonferroni bounds. Patton and Timmermann (2010) introduces a test with the same purpose, but a different null hypothesis of no clear pattern in the returns across portfolios. This ensures that we only reject the null hypothesis of no distinct monotonically increasing returns across portfolios if there is enough evidence to support it, while the abovementioned tests would require the data to show a random pattern instead of a monotonically increasing one. In other words, I prefer to use a test which has no pattern as a prior and relies on the data to prove it wrong.

The hypothesis of the Patton and Timmermann (2010) test is formally written as, using $\Delta_i = \mu_i - \mu_{i-1}$ as the difference between the average return of portfolio i and $i-1$:

$$\begin{aligned} H_0 &: \Delta \leq 0 \\ H_1 &: \Delta > 0 \end{aligned} \tag{4.9}$$

The test statistic is then computed as the minimum of the observed deltas, $J_T = \min(\Delta)$. As there is no preformed distribution of this test statistic on which we can evaluate significance, we will find the distribution through bootstrapping with replacement from the original sample. This is all done in the R package provided by Köstlmeier (2019).

The reason for the inclusion of this test is to identify if there is a monotonically increasing (or decreasing) pattern in the average returns of the portfolios. This will lead to a preliminary conclusion that it might be relevant to look at this signal, given that it when rejecting the null has a significant effect on the expected returns.

This test is, however, based on a view of the entire sample period and does not account for time varying monotonicity or if the pattern persists when corrected for other known risk factors. I will therefore in the further analysis look deeper into the signal and examine how it performs in the cross sectional study when confronted with common risk factors from Fama and French (2015).

4.5 Conditional Analysis

In this subsection I will elaborate upon the different analyses, I deem relevant to evaluate the relation between the signal and stock returns, and the risk premia of the factor. The starting point is the relevant external factors, hereafter they will be evaluated in the factor competition (also known as spanning regressions), the third will be risk premia estimation through the Fama-Macbeth regressions and the three-step procedure of Giglio and Xiu (2021). The pooled cross sectional regression deployed in the analysis, will not be elaborated on in this section, as the implementation is fairly straightforward.

4.5.1 Comparing with Other Factors

When analyzing the portfolios, I will not only base the analysis on a long-short factor calculated from the portfolios, but also incorporate some other factors. This will by no means include all factors available or proved to be significant [5] , but only include the ones relevant. I deem the following three subsets of factors relevant for inclusion in the further analysis.

FF3 The three factors of Fama and French (1992) based on portfolios formed of american stocks are some of the most analyzed (and scrutinised) factors. They were among the first factor models and also include the seminal market factor (introduced by Treynor (1961), Sharpe (1964), Lintner (1965) and Mossin (1966)). The estimated model has the following form:

$$R_{i,t}^{e} = \alpha_i + \beta_{i,MKT} \cdot R_{MKT,t}^{e} + \beta_{i,SMB} \cdot R_{SMB,t}^{e} + \beta_{i,HML} \cdot R_{HML,t}^{e} + \epsilon_{it} \tag{4.10}$$

where $R_{i,t}^{e}$ is the excess return of stock i at time t, α_i is the abnormal return incurred after controlling for the factors (which has been argued to be largely 0 based on no-arbitrage arguments proposed by Ross (1976)). The $R_{k,t}^{e}$ is the factor portfolio excess return of portfolio k and $\beta_{i,k}$ is the corresponding exposure of the excess return of stock i to the factor. ϵ_{it} is the error term assumed to be uncorrelated across stocks and time.

FF5 The five factors of Fama and French (2015) extends the previous model by also including profitability and investment as factors while adapting the SMB (size) factor from FF3 to be influenced by the additional factors included in the model. The model has the following (similar) form:

$$R_{i,t}^{e} = \alpha_i + \beta_{i,MKT} \cdot R_{MKT,t}^{e} + \beta_{i,SMB} \cdot R_{SMB,t}^{e} + \beta_{i,HML} \cdot R_{HML,t}^{e} + \beta_{i,RMW} \cdot R_{RMW,t}^{e} + \beta_{i,CMA} \cdot R_{CMA,t}^{e} + \epsilon_{it} \tag{4.11}$$

With the main difference from the equation above being the two additional factors.

Open Asset Pricing The last subset of relevant factors include subjectively chosen factors from Open

[5] A close to exhaustive list of all factors documented in the literature to have a significant effect on stock returns is comprised by Harvey and Liu (2019) in this google docs document: Google Sheets, Factor Census.

Asset Pricing. The first chosen factor is the liquidity factor of Pástor and Stambaugh (2003). This factor is relevant as it proxies a liquidity risk that might affect shorting constraints, which in turn might be affecting my signal, therefore I deem it relevant to include it for control. The second chosen factor is the volatility risk factor of Ang et al. (2006b), and measures the daily change in the VIX to capture the risk associated with volatility. The third included factor is coskewness factor (CoskewACX) introduced by Ang et al. (2006a), it is formed based on portfolios formed on the coskewness value and using equal-weighted returns. The fourth is also a coskewness factor (Coskewness), introduced by Harvey and Siddique (2000a) and contrary to the other coskewness factor, it is based on a long short investment in portfolios with value-weigthed returns. The fifth and sixth are both optionvolume factors, and the first factor measures the ratio of the offered volume in the option market versus the stock market and the second factor measures the offered option volume to the average, both were introduced by Johnson and So (2012). The seventh factor is a skewness factor and measures the volatility smirk near at-the-money options, introduced by Xing et al. (2010). All of the above-mentioned factors are available at Open Asset Pricing[6] and is only available at a monthly frequency. I have splitted the monthly return into four equal parts (which have a cumulative sum equal to the reported monthly one) resulting in each of these factors being constant over the month. The factors were not available at a daily or weekly frequency. This subset of data will be referred to as OA data.

Other datasets and factors have been proved in the literature to be relevant, but an inclusion of all of them might prove to introduce more noise than explanatory power to the models. I only chose to use simple regressions and a limited use of Principal Components Analysis, leading to overfitting and errors in estimations when the factors are correlated. A different approach using other machine learning techniques or heavier use of Principal Components Analysis, Ridge Regression or Principal Components Regressions would utilize the information in more factors. The focus of this thesis is, however, limited to a portfolio sorting analysis of the implied volatility spread and its relation to equity risk premiums and not an evaluation of the entire factor zoo, see eg. Feng et al. (2017). It should be noted, however, that some of the factors which might provide valuable insight and proxy for some related risk to the one from the implied volatility, includes the PIN factor of Pan and Poteshman (2006) (which describes ratio of the buyer or seller initiated options). The exclusion of relevant factors is, however, due to limitations in data access and not a desire for data reduction.

4.5.2 Factor Competition

To evaluate the portfolios it might be beneficial to isolate the long-short factor formed on the portfolios and relate it to similar factors from the literature. To see if it is merely a proxy for some already discovered risk factor or if it provides some new insights. To do this, I will follow the approach outlined in Fama and French (2015) and perform simple regressions of each factor upon the rest of the factors in the control set, this analysis is also known as spanning regressions. In this case the control set will be FF5, but the rest of the

[6] All data is available at Open Asset Pricing - Data.

factors will be included in the appendix. The following equation outlines the simple regression of estimating the intercept and coefficients of regressing the factors formed on the portfolios upon the remaining factors. In case of a bivariate sorting, both the factor of $n_n - 1_1$ and the diagonal factor of $n_1 - 1_n$ is included:

$$f_{t,k} = a_k + \sum_{j \neq k}^{K} \delta_{k,j} f_{t,j} + \epsilon_{t,k} \tag{4.12}$$

The standard errors are robust against both heteroscedasticity and autocorrelation, as instead of using regular standard errors supplied from the LM function in R, the Newey-West standard errors have been deployed. This will essentially correct the standard errors to be less prone to be affected by any issues with autocorrelation and heteroscedasticity. The autocorrelation in the returns of the individual portfolios are quiet prevalent, and can be seen in the appendix in Figure A3.

The interpretation of this analysis will be to evaluate the intercept. If it is statistically significantly different from zero, then we can conclude that the average return of the factor is not redundant or explained by any of the control factors.

When performing the analysis, only the comparison against FF5 will be included in the main part of the text, however, the analysis of an extended model including both the FF5 and the OA factors will be supplied in the appendix. Furthermore, the regression spans the entire sample. This might provide a skewed picture, as the market for options has matured through the sample period and might provide a better signal of the stock market in recent years compared to the sample period start.

4.5.3 Fama-Macbeth Regressions

In a seminal paper Fama and MacBeth (1973) propose an approach for estimating the risk-premia of factors and variables upon returns was introduced. It is built upon a two-step approach, where we first will estimate the covariance between a factor and the returns over time, and then afterwards use this covariance in estimating the compensation of which the stocks or portfolios provide as price for being exposed towards this factor.

In the analysis, I will include both factors estimated on the portfolios, and the median of the implied volatility spread within each group in time. This is done to evaluate if the signal itself is priced, or if it is merely the tails that are priced through the factors.

Implicitly in this regression we will have an assumption that the exposure towards the factor is constant over time, which I will challenge. Therefore, I will include both a time invariant version of the Fama-Macbeth regression and a time variant version, in which I will assume that the dependence is the same within 52 weeks (approximately a year) and then changing afterwards.

While, of course, I am mostly interested in the risk premia of my signal, it would be beneficial to control for a few factors, and as such, there will also be both a time invariant and time variant version of the

Fama-Macbeth regressions with factors including the signal and the FF5 factors. All returns are excess returns of the portfolios, while any traded factors are excess returns by definition, as they are formed as net-zero-investment long short factors.

The approach of the analysis is as follows:

Step 1: The first step is a simple regression of each individual stock or portfolio upon the factor(s). The purpose of this regression is to estimate the exposure of a certain stock to the factor and then utilize these β's in the next regression:

$$r^e_{t,n} = a_n + \sum_{k=1}^{K} \beta_{n,k} \cdot f_{t,k} + \epsilon_{t,n} \tag{4.13}$$

where $r^e_{t,n}$ is the excess return of portfolio n at time t, a_n is the average return of the stock not explained by the factors, k measures the amount of factors included and $\epsilon_{t,n}$ is the error term. If a time invariant version is estimated, I will estimate this on the entire sample period, if a time invariant version is deployed instead, I will use a rolling window of 52 weeks to estimate the β's for each factor.

Step 2: The second step is then regressing the portfolio returns upon the estimated β's from Equation 4.13. It should be noted here, that I am only assuming a changing exposure to the risk factor and not a changing risk premia, the following equation will therefore be estimated across the entire sample period with either the changing β's of the time variant version or the constant β's from the time invariant version:

$$r^e_{t,n} = \gamma_{t,0} + \sum_{k=1}^{K} \gamma_{t,k} \widehat{\beta}_{n,k} + \epsilon_{t,n} \tag{4.14}$$

where $\gamma_{t,0}$ is the time varying excess return, $\widehat{\beta}_{n,k}$ is the estimated coefficients from Equation 4.13 and $\gamma_{t,k}$ is the price of risk, or the compensation in return for every unit of exposure to factor k.

When doing these regressions, I am essentially estimating the some coefficients whereafter I use those very same coefficients to estimate new coefficients, leading to the errors-in-variables problem. If there is just a slight measurement error in the first regression it will bias the second regression tremendously. Therefore I have chosen to use weekly returns, and not decrease the sampling frequency to monthly observations, in the hope that the factors, returns and signals will be more precisely observed when sampling frequently. Furthermore, this gives me a larger sample of returns, which should make the estimates more reliable.

In addition to this, all included factors are either traded (all the FF5 factors) or it is the median of the implied volatility spread, which is a quite precisely observed signal. This signal used beside the factors formed on the portfolios in the sample, means that I obtain a price for exposure towards the implied volatility spread in itself with a minimal of idiosyncratic effect from individual stocks. This also means that there is very little measurement error in the signal, as it is not a macroeconomic estimate with delayed publication and tons of measurement uncertainty.

To combat this Errors-in-variables problem often discussed in the literature, an alternative approach using GMM to simultaneously estimate both regressions at once is proposed. This can combat both the estimation problems of the coefficients, but also take problems arising from heteroscedasticity and autocorrelation into account. I will argue, based on the sampling frequency, the stationarity of the signal, and the use of portfolios to reduce idiosyncratic noise, that the errors-in-variables does not give rise to an additional use of GMM. The signal is stationary without a trend, as can be seen in Figure A2 in the appendix.

To otherwise combat the issues with variables being difficult to measure, Shanken (1992) has introduced a correction for the standard errors, to make it less likely that coefficients will be concluded to be statistically significantly different from zero when that is not the case. This correction is undoubtedly relevant when using factors, that are difficult to measure, but as this is not the case here, I will not deploy them. Furthermore, it turned out that none of the coefficients were close to being significant, and as such, it would not affect the conclusion to estimate an upwards correction to the standard errors.

While the Fama-Macbeth regressions have been used extensively in the literature, it has not only been faulted for the errors-in-variables problem. Issues with omitted variable bias has, as with any traditional regression models, not been solved, and proved to be a very relevant issue in these estimations. Essentially, it has been highlighted that the estimated risk premia prices are differing according to amount of control factors. To combat this issue, Giglio and Xiu (2021) introduces a three-step procedure utilizing the thoughts from the Fama-Macbeth regressions while accounting for omitted variable bias by deploying principal components analysis across the original set of returns. This will be elaborated upon in the following subsubsection.

4.5.4 Three-step Procedure

Giglio and Xiu (2021) introduces a three-step procedure with a purpose to estimate factor risk premia robust to omitted variable bias. They deploy principal components analysis to identify the underlying directions of variance in the returns, and uses these principal components in the second and third step with Fama-Macbeth regressions to estimate the factor risk premia.

A different approach with a goal of discerning between actual risk factors and spurious factors not related to the cross section of asset returns is proposed by Pukthuanthong et al. (2019), where they focus on a distinction between the factors having both a risk premium and a positive correlation with the returns of assets, and factors without a risk premium and no correlation with the asset returns, called quasi factors. They identify the factors based on principal components analysis and the Fama-Macbeth regressions elaborated above.

Giglio and Xiu (2021) only holds if all factors are pervasive across time and are prevalent in all the different returns, as in the factors are strong. Therefore, I will highlight both the risk premia estimates of the included factors, and also evaluate the factor strength. The three step procedure is structured as follows:

Step 1: PCA The first step is an estimation of the principal components of the set of excess de-meaned returns, $\bar{R}$. The returns are comprised of N portfolios with T periods, leaving us with N principal components, denoted $\hat{V} = T^{1/2}\left(\zeta_1, \zeta_2, \ldots, \zeta_{\hat{p}}\right)'$, where the eigenvalues, ζ, is sorted in descending order. To choose the optimal amount of components, $\hat{p}$, Giglio and Xiu (2021) introduce the following function:

$$\hat{p} = \arg\min_{1 \le j \le p_{max}} \left(\zeta_j + 0.5 \cdot j \cdot (\log(N) + \log(T)) \cdot \left(N^{-0.5} + T^{-0.5}\right) \cdot \operatorname{median}\left(\zeta_1, \zeta_2, \ldots, \zeta_{\hat{p}}\right)\right) \tag{4.15}$$

where p_{max} has been set equal to $N-1$. The purpose of this step is to identify the main directions of variance within the returns, so we have a space upon which we can project our observable factors. This space should in theory not necessarily be spanned by all initial principal components, but only a subset of the ones with highest marginal variance. This choice of $\hat{p}$ relevant components are, however, not entirely robust, and I will therefore conduct all of the analyses across $\hat{p}$ until $\hat{p}+2$ components included. The determination of $\hat{p}$ leads to an indentification of $\hat{V}$, which allows me to estimate the factor loadings as:

$$\hat{\beta} = T^{-1}\bar{R}\hat{V}' \tag{4.16}$$

where $\hat{\beta}$ is a N times $\hat{p}$ matrix of factor loadings.

Step 2: Cross Sectional Regression The first and second steps resemble the Fama-Macbeth regressions. I will start by estimating the cross sectional regression of the de-meaned returns upon the factor loadings, $\hat{\beta}$, but as a slight change, I will add an intercept to the regression, to allow the zero-beta rate to differ from the risk-free rate. This is just to give more flexibility to the model and allow for a potentially significant premium. The intercept is estimated as:

$$\hat{\gamma}_0 = \left(\iota_N' M_{\hat{\beta}} \iota_N\right)^{-1} \iota_N M_{\hat{\beta}} \bar{r} \tag{4.17}$$

and the remaining coefficients for the latent factors are estimated as:

$$\hat{\gamma} = \bar{G}\hat{V}' \left(\hat{V}\hat{V}'\right)^{-1} \left(\hat{\beta}' M_{\iota_N} \hat{\beta}\right)^{-1} \hat{\beta} M_{\iota_N} \bar{r} \tag{4.18}$$

where

$$M_{\iota_N} = I_N - \iota_N \left(\iota_N' \iota_N\right)^{-1} \iota_N'$$

and

$$M_{\hat{\beta}} = I_N - \hat{\beta}\left(\hat{\beta}'\hat{\beta}\right)^{-1} \hat{\beta}'$$

with $\bar{G}$ being the demeaned value of the observed factors, ι_N being a conforming vector of ones with length N and I_N being the identity matrix of size N. This step then leaves us with the estimated risk premium of the latent factors. Which I will then use to estimate the loadings upon the observable factors to find their risk premium in the third and last step.

Step 3: Time-Series Regression The last step will map the principal components of the first step onto the observable factors, $\hat{\eta}$. Taking the product of $\hat{\eta}$ and $\hat{\gamma}$ will then provide us with the risk premiums for the observable factors, $\hat{\gamma}_g$. Therefore, I will start by estimating the mapping:

$$\hat{\eta} = \bar{G}\hat{V}' \left(\hat{V}'\hat{V}\right)^{-1} \tag{4.19}$$

and finally taking the product to get the risk prices:

$$\hat{\gamma}_g = \hat{\eta} \cdot \hat{\gamma} \tag{4.20}$$

which should leave me with consistent estimations of the risk premiums for the observable factors. The robustness will, of course, be taken into account, as the conclusions are only asymptotically valid when T and N goes towards infinity. Therefore, an improvement to the analysis could be done by adding numerous portfolios formed on different asset classes and characteristica to the returns included. The results are, nonetheless, a relevant supplement for the Fama-Macbeth regressions, as they require the model to include all relevant factors.

To test the statistical significance of these results, I use HAC standard errors, to take any autocorrelation and heteroscedasticity problems into account. The statistics are then found using a simple t-statistic. The problems that arose in the Fama-Macbeth regressions with the errors-in-variables are proven to not be an issue here, and therefore I will not deploy anything other than the HAC standard errors.

Besides just inferring from the risk premias and their significance, I will also look at the factor strengths.

Factor Strength The strength of the factors will be estimated through a statistically significance evaluation of η and an assessment of the explanatory power of the regression of the factors upon the principal components. The individual factor strength is evaluated through a Wald test with the following hypothesis:

$$\begin{aligned} H_0 &: \eta = 0 \\ H_1 &: \eta \neq 0 \end{aligned} \tag{4.21}$$

with the test statistic:

$$\hat{W} = T\eta \left(\left(\hat{\Sigma}^v\right)^{-1} \hat{\Pi}_{11} \left(\hat{\Sigma}^v\right)^{-1} \right)^{-1} \hat{\eta}' \xrightarrow{d} \chi^2_{\hat{p}} \tag{4.22}$$

where Σ is the covariance matrix of the factors and Π is the covariance matrix between the error terms of Equation 4.16 and Equation 4.19. And the explanatory power of the regression is calculated as:

$$\hat{R}^2_g = \frac{\hat{\eta}\hat{V}\hat{V}'\hat{\eta}'}{\bar{G}\bar{G}'} \xrightarrow{p} R^2_g \tag{4.23}$$

This three-step procedure provides a way to remove any omitted variable bias. This is pertinent, because I cannot account for all relevant risk factors in a simple model. A model estimated with a clear omission of relevant observable factors, will have biased estimates of the β's and in the second step regression a biased γ, which essentially makes the inference based on the model's estimations wrong and misleading. I would, preferably, have a good variation of strong factors in the model, of which the included signal might be one of them. My a priori assumption is that the signal should be priced, and that there is a positive correlation between the value of the implied volatility spread and the following week's return. This could be either due to short-sale limitations and investor's private information therefore affecting the option market first, or a relevant skewed risk profile of the stock, which is realised in the following week. If I am just estimating the Fama-Macbeth regressions, without accounting for a selected relevant priced (and strong) factors, I would risk assigning some risk premium to my implied volatility spread signal, that instead belonged to some other factor, fx the PIN factor of Pan and Poteshman (2006) or liquidity factors.

After this statistical analysis of significance and factor strength, I deem it necessary to also look at the economic significance of the factor. An investor would understandably be more interested if the signal has an historic effect of providing stable excess returns, and economic arguments as to why it should continue to offer a relevant premium for the investor in the future. Therefore, I will in the following subsections elaborate briefly upon the economic significance of the signal, and how it can compliment an already existing investment portfolio.

4.6 Economic Evaluation

For a more nuanced view on the effect of the implied volatility spread, I will also conduct a brief economic evaluation. The analysis is conducted ex-post and all metrics is based on the entire sample period. Which essentially means, that I am just looking back at how a rational investor might have invested using the portfolios formed on the implied volatility spread. This approach also invites critique regarding the estimation methods, as they are based on assumptions of a constant covariance matrix among the superset of portfolios throughout the entire sample period.

As mentioned above, it is relevant to test if the factor and portfolios are statistically significant, but the gain for a rational investor and their portfolio might be even more relevant in the 'real' world. This is due to the fact, that even though a portfolio might have a statistical significant risk premium, a factor or signal might provide better diversification opportunities for investors. Furthermore, a rational investor might decide that an exposure towards the option markets through the portfolios formed here, will be of interest to them.

I will start by plotting the portfolios' mean and standard deviation and compare this with traditional Fama & French portfolios on size and value. Building on this, the efficient frontier will be estimated and plotted and lastly I will evaluate the allocation to the portfolios estimated earlier given different levels of constraints upon the individual portfolios in the superset of portfolios considered.

Expected utility has been a central aspect of asset pricing and economic evaluation. In this context, however, I see the optimal portfolio allocation as more pertinent to the topic, and utilizing utility frameworks would require me to make a choice regarding a relevant utility function. A relevant utility function would either be the frequently used power utility or similar, which does not take the loss aversion inherently imbedded in the negative value of the implied volatility spread into account. And as a natural consequence of a lack of expected utility evaluation, the certainty equivalent wont be estimated either.

To preface the economic evaluation, it is important to keep in mind that these conclusions are not directly transferrable to the investor's world, as the entire analysis is ex-post, and that no market frictions, such as transaction costs, integer constraints, and shorting constraints, have been taken into account.

I am also only evaluating the investment as a sigle-period investment, with an assumption that all means, standard devations and correlations are constant.

5 Cross Sectional Analysis

The section will present the results of different cross sectional analyses of the signal against the stock returns. Different portfolio sorts will be deployed to give the best possible view of the complete picture. Specifications of the different portfolio sorts are shown in Table B3.

The portfolio sortings will include results from both univariate and bivariate sortings, to illustrate the differences in taken either or both signals into account. Note, that the results in the unconditional analysis will be considering a more diverse set of portfolios to illustrate the choices made in the portfolio formation.

This section is divided into a brief unconditional analysis, which will highlight the descriptive statistics of different portfolio sorts and evaluate the monotonicity in the returns of the portfolios. Following the unconditional analysis is the conditional analysis, which introduces additional external factors into the analysis, with an evaluation of the factor competition, priced risk factors, and explanatory power of the portfolios with the new factors. Lastly, I will illustrate the ex-post economic evaluation through looking at optimal allocation given allocation constraints and forming the efficient frontier.

5.1 Unconditional Analysis

The descriptive statistics of the portfolios are great at providing an initial understanding of the formation. First I would like to compare two univariate portfolio sorts with the two different signals, so as to argue in combination with an economic motivation as to why I choose to have a bivariate sorting with the first sorting variable being the value of the implied volatility spread, and the second sorting variable being the change in the implied volatility spread. In addition to this I find it relevant to relate these two portfolios to a bivariate dependent portfolio sort. All of this is shown in Table 5.1.1.

I notice, first and foremost, that the mean return show a larger difference between the first and last portfolio in the first subplot (the portfolios based on the value of the implied volatility spread) than those of the second subplot (with the portfolios sorted on the change in implied volatility spread). The standard deviation, kurtosis and skewness seem approximately the same, and as a natural effect from this, the Sharpe Ratio shows approximately the same pattern as the returns, with both higher and lower levels obtained in the sorting based on level.

When I look further down at the dependent bivariate portfolio sorting and compare the descriptive statistics of the 'corner'[7] portfolios with two univariate portfolio sets, I note that the returns show more variance, but also that the standard deviation of the returns seem in general to be higher. Furthermore, the skewness is also higher, but the kurtosis seems at the same level. Of course, all the portfolios of the bivariate sorting has not been reported here to provide a more concise overview.

The distributions of the returns are showing a close-to normal distribution. The mean is near 0, the

[7]The 'corners' are just the portfolios in the intersection between the first and last portfolios of the two sorting variables.

TABLE 5.1.1: Descriptive Statistics of 5 Univariate Equal-sized Portfolios Sorted on the two Signals

(A) Signal : Level of Implied Volatility Spread last trading day

stat	1	2	3	4	5
Mean	-0.0254	0.0816	0.1775	0.2841	0.3915
Std	3.2318	2.6529	2.4390	2.5787	3.0877
Skew	-0.5685	-0.6762	-0.2319	-0.1191	0.0802
Kurt	8.0285	7.8741	8.5110	7.9112	9.4463
Sharpe Ratio	-0.0079	0.0307	0.0728	0.1102	0.1268

(B) Signal : Change in Implied Vol Spread during last 7 days

stat	1	2	3	4	5
Mean	0.0435	0.1295	0.2047	0.2407	0.3438
Std	3.1298	2.5518	2.4225	2.5979	3.0649
Skew	-0.3427	-0.4585	-0.2915	-0.2710	-0.0557
Kurt	8.4224	7.2910	7.6734	7.7292	8.7931
Sharpe Ratio	0.0139	0.0508	0.0845	0.0927	0.1122

(C) Bivariate Sorting : first: level, second: change, Dependent

stat	1_1	1_5	5_1	5_5
Mean	-0.1003	0.0419	0.2928	0.4456
Std	3.9400	3.6117	3.4557	3.5492
Skew	-0.2631	-0.2679	0.3267	0.1873
Kurt	9.1387	8.3976	9.2525	8.2882
Sharpe Ratio	-0.0255	0.0116	0.0847	0.1255

Note: Table shows the descriptive statistics measured over the entire sample period. The options fullfill the conditions outlined in Section 3, the implied volatility spread is weighted by open interest in every combination of date and stock, and the portfolios are equal-sized and have value-weighted returns. ScenarioIDs are 1, 10 and 2 respectively.

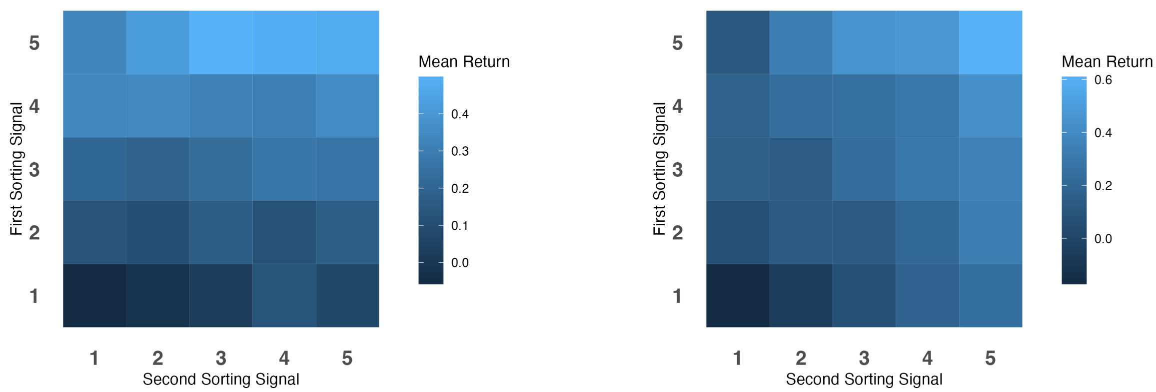
strandard deviation is varying across the portfolios, but seems to be centered around 3, the skewness is within -2 and 2[8], and the kurtosis is around 8, however, this is a little high to fall within the boundaries of a normal distribution.

To illustrate the dynamics in the bivariate portfolio sorting, the mean returns are portrayed in Figure 5.1.1. In subplot (A) there is a clear pattern of the biggest variation in mean return coming from the first sorting which is the level of the implied volatility. Subplot (B) shows a more symmetric pattern, but it still has a tendency of in general having a higher mean return in group 5 of the level sorting. As these scenarios have the level of the implied volatility as the axis of primary variance, it makes sense to focus at the level of the implied volatility spread compared to the change and use this as a signal, as both the economic arguments and patterns of mean returns support this.

[8] According to Byrne (2013) a rule of thumb is that a skewness between -2 and 2, and kurtosis between -7 and 7 means the data is normally distributed.

Figure 5.1.1: Plot of the mean weekly returns of Bivariate Portfolio Sorting

(A) Bivariate Sorting with Level and Change (B) Bivariate Sorting with Change and Level



Note: Figure shows the mean weekly excess return over the entire sample period in percentage. Both sortings are dependent. The scales of the two plots differ, and subplot (B) has a wider scale. The returns are value-weighted. ScenarioID is 2 and 18 respectively.

In the following, I will focus on a bivariate sorting with the level of the implied volatility spread, and the recent change in the implied volatility spread as the second sorting factor. The sorting will be dependent, to ensure that the portfolios will be equal-sized throughout the sample period. All returns are value-weighted[9]. In most cases, a focus will be on the 10 times 10 portfolios as they provide the most sensitive factor, but when it makes sense a smaller set of portfolios with the same choices will be shown. The 10 times 10 portfolios have a ScenarioID of 9, the smaller set of 5 times 5 has a ScenarioID of 2 and the last set of 3 times 3 with a split of 30%, 40% and 30% has a ScenarioID of 14.

Table 5.1.2: Three different Monotonicity Tests

category	monoton_up	monoton_down	Wolak	Bonferroni
increasing	0.0000	1.0000	0.9930	1.0000
decreasing	0.0000	1.0000	0.7400	1.0000

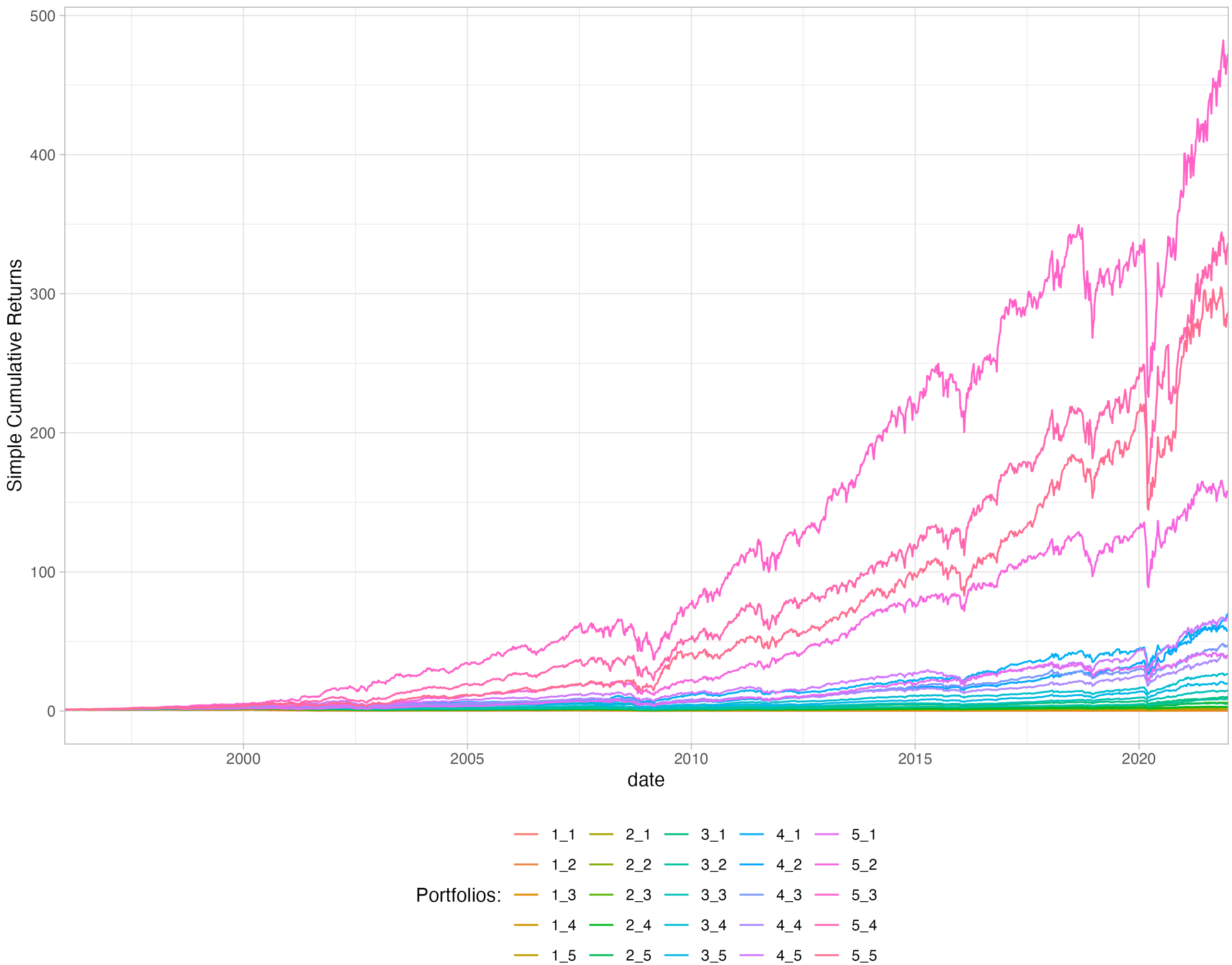
Note: Table shows p-values for three different monotonicity tests, including the Patton and Timmermann (2010) with both a test for increasing and decreasing monotonicity, and two selected others. ScenarioID is 9.

Monotonicity For the monotonicity test of the portfolios, the diagonal set of portfolios are evaluated against eachother[10]. For the abovementioned portfolio set of 10 times 10 bivariate portfolio split, the diagonal portfolios' return are clearly increasing in the value of the level and of the change throughout the entire sample period. This can be seen in Table 5.1.2 through the p-value of monoton-up is approximately 0, which means we reject the null-hypothesis of no monotonicity between the returns, as outlined in Equation 4.9. The remaining two tests have a null hypothesis of a monotonic relation between the portfolios, and they both provide a fairly high p-value, which clearly indicates that they cannot reject the monotonic relationship

[9]The choice of value-weighted vs. equal-weighted is elaborated upon in the following paragraph.

[10]This is due to the function deployed in the scripts are only capable of considering 15 portfolios at a time, so to make the analysis coherent across portfolios formations, this decision was taken.

Figure 5.1.2: Cumulative Simple Weekly Returns, over entire sample period



Note: The cumulative returns are calculated as a simple cumulative product of the return over the entire sample period. The ScenarioID is 2.

between the return of the portfolios. Based on this I can conclude, that not only do I see a positive correlation between the implied volatility spread and the returns, but I also see it across most of the periods and across the portfolio formations.

Cumulative Returns Inspecting the cumulative returns of the portfolios in Figure 5.1.2, and limiting it to scenarioID 2, which has a 5 times 5 bivariate sorting with the same characteristica as scenarioID 9, I can clearly see that the last group in the first sorting dimension provides the highest cumulative returns. The COVID-19 uncertainty hit all of the portfolios, but especially the last portfolio (5_5) seems to recover fast and provide an even steeper climb in the returns afterwards.

Comparing Value-Weighted and Equal-Weighted Returns In this short note on the use of value-weighted returns in the portfolios, I will compare the already shown portfolios with their equivalent equal-weighted portfolios, so any discrepancies or doubts regarding the significance of the results can be disproved.

5.2 Conditional Analysis

Following the unconditional analysis of the portfolios, I will now expand the analysis to include factors from the literature, to evaluate if the long-short zero-net-investment factor formed on the portfolios has a statistical significant risk premium, and if this factor is explained by some common factors from the literature.

The subsection will start with the spanning regressions to evaluate the factor competition, then move on to Fama-Macbeth regressions of both time-invariant and time-variant versions with differing amount of extra factors, to see the effect and evaluate any changes in the coefficients. To supplement the Fama-Macbeth regressions, the three-step procedure is then implemented, and the results of the risk premia estimations as well as the strength of the models are expanded upon. Lastly I will do a brief analysis of a pooled regression with the relevant portfolios regressed upon dummies for the portfolio index and the different factors considered up until then.

5.2.1 Factor Competition

The factor competition is a fairly simple regression of the different factors included regressed upon the remaining factors. I do this to see if each factor is explained by the remaining and therefore considered irrelevant.

As shown in subsection 4.3 in Figure 4.3.1 the intercepts of factor formed on different univariate portfolios show fairly consistently that the remaining factors do not explain the average of the long-short factor formed on the level of the implied volatility spread. It is, however, interesting to see if both the long-short factor and the diagonal version have intercepts different from zero.

Table 5.2.1: Factor Competition of Long-Short Portfolios against FF5 factors

Dep. Variable	Intercept	IMPVOL	IMPVOL_D	MKT_RF	SMB	HML	RMW	CMA
IMPVOL	1.04 (***)		0.04	0	-0.07	0.12	-0.08	-0.08
IMPVOL_D	0.54 (***)	0.03		0.06	0.11	-0.04	0.22 (*)	0.14
MKT_RF	0.22 (***)	0	0.02		0.14 (**)	0.27 (***)	-0.55 (***)	-0.5 (***)
SMB	0.03	-0.01	0.01	0.05 (**)		-0.08 (**)	-0.34 (***)	0.06
HML	-0.05	0.02	-0.01	0.11 (***)	-0.09 (**)		0.2 (***)	0.6 (***)
RMW	0.1 (**)	-0.01	0.02 (*)	-0.12 (***)	-0.22 (***)	0.1 (***)		0.2 (***)
CMA	0.04	0	0.01	-0.07 (***)	0.03	0.2 (***)	0.13 (***)	

Note: Table shows the coefficients of the regressions according to Equation 4.12. The regression span the entire sample period and consists of weekly returns. The significance of the coefficients are coded according to the p-value: $0 < (***) < 0.001 < (**) < 0.01 < (*) < 0.05$. The scenarioID is 9.

In Table 5.2.1 the factor competition coefficients are shown. The FF5 factors are clearly having an effect upon eachother, but their relation to the estimated factors here are statistically not different from zero. Furthermore, a surprising observation is that the coefficients between the implied volatility spread factor and the diagonal version is not significant, which means that the two sorting factors are complimenting eachother in the formation. This will be explored further in the following sections.

Note that the results in this subsection is not directly comparable to the ones reported in Fama and French (2015), as the frequency is different. These results are based on a slightly higher frequency of weekly returns compared to the monthly returns which constituents the main focus of their article.

To evaluate the factors against more than just the FF5 factors, I have included results in Table B1 of the factor competition with all factors outlined in Section 4 and this provides the same picture as above, with the intercept of the implied volatility spread factors being statistical significant different from zero at a fairly high degree and only a few of the added factors have coefficients different from zero at 10% certainty. Based on this, I deem it relevant to use this factor in the further analysis, as it provides a new dimension of variability not captured by the remaining factors.

5.2.2 Fama-Macbeth Regressions

The Fama-Macbeth regressions are made to get an estimate for the risk premium. LS denotes the long-short factor, and LSD denotes the diagonal version. The first two sets of regressions will be based on the portfolios from ScenarioID 9 with a bivariate dependent portfolio sort of 10 times 10 portfolios sorted on level and then change in the implied volatility spread. The factors are limited to the long-short factors of the portfolios and on the median level of the implied volatility spread within each portfolio at each point in time.

Table 5.2.2: Fama-Macbeth Regressions, using Median of the signal as Factor

(A) Time-Invariant

term	gamma_hat	gamma_hat_se	tstat
(Intercept)	0.1584	3.5398	0.0448
LS	1.3958	4.5724	0.3053
LSD	0.7398	4.1664	0.1776
median	-0.0006	0.0621	-0.0090

(B) Time-Variant, rolling window = 52 weeks

term	gamma_hat	gamma_hat_se	tstat
(Intercept)	0.1747	2.7459	0.0636
LS	0.6457	4.7647	0.1355
LSD	0.4632	4.4420	0.1043
median	0.0007	0.0156	0.0429

Note: The regression span the entire sample period and consists of weekly returns. The rolling window used for subtable (B) has a lenght of 52 weeks. The scenarioID is 9.

Simple and Time Invariant Starting with the time-invariant version, in which I assume that the exposure of the portfolios against the included factors are constant in the sample period and the risk premium is also constant. The results are shown in Table 5.2.2. The coefficients are all insignificant, as the t-statistics for each and every one of them are less than 0.5 in absolute value. Furthermore, the median actually has a negative coefficient, which seems at odds with the conclusions from above, where I identified a clear positive correlation between the mean returns of the portfolios and the sorting variable in both the monotonicity tests and the descriptive statistics. This could, however, be due to the LS factor catching the increasing effect and then the median correcting for the less extreme variations in the middle portfolios.

Simple and Time Variant Shifting the focus to the second part of Table 5.2.2, the coefficients for the factors have nearly halfed in size and are even further from being significantly different from zero. The median has, however, now a more intuitive positive coefficient.

The Fama-Macbeth regressions do require, however, that all priced factors are included in the regression. Therefore, I will expand the analysis to take FF5 factors into account, and hope to gain some more meaningfull coefficients (and risk premiums) from this[11].

TABLE 5.2.3: Fama-Macbeth Regressions, using FF5

(A) Time-Invariant

term	gamma_hat	gamma_hat_se	tstat
(Intercept)	0.2401	4.3444	0.0553
CMA	0.2344	2.7320	0.0858
HML	1.1102	5.8846	0.1887
LS	1.3198	4.5334	0.2911
LSD	0.7027	4.0196	0.1748
median	0.0015	0.0315	0.0483
MKT	-0.0644	7.3994	-0.0087
RMW	0.1573	2.1711	0.0724
SMB	0.0753	4.1812	0.0180

(B) Time-Variant, rolling window = 52 weeks

term	gamma_hat	gamma_hat_se	tstat
(Intercept)	0.1200	1.8007	0.0666
beta_CMA	0.0372	1.1131	0.0335
beta_HML	0.0378	2.0016	0.0189
beta_LS	663360244.9724	24137563699.8782	0.0275
beta_LSD	-98721798.5894	3592171423.8114	-0.0275
beta_median	0.0013	0.0147	0.0889
beta_MKT	0.0605	3.0270	0.0200
beta_RMW	0.0115	1.4364	0.0080
beta_SMB	0.0355	1.9514	0.0182

Note: The regression span the entire sample period and consists of weekly returns. The ScenarioID is 9.

FF5 and Time Invariant To accommodate the requirements of the Fama-Macbeth regression framework, I have added five additional factors to the regression. The results are shown in Table 5.2.3, and as with the results of the more limited regression, none of the coefficients are statistically significantly different from zero. And additionally, the market factor seems to have an estimated negative risk premium. This is at odds

[11] The reason for the exclusion of the Open Asset Pricing Factors is the lower frequency of their availability, as they are only reported on a monthly basis.

with the expectation, as the market risk premium should be positive as the market excess return has been positive on a historic basis and is positive in the estimation made by Giglio and Xiu (2021).

FF5 and Time Variant Shifting the focus to the time-variant version of the Fama-Macbeth regressions, two very large risk premium estimates stand out. It seems that the coefficients have been affected by some of the data missing or being very alike in value resulting in the OLS estimator finding an unlikely high number. The standard errors are equally as large and the resulting t-statistic is within the same range as the remaining. This might also be a sign of colinearity issues between the factor and its diagonal version, which also makes sense given that the risk premium of the diagonal version is suddenly estimated to be negative. The remaining coefficients have the expected sign, as all factors have a positive risk premium.

All-in-all the results from the Fama-Macbeth regressions left no clear conclusion, as the risk premiums were insignificant, which is in contrast to results from the literature. I will go forward with the three-step procedure to see if this approach can identify the underlying variations of the portfolios and map this to the factors included.

5.2.3 Three-Step Procedure

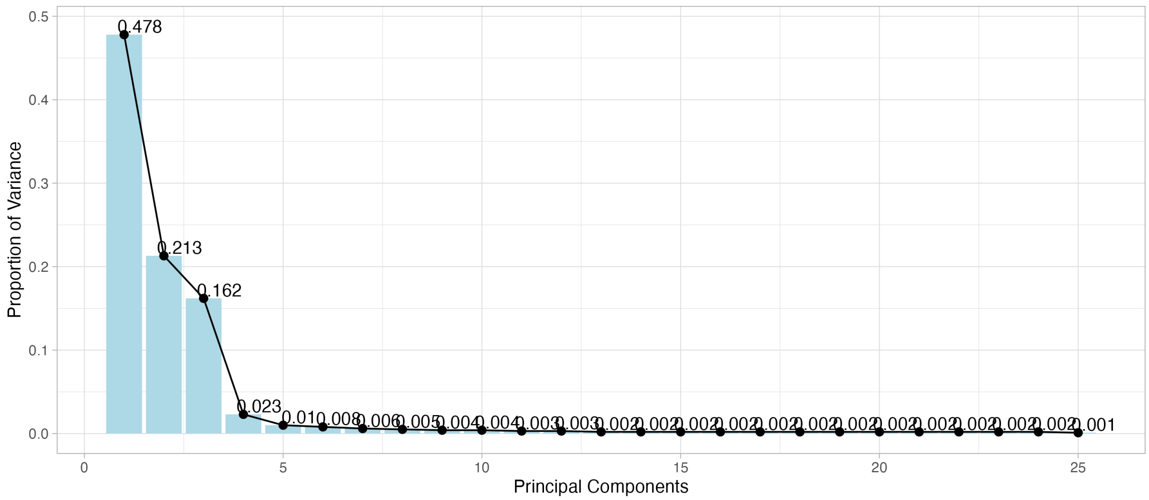
The three-step procedure was introduced to remove the omitted variable bias and to exploit the deployment of principle components analysis to find the main dimensions of the variance of the returns.

The returns set I am deploying here is the 100 portfolios estimated in Scenario 9, with bivariate dependent portfolio sorting and the set of 100 portfolios from Kenneth French's Library based on size and value. Note that no other portfolios are considered, even though the authors argue for the inclusion of portfolios exposed to all asset classes.

The first step is to estimate the main points of variance in the returns through principal components analysis. Using the cost function outlined in Section 4 in Equation 4.15 and results in an optimal choice of relevant dimensions to be three for this analysis. The explained variance of each of the first 25 principal components are shown in Figure 5.2.1, where the first principal component clearly outperforms the remaining. The choice of the first three principal components to be used further in the analysis is, of course, subject to critique and I will therefore include the results of 1 principal component up until the optimal plus two. This way we can see the robustness in the coefficients across different choices for the principal components.

In Table 5.2.4 the risk premia estimates are reported along with their statistical significance for each choice of principal components. As the optimal value for $\hat{p}$ is 7, I note that only the long-short factor based on the portfolios is significant, and that only at a 5% level. The factor stays somewhat significant through the robustness versions, as can be seen in column 8 and 9. The results from the regression seems robust to including more principal components, as the coefficients and significance levels are fairly constant across the different runs. None of the remaining included factors besides the long-short factor based on the portfolios are significant, which could be explained by the limited set of portfolios used in the regression, as Giglio and

Figure 5.2.1: Proportional Explained Variance of Principal Components



Note: the figure shows the explained proportion of variance of each principal component of the set of returns of 100 portfolios from ScenarioID 9 and 100 size-value portfolios from Kenneth French.

Xiu (2021) argues that a variety of different portfolios should be included in the original estimation.

Table 5.2.4: Coefficients from the Three-Step Procedure, using FF5

Factor	1	2	3	4	5	6	7	8	9
Intercept	0.218 (**)	0.177 (**)	0.200 (**)	0.305 (***)	0.384 (***)	0.219 (***)	0.230 (***)	0.228 (***)	0.259 (***)
MKT_RF	-0.008	0.01	-0.013	-0.074	-0.244	-0.157	-0.162	-0.179	-0.214
SMB	-0.001	-0.017	-0.015	-0.008	-0.149	-0.17	-0.169	-0.183	-0.184
HML	0.001	0.007	0.022	0.036	-0.094	-0.1	-0.099	-0.111	-0.113
LS	0.001	-0.003	-0.009	0.004	0.531 (*)	0.57 (**)	0.572 (**)	0.612 (**)	0.701 (**)
LSD	0	0	0	-0.003	0.112	0.079	0.076	0.094	0.046
RMW	0.004	0.003	0.007	0.018	-0.054	-0.091	-0.09	-0.097	-0.099
CMA	0.002	-0.001	0.009	0.028	-0.103	-0.13	-0.128	-0.136	-0.138

Note: Table shows the coefficients of the regressions according to Equation 4.20. The columns denote the amount of principal components used in the estimation, the optimal $\hat{p}$ is 7, columns higher than 7 therefore constitutes the robustness checks. The regression span the entire sample period and consists of weekly returns. The significance of the coefficients are coded according to the p-value: $0 < (***) < 0.001 < (**) < 0.01 < (*) < 0.05$. The standard errors are reported in Table B2 in the Appendix. The ScenarioID is 9.

The coefficients of this regression seem quite unintuitive, as all FF5 factors are negative across the robustness tests, which is in contrast to the reported risk premia estimates of Giglio and Xiu (2021) in their Table B2. They clearly report positive risk premias and a slight significance in the test of the risk premia being zero. The only difference between their approach and mine is their exclusion of the intercept and instead assuming the zero-beta rate (which is the intercept) being equal to the risk free rate.

Moving on to evaluating the factor strength, it seems intuitive that none of the models should provide a high explanatory power, as only the intercept and one of the factors are remotely significant. Therefore, I would expect the R_g^2 to be low and thus the test of the strength of the factors is not relevant, given the risk premiums are close to zero. Examining the p-values of the Wald test of the factors being weak, I reject the

Table 5.2.5: Explanatory Power and Wald tests from the Three-Step Procedure, using FF5

term	1	2	3	4	5	6	7	8	9
R2_G	0	0	0	0	0	0	0	0	0
MKT_RF	0	0	0	0	0	0	0	0	0
SMB	0.3422	0	0	0	0	0	0	0	0
HML	0.3569	0	0	0	0	0	0	0	0
LS	0.9561	0.363	0.572	0.0475	0	0	0	0	0
LSD	0.9955	0.9102	0.3642	0.7695	0.0022	0	0	0	0
RMW	0	0	0	0	0	0	0	0	0
CMA	0.0395	0.0028	0	0	0	0	0	0	0

Note: The regression span the entire sample period and consists of weekly returns. First row of data is the R^2 of the factors and their explanatory power over the pincipal components. The remaining rows are p-values for the test of the factor being weak. The ScenarioID is 9.

null hypothesis of the factor being weak in most cases, and all cases close to the optimal $\hat{p}$.

As a final conclusion, the risk premiums of the factors are statistically insignificant different from zero, except the long-short factor of the 10 times 10 portfolio formation. As Fama-Macbeth and this three-step procedure reaches the conclusion of only very low significance of the long short factor with a positive risk premium, I will assume it is reasonably correct. Robustness checks against fewer portfolios in the sorting and a univariate sorting is carried out in the Appendix.

5.2.4 Pooled Cross Sectional Regression

A pooled cross sectional regression is a simple multivariate regression without any time index and only dummies to indicate which portfolio the returns belong to. The results of such a regression allow me to conclude on the general effect of a factor on the returns and in particular how the portfolios mean return differ from one another.

Note that the long-short factor and its diagonal version will not be included here, and that the scenarioID used in this case is 2. The only difference between scenarioID 9 and 2 is the amount of portfolios, as 9 has a 10 times 10 split, and 2 has a 5 times 5 split.

In Table 5.2.6 the coefficients and their significance is displayed for the pooled cross section across the entire sample period. The simple version of the pooled regression (the rightmost column) shows clearly that the main significance is achieved in the first sorting dimension, which is the level of the implied volaitlity. Furthermore, only the first portfolio (1_1) has the lowest expected return, which is equal to the intercept, and then as we get further out from the first portfolio, the average returns are increasing (as seen in the coefficients). The explanatory power of the model is very low at 0.18%, which is to be expected, given that we are only testing if the average portfolio return differs across portfolios over the sample period.

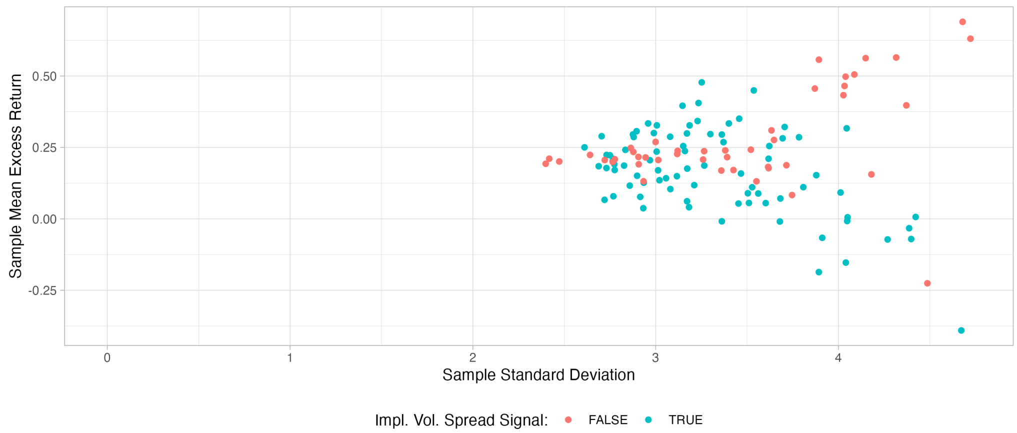
The second right-most column shows the coefficients for an extended pooled regression with the FF5 factors included. The coefficients for the dummies of each individual portfolio (except the intercept) is equal to those of the simple model, but the standard errors have decreased, which leads the most of the coefficients being significantly different from zero at a 1% level. Here it is clear that almost all portfolios has a significantly higher average return than the first portfolio, and the average returns is again increasing in the sorting variables. All the included factors are significant, except the risk free rate. So the portfolio returns are trending with the factors formed on the characteristica, which indicates that the portfolio are positively correlated with the HML, SMB and market factor. As the portfolios both have a clear influence from the market factor, the remaining factor's significance shows it is not only the general tendencies of the market affecting the portfolios, but also the particular exposure towards the already known risk premias. The explanatory power of this model vastly improves from the previous one, as it reaches 79.07%. This makes sense, as the new factors changes across time and includes the general market movements.

The first column denotes the entire model with all available factors included. The coefficients seem mostly unchanges, with the significance, sign and magnitude of the coefficients fairly equal to the ones reported in the 'middle' model. Of the few new factors included, only three of them show significance. All of the new factors have a positive coefficient, meaning that they are positively correlated with the portfolios. In general all the coefficients of the factors are in absolute values lower than or equal to 0.1 except the market factor. These new factors added in this model have even lower coefficients than the other factors. CoskewACX and OptionVolume2 are the two factors with significant coefficients, which indicates that the coskewness of the equal-weighted returns and the ratio of offered option volume has a positive effect on the portfolio returns. Thus the portfolios has on average a positive correlation with the exposure towards coskewness in the markets, which indicates that the portfolios catch some of the risk premia incurred from skewness in returns. This makes intuitive sense, as some of the economic reasons for using the implied volatility spread as a signal is based on the tail risk of the stock market which is priced in the inherently forward looking option market. The other factor, which coefficient is also rather significant, is the optionvolume2. The significance and positive coefficient of this factor might be because of the selection bias, in that only stocks with traded options are included in the formation of this factor as well as in this project's data, leading to a natural correlation between the returns. Note, that all the additional factors included in the biggest model is only observed at a rather low frequency of monthly returns, and therefore, they are constant across 3 to 4 observations of the weekly returns. This might affect the results, if compared against a similar analysis of monthly returns, as I cannot account for weekly fluctuations in the factors. Lastly, I note that the explanatory factor only shows a slight improvement, as the value has increased from 79.07% to 79.1%.

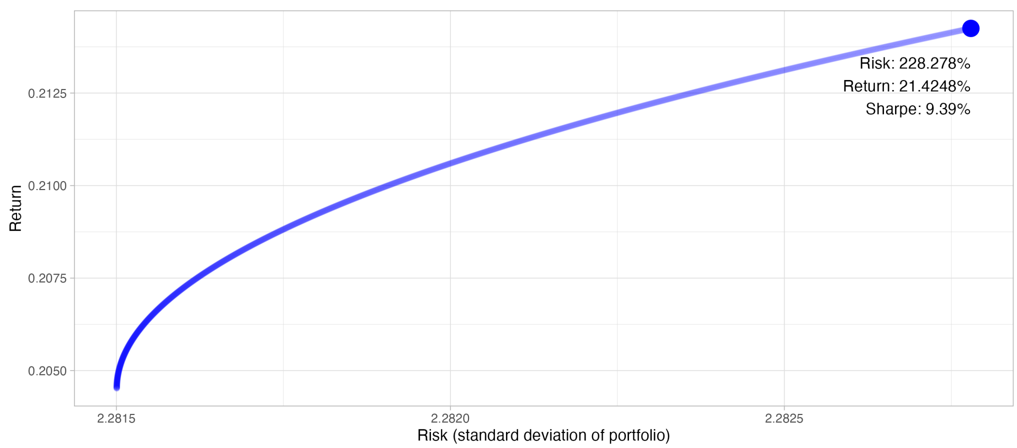
The results of the conditional analysis shows that there is a small priced risk premia as shown in the three-step procedure, and the pooled cross sectional regression show a clear positive correlation between the portfolios and the FF5 factors, and the coskewness and option volume ratio. Through the robustness tests shown in the appendix of the different analyses with adapted number of portfolios and other choices, I can see that the results are stable across the choices made.

FIGURE 5.3.1: Efficient Frontier based on the portfolios and FF5x5 Portfolios

(A) Mapping of Returns and Standard Deviation



(B) Efficient Frontier



Note: The returns are reported as the mean and standard deviation of each portfolio across the entire sample period. The x-axis denotes the standard deviation, and the y-axis denotes the excess weekly return. The first subfigure is colored according to the origin of the portfolios, turquoise for the implied volatility spread and orange for the Fama-French 5 times 5 portfolios on size and value. The ScenarioID is 9.

5.3 Optimal Allocation

Moving on to the economic analysis of the portfolios, I am implicitly assuming a mean-variance optimising investor. Moving forward, this will be the optimal choice as long as the returns are assumed gaussian distributed.

When plotting the portfolios' mean and standard deviation within the sample period against the one incurred by the Fama-French portfolios in Figure 5.3.1 subfigure (A) the portfolios form a distinct pattern. The Fama-French portfolios show an upwards trend of a higher standard deviation warrants a higher return, whereas the implied volatility spread portfolios show a sligthly more blurred picture, with a slight downward trend.

A rational investor would, however, not only let themselves be guided by the isolated mean and standard deviation of a portfolio, but instead evaluate it in the context of their existing portfolios. So in Figure 5.3.1 subfigure (B) I have estimated the efficient frontier upon these 125 portfolios from above, and found identified the point upon the efficient frontier which provides the highest Sharpe Ratio.

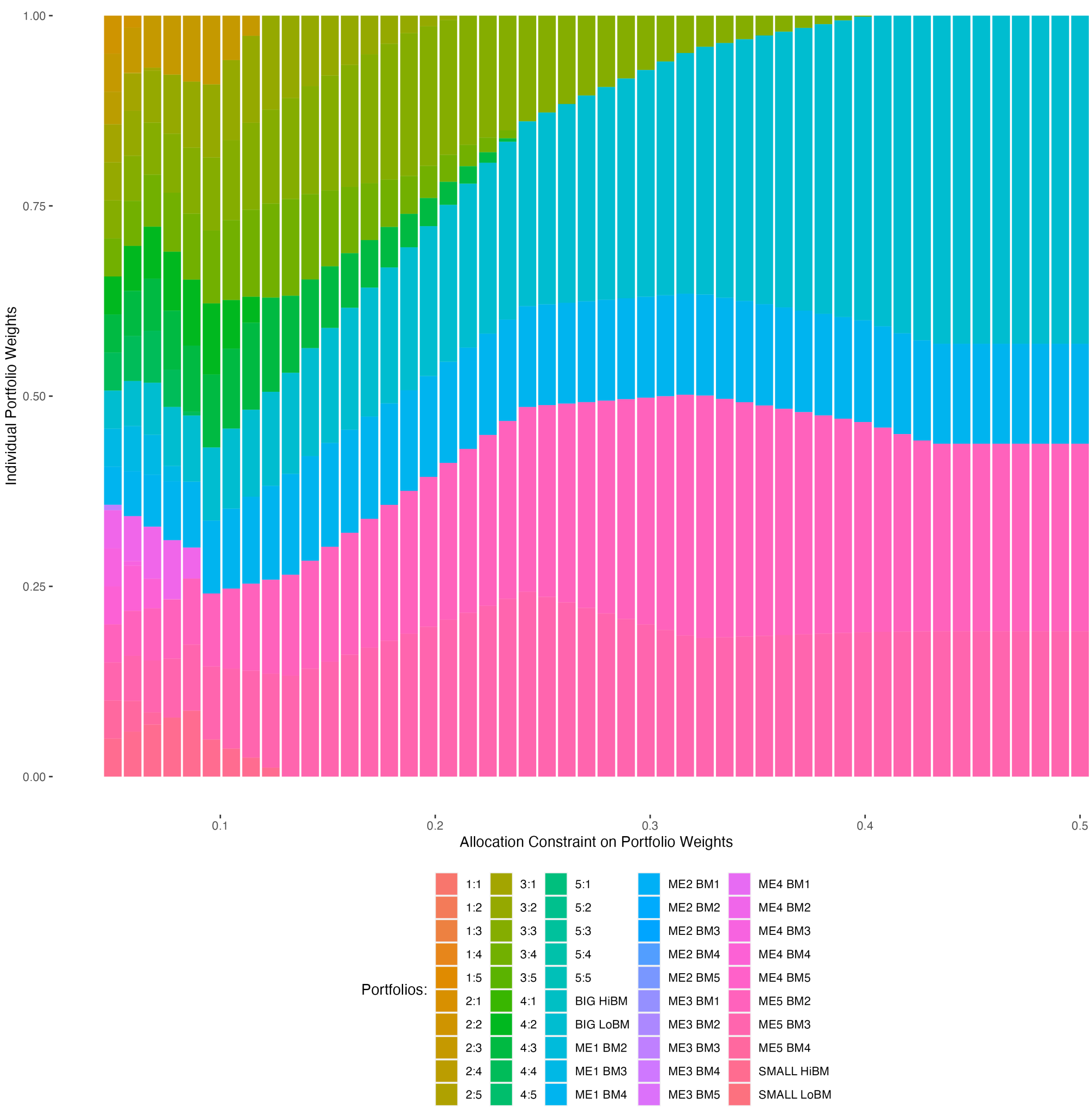
Figure 5.3.1 does not, however, show how much to invest in the individual portfolios, which gives rise to the following analysis of the allocation to each portfolio given differing allocation constraints.

The argument as to why the allocation constraints follow from a rational investor, who of course will only invest in the portfolio upon the efficient frontier with the highest Sharpe ratio, and then allocate their funds between the risky portfolio of assets and the risk free rate to satisfy their risk aversion.

A rational investor might argue against forming the portfolios without any allocation constraints towards the individual portfolios. They could assume that exposure to only one particular portfolio on a subset of assets would not provide them with the diversification they desire. Therefore, allocation constraints raging from 0.05 to 0.5 must be evaluated, and the investor should then choose which of these combinations that provide them with sufficient diversification in their eyes.

This is the reason as to a consideration of the different optimal allocation splits of the portfolio with the highest Sharpe Ratio given a range of allocation constraints for the individual portfolio. The results are shown in Figure 5.3.2.

FIGURE 5.3.2: Optimal Allocation given Allocation Constraints, in combination with FF5x5



Note: The figure shows different investment structures given allocation constraints towards investment in the individual portfolio. All optimal portfolios are the combined portfolio with the highest Sharpe Ratio. The optimal allocation is calculated using shorting constraints. The ScenarioID is 2.

TABLE 5.2.6: Cross Sectional Regression

term	coefficient_all	coefficient_some	coefficient_simple
(Intercept)	-0.3205 (***)	-0.3125 (***)	-0.1003
Portfolio_1_2	0.046	0.0417	0.0417
Portfolio_1_3	0.0885	0.0835	0.0835
Portfolio_1_4	0.2115 (***)	0.1899 (***)	0.1899
Portfolio_1_5	0.1681 (**)	0.1422 (**)	0.1422
Portfolio_2_1	0.1762 (**)	0.1945 (***)	0.1945
Portfolio_2_2	0.1654 (**)	0.1695 (**)	0.1695
Portfolio_2_3	0.2312 (***)	0.2349 (***)	0.2349 (*)
Portfolio_2_4	0.1985 (***)	0.1864 (***)	0.1864
Portfolio_2_5	0.2447 (***)	0.2376 (***)	0.2376 (*)
Portfolio_3_1	0.2771 (***)	0.2738 (***)	0.2738 (*)
Portfolio_3_2	0.2408 (***)	0.2545 (***)	0.2545 (*)
Portfolio_3_3	0.291 (***)	0.2893 (***)	0.2893 (*)
Portfolio_3_4	0.3381 (***)	0.3365 (***)	0.3365 (**)
Portfolio_3_5	0.3284 (***)	0.3218 (***)	0.3218 (**)
Portfolio_4_1	0.4295 (***)	0.4045 (***)	0.4045 (***)
Portfolio_4_2	0.4015 (***)	0.415 (***)	0.415 (***)
Portfolio_4_3	0.3837 (***)	0.3786 (***)	0.3786 (**)
Portfolio_4_4	0.3754 (***)	0.3719 (***)	0.3719 (**)
Portfolio_4_5	0.4311 (***)	0.4233 (***)	0.4233 (***)
Portfolio_5_1	0.4148 (***)	0.3931 (***)	0.3931 (***)
Portfolio_5_2	0.5148 (***)	0.4847 (***)	0.4847 (***)
Portfolio_5_3	0.5873 (***)	0.5619 (***)	0.5619 (***)
Portfolio_5_4	0.5815 (***)	0.5558 (***)	0.5558 (***)
Portfolio_5_5	0.5737 (***)	0.5459 (***)	0.5459 (***)
RF	-0.038	0.2567	
CMA	-0.0809 (***)	-0.0943 (***)	
HML	0.0739 (***)	0.0984 (***)	
MKT	1.0596 (***)	1.0612 (***)	
RMW	-0.1168 (***)	-0.1116 (***)	
SMB	0.1099 (***)	0.1271 (***)	
skew1	0.0035		
betaVIX	0.0021		
CoskewACX	0.0124 (***)		
Coskewness	0.0041		
OptionVolume1	0.0032		
OptionVolume2	0.0162 (***)		
BetaLiquidityPS	0.0063 (*)		
Rsquared	0.791	0.7907	0.0018

Note: The table shows a pooled cross sectional regression of all portfolios formed on bivariate sorting of 5 by 5 dependent sort with the first sorting being on the value of the implied volatility spread and the second sorting being the change. The ScenarioID is 2.

6 Discussion and Further Research

A few points of clarifications and considerations for further research is outlined in this section.

The choice of the singal fell on the value of the implied volatility spread to proxy the mispricings in the option markets, and evaluate how this affected the stock markets. Cremers and Weinbaum (2010) investigated how this relation was a consequence of the informed investors trading in the option markets before moving over to the stock market through the use of the PIN factor of Pan and Poteshman (2006). Furthermore, I wanted to investigate how this relation held when changing the choices made during the portfolio formation and signal estimation. In addition to this, the other relevant factor was included to proxy this change in investor behaviour, as the information in the option markets changed and how it affected the stock markets in the following week.

Additional variations of the implied volatility spread could have been used. Both in this format of portfolios, but also in a more classic asset pricing exercise with an investigation into individual assets. For a different signal estimation in this setting, I could have investigated only the strictly negative implied volatility spreads, and changed all the positive spreads equal to zero. This would have highlighted the emphasis of cheap calls and expensive puts, and shown the effects of loss aversion, assumedly, more in depth. The opposite signal, of only looking at the positive implied volatility spreads, could have been used as a proxy for private information leaking to the option markets first, or investors being interested in exposure towards the volatility of an asset, because of some scheduled news.

Another choice would be the absolute value or the squared value of the implied volatility spread. I assume, however, that there is no risk in stock returns associated with a general mispricing, but more of a directional mispricing.

To supplement these alternative choices of signals, one could also consider state dependency. That is, as the option market has matured throughout the last three decades, the implied volatility spread might provide more information towards the stock market in recent years, as it is more reflective of the market's view. This allows for a state-dependency split of the sample into several subperiods or samples based on the liquidity factor as a proxy for the interest in the option markets.

Regarding the mean-variance model of Markowitz (1952), where we assume that the investors are onlu concerned with the risk associated with the variance of a portfolio return, one could challenge this aspect. Investors have introduced several other risk measures, which might not provide the same analytical formulas for solving the problem, but are able to take these other kind of risk measures into account. Two such risk measures are commonly chosen to be either Value-at-Risk, or the slightly adapted Conditional-Value-at-Risk. To do optimal portfolio allocation, a slightly more sophisticated approach is needed (see for example of CVaR: Rockafellar et al. (2000)). Other approaches could entail the cumulative prospect theory investor, which still operates in the mean-variance space, but takes central concepts from prospect theory into account.

I have mainly based this entire project on simple regressions, and only a marginal use of principal components analysis, which seems at odds with the hype surrounding machine learnings techniques and complex models. I have decided against using neural networks and random forests, based on the complexity and black-box impression they provide. Even though an increasing amount of literature is researching the added explanatory power of such models, see e.g. Weigand (2019).

It could, however, be a relevant consideration if I would need to estimate the Stochastic Discount Factor. As I mentioned in Section 2, the stochastic discount factor is formed in the space between the $\mathbb{P}$ and $\mathbb{Q}$ measure, and an estimation hereof would have been a natural extension of this project.

A last comment, related to the sample period, and inclusion of all data in the analyses. A different approach would have been relevant if looking at predictability, as the present project investigated the whole sample in as one, a different approach would be to test this predictability abilities of the signal within the last few years. This would require a training and test sample split, and an evaluation of how these portfolios performed out of sample, given different model specifications. The results from such an analysis would have been more indicative of a future profit from investing according to the implied volatility spread.

7 Conclusion

Setting out to investigate the relation between the implied volatility spread and stock returns, I expected to find the signal to be priced in the cross section, and the portfolios to provide a clear pattern in the returns. I expected the returns of the tail portfolios to be more extreme and the returns of the middle portfolios to be fairly similar.

Stock returns and option prices are available at a daily granularity through CRSP and OptionMetrics. The sample spans from 1996 to 2021, and consists of all stocks registered at American exchanges which have an option written upon them and traded within the last 7 days. Combining this data into portfolios allows me to remove idiosyncratic noise from the weekly returns. Choices regarding the formation of the portfolios are all highlighted, and alternative results are shown in the appendix to show the robustness of the analysis. The implied volatility spread is mutated into two distinct signals. First is the value of the implied volatility spread of the latest trading day before portfolio formation, and the second is the change within the last 7 days in the implied volatility spread. The first signal estimated as a proxy for the beliefs of the market towards the future distribution of the stock, and the second signal a proxy for new information affecting a change in the expectations to the future stock price.

The cumulative returns of the portfolios are clearly positively correlated with the value of the implied volatility spread. In the unconditional analysis, I find the portfolios' returns to show a clear monotonic relationship, and the long-short factors not to be fully spanned by other external factors. This indicating that informed investors participate in the option markets before moving over to the stock market, resulting in private information affecting the prices in the option markets first.

Alternating the choices made for the portfolio formation, I find that across the amount of portfolios, the amount of sortings, and the choice of signal, the monotonic relationship persists. The cumulative returns of the portfolios increases in the value of the implied volatility spread. Intuitively, this means that a high call price indicates higher probability of the stock increasing in value. This makes intuitively sense, as a high implied volatility spread means a higher priced call, insinuating that the investors believe the stock to increase in value.

In the conditional analysis, the factor competition shows that the long-short net-zero investment factors in the portfolios are not redundant when comparing with Fama & French's five factor model, and when correcting for a subjectively chosen list of factors from Open Asset Pricing. This could be due to the factor proxying investors' private information or it could be due to shorting limitations in the stock market, resulting in investors taking positions in the option markets to get their desired exposure. I am also using both Fama-macbeth regressions and the three-step procedure to evaluate the price of risk. The Fama-Macbeth regressions provide inconclusional results, using either time invariant or time varying exposure, and accounting for just the factor or including all five factors from Fama and French (2015). The three-step procedure did provide a small risk premia of the factor, but the explanatory power of the model was very low, so the evaluation of the factor strength was not relevant. A pooled cross sectional regression provides only the expected relation between the portfolios, and intuitive coefficients for the included factors.

The economic analysis of the portfolios are based on the optimal portfolio allocation of a rational investor, without using the utility framework. When forming the efficient frontier and the optimal allocation given allocation constraints, the portfolios formed on the implied volatility spread were excluded in favor of including the size and value portfolios of Fama & French. Only in the case of imposed allocation constraints were the portfolios allocated a positive weight. Thus in a mean-variance framework, the portfolios seem to be deprecated in favor of other external portfolios.

Using the portfolio optimisation framework, we are assuming the primary risk factor of concern to the investor to be the variance. The implied volatility spread does, however, catch the loss aversion of the investors, as puts are often overpriced. It would make sense to adapt the portfolio formation to take this factor into account, and include thoughts from prospect theory to evaluate the optimal portfolio.

The approach deployed here is subject to intense data snooping, as there is no formal and correct way prespecified. Therefore, I have strived to illuminate all choices made, and provide sufficient alternative results throughout the analysis and in the appendix.

I believe that further research into the maturing of the option markets would yield more nuanced results, compared to the results supplied in this project spanning the entire sample period. Furthermore, it could provide relevant insights, if the signal were adapted to incorporate only the negative spreads, so an analysis was conducted upon the signal of $(IV_{call} - IV_{put})^+$, to proxy for the risk associated with a put being overpriced. This could also be proxied by including the PIN factor, which measures the ratio of buyer-initiated calls and puts, to identify the informed investors and their trading.

In addition to the risk premia estimated through Fama-Macbeth regressions and the three-step procedure, a machine learning appproach for estimating the SDF on the basis of this new factor, would be relevant. This would allow for a mapping between the $\mathbb{P}$ and $\mathbb{Q}$ measure as discussed in the option pricing section.

All in all, I find that the rational investor would not invest in the portfolios, if it were among the ones considered for the efficient frontier, but it does provide a small weekly premium according to the three-step procedure. In general, the results are stable across specific choices made in the signal and portfolio formation, and it seems to be providing some variation not explained by other external factors, which leads me to assume that I have found a somewhat relevant factor of the cross section of stock returns.

References

K. Amin, J. D. Coval, and H. N. Seyhun. Index option prices and stock market momentum. *The Journal of Business*, 77(4):835–874, 2004.

A. Ang, J. Chen, and Y. Xing. Downside risk. *The review of financial studies*, 19(4):1191–1239, 2006a.

A. Ang, R. J. Hodrick, Y. Xing, and X. Zhang. The cross-section of volatility and expected returns. *The journal of finance*, 61(1):259–299, 2006b.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81 (3):637–654, 1973.

D. T. Breeden and R. H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651, 1978.

B. M. Byrne. *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* routledge, 2013.

M. D. Cattaneo, R. K. Crump, M. H. Farrell, and E. Schaumburg. Characteristic-sorted portfolios: Estimation and inference. *Review of Economics and Statistics*, 102(3):531–551, 2020.

A. Y. Chen and T. Zimmermann. Data (march 2022 release). URL `https://www.openassetpricing.com/data/`.

J. C. Cox, S. A. Ross, and M. Rubinstein. Option pricing: A simplified approach. *Journal of financial Economics*, 7(3):229–263, 1979.

M. Cremers and D. Weinbaum. Deviations from put-call parity and stock return predictability. *Journal of Financial and Quantitative Analysis*, 45(2):335–367, 2010.

F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische annalen*, 300(1):463–520, 1994.

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

E. F. Fama. Term premiums in bond returns. *Journal of Financial economics*, 13(4):529–546, 1984.

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

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22, 2015.

E. F. Fama and J. D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636, 1973.

G. Feng, S. Giglio, and D. Xiu. Taming the factor zoo. Technical report, Working paper, 2017.

S. Figlewski and G. P. Webb. Options, short sales, and market completeness. *The Journal of Finance*, 48 (2):761–777, 1993.

S. Giglio and D. Xiu. Asset pricing with omitted factors. *Journal of Political Economy*, 129(7):1947–1990, 2021.

I. V. Girsanov. On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability & Its Applications*, 5(3):285–301, 1960.

C. R. Harvey and Y. Liu. A census of the factor zoo. *Available at SSRN 3341728*, 2019.

C. R. Harvey and A. Siddique. Conditional skewness in asset pricing tests. *The Journal of finance*, 55(3): 1263–1295, 2000a.

C. R. Harvey and A. Siddique. Time-varying conditional skewness and the market risk premium. *Research in Banking and Finance*, 1(1):27–60, 2000b.

J. Isakas and B. Melenberg. The option implied risk-neutral distribution. *Tilburg School of Economics and Management*, 2020.

R. Jarrow and A. Rudd. Approximate option valuation for arbitrary stochastic processes. *Journal of financial Economics*, 10(3):347–369, 1982.

T. L. Johnson and E. C. So. The option to stock volume ratio and future returns. *Journal of Financial Economics*, 106(2):262–286, 2012.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, pages 99–127. World Scientific, 2013.

S. Köstlmeier. *monotonicity: Test for Monotonicity in Expected Asset Returns, Sorted byPortfolios*, 2019. URL `https://CRAN.R-project.org/package=monotonicity`. R package version 1.3.1.

J. Lintner. Security prices, risk, and maximal gains from diversification. *The journal of finance*, 20(4): 587–615, 1965.

H. Markowitz. Portfolio selection, the journal of finance. 7 (1), 1952.

H. M. Markowitz. Efficient diversification of investments–new york, 1959.

R. C. Merton. Theory of rational option pricing. *The Bell Journal of economics and management science*, pages 141–183, 1973.

J. Mossin. Equilibrium in a capital asset market. *Econometrica: Journal of the econometric society*, pages 768–783, 1966.

OptionMetrics. Ivydb us, file and data reference manual. Technical report, OptionMetrics, December 2015. Version 5.0.x.

J. Pan and A. M. Poteshman. The information in option volume for future stock prices. *The Review of Financial Studies*, 19(3):871–908, 2006.

L. Pástor and R. F. Stambaugh. Liquidity risk and expected stock returns. *Journal of Political economy*, 111(3):642–685, 2003.

A. J. Patton and A. Timmermann. Monotonicity in asset returns: New tests with applications to the term structure, the capm, and portfolio sorts. *Journal of Financial Economics*, 98(3):605–625, 2010.

K. Pukthuanthong, R. Roll, and A. Subrahmanyam. A protocol for factor identification. *The Review of Financial Studies*, 32(4):1573–1607, 2019.

R. T. Rockafellar, S. Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

S. A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976.

D. Shang. *Option markets and stock return predictability.* PhD thesis, The University of Arizona, 2016.

J. Shanken. On the estimation of beta-pricing models. *The review of financial studies*, 5(1):1–33, 1992.

W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

J. L. Treynor. *Toward a theory of market value of risky assets.* 1961.

J. Von Neumann and O. Morgenstern. Theory of games and economic behavior. In *Theory of games and economic behavior*. Princeton university press, 2007.

A. Weigand. Machine learning in empirical asset pricing. *Financial Markets and Portfolio Management*, 33: 93–104, 2019.

F. A. Wolak. An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association*, 82(399):782–793, 1987.

F. A. Wolak. Testing inequality constraints in linear econometric models. *Journal of econometrics*, 41(2): 205–235, 1989.

Y. Xing, X. Zhang, and R. Zhao. What does the individual option volatility smirk tell us about future equity returns? *Journal of Financial and Quantitative Analysis*, 45(3):641–662, 2010.

Appendices

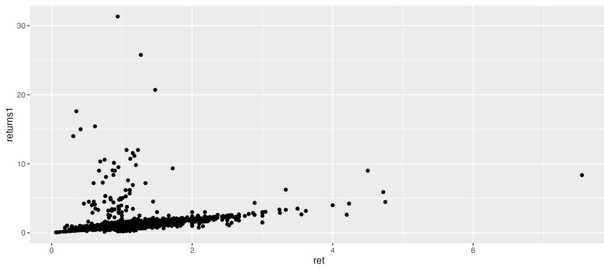
List of Figures

List of Tables

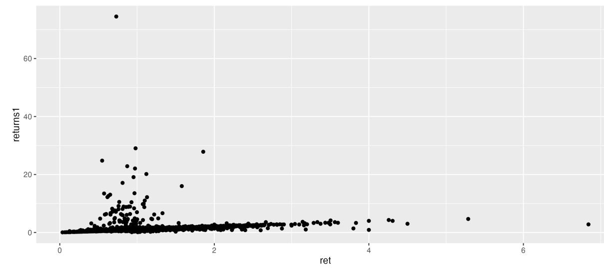
A Figures

Figure A1: Plots of CRSP reported holding period returns against daily close/open retuns

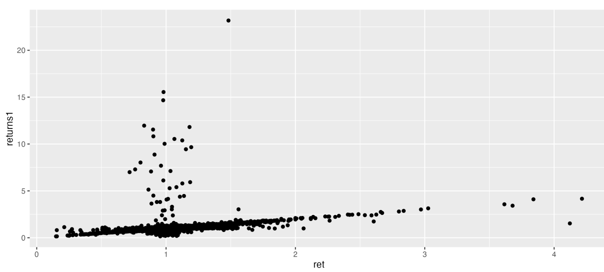
(a) Year 1997



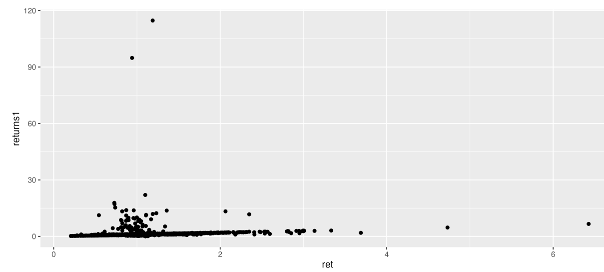
(b) Year 2002



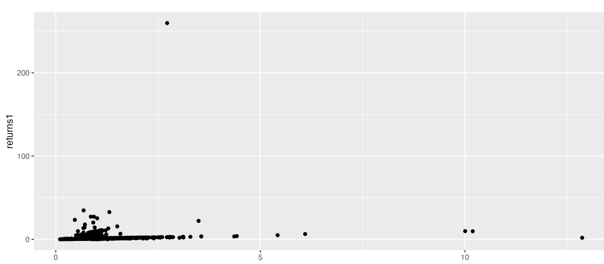
(c) Year 2006



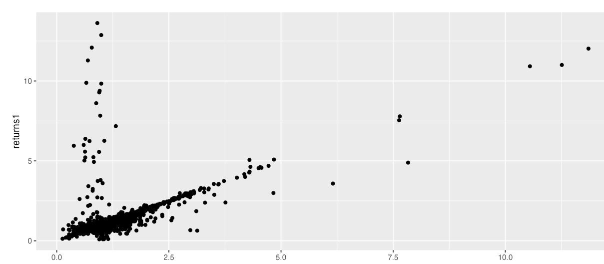
(d) Year 2010



(e) Year 2015



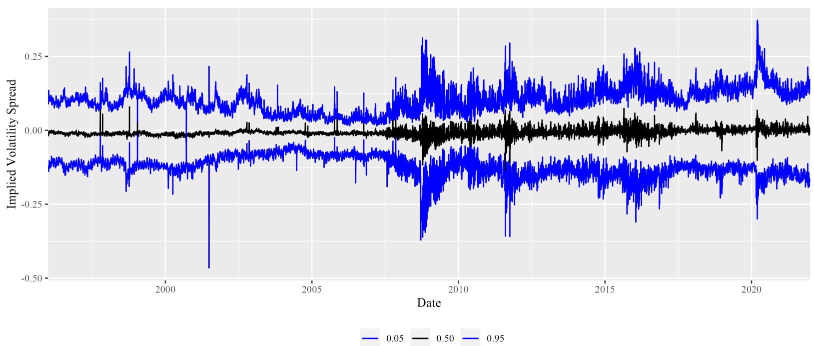
(f) Year 2021



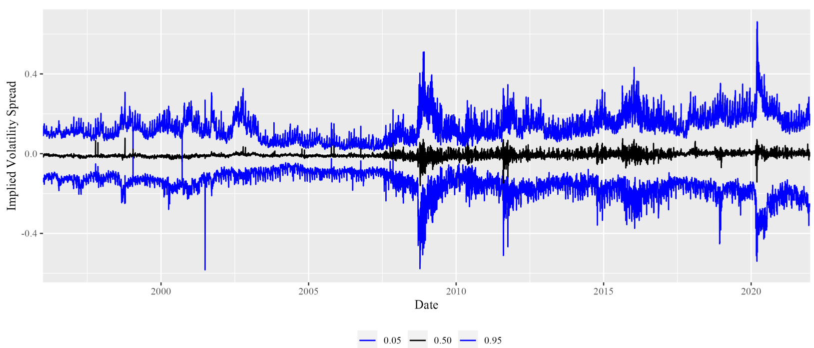
Note: Figure shows the difference between using the daily holding period return supplied by CRSP or calculating the returns as open/close every trading day without accounting for stock splits or dividends, of all returns in the sample across the particular year. The x axis depicts (ret) daily holding return, and the y axis depicts (returns1) open/close. This is just 6 representative years of data evenly spaced out throughout the sample period.

FIGURE A2: Plots of the distribution of the Implied Volatility Spread, across modifications to the signal

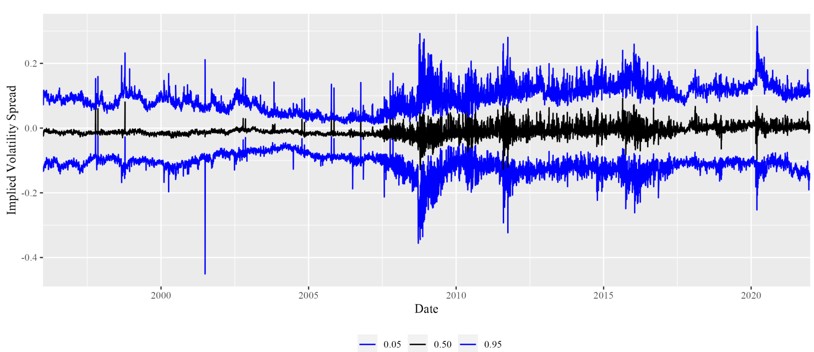
(A) Filtered & weighted signal



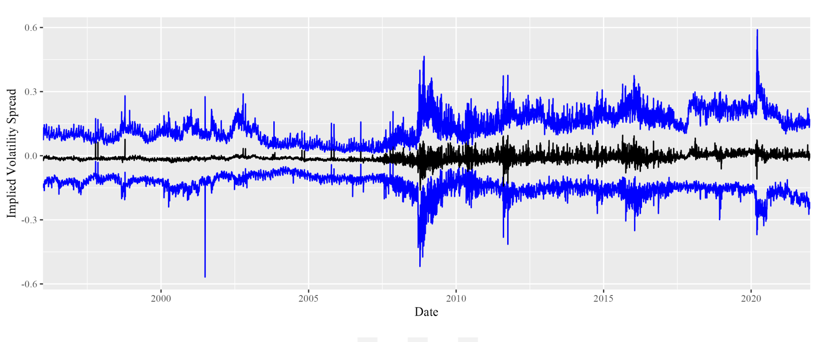
(B) Raw & weighted signal



(C) Filtered & simple signal

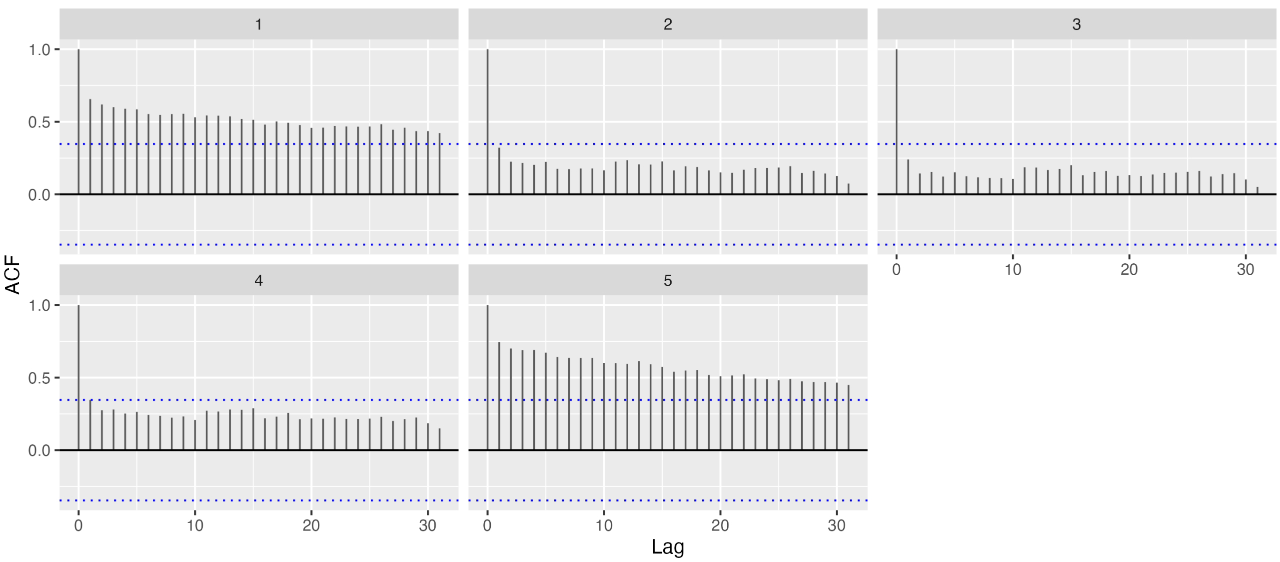


(D) Raw & simple signal



Note: Figure shows the distribution of the implied volatility spread across the entire sample period. Filtered and raw dictates if the conditions outlined in section 3 are implemented. Simple or weighted dictates if the implied volatility spread is weighted by the open interest for each pair. The blue lines are the 0.05 and 0.95 percentiles in the distribution, with the black line equal to the median.

Figure A3: Autocorrelation Function of Univariate Portfolio Sort, 5



Note: The five equal-sized portfolios are sorted on the implied volatility spread across the entire sample period with value-weighted weekly returns, with a signal observed on the last trading day before the next week (in which the return is observed) incurred. The returns have been demeaned across t before calculated the ACF.

B Tables

TABLE B1: Factor Competition Coefficients, Bivariate Dependent Sort 10x10

Dep. Variable	Intercept	Factor	Factor_D	MKT_RF	SMB	HML	RMW	CMA
Factor	0.97 (***)		0.04	0	-0.02	0.09	-0.11	-0.04
Factor_D	0.49 (***)	0.03		0.06	0.08	-0.04	0.25 (**)	0.14
MKT_RF	0.23 (***)	0	0.02		0.13 (**)	0.28 (***)	-0.52 (***)	-0.48 (***)
SMB	0.04	0	0.01	0.04 (**)		-0.07 (**)	-0.31 (***)	0.04
HML	-0.05	0.01	-0.01	0.11 (***)	-0.09 (**)		0.16 (***)	0.53 (***)
RMW	0.05	-0.01	0.02 (**)	-0.11 (***)	-0.2 (***)	0.09 (***)		0.17 (***)
CMA	0.05	0	0.01	-0.07 (***)	0.02	0.18 (***)	0.11 (***)	
BetaLiquidityPS	0.06	-0.05 (*)	0.03	0.06	0.16 (**)	-0.06	-0.04	0.2 (*)
betaVIX	0.04	0.05 (*)	-0.04	0.01	-0.33 (***)	-0.09	0.31 (***)	-0.16
CoskewACX	0.26 (**)	0.01	-0.02	0.07 (*)	0.03	0.08	0.11	-0.03
Coskewness	0.1	0.03 (*)	-0.02	0.03	-0.15 (***)	0	-0.07	-0.26 (***)
OptionVolume1	0.46 (***)	0.01	0.01	-0.05	0	0.23 (***)	0.29 (***)	0.26 (***)
OptionVolume2	-0.1	0.01	0.04 (*)	0.03	0.05	-0.14 (**)	-0.23 (***)	-0.19 (**)
skew1	0.45 (***)	0.01	0.01	-0.03	-0.03	-0.04	0.1 (*)	-0.13 (*)

Dep. Variable	BetaLiquidityPS	betaVIX	CoskewACX	Coskewness	OptionVolume1	OptionVolume2	skew1
Factor	-0.1 (*)	0.07 (*)	0.03	0.12 (*)	0.03	0.04	0.05
Factor_D	0.05	-0.04	-0.03	-0.05	0.01	0.1 (*)	0.04
MKT_RF	0.04	0	0.04 (*)	0.03	-0.05	0.03	-0.07
SMB	0.04 (**)	-0.05 (***)	0.01	-0.05 (***)	0	0.02	-0.02
HML	-0.02	-0.01	0.02	0	0.1 (***)	-0.05 (**)	-0.03
RMW	-0.01	0.03 (***)	0.01	-0.02	0.06 (***)	-0.05 (***)	0.04 (*)
CMA	0.02 (*)	-0.01	0	-0.04 (***)	0.04 (***)	-0.03 (**)	-0.04 (*)
BetaLiquidityPS		0.06 (**)	0.02	0.32 (***)	-0.02	0.05	0.62 (***)
betaVIX	0.09 (**)		-0.09 (**)	-0.1 (*)	0.49 (***)	-0.16 (***)	0.34 (***)
CoskewACX	0.02	-0.06 (**)		0.07	-0.11 (**)	0.09 (*)	0.15 (**)
Coskewness	0.18 (***)	-0.04 (*)	0.04		0.04	0.25 (***)	-0.3 (***)
OptionVolume1	-0.01	0.2 (***)	-0.06 (**)	0.04		0.57 (***)	-0.56 (***)
OptionVolume2	0.03	-0.07 (***)	0.05 (*)	0.28 (***)	0.59 (***)		0.17 (***)
skew1	0.19 (***)	0.07 (***)	0.04 (**)	-0.16 (***)	-0.27 (***)	0.08 (***)	

Note: The regression span the entire sample period and consists of weekly returns. The significance of the coefficients are coded according to the p-value: $0 < (***) < 0.001 < (**) < 0.01 < (*) < 0.05$.

Table B2: Three-Step Procedure, coefficients and standard errors

Factor	1	2	3	4	5	6	7	8	9
CMA:estimate	0.0024	-6e-04	0.0092	0.0278	-0.1026	-0.13	-0.1277	-0.1359	-0.138
CMA:std.error	0.0445	0.0477	0.0515	0.0547	0.0724	0.0729	0.0733	0.0733	0.0747
HML:estimate	0.0011	0.0071	0.0217	0.0361	-0.0936	-0.0998	-0.0992	-0.1113	-0.1128
HML:std.error	0.0208	0.0476	0.0654	0.0701	0.0906	0.0902	0.0901	0.0913	0.092
Intercept:estimate	0.2182	0.1775	0.1998	0.3052	0.3836	0.2189	0.2301	0.2282	0.2594
Intercept:std.error	0.0706	0.0688	0.0668	0.0627	0.0662	0.064	0.061	0.0578	0.0547
LS:estimate	5e-04	-0.0027	-0.0093	0.0039	0.5307	0.5699	0.5716	0.6124	0.7007
LS:std.error	0.0106	0.01	0.0209	0.0368	0.2315	0.2338	0.2288	0.242	0.2787
LSD:estimate	4e-04	1e-04	0	-0.0034	0.1125	0.0786	0.0764	0.0936	0.0462
LSD:std.error	0.0074	0.0092	0.0094	0.0145	0.0766	0.0857	0.0798	0.0824	0.1155
MKT_RF:estimate	-0.0075	0.0098	-0.0133	-0.0741	-0.2444	-0.1573	-0.1621	-0.1789	-0.2135
MKT_RF:std.error	0.142	0.1442	0.1511	0.1672	0.1609	0.1602	0.1609	0.1674	0.1671
RMW:estimate	0.0041	0.0034	0.0069	0.0177	-0.054	-0.0912	-0.0898	-0.0971	-0.099
RMW:std.error	0.0763	0.0907	0.09	0.0938	0.0967	0.0983	0.0983	0.0974	0.0997
SMB:estimate	-9e-04	-0.0169	-0.0152	-0.0081	-0.1486	-0.1703	-0.1693	-0.1828	-0.1836
SMB:std.error	0.0175	0.0936	0.0937	0.0938	0.1135	0.1132	0.1131	0.1148	0.1147

Table B3: Overview of ScenarioIDs

ScenarioID	Conditions	Signal weight	by	doublesort	Returns	Signal1	Signal2	splits_1	splits_2	splits_number
1	filtered	weighted	open interest		value	Level		5		1
2	filtered	weighted	open interest	dependent	value	Level	Change	5	5	2
3	filtered	weighted	open interest	dependent	value	Level	Change	3	3	2
4	filtered	weighted	open interest	independent	value	Level	Change	5	5	2
5	filtered	weighted	open interest	independent	value	Level	Change	10	10	2
6	filtered	simple			value	Level		10		1
7	raw	weighted	open interest		value	Level		10		1
8	raw	simple			value	Level		10		1
9	filtered	weighted	open interest	dependent	value	Level	Change	10	10	2
10	filtered	weighted	open interest		value	Change		5		1
11	filtered	weighted	open interest	dependent	value	Change	Level	10	10	2
12	filtered	weighted	open interest		value	Level		custom		1
13	filtered	weighted	open interest		value	Level		10		1
14	filtered	weighted	open interest	dependent	value	Level	Change	custom	custom	2
15	filtered	weighted	open interest		simple	Level		5		1
16	filtered	weighted	open interest	dependent	simple	Level	Change	5	5	2
17	filtered	weighted	open interest		simple	Change		5		1
18	filtered	weighted	open interest	dependent	value	Change	Level	5	5	2