Redundancy Reimagined : A study of reverse causality between the stock and options market

Name of student: Devin Ahluwalia

Student ID number: 680189

Master Thesis: MSc Financial Economics

Supervisor: Dr. PJPM Versijp

Second assessor: Dr.

Date: 9th August, 2024

Erasmus School of Economics

Erasmus University Rotterdam

Acknowledgements

I would like to sincerely extend my gratitude to my supervisor and professor, Dr PJPM Versijp, whose guidance and feedback along the way enabled me to shape my thesis into a respectable study on a topic I find truly interesting. Having studied the module of Advanced Investments under Dr PJPM Versijp, I was highly intrigued by factor models and the way they can influence investment strategies in practice which motivated me to pursue my thesis on the topic of reverse causality between the stock and options market. Finally, the knowledge I gained in the Advanced Investments seminar studied under Dr Esad Smajlbegovic played a huge role in ensuring the robustness of my results and arriving at the conclusions in my study.

Abstract

The paper explores the possibility of a reverse causality affect running from the options market to the stock market which would result in the stock return and price being influenced or to an extent, determined, by options market metrics such as volume, open interest, implied volatility, moneyness and premium. The study yields positive results in the support of our reverse causality hypothesis pointing towards the possibility of stock prices being an endogenous variable in the formula for calculating option prices. A factor model approach is also a part of our study to construct long-short portfolios akin to the fama-french 5 factors with the intention to determine predictability of stock returns.

Contents

1 Introduction **4**

2 Literature Review **8**

3 Data **13**

3.1 All-Options Approach 13

3.2 Liquidity and Volatility 14

3.3 Factor Model 14

3.4 Descriptive Statistics 14

4 Methodology **17**

4.1 The Model 17

4.2 Variable Construction 17

4.3 Factor Construction 20

5 Results **22**

5.1 Primary Regression 22

5.2 Robustness Checks 24

5.3 Factor Regression 28

6 Conclusion **31**

7 Appendix **33**

7.1 Tables 33

7.2 Tests 39

8 References **41**

1 Introduction

As more and more asset classes are characterised by sophisticated and developed markets, it would be logical to assume that relationships between different asset classes may not be as simple as previously understood. Considering the equity derivatives market, the leading school of thought on stock options is that their value – price, return, volume, volatility – is derived from the price of the underlying stock with the Black-Scholes-Merlon model of option valuation being a foundational pillar of Financial Economics (Black & Scholes, 1973). The Black-Scholes model uses options market parameters, namely the present value of the strike price using the annualised risk-free rate and the time to expiration combined with the prevailing stock price and the volatility of its returns to work out the price of the call or put option. The paper does not aim to argue with the notion of option prices being derived entirely or majorly from the underlying stock, but rather seeks to explore the possibility of the existence of a reverse causality effect on the stock market of the options market. The paper seeks to explain the return of stocks through a simplistic model using options market metrics such as volume, implied volatility and open interest to construct a six-factor model akin to the Fama-French factor asset pricing models *(Fama & French, 2015, Fama & French, 1993).* The basis of the argument is built upon the relative information efficiency of the two markets in question - the options market and the stock market *(Chakravarty et al. 2004).* The options market argument is based upon the idea of informed investors potentially possessing private information regarding the price movement of the stock choosing to trade in the options market to be able to exploit the leverage provided by options and amplify their gains resulting in the options market being quicker in reflecting the true value of the underlying stock's price *(Chakravarty et al. 2004).* In contrast, the stock market generally exhibits higher liquidity and volume due to a larger number of participants and greater popularity as a venue for investment and trading supporting the idea that by accounting for and taking into consideration a greater number or data of deterministic factors, the stock market more closely and efficiently reflects the true value of the stock *(Goncalves-Pinto et al. 2020).*

The argument of reverse causality is backed by strong economic reasoning for its existence based largely on the idea of informed trading, with renowned authors studying and supporting the same argument *(Easley et al. 1998, Pan and Poteshman 2004).* An investor possessing private information regarding the directional price movement of a stock will aim to earn maximum profit by exploiting this information. The options market offers leverage, with option contract sizes usually being 100 shares of a stock, trading at a fraction of the underlying stock's price and lower transaction costs *(Ho et al., 2012).* Hence, informed traders should be expected to trade in the options market as opposed to the stock market in order to amplify their gains *(Chan et al. 1999)*. Assuming that this argument is a reasonably accurate representation of reality, it would not be too far off to believe that the options market may lead the stock market for short time periods, such as at a daily frequency. For longer time periods such as monthly intervals, it would be logical to expect the information to spill over to the equity market *(Boluch and Chamberlain, 1997)*. Thus, our study is primarily limited to daily frequency of asset returns while also incorporating the results for the same model using monthly data averaged from daily data, yielding results which support the hypothesis of the reverse causality offering strong explanatory power for daily data and weakening with longer time horizons. The results also extend the argument suggesting that the reverse causality phenomena may also hold for other assets such as indices, commodities, currencies and cryptocurrency.

The study inherently contradicts the efficient market hypothesis as it is built upon the informed or asymmetric information trading argument. However, it can still be perceived as within the scope of the semi-strong efficient market, reflecting past and present publicly traded information, however unable to maintain its integrity under the presence of private information held by certain investors which results in a lead or lag relationship between the options market and the stock market *(Patel et al. 2020)*.

The model offers a mispricing based explanation of stock returns, arguing that due to a higher level of informed trading, options market metrics should more accurately predict the true value of the stock than the stock market which is more sensitive to market sentiment due to its inherently higher liquidity and is hence more prone to mispricing the stock *(Goncalves-Pinto et al. 2020)*. The stock market is the most popular form of investing, attracting various profiles of investors ranging from an employee investing 20% of his salary to an institution investing billions. On the contrary, the options market requires a relatively higher level of nuanced and sophisticated knowledge of investment vehicles which are more complex to understand and trade than equity securities. These discrepancies between the two markets reinforces the notion that due to a higher number of participants and also a lower proportion of informed or professional investors, the stock market is exposed to higher liquidity, which leads to increased sensitivity to market forces, sentiment and shocks as compared to the options market, resulting in higher price volatility which leads to deviation of the price of a stock from its true value as governed by the put-call parity concept of the options market. The potential mispricing aspect is further reinforced by virtue of the informed trading argument, the reasoning being that due to a lag in the spill over of information from the options market to the stock market, the options market represents the true value of the stock before the stock's price reflects the same. *Goncalves-Pinto et al. (2020)* offer a more detailed explanation of both the liquidity and informed trading argument through the concept of price pressure resulting in the deviation of the stock price from its put-call parity implied value. The return on equity securities is influenced by a multitude of factors such as macroeconomic news, geopolitical developments, pandemics and individual investor behaviour, all of which cannot be quantified, much less modelled in a sophisticated way. Hence, the paper aims to achieve some explanatory power with regards to daily stock returns while being primarily focused on establishing the significance of the options market variables in explaining stock market returns in line with our economic reasoning. Even in the presence of strong reverse causality it would be imprudent to expect the options market to be the primary predictor of stock returns.

The paper uses factors constructed from the the volume, open interest, implied volatilities and prices of options to explain stock returns on a daily basis. The idea behind certain variable constructions is inspired by prior literature, such as the put-call ratio construction by *Pan and Poteshman (2004)* and the implied volatility spread measure proposed by *Xing et al. (2010)* wherein they calculate the implied volatility spread (IVS) as the difference between the implied volatilities of out-of-the-money (OTM) put options and at-the-money (ATM) call options. However, our paper extends these factors by constructing them in a unique or slightly different way so as to ensure consistency across our data while incorporating other factors. As such, all variables individually remain based upon the papers from which they are derived, however, in order to establish the relation between the return of a stock and the independent variables used in our regression, it should be noted that the volume, open interest, implied volatility and prices of all the option contracts of a stock traded on a given day are averaged to form a single option observation per stock per day, however, differentiated by whether the option is a put or a call.

Another mechanism which has important implications in terms of the influence of options on the underlying equity security's price, however outside the scope of this paper, is the hedging activity traders engage in with the intention of offsetting the risk of their positions in the stock market. Investors holding a stock may want to offset the risk of this stock falling in price through the purchase of a put option, which will generate a profit in the case that the stock price falls below the strike price stated in the put contract. The demand for put options theoretically drives their price up, increasing the implied volatility of put options. The opposite side of this trade would involve sellers of the put option hedging their own position by short-selling the underlying stock. This could create downward pressure on the stock, leading to a fall in its price. Hence, the study of hedging activity in the options market to investigate its effect on stock returns requires a more nuanced approach with factors different from the ones involved in this paper *(Easley et al. 1998)*.

The main model of the paper uses daily data to explain stock return movements through its options' volume, open interest, implied volatility, moneyness and price or premium and establish reverse causality between the two markets. Following the primary regression around which the paper is cantered, the study undertakes robustness checks by isolating time periods and finding consistency in coefficients and their significance while maintaining an expected level of explanatory power. Finally, the daily data is transformed to monthly data as a robustness check in addition to investigating whether stock returns are explained by options market parameters when using data with lower frequency. As theorised, the explanatory power of the model falls when using monthly data with the coefficients maintaining their magnitude, signs and significance. Since our argument is based upon the prevalence of informed trading in the options market and the resultant lag occurring in information spill over to the stock market the former more accurately reflects the true value of the stock, we expected the spill over lag to disappear to a large extent when considering monthly data which is what the results of our model also supports.

Finally, the paper also includes the construction of long-short portfolios constructed as factors built from our 6 primary independent variables. The long-short portfolios while mostly significant in our regression with ret (RET - rf) as the dependent variable, return generated per year by investing in accordance with each factor primarily yields negative results indicated by the yearly means of each factor. Additionally, including the fama-french 5 factors and momentum as control variables to the regression renders a majority of the factors insignificant. This indicates that while our options market independent variables are useful in explaining stock return movements on the same day, long-short portfolio factors constructed from said variables using one-day lagged values (formation period being t-1) are insignificant in explaining stock returns. This conclusion supports the assumption that since option trading or speculation is primarily conducted intra-day, a formation period of the previous day fails to deliver significant results in our regression.

2 Literature Review

Academic literature has heavily focused on the price discovery mechanism provided by the options market following from the informed trading argument. The core argument which our paper considers in establishing the presence of reverse causality is the informed trading argument which states that investors possessing private information regarding the future movement of stock prices will choose to trade in the options market so as to exploit the benefit of leverage - that is, pay a small cost upfront and possibly earn high rewards essentially amplifying one's potential gains from the trade.

The basic concept upon which the informed trading argument is built is reflected in the work of many papers, such as *Chakravarty et al. (2004)*, *Pan and Poteshman (2004)* and *Easley et al. (1998)*, which reinforce each others general argument of informed investors choosing to trade in the options market aiming to capitalise on the leveraging aspect offered by options. *Patel et al. (2020)* and *Mayhew and Stivers (2003)* found evidence suggesting that option volume and liquidity are heavily and positively linked to the informational content embedded in options and their significance in explaining future stock returns and stock return volatility. *Pan and Poteshman (2004)*, *Easley et al. (1998)*, *Chan et al. (1999)* and *Boluch and Chamberlain (1997)* focus on the use of option volumes to study its effect on stock returns and establishing lead-lag relationships between the two markets with their studies yielding mixed results regarding the idea of the options market signaling future stock price movements, varying with the frequency of data used and the period under consideration. *Xing et al. (2010)* and *Gao et al. (2019)* utilise the implied volatility of options to construct their measure and predict future stock returns using call-put implied volatility spreads with both of their studies yielding significant results in the favour of the options market leading the stock market through informed trading. *Goncalves-Pinto et al. (2020)* offers an argument in favour of the options market leading stock price movements based upon higher liquidity and noise in the stock market which results in price pressure and hence deviation from the stock's true value as implied by the put-call parity in contrast to the informed trading argument.

Chakravarty et al. (2004) investigate the extent of the options markets' price discovery mechanism through stock volatility, and trading volume and spreads in both markets. The authors follow an 'information share' approach aiming to establish how much either market contributes to the price discovery mechanism. The authors find significant evidence in support of the idea that the options market facilitates price discovery. The study finds that the extent of the price discovery mechanism varies across stocks, with the options market facilitating stronger price discovery in the case when the option volumes traded are large and the option effective spreads are narrow signifying high liquidity in the options market with the stock market in contrast exhibiting a lower trading volume and liquidity. The information share of the options market was found to be negatively related to the volatility of the underlying stock with moderate significance. Their paper also accounts for how different strike prices and moneyness influences price discovery in the options market, finding evidence in favour of out-of-the-money options displaying a higher information share as compared to options with strike prices resulting in at-the-money or in-the-money contracts. However, emphasizing that this result is subject to trading volume and spreads. The study comes to the conclusion that the extent of leverage and liquidity prevalent in the options market at the time are primary factors facilitating the price discovery mechanism. The authors reinforce the idea that their study does not suggest that informed traders should strictly be expected to trade in the options market, however suggesting that it should be expected for there to be higher traffic in terms of informed traders in the options market close to the announcement of significant corporate disclosures and events such as earnings announcements and merger and acquisition announcements.

Pan and Poteshman (2004) investigate the role of option trading volumes in predicting stock returns and future stock prices. Their paper also serves as an inspiration to our study, utilising the put-call ratios derived from daily option volumes to explain stock returns in our model. They find significant evidence in favour of option volumes carrying information regarding stock prices, supporting the argument that the options market attracts informed traders. The paper points towards private information based upon which traders in the options market take positions being the basis for the price discovery mechanism in the options market as opposed to market inefficiencies with price predictability in the options market being positively related to the leverage offered by options and the number of informed traders. The study does not find significant price discovery in the options market for indices, coming to the conclusion that private information based upon which trades are made in the options market is likely to be firm specific than market or industry specific.

Goncalves-Pinto et al. (2020) contribute to existing literature by studying the role of price pressure in the stock market through the lens of the options market. Their study involves analysing stock price deviation from its put-call parity implied value to explain returns through the DOTS (Difference between the Option-implied Stock value and the Traded Stock price). One of the explanations the paper offers follows the hypothesis of informed traders choosing to trade in the options market due to which the put-call imparity implied value of the stock signals the price to which the stock will rise or fall once the information spills over. Their study follows the belief that the options market can lead the stock market under certain conditions. *Goncalves-Pinto et al. (2020)* also offers an interesting alternative explanation of the discrepancy between option-implied stock prices and prices prevailing in the stock market. The explanation is built upon the reasonable assumption that the stock market is more liquid than the options market, due to which information is first reflected in the traded stock price through price pressure, and the option-implied put-call parity based value of the stock signals the level to which the price will rise or fall once all information is fully reflected in the price.

Easley et al. (1998) was one of the pioneering papers questioning the redundancy of options. Their study argues that while stock prices lead option volumes by inducing hedging focused trades on the stock, certain option volumes lead stock prices, supporting the notion that the options market is the venue where informed traders trade. The authors support the theory that the information embedded in option volumes of certain stocks can predict future stock prices. The study emphasises that while their results support the idea of informed traders participating in the options market, it is likely that this effect also includes the influence of a multitude of factors not considered in the study. An important implication of their study is that as option volume carries signals for future stock price movements, under asymmetric information and imperfect markets it would not be possible to consider the stock rice as an exogenous determinant in the calculation of the option's price.

Patel et al. (2020) builds upon *Easley et al. (1998)* finding that options lead stock prices one-quarter of the time. Their findings are backed by the prevalence of insider trading in the options market due to its leveraging characteristic. Hence, private information possessed by said insider traders is first reflected in the options market causing them to lead stock prices. Their paper utilises data on insider trading prosecutions and a measure representing the speed at which prices reflect the presence of information-based trades. Their research strongly supports the argument that leverage is a key component of the options market which attracts informed traders. Their study argues that liquidity of options plays an important role in contributing to the price discovery mechanism as they found evidence that actively traded options of large cap stocks have a higher option information share than illiquid options of relatively smaller stocks. Based on the study, the authors also expect the option information share facilitating the price discovery mechanism will be higher prior to the announcement of important corporate events.

Xing et al. (2010) utilise the implied volatility spread measure using out-of-the-money (OTM) put options and at-the-money (ATM) call options to suggest that the options market is the venue for informed trading. They also extend the linkage of options to the equity market by concluding that the volatility smirk is informative of firm earnings, implying that there exists a relation between firm fundamentals and the implied volatility of options. Their study argues that informed traders trade in the options market resulting in the options market leading the stock market as the latter is slower in reflecting information.

Chan et al. (1999) found evidence that the low liquidity in option trading relative to stock trading limits the potential of exploiting the financial leverage offered by the options market resulting in reduced informational efficiency provided by the options market. The authors found that informed traders initiate trades in the stock market through a study of option and stock net trade volume which indicated faster and more efficient stock quote revisions compared to option quote revisions suggesting that the stock market is faster and more efficient in processing and reflecting information regarding the underlying stock's price. An important implication of their study was that it yielded little significant evidence in terms of hedging activity in the options market as they found that option volumes were not significantly influenced by stock price movements. They argue that although the options market attracts informed investors, prudence in terms of limit order submitted for call and put options limits the information conveyed by these trades.

Mayhew and Stivers (2003) study to what extent the volatility of stock returns can be explained through the implied volatility of stock options. They found evidence supporting the conclusion that information regarding future stock volatility of actively traded options are comprehensively captured by the implied volatility of these stock's options. As for options that are not as actively traded and have lower trading volumes, the informational content embedded in the implied volatility of these options in the context of explaining future stock return volatility is inferior in comparison to models analysing return shocks of the stocks. Hence, illiquidity implied by lower option trading volume negatively affects the ability of the options market in explaining stock returns. Additionally, the authors' study suggests that equity index options can provide a reasonably good estimate of implied volatility at the firm level when individual stock options are unavailable or the stock options market is exhibiting low liquidity.

Boluch and Chamberlain (1997) study the relationship between stock price movements and the option volumes of the underlying stock. They find bi-directional or bi-causal evidence suggesting that option volumes or transaction activity influence stock returns and vice-versa using intraday data for their study. Using intraday frequency of data, the authors found that the feedback or lead-lag relationship between the two markets, if any, remains for a very short period of time as both markets are fast to accurately reflect any information conveyed by the other. Their study contradicts the idea that the options market leads the stock market due to the presence of informed trading in the options market and offer the explanation that such a contradictory result arises due to inconsistencies between the frequency of data as they have used relative to studies which primarily utilise daily data.

Bergsma et al. (2019) focus on the importance of option moneyness in indicating the direction and intensity of stock price movements. Their study is based upon the idea that different levels of moneyness of options should convey relatively stronger or weaker signals regarding the expected return movement of the stock. Their argument seeks to explore whether out-of-the money or at-the-money options are indicative of stronger information possessed by participants in the options market compared to in-the-money options as the former are a more risky bet and speculating on them should reflect high quality private information possessed by the trader. They use a measure which captures the change in option trading activity or option volume as its moneyness (K/S) increases. Their study finds evidence that informed traders tend to use out-of-the-money options due to a higher leverage offered by these options and hence that higher option volumes, especially call option volumes, for non in-the-money options are more indicative of stock returns. Furthermore, they also find evidence in support of the notion that options with a higher level of implied volatility attract informed traders due to the ability to earn higher rewards in terms of more favourable leverage.

Gao et al. (2019) focus on constructing implied volatility spreads between call and put options and how studying information embedded in these spreads as investor attention on stocks increases. They found significant evidence supporting the notion that the informational content of volatility spreads and subsequently return predictability of options is higher for stocks which are receiving higher investor attention. Their results support the informed trading argument and generate a fama-french 5 factor significant alpha of more than 2%.

3 Data

Daily stock price and return data traded in the US on the New York Stock Exchange (NYSE), the National Association of Securities Dealers Automated Quotations (NASDAQ) and the American Stock Exchange (AMEX; exchange codes 1, 2, and 3) are extracted from the OptionMetrics security market data. Daily stock data is merged with options data extracted from OptionMetrics. The period under study is from 01/01/2010 to 31/12/2020. OptionMetrics has been chosen instead of CRSP to extract daily stock data as it enables the use of SECID as a common identifier to merge option data with stock data, while CRSP uses PERMNO as the stock identifier.

The total number of stocks or firms in our regression sample representing the panel variable (SECID) are 981 and the total number of observations in our sample are 271,968 with daily data as our time identifier extending from 04/01/2010 to 31/01/2020, while the overall sample includes 995 unique stocks and 2,537 days. Total number of options per stock, differentiated by whether it is a call or a put option, are averaged for each day to form the independent variables for our regression.

3.1 All-Options Approach

Contrary to prior literature on the subject of our paper, we utilise 6 different factors to explain stock returns through the options market while most papers such as have focused on the sophisticated construction of a single factor aiming to entirely encapsulate the effect of the options market on the equity market such as implied volatility spreads *(Gao et al. 2019)* and option volume *(Pan and Poteshman 2004)*. As a result of our ambition to utilise several options market metrics such as volume, open interest, implied volatility, moneyness and premiums, it was deemed important to keep the construction of our independent variables relatively simplistic. Hence, in order to facilitate our study's primary regression using all 6 independent variables, each independent variable is constructed by averaging its value for the day for a stock. Our study is focused on ensuring that all option contracts, following the standard data cleaning procedure, contribute to the construction of our factors and variables. Barring the ITM diff variable, the implication of such an approach is that the intensity of moneyness as determined by the moneyness of options on the same stock with different strike prices and the same expiration date is not considered. Instead, we choose to construct each variable with a similar approach and average the result across the day for a given stock. While employing a more sophisticated approach to the construction of our variables on an individual basis is an attractive avenue for further study, our current approach enables all option observations on a stock for a day to be considered while constructing the variables, providing a more comprehensive outlook on the effects of options market parameters on the stock market.

3.2 Liquidity and Volatility

Penny stocks (Price less than or equal to $5) are excluded from our dataset to control for their unreliable volatility and maintain the integrity of the sample set. Stocks with a daily trading volume of less than or equal to 1000 are dropped to ensure the exclusion of illiquid stocks from our sample.

Option contracts with a daily price less than or equal to 100 are dropped from our sample and options with daily volume of less than or equal to 100 are dropped to control for illiquid options which are not traded actively.

3.3 Factor Model

Following the use of our 6 independent variables to explain stock return movements through activity in the options market, we extend our study by constructing factors based upon our independent variables to analyse portfolio returns generated as a result of investing in accordance with these factors, and also aiming to establish and rationalise whether the returns generated carry a mispricing or risk premium. It is important to note that the number of observations fall significantly when the factors are constructed as these portfolios are built using lagged values of our options market variables resulting in observations with gaps in their periods being dropped.

3.4 Descriptive Statistics

Table 1: Variable Summary

Variable	Obs	Mean	Std. Dev.	Min	Max
RET	274,249	0	0.026	-0.517	1.194
DAILY PC RATIO	274,249	0.452	0.247	0	0.999
DAILY OI RATIO	274,249	0.433	0.281	0	1
DAILY IVS	274,249	0.154	1.597	-74.27	42.45
IVS avg	271,968	-0.006	0.071	-2.445	2.472
ITM diff	274,249	0.39	2.449	-124	57
price spread	274,249	17.177	251.738	-23,890	16,321.285

Table 2: Correlation Matrix of Independent Variables

Variables	RET	DAILY_PC_RATIO	DAILY_OI_RATIO	DAILY_IVS	IVS_avg	ITM_diff	price_spread
RET	1.000						
DAILY_PC_RATIO	-0.241	1.000					
DAILY_OI_RATIO	-0.233	0.575	1.000				
DAILY_IVS	0.271	-0.400	-0.389	1.000			
IVS_avg	-0.045	-0.069	-0.059	0.211	1.000		
ITM_diff	0.287	-0.327	-0.362	0.601	0.104	1.000	
price_spread	0.165	-0.185	-0.204	0.213	0.246	0.408	1.000

Table 3: Monthly Interval

Variable	Obs	Mean	Std. Dev.	Min	Max
RET	35,152	0	0.02	-0.313	0.392
DAILY PC RATIO	35,152	0.463	0.169	0	0.997
DAILY OI RATIO	35,152	0.431	0.227	0.001	0.999
DAILY IVS	35,152	0.106	0.755	-26.457	20.086
IVS avg	34,881	-0.007	0.069	-2.14	1.012
ITM diff	35,152	0.238	1.057	-31.421	20.381
price spread	35,152	12.587	181.746	-3,985.833	6,073.75

Table 4: Factor Summary

Variable	Obs	Mean	Std. Dev.	Min	Max
put call factor	1985	-0.00064	0.00757	-0.03721	0.0424
open interest factor	1985	-0.00085	0.00801	-0.04409	0.0475
agg IVS factor	1985	-0.00051	0.00834	-0.05906	0.04662
avg IVS factor	1985	0.00048	0.00767	-0.04204	0.04248
itm factor	1985	-0.00161	0.0133	-0.09848	0.06191
price spread factor	1985	-0.00036	0.00749	-0.04918	0.04103

Table 5: Fama and French Control Factors and Momentum

Variable	Obs	Mean	Std. Dev.	Min	Max
mktrf	1985	0.00062	0.00946	-0.0504	0.0497
smb	1985	-0.00001	0.00512	-0.0169	0.0363
hml	1985	-0.00008	0.00497	-0.019	0.0244
umd	1985	0.00009	0.00687	-0.0316	0.0361
rmw	1985	-0.00007	0.00342	-0.0183	0.0147
cma	1985	0.00003	0.00293	-0.0122	0.0199

4 Methodology

4.1 The Model

The paper utilises a fixed-effects panel regression model deeming it the most appropriate as it controls for firm or entity-specific effects which may remain constant across time for a stock but differ between stocks. Theoretical examples of effects which could be controlled for through our fixed-effects model would be the management's capability, industry or product-specific market characteristics and age of the firm. The use of time fixed-effects could theoretically control for general macroeconomic trends such as interest rate changes, geopolitical shocks and overinflated investor sentiment during certain periods in history, however the use of daily data in our sample complicates a time fixed-effects regression due to which the paper does not consider the same.

$$\text{RET} = \alpha_i + \beta_1 \,\text{Daily PC ratio}_{it} + \beta_2 \,\text{daily OI ratio}_{it} + \beta_3 \,\text{AggIVS}_{it} + \beta_4 \,\text{AvgIVS}_{it} + \beta_5 \,\text{ITM diff}_{it} + \beta_6 \,\text{price spread}_{it} + \epsilon_{it}$$

where:

$$\begin{aligned}
\text{RET} &= \text{Daily stock return} \\
\text{daily PC ratio} &= \text{Daily Put-Call volume ratios of the stock} \\
\text{daily OI ratio} &= \text{Daily Put-Call open interest ratios of the stock} \\
\text{AggIVS} &= \text{Aggregate Implied Volatility Spread of the stock options for the day} \\
\text{AvgIVS} &= \text{Average Implied Volatility Spread of the stock options for the day} \\
\text{ITM diff} &= \text{In-The-Money spread of call and put options for the day} \\
\text{price spread} &= \text{Average call-put price spread of a stock for the day}
\end{aligned}$$

4.2 Variable Construction

Daily PC Ratio

$$\text{Daily PC ratio} = \frac{\text{total daily put volume}}{\text{total daily option volume}}$$

The *daily PC ratio* constructed from the daily put and call volumes aims to classify the overall sentiment in the market as bullish or bearish, with a high (low) put-call volume ratio indicating a larger number of puts (calls) trading, indicating that market participants expect the value of the underlying stock to go down (up). The call and put volumes of a stock are aggregated across the day to calculate the overall sentiment regarding the stock's price movement as opposed to differentiating between call or put options having different strike prices and days to maturity. In our dataset, volume calculation based upon differentiated call and put volumes was not indicative of the prevailing market sentiment, lacking any significant explanatory power.

daily OI ratio

$$\text{daily OI ratio} = \frac{\text{total daily put open interest}}{\text{total daily option open interest}}$$

The *daily OI ratio* is constructed similarly to the *daily PC ratio* factor, built from the daily open interest values of put and call options, aggregated across the day. Open interest of options differs from the volume in that it represents open positions taken by investors, indicating active interest in an option, while volume measures the number of times the contract essentially exchanged hands across a specific time period. Open positions measured by open interest are theorised in our paper to represent the belief of an investor that his current position will be profitable and can hence be indicative of an upward price movement if the put-call open interest ratio is low, and a downtrend in price in case the put-call open interest ratio is high represent a higher number of open positions betting on the price to go down.

AggIVS

$$\text{AggIVS} = \frac{\text{Aggregate Call IV - Aggregate Put IV}}{\text{Total number of options traded on the day}}$$

AvgIVS

$$\text{AvgIVS} = \text{Average daily Call IV - Average daily Put IV}$$

The *Agg IVS* and the *AvgIVS* factor differ only in the fact that *AggIVS* is calculated by averaging aggregate implied volatility spreads of call and put options across the day, while *AvgIVS* is calculated by averaging average implied volatility spreads of call minus put options across the day, with both measures or factors remaining significant in our model and having a correlation of only 0.1839. *AvgIVS* is likely to be relatively more consistent with prior literature *(Gao et al. 2019)* covering the use of the IVS measure, however *AggIVS* aims to capture the fact that a higher implied volatility of a call (put) option owing to a higher number of call (put) options being traded compared to put (call) options signifies that the market has higher demand for call (put) options of a stock compared to put (call) options on it. More generally speaking, the Implied Volatility Spread itself as an indicator of stock price movement or returns is based on the fact that higher demand for an option increases its volatility. Hence, a positive IVS measure in our regression would indicate that call options are being perceived as more profitable suggesting a positive return from holding the stock. Both factors are based on demand resultant implied volatility, while *AggIVS* amplifies this concept by not adjusting for the number of call or put options traded on the day.

ITM diff

$$\text{ITM diff} = \text{number of ITM calls in the day} - \text{number of ITM puts in the day}$$

The *ITM diff* measures the difference between the number of in-the-money (ITM) call options and in-the-money (ITM) put options traded on a given day. In-the-moneyness is defined as having a (strike price)/(stock price) ¡ 0.97 for a call option and (strike price)/(stock price) ¿ 1.03 for a put options. This factor is unique in the sense that it uses the difference between the number of ITM options as opposed to option metrics such as volume, open interest and implied volatility. The intuition behind the inclusion of this factor is rooted in the logic that a positive value for this measure resulting from a higher number of ITM call options being traded on a given day than ITM put options on the stock suggests that call options are currently more profitable on aggregate, which should result in a higher demand for call options, further conveying a bullish sentiment regarding the underlying's price. This loop-based explanation can be thought of as momentum anomaly manifesting through the options market.

Price Spread

$$\text{Price Spread} = \text{Average daily call price - Average daily put price}$$

The *price spread* factor measures the price spread between the average call option contract price and average put option contract price for a given day. This factor also follows from a demand-based explanation, theorised as the idea that a positive value for this measure implies a higher demand for call options as opposed to put options, which may convey investor expectation of positive returns from holding the stock, i.e an upward trend in the underlying stock's price.

An important aspect to note would be the assumption that the inclusion of the *ITM diff* and *price spread* factors would result in multicollinearity between the two and also point towards endogeneity with the dependent variable. While all three variables, RET, ITM diff and price spread are partially or majorly functions of the stock price, the correlation matrix of the variables as shown in Table 4.1 suggests low correlation across all three variables. Additionally, studying the concept of reverse causality itself warrants the use of the two independent variables. Hence, while the VIF values allow us to proceed with the assumption that our concerned independent variables do not exhibit multicollinearity, since option values are derived from the stock's price, we must take into account the presence of endogeneity in our regression, even though the correlation matrix does not point towards a high level of correlation between RET, ITM diff and price spread . While it would be ideal to tackle this issue through a two-stage least squares regression, finding instrumental variables which are a suitable proxy for our options market metrics while being uncorrelated with the stock return is a task which cannot be solved within the scope of this paper. Furthermore, due to the number of stocks considered and the lengthy time horizon under study given daily data, the heavily unbalanced nature of our dataset also renders a causal analysis model unfeasible.

Table 6: Correlation Matrix of Variables

Variables	(1) RET	(2) ITM diff	(3) price spread
(1) RET	1.000	0.285	0.152
(2) ITM diff	0.285	1.000	0.384
(3) price spread	0.152	0.384	1.000

ITM calls are defined as having a moneyness ((strike price)/(stock price)) of less than 0.97 (moneyness<0.97)
ITM puts are defined as having a moneyness ((strike price)/(stock price)) of greater than 1.03 (moneyness>1.03)

4.3 Factor Construction

The factors namely - put call factor, open interest factor, aggregate IVS factor, average IVS factor, ITM factor and price spread factor - are constructed as long-short portfolios using one-period (one-day) lagged values of our 6 independent variables. The lagged values of the independent variables are ranked and subsequently divided into quintiles.

For the put call factor and open interest factor, the top quintile represents the short leg of the portfolio as it is characterised by the highest put-call ratios, while the bottom quintile represents the long leg of the portfolio characterised by the lowest put-call ratios or alternatively, highest call-put ratios. The resultant factors are hypothesized to yield positive returns in line with the directional movement indicated by our independent variables in the primary regression.

As for the Agg IVS factor, Avg IVS factor, ITM factor and price spread factor, the top quintile represents the long leg of the portfolio characterised by the highest implied volatility spread, in-the-money difference or call-put price spread, while the bottom quintile represents the short leg of the portfolio characterised by the lowest implied volatility spread, in-the-money difference or call-put price spread

The factor construction and the regression are included as a part of our study purely from a theoretical perspective to determine whether the options market metrics can be robust in predicting stock returns as daily rebalancing in practice would result in extremely high transaction costs which would make it unfeasible for one to invest with this strategy.

5 Results

5.1 Primary Regression

Table 7 represents an entity fixed effects panel regression with robust standard errors. Each column represents the addition of an independent variable to assess the significance of the variables in isolation and together with the other variables.

Table 7: Primary Regression

	(1)	(2)	(3)	(4)	(5)	(6)
	RET	RET	RET	RET	RET	RET
DAILY_PC_RATIO	-0.0277***	-0.0187***	-0.0133***	-0.0134***	-0.0127***	-0.0122***
	(0.000839)	(0.000655)	(0.00133)	(0.00147)	(0.00167)	(0.00154)
DAILY_OI_RATIO		-0.0146***	-0.00960***	-0.00921***	-0.00754***	-0.00714***
		(0.000508)	(0.000949)	(0.00109)	(0.00144)	(0.00129)
DAILY_IVS			0.00367***	0.00421***	0.00276***	0.00299***
			(0.000828)	(0.000973)	(0.000654)	(0.000676)
IVS_avg				-0.0507***	-0.0475***	-0.0558***
				(0.00487)	(0.00439)	(0.00386)
ITM_diff					0.00147***	0.00106**
					(0.000426)	(0.000397)
price_spread						0.00000944***
						(0.00000154)
_cons	0.0127***	0.0149***	0.00978***	0.00930***	0.00793***	0.00745***
	(0.000379)	(0.000420)	(0.00108)	(0.00125)	(0.00160)	(0.00143)
r2	0.0630	0.0788	0.115	0.131	0.140	0.146
N	274249	274249	274249	271968	271968	271968

Section 7.2.1 conducts the Durbin-Watson test and yields the result of no autocorrelation present in our data

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In the first regression, RET is regressed upon the daily put-call ratio of the stock, yielding a significant negative coefficient of -0.0277. While the absolute magnitude of the coefficient for the daily put-call ratio may not be high, as is the case for the other explanatory variables in our regression, it is important to note that the negative coefficient is in line with our hypothesized economic reasoning and the coefficient continues to remain negative even when additional explanatory variables are added to our regression. Hence, for our sample, a higher put-call ratio of a stock for the day explains the negative or lower returns that the stock yields.

In our second regression, RET is regressed upon the daily put-call ratio and daily open interest ratio of the stock, yielding negative and significant coefficients for both independent variables maintaining the integrity of our economic reasoning behind the use of the two variables. The addition of the Open interest variable reduces the absolute value of the put-call variable; however, it does not lead to insignificance of the latter and adds to the explanatory power of our model.

The third regression is characterized by the addition of our DAILY IVS variable which is constructed as the difference between the aggregate implied volatility of call and put options averaged for the total number of options for the given stock on the day. A positive and significant measure of the coefficient of 0.0037 is in line with the proposition that a positive value of the variable, representing higher call option volatility relative to put option volatility on the same stock for the day, explains positive returns on that stock for the day. The addition of the DAILY IVS variable significantly improves the explanatory power of the model with an r-square measure of 11.5% while lowering the absolute value of the put-call ratio and open interest ratio measures.

The IVS AVG, calculated as the difference between the average implied volatility of call options and put options on the same stock for a day, has a negative coefficient of -0.0507 which contradicts our postulated theory of a positive IVS AVG measure explaining positive returns on a stock. A possible and reasonable explanation behind the negative coefficient of the IVS AVG measure could be the hedging activity of traders. A positive IVS AVG measure representing high call option volatility relative to put option volatility indicating higher demand for call options could induce hedging activity in the form of shorting the stock so as to lower the risk of the call option expiring out of the money resulting in the loss equal to the premium paid on the call option. Such hedging activity could drive bearish sentiment for the stock resulting in lower returns in the presence of a positive IVS AVG measure.

The addition of the ITM diff variable, which aims to establish a positive relation between the return of the stock and the difference between the number of in-the-money call options and in-the-money put options of a stock for a given day to establish whether a call or a put position is more profitable for that day on an average in our 5th regression, increases the r-square of the model and itself has a positive and significant coefficient of 0.0015 which is in accordance with the concept upon which the variable is constructed.

Our 6th and final regression of the main model includes all 6 independent variables to explain the return on the stock for a day with the addition of the price spread measure which measures the difference between the average price of call options and put options on the stock for the day. Based on the reasoning that a higher demand for call options relative to put options for that stock would result in a positive price spread measure through a higher average call premium relative to put premium conveying bullish sentiment regarding the stock's return, the price spread measure has a coefficient of 0.000006 which is positive and significant, however, the magnitude of the measure is too extremely minute and hence should be expected to barely contribute to explaining stock returns. The models boasts an explanatory power of 14.6% which is in line with our expectations of options market metrics driving or explaining stock returns to a conservative extent, with the primary causality running in the opposite direction from the stock market to the options market. In our final regression and complete model, all variables remain significant and the only variable which contradicts our economic reasoning remains to be the IVS AVG measure.

5.2 Robustness Checks

Table 8, 9 and 10 display the year-wise regression results for the model with all 6 independent variables.

Table 8: Year-Wise Robustness Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	RET	RET	RET	RET	RET	RET	RET	RET	RET	RET	RET
DAILY_PC_RATIO	-0.0155***	-0.0187***	-0.0104***	-0.00965***	-0.0103***	-0.0120***	-0.0125***	-0.00827***	-0.0128***	-0.0103***	-0.0150***
	(0.00121)	(0.00145)	(0.00114)	(0.00105)	(0.00116)	(0.00166)	(0.00189)	(0.00128)	(0.00210)	(0.00251)	(0.00267)
DAILY_OI_RATIO	-0.0107***	-0.00496***	-0.00804***	-0.00737***	-0.00549***	-0.00929***	-0.00638***	-0.00665***	-0.00950***	-0.00893***	-0.00806**
	(0.000926)	(0.00121)	(0.000847)	(0.000856)	(0.00105)	(0.00138)	(0.00157)	(0.00127)	(0.00170)	(0.00182)	(0.00244)
DAILY_IVS	0.00334***	0.00305***	0.00425***	0.00437***	0.00447***	0.00447***	0.00512***	0.00477***	0.00195	0.00319*	0.00188*
	(0.000416)	(0.000684)	(0.000735)	(0.000549)	(0.000595)	(0.000811)	(0.000843)	(0.000819)	(0.00103)	(0.00142)	(0.000729)
IVS_avg	-0.0645***	-0.0920***	-0.0555***	-0.0615***	-0.0717***	-0.0699***	-0.0790***	-0.0556***	-0.0403***	-0.0347***	0.00432
	(0.00633)	(0.00878)	(0.00619)	(0.00616)	(0.00829)	(0.00619)	(0.0101)	(0.00715)	(0.00795)	(0.00707)	(0.0217)
ITM_diff	0.00113**	0.00170**	0.00139**	0.00120**	0.000917	0.000362	0.000389	0.000715	0.000489	0.000842*	0.00123**
	(0.000372)	(0.000652)	(0.000470)	(0.000379)	(0.000516)	(0.000369)	(0.000465)	(0.000383)	(0.000324)	(0.000402)	(0.000397)
price_spread	0.0000136***	0.0000195***	0.00000822***	0.00000766**	0.00000751	0.0000101***	0.0000138***	0.00000648**	0.0000110***	0.00000545**	0.00000526
	(0.00000215)	(0.00000311)	(0.00000182)	(0.00000238)	(0.00000384)	(0.00000176)	(0.00000235)	(0.00000206)	(0.00000188)	(0.00000165)	(0.00000417)
_cons	0.00998***	0.00730***	0.00728***	0.00639***	0.00547***	0.0103***	0.00792***	0.00509***	0.00851***	0.00745***	0.00535***
	(0.000779)	(0.00103)	(0.000702)	(0.000681)	(0.000925)	(0.00133)	(0.00157)	(0.00111)	(0.00176)	(0.00215)	(0.00141)
r2	0.174	0.189	0.165	0.171	0.162	0.162	0.177	0.170	0.119	0.140	0.128
N	29767	32896	29345	30193	25301	22672	23274	21304	28573	26288	2355

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Through every year, from 2010 to 2020, the model consistently maintains its explanatory power and all 6 independent variables continue to be significant for each year individually. Furthermore, the signs of the explanatory variables remain consistent across all years implying that the conceptual explanation upon which the variables are constructed maintains its integrity. The regression results for each year add to the robustness of our model and study by conveying that the relationship between the independent and dependent variables remains consistent throughout our sample period when isolated by each year. Hence, the results of the model and its implications for the relationship between the return of the stock and its options market metrics is consistent and robust to different time periods. Additionally, the variable – IVS AVG – remains the only one contradicting our hypothesised relationship between our independent variables and the return of the stock. Thus, the negative sign for the coefficient of IVS AVG for each year reinforces the alternative explanation of hedging activity which results in a lower return on the stock through downward price pressure caused by shorting activity on the stock in response to a higher demand for call options.

Table 9 includes the 5 Fama and French factors along with momentum as control variables to test the robustness of our independent variables in the presence of established factors which have been shown to explain stock returns in the past *(Fama & French, 2015)*

Table 9: Fama-French Control Variable Regression

	(1) RET	(2) RET	(3) RET	(4) RET	(5) RET	(6) RET
DAILY_PC_RATIO	-0.00880***	-0.00878***	-0.00878***	-0.00876***	-0.00876***	-0.00874***
	(0.00146)	(0.00146)	(0.00146)	(0.00146)	(0.00146)	(0.00145)
DAILY_OI_RATIO	-0.00559***	-0.00558***	-0.00557***	-0.00554***	-0.00554***	-0.00554***
	(0.00124)	(0.00124)	(0.00124)	(0.00124)	(0.00124)	(0.00123)
DAILY_IVS	0.00276***	0.00275***	0.00276***	0.00275***	0.00275***	0.00275***
	(0.000653)	(0.000653)	(0.000653)	(0.000652)	(0.000652)	(0.000652)
IVS_avg	-0.0437***	-0.0438***	-0.0438***	-0.0438***	-0.0438***	-0.0437***
	(0.00431)	(0.00431)	(0.00431)	(0.00430)	(0.00430)	(0.00430)
ITM_diff	0.000787*	0.000788*	0.000785*	0.000783*	0.000783*	0.000781*
	(0.000389)	(0.000387)	(0.000387)	(0.000387)	(0.000387)	(0.000386)
price_spread	0.00000782***	0.00000782***	0.00000784***	0.00000787***	0.00000787***	0.00000788***
	(0.00000150)	(0.00000149)	(0.00000149)	(0.00000149)	(0.00000149)	(0.00000149)
mktrf	0.910***	0.881***	0.876***	0.884***	0.881***	0.896***
	(0.0483)	(0.0458)	(0.0449)	(0.0449)	(0.0430)	(0.0431)
smb		0.165***	0.163***	0.148***	0.143***	0.136***
		(0.0289)	(0.0288)	(0.0287)	(0.0276)	(0.0276)
hml			0.0677	0.0142	0.0108	-0.0862**
			(0.0347)	(0.0323)	(0.0322)	(0.0307)
umd				-0.108***	-0.109***	-0.122***
				(0.0204)	(0.0206)	(0.0210)
rmw					-0.0284	-0.0374
					(0.0403)	(0.0404)
cma						0.265***
						(0.0360)
_cons	0.00518***	0.00519***	0.00519***	0.00517***	0.00517***	0.00514***
	(0.00137)	(0.00137)	(0.00137)	(0.00137)	(0.00137)	(0.00137)
r2	0.270	0.271	0.271	0.272	0.272	0.272
N	271968	271968	271968	271968	271968	271968

Section 7.2.1 conducts the Durbin-Watson test and yields the result of no autocorrelation present in our data

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

It is interesting to observe that all our independent variables based on options market parameters remain significant with *ITM diff* being the only exception as despite remaining significant, it does drop in terms of significance level. In all 6 regressions the alpha remains highly significant implying that there are factors influencing stock returns which remain unexplained by the included independent variables. Furthermore, in addition to upholding their significance in the presence of the fama-french plus momentum factors, the coefficients of the options market variables remain roughly the same in terms of magnitude. All our coefficients for the options market variables also maintain their direction in terms of a positive or a negative sign keeping the result in line with our economic reasoning, with *IVS avg* remaining the only explanatory options market variable with exhibits a relation with stock return that contradicts our economic hypothesis for the same.

It is important to note that the Durbin-Watson Test mentioned in appendix 7.2.1 conducted on our 6 independent options market variables rejects the hypothesis of autocorrelation being present in our data and model. The regression in Table 9 also follows a stock-fixed effects regression with clustered standard errors.

Table 10 displays the output of the regression of our main model using monthly data, with monthly averaged values for return and the options market metrics which are used as the independent variables.

Table 10: Comparison of the model with monthly and daily data

VARIABLES	RET (Monthly)	RET (Daily)
DAILY_PC_RATIO	-0.0103***	-0.0122***
	(0.0013)	(0.0015)
DAILY_OI_RATIO	-0.0049***	-0.0071***
	(0.0009)	(0.0013)
DAILY_IVS	0.0016***	0.0030***
	(0.0006)	(0.0007)
IVS_avg	-0.0368***	-0.0558***
	(0.004)	(0.0039)
ITM_diff	0.0015***	0.0011***
	(0.0004)	(0.0004)
price_spread	0.000011***	0.0000***
	(0.000001)	(0.0000)
Constant	0.0062***	0.0075***
	(0.0009)	(0.0014)
Observations	34,881	271,968
Number of secid	981	981
R-squared	0.063	0.146

An important thing to note is that is that the explanatory power of the model drops significantly following the transformation of daily data to monthly data which is in line with our expectations of the model's and options market metrics' ability to explain reverse causality running from the options market to the stock market. However, all 6 independent variables continue to remain significant at the monthly level, reinforcing the suitability of our variables in explaining stock market returns.

Additionally, the signs of the coefficients continues to remain the same when using monthly data and the absolute value or magnitude of the coefficient is also roughly around the same level. Furthermore, using the VIF test to detect multicollinearity rejects the hypothesis supporting significant correlation between our independent variables for daily and monthly data warranting no further investigation in terms of multicollinearity concerns in our model.

5.3 Factor Regression

Table 11 uses factors constructed from our independent variables to explain returns of the stocks in our sample. The first model boasts an R-square of 4.07% which is impressive given the lagged nature of the variables from which our long-short portfolio factors are constructed. The explanatory power as expected jumps significantly following the addition of the mktrf factor in particular, followed by the rest of the control variables.

In the first regression, all factors with the exception of the price spread factor are observed to be significant. However, barring the Agg IVS factor, the coefficient of all other factors comes out to be negative, indicating that the the stocks in our data have negative exposure to these factors. Furthermore, it is possible that due to hedging activities, the pressure on the stock's return moves in the opposite direction to which the factors insinuate. High put-call values which represent the short leg of the portfolio imply that there are a large number of put options traded on the stock, and the hedging focused trade on the stock which would follow would be to buy or hold the stock so as to lower or eliminate the risk of loss arising due to a rise in the price of the stock by investors holding these put options. Similarly, It is particularly interesting to note the difference in the sign of the coefficients of the Agg IVS and Avg IVS factors, however, when controlled for the market return (mktrf), size (smb), value (hml), momentum (umd), profitability (rmw) and investment (cma), half of our factors - put call factor, open interest factor and Agg IVS factor - lose their significance in the second regression. However, the remaining three factors - Avg IVS factor, ITM factor and price spread factor - remain highly to moderately significant while possessing coefficients with reasonably strong absolute magnitudes. The three significant factors display a negative sign on the coefficient even in the presence of our control variables indicating the robustness of the result.

The negative relation of our constructed factors with the returns could imply that it is the result of a premium which investors are paying i order to bet in the direction of the market as opposed to against the market. It is important to note that the first regression yields a highly significant yet marginal alpha which persists even in the presence of our control variables.

Table 11: All-Factor Regression

	(1)	(2)
	ret	ret
put_call_factor	-0.0666**	0.00840
	(0.0204)	(0.0186)
open_interest_factor	-0.385***	-0.0209
	(0.0339)	(0.0192)
agg_IVS_factor	0.0438***	-0.00967
	(0.00973)	(0.00930)
avg_IVS_factor	-0.355***	-0.0884***
	(0.0269)	(0.0180)
itm_factor	-0.141***	-0.0237**
	(0.0122)	(0.00828)
price_spread_factor	-0.000155	-0.0528**
	(0.0209)	(0.0186)
mktrf		1.023***
		(0.0489)
smb		0.0670
		(0.0358)
hml		-0.0939*
		(0.0414)
umd		-0.127***
		(0.0253)
rmw		-0.0183
		(0.0518)
cma		0.315***
		(0.0521)
_cons	-0.000182***	-0.000340***
	(0.0000287)	(0.0000333)
r2	0.0407	0.182
N	152018	152018

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12 displays the results of our regression employed with the intention of finding an alpha for each of our 6 long-short portfolio factors and investigate whether the return generated by these factors can be explained by the 5 fama-french factors and the momentum anomaly.

Table 12: Alpha Regression

	(1) put_call_factor	(2) open_interest_factor	(3) agg_IVS_factor	(4) avg_IVS_factor	(5) itm_factor	(6) price_spread_factor
mktrf	-0.198***	-0.239***	-0.0515*	-0.0345	-0.371***	-0.0187
	(0.0238)	(0.0257)	(0.0248)	(0.0221)	(0.0503)	(0.0221)
smb	-0.0657	-0.0821	-0.0123	-0.0611	0.0213	0.0397
	(0.0375)	(0.0441)	(0.0403)	(0.0416)	(0.0776)	(0.0364)
hml	-0.0548	-0.0989	0.0846	0.202***	0.0391	0.105*
	(0.0496)	(0.0508)	(0.0588)	(0.0521)	(0.0890)	(0.0508)
umd	0.0426	0.124***	0.0816*	0.0473	0.0153	0.0978**
	(0.0320)	(0.0330)	(0.0341)	(0.0344)	(0.0604)	(0.0311)
rmw	-0.0156	0.0814	0.0747	0.219***	0.130	0.0759
	(0.0611)	(0.0651)	(0.0723)	(0.0609)	(0.111)	(0.0671)
cma	-0.0367	0.108	-0.0336	0.0938	-0.0985	0.0730
	(0.0779)	(0.0827)	(0.0902)	(0.0788)	(0.141)	(0.0885)
_cons	-0.000527**	-0.000719***	-0.000475*	0.000523**	-0.00137***	-0.000349*
	(0.000165)	(0.000170)	(0.000188)	(0.000170)	(0.000291)	(0.000168)
r2	0.0750	0.128	0.00934	0.0355	0.0750	0.0115
N	1985	1985	1985	1985	1985	1985

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As can be observed from the regression, we find that 4 of our factors with the exception of the Agg IVS factor and price spread factor maintain a high level of significance implying that the positive or negative returns generated by these portfolios are not fully explained by the fama-french 5 factor model along with momentum. While the absolute magnitudes of the constants or alphas are minute, their robust significance indicates that our portfolios generate positive or negative returns which could be classified as an anomaly. However, it also presents an avenue for further research which would involve the analysis of the return of our long-short portfolios using hedging based indicators as control variables to determine whether the alphas persist.

6 Conclusion

The paper aimed to investigate the reverse causality between the stock market and the options market, essentially focusing on determining the presence of a feedback effect. The study is built upon the informed trading argument in the options market which implies that investors possessing private information regarding the stock's returns should be expected to take advantage of such information through investing or 'informed speculation' in the options market due to the leverage available in said market, which can significantly amplify the gains arising out of changes in the underlying stock's price. The paper yields results which are mostly analogous with literature mentioned in the paper and of the past primarily adding weight to the support of the options market explaining stock returns while being dissimilar to an extent in terms of long-short portfolio returns which can be majorly explained due to the use of an all-options approach and using an identical dataset for the construction of all our variables.

Using daily options market metrics based upon the put-call volume and open interest ratios, implied volatility spreads of call and put options on a stock, its moneyness, and the difference between the average call premiums and put premiums of the stock, our model yields significant results with an economically reasonable level of explanatory power supporting the hypothesis that a high put-call volume and open interest ratio explains negative return movement, while an aggregated implied volatility spread of a stock averaged for the total number of call and put options on the stock for the day is positively linked to the underlying's return, as is the the difference between the number of ITM call and put options for a given day of a stock and the price spread of call and put options. A finding contradictory to our initial hypothesis is the relation between the spread of average implied call volatility and average implied put volatility on the stock for a given day, with the coefficient possessing a negative return indicating that a higher magnitude of average call option volatility relative to average put option volatility is negatively linked to stock returns. The finding also contradicts prior literature on the study of Implied Volatility Spreads and its relationship with stock market returns such as *(Gao et al. 2019)* and *(Xing et al. 2010)*, however, their studies also differ from our paper in the construction the Implied Volatility Spread measure in that their IVS factor was constructed utilising at-the-money put and call options with the same expiration date and strike price, while our study averages all put and call options on the stock irrespective of moneyness to arrive at the IVS AVG measure. Additionally, our paper extends a hedging activity based explanation for the contradictory finding, based upon shorting activity in the stock market to hedge the risk of a call option expiring out of the money or investors engaging in a protected put strategy resulting in a positive return on the stock driven by higher demand to hold the stock as a hedge for an existing long put position on the stock.

The factor regressions point towards low significance overall as for our factors, with only three of them remaining significant in the presence of control variables when the risk-free daily stock return is regressed upon our long-short portfolio factors and the fama-french plus momentum control variables. The conclusion that the long-short portfolios do not yield the desired positive results when investments are made in accordance with their long and short legs is one which can be logically explained as trading in the options market is predominantly intr-day, hence, while return movements can be explained by options market metrics at a daily frequency, the factors constructed from one-day lagged values fail to keep up with the intraday movements in options markets which could be expected to primarily predict stock returns. Furthermore, the study of long-short factors is a purely theoretical approach as daily rebalancing demanded by the factors constructed would undoubtedly result in exorbitant transaction costs which would make it unfeasible to invest in accordance with these factors. Additionally, our study found significant alphas for most of our portfolio factors when using fama-french plus momentum factors as controls pointing towards a course for further research on the topic which could incorporate hedging-based trades made in the options market as control variables to determine the robustness of the significance of the alphas we found.

The study concludes with significant evidence for the presence of reverse causality running from the options market to the underlying equity securities. An important result of our study is its implication for option pricing models which utilise the underlying stock's price to calculate the price of the call or put option. In the presence of reverse causality, the stock price should no longer be assumed as exogenous in the process of determining the option's price. Instead, it should be understood as an endogenous input in the determination of an option's value. The implications of the study opens doors for further research across asset classes such as equity indices and commodities which may also be exhibiting a feedback or reverse causality loop with their derivative market counterparts. Additionally, the same study carried out o=using intraday frequency of data for a shortened sample size and time horizon under consideration would certainly extend the concepts explored in our study and would provide another dimension to the lead lag relationship. Another area for further research could be of constructing more sophisticated variables by using differing data sets or limiting the study to specific variables inspired by our paper's variable construction approach and analysing the implications. The long-short portfolio factors could also be constructed as double sorted portfolios when limiting the study to fewer explanatory variables and would provide a more in-depth analysis of whether options market movements can predict future stock returns.

7 Appendix

7.1 Tables

Table 13: Daily PC Ratio

Year	Obs	Mean	Std. Dev.	Min	Max
2010	29,931	0.436	0.231	0	0.995
2011	33,074	0.443	0.238	0	0.997
2012	29,603	0.458	0.234	0	0.998
2013	30,413	0.442	0.241	0	0.999
2014	25,549	0.452	0.255	0.001	0.998
2015	22,903	0.488	0.264	0.002	0.999
2016	23,470	0.472	0.255	0.002	0.999
2017	21,501	0.450	0.254	0.001	0.999
2018	28,837	0.447	0.252	0	0.999
2019	26,592	0.444	0.250	0	0.999
2020	2,376	0.411	0.249	0.003	0.998

Table 14: Daily OI Ratio

Year	Obs	Mean	Std. Dev.	Min	Max
2010	29,931	0.416	0.263	0	0.998
2011	33,074	0.445	0.269	0.001	0.999
2012	29,603	0.441	0.272	0.001	0.999
2013	30,413	0.402	0.274	0	0.999
2014	25,549	0.429	0.293	0.001	0.999
2015	22,903	0.489	0.301	0.001	0.999
2016	23,470	0.440	0.284	0	0.999
2017	21,501	0.417	0.288	0	1
2018	28,837	0.440	0.288	0.001	1
2019	26,592	0.419	0.278	0	1
2020	2,376	0.383	0.279	0	0.998

Table 15: Daily IVS

Year	Obs	Mean	Std. Dev.	Min	Max
2010	29,931	0.228	1.357	-37.441	22.749
2011	33,074	0.251	1.906	-33.252	39.453
2012	29,603	0.094	1.074	-16.606	14.327
2013	30,413	0.144	1.169	-55.341	21.715
2014	25,549	0.116	1.330	-18.492	31.157
2015	22,903	-0.081	1.646	-32.617	20.735
2016	23,470	0.080	1.617	-26.996	34.517
2017	21,501	0.247	1.439	-25.823	42.450
2018	28,837	0.159	1.978	-74.270	24.511
2019	26,592	0.214	1.997	-60.905	36.433
2020	2,376	0.566	2.388	-31.686	23.922

Table 16: IVS Average

Year	Obs	Mean	Std. Dev.	Min	Max
2010	29,767	0.002	0.059	-1.019	1.258
2011	32,896	-0.003	0.068	-1.811	1.142
2012	29,345	-0.005	0.073	-2.291	1.145
2013	30,193	-0.004	0.056	-1.201	0.723
2014	25,301	-0.005	0.069	-1.866	0.914
2015	22,672	-0.011	0.082	-1.838	1.092
2016	23,274	-0.013	0.084	-1.640	2.303
2017	21,304	-0.005	0.072	-2.445	1.704
2018	28,573	-0.005	0.072	-1.596	0.885
2019	26,288	-0.009	0.076	-1.691	2.472
2020	2,355	-0.005	0.070	-0.718	1.318

Table 17: ITM Difference

Year	Obs	Mean	Std. Dev.	Min	Max
2010	29,931	0.561	2.206	-31	30
2011	33,074	0.529	2.560	-45	45
2012	29,603	0.397	2.034	-19	41
2013	30,413	0.475	2.195	-32	54
2014	25,549	0.315	2.321	-35	57
2015	22,903	0.064	2.548	-96	42
2016	23,470	0.344	2.348	-30	39
2017	21,501	0.537	2.442	-24	37
2018	28,837	0.160	3.000	-124	46
2019	26,592	0.405	2.622	-52	57
2020	2,376	0.766	2.892	-15	28

Table 18: Price Spread

Year	Obs	Mean	Std. Dev.	Min	Max
2010	29,931	32.942	183.583	-4,435	3,702.5
2011	33,074	32.646	202.334	-4,084.542	4,493.929
2012	29,603	19.559	212.226	-20,891.25	3,062.5
2013	30,413	28.634	245.145	-23,890	12,107.5
2014	25,549	15.637	248.496	-4,601.25	16,321.285
2015	22,903	-15.754	275.975	-6,350	9,475
2016	23,470	0.849	228.578	-5,275	6,955
2017	21,501	27.871	268.506	-10,345	7,410.398
2018	28,837	-0.559	306.720	-8,657.5	8,817.5
2019	26,592	14.854	320.743	-7,981.625	5,052.5
2020	2,376	66.660	318.822	-1,035	7,962.133

Table 19: Variance Inflation Factor for Daily Data

Variable	VIF	1/VIF
ITM diff	1.853	0.540
Daily IVS	1.793	0.558
Daily OI Ratio	1.597	0.626
Daily PC Ratio	1.586	0.631
Price spread	1.285	0.778
IVS avg	1.112	0.900
Mean VIF	1.538	

Table 20: Variance Inflation Factor for Monthly Data

Variable	VIF	1/VIF
ITM diff	1.657	0.604
Daily IVS	1.578	0.634
Daily OI Ratio	1.578	0.634
Daily PC Ratio	1.570	0.637
Price spread	1.240	0.806
IVS avg	1.108	0.902
Mean VIF	1.455	

Table 21: Put Call Factor

Year	Obs	Mean	Std. Dev.	Min	Max
2010	199	-0.00022	0.00685	-0.03254	0.01849
2011	199	-0.00069	0.00756	-0.02610	0.02479
2012	194	0.00030	0.00632	-0.01924	0.01643
2013	196	-0.00047	0.00606	-0.01781	0.01598
2014	197	-0.00059	0.00716	-0.03234	0.01676
2015	198	-0.00141	0.00911	-0.02845	0.02887
2016	199	-0.00089	0.00890	-0.03132	0.02300
2017	197	-0.00064	0.00811	-0.03721	0.02345
2018	194	-0.00062	0.00711	-0.02155	0.02259
2019	196	-0.00133	0.00791	-0.02592	0.04240
2020	16	0.00139	0.00743	-0.01703	0.01180

Table 22: Open Interest Factor

Year	Obs	Mean	Std. Dev.	Min	Max
2010	199	-0.00050	0.00754	-0.03151	0.02504
2011	199	-0.00083	0.00947	-0.03581	0.04750
2012	194	0.00026	0.00615	-0.01718	0.01247
2013	196	-0.00092	0.00634	-0.02166	0.01320
2014	197	-0.00088	0.00791	-0.02509	0.01909
2015	198	-0.00140	0.00892	-0.03041	0.02235
2016	199	-0.00147	0.00906	-0.02246	0.02369
2017	197	-0.00077	0.00824	-0.04409	0.02249
2018	194	-0.00076	0.00742	-0.02146	0.02004
2019	196	-0.00134	0.00825	-0.02334	0.03845
2020	16	0.00104	0.00893	-0.01759	0.01055

Table 23: Agg IVS Factor

Year	Obs	Mean	Std. Dev.	Min	Max
2010	199	-0.00033	0.00608	-0.02067	0.01494
2011	199	-0.00029	0.00684	-0.02670	0.02408
2012	194	0.00051	0.00639	-0.01815	0.02212
2013	196	-0.00054	0.00630	-0.03171	0.01746
2014	197	-0.00144	0.00833	-0.02805	0.02123
2015	198	-0.00010	0.01089	-0.03870	0.04506
2016	199	-0.00119	0.00977	-0.03783	0.02073
2017	197	-0.00058	0.00885	-0.05906	0.02169
2018	194	-0.00048	0.00892	-0.02622	0.04662
2019	196	-0.00078	0.00948	-0.04095	0.04181
2020	16	0.00072	0.00847	-0.01338	0.01399

Table 24: Avg IVS Factor

Year	Obs	Mean	Std. Dev.	Min	Max
2010	199	-0.00022	0.00685	-0.03254	0.01849
2011	199	-0.00069	0.00756	-0.02610	0.02479
2012	194	0.00030	0.00632	-0.01924	0.01643
2013	196	-0.00047	0.00606	-0.01781	0.01598
2014	197	-0.00059	0.00716	-0.03234	0.01676
2015	198	-0.00141	0.00911	-0.02845	0.02887
2016	199	-0.00089	0.00890	-0.03132	0.02300
2017	197	-0.00064	0.00811	-0.03721	0.02345
2018	194	-0.00062	0.00711	-0.02155	0.02259
2019	196	-0.00133	0.00791	-0.02592	0.04240
2020	16	0.00139	0.00743	-0.01703	0.01180

Table 25: ITM Factor

Year	Obs	Mean	Std. Dev.	Min	Max
2010	199	-0.00058	0.01258	-0.05791	0.03680
2011	199	-0.00045	0.01609	-0.07554	0.05444
2012	194	0.00016	0.00882	-0.02399	0.03076
2013	196	-0.00224	0.01106	-0.03899	0.03907
2014	197	-0.00242	0.01458	-0.09848	0.04487
2015	198	-0.00228	0.01506	-0.06052	0.04211
2016	199	-0.00161	0.01604	-0.08488	0.06191
2017	197	-0.00160	0.01301	-0.09515	0.03317
2018	194	-0.00238	0.01094	-0.05434	0.02111
2019	196	-0.00301	0.01273	-0.04694	0.03444
2020	16	0.00158	0.01217	-0.02932	0.02223

Table 26: Price Spread Factor

Year	Obs	Mean	Std. Dev.	Min	Max
2010	199	0.00015	0.00551	-0.01834	0.01616
2011	199	0.00069	0.00597	-0.02022	0.01960
2012	194	-0.00025	0.00625	-0.01665	0.02285
2013	196	-0.00026	0.00646	-0.01696	0.01780
2014	197	-0.00127	0.00754	-0.02811	0.01909
2015	198	-0.00104	0.00915	-0.02700	0.02645
2016	199	0.00004	0.00908	-0.02798	0.04103
2017	197	-0.00032	0.00841	-0.03396	0.02708
2018	194	-0.00112	0.00736	-0.02863	0.02215
2019	196	-0.00034	0.00830	-0.04918	0.02029
2020	16	0.00043	0.00429	-0.00640	0.00702

7.2 Tests

7.2.1 Wooldridge test for autocorrelation in panel data

H0: no first order autocorrelation
F(1, 559) = 0.326
Prob > F = 0.5685

The test fails to reject the hypothesis of the absence of serial autocorrelation indicating that the series is not AR(1), allowing us to proceed with our primary regression being assured of the fact that autocorrelation is not present in our regression.

7.2.2 Hausman (1978) specification test

Coef.
Chi-square test value = 73.781
P-value = 0

The Hausman test indicates that the fixed effects regression is more suitable for our data reinforcing the belief that there would be systematic time-invariant effects across stock returns.

7.2.3 Ramsey RESET test for omitted variables

Omitted: Powers of fitted values of RET
H0: Model has no omitted variables
F(3, 271958) = 456.12
Prob > F = 0.0000

The test points towards omitted variable bias in our primary regression which is expected given that primary stock market movers are not included in this regression such as market sentiment, macroeconomic changes and other factors.

7.2.4 Ramsey RESET test for omitted variables

Omitted: Powers of fitted values of RET
H0: Model has no omitted variables
F(3, 271952) = 533.07

Prob > F = 0.0000

The test continues to indicate the presence of omitted variable bias when the fama-french 5 factors and momentum are included as control variables. While it may point towards biased estimates, it is natural for this to be the case when working with financial data, in particular stock returns, as all factors influencing stock returns are impossible to quantify, much less include in our regression.

8 References

Black, F., Scholes, M. (1973). The pricing of options and corporate liabilities. Journal of Political Economy, 81(3), 637–654. https://doi.org/10.1086/260062

Fama, E. F., French, K. R. (2015). A five-factor asset pricing model. Journal of Financial Economics, 116(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

Fama, E. F., French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3–56. https://doi.org/10.1016/0304-405x(93)90023-5

Chakravarty, S., Gulen, H., Mayhew, S. (2004). Informed trading in stock and options markets. the Journal of Finance/the Journal of Finance, 59(3), 1235–1257. https://doi.org/10.1111/j.1540-6261.2004.00661.x

Ho, K., Zheng, L., Zhang, Z. (2012). Volume, volatility and information linkages in the stock and options markets. Review of Financial Economics, 21(4), 168–174. https://doi.org/10.1016/j.rfe.2012.06.001

Pan, J., Poteshman, A. (2004). The information of option volume for future stock prices. https://doi.org/10.3386/w10925

Goncalves-Pinto, L., Grundy, B. D., Hameed, A., Van Der Heijden, T., Zhu, Y. (2020). Why do option prices predict stock returns? The role of price pressure in the stock market. Management Science, 66(9), 3903–3926. https://doi.org/10.1287/mnsc.2019.3398

Easley, D., O'Hara, M., Srinivas, P. (1998). Option volume and stock prices: evidence on where informed traders trade. the Journal of Finance/the Journal of Finance, 53(2), 431–465. https://doi.org/10.1111/0022-1082.194060

Xing, Y., Zhang, X., Zhao, R. (2010). What does the individual option volatility Smirk tell us about future equity returns? Journal of Financial and Quantitative Analysis, 45(3), 641–662. https://doi.org/10.1017/s0022109010000220

Patel, V., Putniņš, T. J., Michayluk, D., Foley, S. (2020). Price discovery in stock and options markets. Journal of Financial Markets, 47, 100524. https://doi.org/10.1016/j.finmar.2019.100524

Chan, K., Chung, Y. P., Fong, W. M. (1999). The informational role of stock and option volume. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.170356

Mayhew, S., Stivers, C. (2003). Stock return dynamics, option volume, and the information content of implied volatility. Journal of Futures Markets, 23(7), 615–646. https://doi.org/10.1002/fut.10084

Boluch, M. J., Chamberlain, T. W. (1997). Option volume and stock price behavior: Some evidence from the Chicago board options exchange. Atlantic Economic Journal, 25(4), 358–370. https://doi.org/10.1007/bf02298346

Bergsma, K., Csapi, V., Diavatopoulos, D., Fodor, A. (2019). Show me the money: Option moneyness concentration and future stock returns. Journal of Futures Markets, 40(5), 761–775. https://doi.org/10.1002/fut.22071

Gao, X., Wang, X., Yan, Z. (2019). Attention: implied volatility spreads and stock returns. Journal of Behavioral Finance, 21(4), 385–398. https://doi.org/10.1080/15427560.2019.1692846