BI

WISEflow
Europe/Oslo(CEST)
24 Jun 2024

Handelshøyskolen BI

GRA 19703 Master Thesis

Final Thesis Master of Science W 100%

Predefinert informasjon

Startdato:	08-01-2024 09:00 CET
Sluttdato:	01-07-2024 12:00 CEST
Eksamensform:	T
Termin:	202410
Vurderingsform:	Norsk 6-trinns skala (A-F)
Flowkode:	202410||11436||IN00||W||T
External assessor:	External assessor 1
Internal assessor:	Internal assessor 1

Deltaker

Navn:	Martinus Teige og Raymon Meede Gjuvsland

Informasjon fra deltaker

Tittel *:	Options Trading and its Predictive Power on Short-Term Stock Market Movements
Navn på veileder *:	Adam Winegar

Inneholder besvarelsen konfidensielt materiale?:	Nei	**Kan besvarelsen offentliggjøres?:**	Ja

G r u p p e

Gruppenavn:	(Anonymisert)
Gruppenummer:	1
Andre medlemmer i gruppen:	

Options Trading and its Predictive Power on Short-Term Stock Market Movements

Master Thesis

by

Martinus Olavius Teige and Raymon Meeder Gjuvsland

MSc in Business with Major in Finance

Name of Supervisor:

Adam Winegar

Oslo, 01 July 2024

Abstract:

Investigating options trading from 2021-2023, we find that the predictive power of the options market on the underlying stock market has changed relative to previous studies. Prior studies examine market data from 1995-2010, before the introduction of short-term options and the rise of retail options trading. We find, in contrast to previous studies, that call options are more statistically significant predictors of share price movements than put options. Moreover, we find that heightened call options trading predicts negative next-day returns, while heightened put options trading predicts positive returns, contrary to common intuition. These relationships change across different years and market structures but are consistent in the direction of their prediction. Additionally, we find no evidence that option traders within the longer dated options market are more sophisticated than traders in the shorted dated options market.

Acknowledgements

We want to express our gratitude to our supervisor Adam Winegar, who has helped us throughout the thesis with support, ideas, and execution. Through discussions and constructive feedback his help has been instrumental in forming the thesis.

AI tools, such as ChatGPT, have been used improve upon code in the computer programs which we use. However, it has not independently suggested methods, statistical tests, or otherwise worked without clear user input.

Table of Contents

1. Introduction

Several papers have been written on options and options trading, particularly with regards to pricing. There is also a large literature that deals with how options trading affect, or predict, stock market returns. The literature has, for the most part, agreed that higher option volume relative to the stock volume reflects negative news, and subsequent lower future returns for those shares. The literature has mainly attributed this to trading in put options. However, the vast majority of this literature is now more than 10 years old and the data is usually from the 1990's to the very early 2010's. Since then, there has been an increase in option volumes, with shorted dated options becoming more popular (see Figure 1), in addition to an influx of retail traders (Bryzgalova et al., 2023). These changes have the potential to affect the dynamics of the options market and may yield different conclusions than what previous literature has found.

Our thesis aims to answer the following questions: Do call and put options trading exhibit the same characteristics as they did historically in terms of their predictability of stock returns? Does this effect vary over time, and if so, how does this behavior differentiate itself in markets that are uptrending or downtrending? Are there differences when examining short-term dated option trading vs. longer term-dated option trading? Will the options on a subsample of 15 large stocks, be able to predict the returns of the S&P 500 Index?

We examine a sample of options trading on the 15 largest stocks in the US as of December 1, 2023, over the period from January 1, 2021 to December 1, 2023. Utilizing detailed option data, we examine how the options volume influences underlying returns. This analysis includes logarithmic and excess returns for both individual stocks, and the combined dataset. As our sample involve both "bull" and "bear" years for the S&P 500, we see how these effects change when looking at the separate years. To remove the influence of external factors unrelated to the relation between the options market and stock market, we include several controls, such as CPI announcement and earnings announcement dates.

First, we show that the predictive ability of call and put options has changed compared to what large parts of the literature find. We find that high levels of call options volume, relative to the volume of the underlying security, predict negative next-day returns, and that heightened volume of put options predict positive returns the next day. This effect varies depending on the exact regression specification, but call options volume tends to be statistically significant at the 5% level for predicting logarithmic returns, while put options volume tends to be less statistically significant.

Secondly, we run separate regressions on the individual years to account for different market structures in 2021, 2022 and 2023. We show that while the direction in which call and put options affects next-day returns is constant over these years, the size of this effect changes. The statistical significance varies across individual years, which could be due to different market structures.

Thirdly, building on the recent uptick of retail trading within the options market, especially in short-dated options (Bryzgalova et al., 2020), we show that when using volume from short-dated options, there are minimal changes in the predictability of next-day returns, when compared to longer-dated options. This would suggest that information from market participants in the longer dated options market is not more informative in predicting returns than information from the short-dated options market.

Fourthly, we do not find any statistically significant results from our regression on the S&P 500 index, in addition, the coefficient values and signs are inconsistent, and present no discernible pattern across specifications. This could suggest that our sample of stocks is not a good proxy for the broad market, or that the broad markets response to options volume differs from the response for individual stocks.

2. Background

The motivation for this thesis stems from the significant increase in options volume observed in American stocks and indices in recent years. There has been a strong increase in options trading, both in the short-term as well as the long-term (see Figure 1), and they are becoming an increasingly important part of US financial markets. Given the record-high interest in options, investigating how options volume affects stock returns could yield new insights and provide a better understanding of the relationship in the current market. As option volumes relative to the underlying stocks' volume increase, we may observe irrational price movements in response to changes in the underlying stock prices due to hedging activities initiated by the option traders. To examine this, we collect several data points on options and shares, such as prices, volumes, and option-specific variables.

Options trading has grown in popularity throughout the 2010's, with a sharp uptick in 2020 as the pandemic hit. From 2000 to 2010 the number of options traded daily grew from around 5 million to 15 million. This number was relatively stable up until 2018, but in 2022 the number had grown to more than 40 million (see Figure 1). This trend is believed at least partially fueled by zero-commission trading platforms such as Robinhood (Mark, 2021). Additionally, there has been a recent trend towards shorter time to expiry options, with <30 days to expiry options increasing their share of options volume from 11,9% in November 2019 to 30,7% in September 2023 (see Figure 2).

We find these trends fascinating, and it is evident that even though there was a spike in equity and option trading in 2021, it seems that options trading, and perhaps especially amongst retail investors, is a trend that is here to stay. With more competitors offering zero-based commission trading, such as Webull[1] & Moomoo[2], we see these trends as likely to continue. Both from a purely financial and economical standpoint, as well as from a policy and regulatory point of view, it will be helpful to understand how these trends impact the financial markets.

Traditionally, options have been traded primarily by professional investors, who are assumed to make more rational trading decisions than retail investors. As the option market has become more accessible to retail investors due to the influx of the mentioned trading platforms, combined with the recent shifts towards shorter-dated options, we believe it is time to reexamine if these changes have impacted the predictability of options trading on the returns of the underlying stocks, as previously identified in historical studies.

[1] https://www.webull.com/

[2] https://www.moomoo.com/us/

3. Literature Review

This thesis builds upon existing literature examining the relationship between options volume and the underlying share price movements. There is a wide literature on this subject, but most of the papers use data from the 1990's to the early 2010's, potentially limiting their insights into how the option and stock markets behave in the post-pandemic period.

The Options/Stock Trading Volume Ratio and Option Markets

Roll, Schwartz and Subrahmanyam (2010) first define the options-to-stock ratio (O/S ratio) and use it as a predictive tool. This ratio is calculated by dividing the trading volume of options by the trading volume of the shares of a given firm, both given in US dollars

$$\frac{O}{S} = \frac{Volume_{Options}}{Volume_{Stocks}}$$

Roll, Schwartz and Subrahmanyam construct this ratio for a period spanning almost 3,000 trading days, using data from 1996-2007, with the average number of firms ranging from 752 and 2007 during these years. The authors' primary objective in examining the O/S ratio is to determine what drives higher or lower relative options volume. They find that the O/S ratio depends on various factors, including trading costs, firm size, the degree of leverage available in options, institutional holdings, and analyst coverage. In incomplete markets, options cannot be dynamically replicated with stocks and bonds, and they argue options may help complete markets and enhance welfare. The options market is therefore believed to be a more efficient trading venue for the informed agents, compared to trading the underlying shares. Ni, Pan and Poteshman (2008) show that option markets attract traders informed about future volatility. Easley, O'hara and Srinivias (1998) also conclude that options are attractive to informed traders if they do not believe the current, or new information, is incorporated into the stock prices. As the share of retail trading has increased substantially since these papers have been published, these relations may not hold anymore.

During the 2008 short sale ban on certain stocks, informed investors were the primary traders of derivatives in this period, leading to options being more informationally efficient compared to stock prices (Ni & Pan, 2020). The findings from Roll et al. (2010) show that factors like institutional holdings are negatively associated with the O/S ratio. It implies that stocks which are well known and held by institutions have a higher level of sophisticated individual investors, less "private" information to be uncovered, and therefore less traders acting on informed information. On the other hand, Roll et al. (2010) argue that analysts' dispersion on earnings forecast should be positively correlated with O/S, while in fact the opposite is true, and there is a negative correlation. Although there are conflicting views, the key finding is that it is likely that trading in the options markets depends on external factors. Our thesis does not aim to investigate why some informed traders prefer options to stocks, or how we can predict the option-to-stock volume ratio, but we rather use this as a foundation in building the model we use to predict stock returns.

When examining the predictability of the O/S ratio on post-earnings-announcement cumulative absolute returns (CAR), Roll et al. (2010) make two findings. A high O/S ratio ahead of positive earnings announcements implies bigger positive CAR in the following period, but high O/S ahead of negative earnings announcements imply a larger negative CAR. This indicates that informed traders are informed regardless of the direction, but given that they find a larger absolute coefficient prior to negative announcements, this implies that option markets are more attractive to traders with negative views. They argue that this is due to constraints and imperfect markets, such as short-selling constraints. We contribute to the prior literature on this topic by extending the number of variables, running call and put volume separately instead of pooled together, and lastly using a sample period which includes the recent changes in the options market that plausibly would affect the conditions and assumptions underlying the prior findings. As previously noted, most of the literature use data which is a decade or more old, despite the changing market dynamics. This warrants new research, and potentially new conclusions drawn, compared to the findings of previous literature.

Examining Put and Call Options Separately

Many papers conclude that volume on put options is more important in predicting future returns (Roll et al., 2009 & Johnson and So, 2012). However, Ge et al. (2016) find that examining put and call volumes separately reveals that, contrary to the conclusions of many other studies, call volumes are more significant in explaining future returns. Collecting data from 2005-2012, they show that the purchase volume of put options negatively predicts returns, while the purchase volume of call options positively predicts returns with a larger coefficient. For both parameter estimates they exclude option expiration weeks from their sample, and find that both are statistically different from zero at the 1% significance level. These findings are also consistent with Bergsma et al. (2020), who find predictive power of options in predicting same day returns from bullish option trades (buy calls or sell puts) and bearish option trades (sell calls or buy puts) within the first thirty minutes of the trading day. We build upon Ge et al. (2016) and Bergsma et al. (2020) by examining put and call options separately, but with newer data. Our findings show that call options are more statistically significant than put options in predicting returns, but that call options predict negative future returns, while put options predict positive future returns.

Retail Trends in the Options Market

Retail trading in options has become very popular in recent years, especially among young investors. Bryzgalova et al. (2023) estimate that more than half of the total trading volume in options consisted of retail trading in 2021, with the total retail volume close to doubling from 2020 to 2021. Statistics from the NYSE (Poser, 2023b) show that this share has remained constant from 2021 until 2023. These statistics also show that the share of total short-term options (less than one month to expiry) being traded by retail investors has increased substantially from 2020 to 2023 (see Figure 2 and 3). This is consistent with Bryzgalova et al. (2023), arguing that the options market for these types of investors is mostly speculation-driven, making the short-term options more suitable for these purposes. They also find that retail traders have a positive tilt toward call options, with call and put options representing 69% and 31% of the retail options volume respectively. We build on parts of these findings by examining the significance level of call options volume versus put options volume. We investigate if retail traders create additional noise in call option volumes, and if this renders these volumes with little to no predictive power. We also examine whether there is a clear difference in statistical significance and predictability between shorter-dated options trading volume (7 days or less) and longer-dated options volume (more than 7 days), as retail traders prefer shorter-dated options more than the general market does.

4. Hypothesis

In the following section, we outline our key hypotheses and the variables of interest. Our thesis aims to investigate the following four hypotheses.

Hypothesis 1:

Retail trading makes the options market less informational, and should lead to less statistically significant results on the volume of options, especially on calls. Options trading for these types of market participants are to a large extent a purely speculative vehicle (Bryzgalova et al., 2020). We therefore expect that put option volumes will be more informational, and more statistically significant, than call option volumes. We also believe that the coefficient is economically significant, which means that the coefficient is large enough in comparison to other variables such that it has a real economic impact.

Hypothesis 2:

The three years 2021, 2022, and 2023, have different characteristics in the parameter estimates of the effect of call- and put-volume on the share prices. For the years 2021, 2022 & 2023, the S&P 500 returned 26.9%, -19.4% and 24.2% respectively (Macrotrends, n.d.). A change in the structure of the market, we argue, will change who trades options, what direction they trade, how they trade, and the predictability that the options volume has on future stock market returns. In their paper, Roll et al. (2010) also find structural changes in the option markets over time, with increasing volume every year leading up to earnings announcements. Our hypothesis is that when running separate regressions on the three years, the coefficients for the call- and put-volumes will be different from each other.

Hypothesis 3:

Supplementing the previous regressions, we run two additional regressions; one on options traded with 7 days or less to expiry, and one on options traded with more than 7 days to expiry. We believe that due to a higher proportional noise from retail traders in short-dated options, as shown by Bryzgalova et al. (2020), in addition to other noise in short-term dated options because they are so close to expiry, we should see different results. Our hypothesis is that the sample containing more than 7 days to expiry will be more reliable (statistically significant), while the sample with less than 7 days to expiry will be less reliable (statistically insignificant).

Hypothesis 4:

The sample of the summed call, option, and share volume, is a good representation of the general market and can be used in regressions to predict future S&P 500 returns, which are proxied through the SPDR S&P 500 Trust ETF (SPY). We believe that the 15 stocks, which represent a significant part of the SPY assets, are a sufficiently good proxy for the index. These values can, as such, be used to form a regression to form broader market hypotheses.

5. Research Methodology

In this section, we discuss the research methodology as well as the regression models we use. We go over the type of data which we use and its characteristics, the various regressions that we run to test for predictability of call and put options volume on stock returns, as well as model assumptions and diagnostics checking to ensure that our models are robust. In the regression models we will present, the variables of interest are β_2 and β_3, which measure to what extent call and put option volumes have a predictive ability on next-day returns. The description and calculation of these variables will follow below, after we introduce the main regression model.

Panel Data

Our data follows a panel structure, with both time series spanning daily observations from 2021 to 2023, and cross-sectional dimensions including 15 different stocks. A key advantage of panel data is that we can examine a broader range of issues, and better handle the complexity and size of the data. In some cases, we also want to examine how variables and relationships change over time, and panel data is therefore a good fit in our case. Structuring the model in a correct way may also remove the impact of omitted variables in our regression results. The regression models we use are pooled regressions and fixed effects regressions, and the returns we use are both logarithmic returns based on the stock prices as well as excess adjusted returns.

Data Description and Calculation

We compute the returns from the stock prices in order to induce stationarity. Stationary can be defined as a series "with a constant mean, constant variance and constant autocovariances for each given lag" (Brooks, 2019). The stationarity assumption for data is crucial, as non-stationary series can provide a wide range of problems. Non-stationarity can affect the behaviour and properties of the parameters in our models, for example that shocks will stay persistent over time. We may also have the problem of spurious regressions, where two non-stationary variables trending over time that are regressed on each other could yield a high R^2. This indicates that one variable explains variations in the other variable well, even though the series may actually be completely unrelated to each other. Additionally, t-tests and F-tests will not follow their respective distribution, as the assumptions for those tests are not met with non-stationary data, and any inferences we draw may therefore be incorrect.

We face this issue in our data set with the prices of the individual stocks and the SPY, as well as the volume of both the shares and the options. To deal with these, we first construct log returns of the series, where the logarithmic return of asset i in period t can be defined as

$$\mathrm{r_{i,t}} = \ln\left(\frac{S_{i,t}}{S_{i,t-1}}\right)$$

, where $S_{i,t}$ is the price of security i at time t, and $S_{i,t-1}$ is the price of security i at time $t-1$.

Excess adjusted returns build upon the logarithmic returns and represent the abnormal return for each individual security. These returns capture the returns of security i in period t in excess of what the CAPM predicts, using a 6-month rolling beta as shown in Figure 4. The excess adjusted returns of asset i in period t is computed as

$$R_{i,t} = \left(r_{i,t} - r_{f,t}\right) - R_{M,t} \times \beta_{i,t}$$

, where $r_{i,t}$ is the log return of asset i in period t, $r_{f,t}$ is the risk-free rate in period t, $R_{M,t}$ is the market return in period t in excess of the risk-free rate and $\beta_{i,t}$ is the 6-month rolling beta for security i in period t.

Second, in order to normalize the stock and options volume, and induce stationarity in our option parameters, we create a ratio between the options volume and the share volume

$$\left(\frac{C}{S}\right)_t = \frac{\sum_{i=1}^{15} CV\$_{i,t}}{\sum_{i=1}^{15} SV\$_{i,t}} \qquad \left(\frac{P}{S}\right)_t = \frac{\sum_{i=1}^{15} PV\$_{i,t}}{\sum_{i=1}^{15} SV\$_{i,t}}$$

, where $CV\$_{i,t}$ represents the call volume of security i at period t, $PV\$_{i,t}$ represents the put volume of security i at period t, and $SV\$_{i,t}$ represents the stock volume of security i at period t, in US dollars. $\left(\frac{C}{S}\right)_t$ and $\left(\frac{P}{S}\right)_t$ represents the call-to-stock-volume and put-to-stock-volume ratio respectively.

Pooled Regressions

If there is no heterogeneity in the data, we could use pooled regression methods, which are the first sets of regression we run. Through a pooled regression we estimate a single equation for the whole data set, and assume that the relationship over time, as well as cross-sectionally, is constant.

With these regressions we try to capture the predictive ability of call and put option volumes on future returns, aiming to test our first hypothesis. As earnings announcements affect a company's return on the first trading day after the announcement, we want to exclude and isolate this effect by including dummy variables into our regressions. We also believe that CPI release dates have a large impact on returns, and thus we also incorporate indicator variables for these into our regressions.

However, as we believe the relationships may differ over time, we incorporate time-varying effects by running several regressions; two regressions on the full sample data from 2021-2023, both with and without dummy variables, and then additional regressions which only uses data from the individual years 2021, 2022 and 2023, all of which includes dummy variables. This will enable us to test our second hypothesis.

Pooled Regression on Logarithmic Returns

$$r_{i,t} = \alpha_0 + \beta_1 r_{i,t-1} + \beta_2\ CallToStock + \beta_3 PutToStock + u_{i,t}$$

$$r_{i,t} = \alpha_0 + \beta_1 r_{i,t-1} + \beta_2\ CallToStock_{i,t-1} + \beta_3 PutToStock_{i,t-1} + \beta_4 D1_{i,t} + \beta_5 D2_t + u_{i,t}$$

, where $r_{i,t-1}$ is the log return of asset i in period $t-1$. $CallToStock$ is the ratio of call volume to stock volume, and $PutToStock$ is the ratio of put volume to stock volume, both expressed in percentages. A value of 1 indicates that the call-or-put volume relative to the stock volume is 1%. $D1_{i,t}$ is a dummy variable which takes the value 1 if stock i in period t has its first trading day after the earnings announcement in t, and takes on the value of 0 otherwise. $D2_t$ is a dummy which takes the value 1 if the CPI is released during period t, and takes on the value of 0 otherwise.

Pooled Regression on Excess Adjusted Returns

$$R_i = \alpha_0 + \beta_1 R_{i,t-1} + \beta_2\ CallToStock + \beta_3 PutToStock + u_{i,t}$$

$$R_{i,t} = \alpha_0 + \beta_1 R_{i,t-1} + \beta_2\ CallToStock_{i,t-1} + \beta_3 PutToStock_{i,t-1} + \beta_4 D1_{i,t} + \beta_5 D2_t + u_{i,t}$$

, where $R_{i,t-1}$ is the excess adjusted returns for security i in period $t-1$.

Fixed Effects Regressions

We have also performed fixed effects regressions, as we believe the assumptions of constant cross-sectional relationships may not hold, with heterogeneity being present in the data. While we could have included a time-varying intercept λ_t, which would take different values if the observation was in 2021, 2022 or 2023, we have decided to proceed with the method discussed earlier where we run separate regressions on the individual years.

Fixed Effects Regression on Logarithmic Returns

$$r_{i,t} = \alpha + \beta_1 r_{i,t-1} + \beta_2\ CallToStock_{i,t-1} + \beta_3 PutToStock_{i,t-1} + \mu_i + v_{i,t}$$

$$r_{i,t} = \alpha + \beta_1 r_{i,t-1} + \beta_2\ CallToStock_{i,t-1} + \beta_3 PutToStock_{i,t-1} + \beta_4 D1_{i,t} + \beta_5 D2_t + \mu_i + v_{i,t}$$

, where μ_i controls for security-specific effects from security i, one for each of the individual shares (14 total, where one of the intercepts are dropped to avoid the dummy variable trap).

Fixed effects Regression on Excess Adjusted Returns

$$R_{i,t} = \alpha + \beta_1 R_{i,t-1} + \beta_2\, CallToStock_{i,t-1} + \beta_3 PutToStock_{i,t-1} + \mu_i + v_{i,t}$$

$$R_{i,t} = \alpha + \beta_1 R_{i,t-1} + \beta_2\, CallToStock_{i,t-1} + \beta_3 PutToStock_{i,t-1} + \beta_4 D1_{i,t} + \beta_5 D2_t + \mu_i + v_{i,t}$$

Regressions on Different Option Maturities

Additionally, to test our third hypothesis, we run separate regressions on different option maturities. The reason for this is that previous findings show that retail traders to be more present in the short-dated options market (Poser, 2023b) and that this market could therefore be less informational than the long-dated options market. In order to do this, we create two separate data sets: one containing options with 7 days or less to expiry only and the other containing options 7 days or more to expiry. These regressions are run through both pooled and fixed effects regression frameworks, but only on logarithmic returns. The reason for which we do not perform these regressions on excess adjusted returns will be discussed later. These regressions are run on the full sample from 2021-2023, as well as for the individual years 2021, 2022 & 2023.

Regressions on the Market - S&P 500

The final regression tests for the broad market as proxied by the SPY. The total options volume is calculated by summing all 15 individual stocks' call and option volumes. The call-to-stock and put-to-stock volume ratio is calculated by dividing these numbers on the summed value of all stocks' share volumes. These ratios are used in a regression involving the returns of SPY. These regressions are displayed as follows:

$$r_{m,t} = \alpha_0 + \beta_1 r_{m,t-1} + \beta_2\, CallToStock + \beta_3 PutToStock + u_t$$

$$r_{m,t} = \alpha_0 + \beta_1 r_{m,t-1} + \beta_2\, CallToStock_{i,t-1} + \beta_3 PutToStock_{i,t-1} + \beta_4 D2_t + u_t$$

, where $r_{m,t}$ and $r_{m,t-1}$ are the logarithmic returns of the SPY in period t, and period $t-1$ respectively.

The excess adjusted returns of the SPY would be zero, as the market return can't differ from itself. Additionally, this regression does not contain cross-sectional data. Given this, regressions are only run on logarithmic returns, and through a standard multiple linear regression framework. These regressions are run on the full sample from 2021-2023, as well as for the individual years 2021, 2022 & 2023.

Model Assumptions and Diagnostic Tests

Heteroscedasticity in errors is a common problem faced when working with panel data and means that the variance of the errors is not constant, but depends on the parameters, the squared parameters, and cross-products of these. For a simple model with three betas, the White's test for heteroscedasticity is:

$$u_t^2 = \alpha_1 + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \alpha_4 x_{2t}^2 + \alpha_5 x_{2t}^2 + \alpha_6 x_{2t} x_{3t} + v_t$$

$$H_0: \alpha_2 = 0 \text{ and } \alpha_3 = 0 \text{ and } \alpha_4 = 0 \text{ and } \alpha_5 = 0 \text{ and } \alpha_6 = 0$$

$$H_1: \alpha_2 \neq 0 \text{ or } \alpha_3 \neq 0 \text{ or } \alpha_4 \neq 0 \text{ or } \alpha_5 \neq 0 \text{ or } \alpha_6 \neq 0$$

Heteroscedasticity in regression models will lead to incorrect estimates of the standard errors of the parameter estimates, which may lead to wrong inferences being made, as the standard errors are presented to be lower than what they actually are. In the regressions we run, we find evidence of heteroscedasticity.

Our model builds upon autocorrelation, as our regressions are run on lagged values of the dependent variable itself ($y_{i,t}$), as well as all the independent variables (call ratios, put ratios, earnings announcement & CPI). Autocorrelation states that there are patterns in the estimated residuals from the regression. The Breusch Godfrey Test (Breusch, 1978; Godfrey; 1978), for autocorrelation in the r^{th} residuals, is:

$$\hat{u}_t = \gamma_1 + \gamma_2 x_{2,t} + \gamma_3 x_{3,t} + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \cdots + \rho \hat{u}_{t-r} + v_t$$

$$H_0: \rho_1 = 0 \text{ and } \rho_2 = 0 \text{ and } \ldots \text{ and } \rho_r = 0$$

$$H_1: \rho_1 \neq 0 \text{ or } \rho_2 \neq 0 \text{ or } \ldots \text{ or } \rho_r \neq 0$$

The data and the regressions we work with show evidence of autocorrelation, indicated through the Breusch Godfrey Test. For example, the options-to-volume ratio today is correlated with the options-to-volume ratio yesterday, and so forth, as illustrated by Figure 5 in the appendix. In the presence of autocorrelation, the standard errors of the parameter estimates could be wrong, which could lead to incorrect inferences. Additionally, the R^2 may be too high, although R^2 is not something which this paper studies and is therefore not seen as relevant.

To correct for both heteroscedasticity and autocorrelation, we employ Newey West Heteroscedasticity and Autocorrelation Consistent Standard Errors (as of here on NW). NW developed, in 1987, a variance-covariance estimator which is consistent in the presence of heteroscedasticity and autocorrelation (Brooks, 2019). Using this, we retrieve the correct standard errors, and are thus able to make the correct inferences.

Our data, including various factors such as share price returns and option-to-stock volume, is most likely non-normal in its distribution. There is a high chance that there is skewness and/or excess kurtosis, violating the fifth Classical Linear Regression Model (CLRM) assumption. However, given that OLS estimators are still unbiased, in addition to our large sample size, we believe this will not pose problems for our models, and the issue has therefore not been examined closer.

6. Data Collection and Analysis

Data Content

Our data collection includes collection on share prices and share volumes for the 15 largest American stocks as of 1. December 2023, in the period between 1. January 2021 – 1. December 2023. The equivalent has been done for the S&P 500 index proxied through the SPY. We have also collected detailed options data on these 15 stocks. To control for additional variables, we have extracted the dates for which the individual stocks had earnings announcements, and the dates in which CPI numbers were released by the Bureau of Labor Statistics.
The 15 largest companies as of 1. December 2023 were Apple, Microsoft, Alphabet, Amazon, Nvidia, Meta, Berkshire Hathaway, Tesla, Eli Lilly, Visa, United Health Group, J.P. Morgan, Broadcom, Walmart and Exxon Mobil.
Daily share prices and volumes of these companies, adjusting for splits, has been

retrieved from Bloomberg. The main variables retrieved were; Open, Low, High, and Close, of the individual share prices, as well as the number of shares traded of that company during that day. For companies which have two share classes, such as Alphabet with both "GOOGL" and "GOOG" shares, corresponding to its class A and B shares respectively, the most liquid one has been chosen. For Alphabet specifically this would be their class B shares with the corresponding ticker "GOOG". As a proxy for the S&P 500 Index the SPDR S&P 500 ETF trust has been chosen. This ETF is the largest ETF in the U.S. market, and we believe that the tracking error should be minimal and not a source for errors. We retrieve this data from Bloomberg and believe it to be factual and correct.

Our option data comes from a provider called "Optionistics", which provides subscriptions to access detailed historical options data on U.S. stocks, as well as other exchange traded instruments. The data extracted from Optionistics involves detailed daily option chains, volumes, and prices, on the 15 stocks that we examine. The data is structured in a way where each individual options contract, differentiated on factors such as the type of option (call or put), strike price, the current date, and the expiration date, has its own row of data. The information retrieved from these individual options include, but are not limited to, last traded price, bid, ask, and the number of options traded on that individual contract. We have only extracted data on the options which Optionistics determines to be "Near-The-Money", but this band is still rather wide and can range more than 10% away from the current price of the underlying share. Nevertheless, the data as such is complicated and large, with many individual shares having observations in the hundreds of thousands for one year of data. Given this data limitation, as well as our keen interest in studying how options behaved during the pandemic, and the aftermath of it, the data is limited to 2 years and 11 months. The number of dates in which we have options data for all stocks is 727 days. There are some days in which we have returns on the underlying shares, but no options data, and these dates have been excluded in our code. Data on the S&P 500 Index, through a proxy such as the SPY, has not been performed mainly due to the extremely large data sets and computational requirements to work with these data sets.

The dates of earnings announcement are also from Bloomberg, and on an individual share basis. We have extracted 11-12 dates for each of the shares, on the day that they released their earnings, totaling 176 observations. The CPI data only concerns the date of the release, and not the underlying numbers themselves, and is from the Bureau of Labor Statistics website.

Data Processing

The data processing for the individual share prices, volumes, as well as for the SPY, is relatively straightforward. The data is easy to process, contains limited variables, and volumes and returns can be easily constructed. Given that we have no data on the volume traded on these shares in dollars, where applicable, our approximation for traded volume any given day is:

$$Volume_{Dollars} = \frac{Open + High + Low + Close}{4} \times Volume_{in\ shares}$$

The options data is more complex. The volume is only stated in number of options traded, and unlike the share data, we only have bid and ask prices at the end of each day, as well as the last-traded-price. As such, our approximation, in most cases, has been calculated as follows

$$Volume_{Dollars} = \frac{\left(\frac{Bid_{t-1} + Ask_{t-1}}{2} + \frac{Bid_t + Ask_t}{2}\right)}{2} \times Volume_{in\ options}$$

This provides us an approximation for the dollar volume of traded options based on the bid and ask price on that day, as well as for the previous day. The reasoning behind this is that option prices are volatile, especially when the contract is near-the-money and/or close to expiration, and we would therefore like to rely on more than just one observation or data point. Previous papers, Roll et al. (2012), have used similar approximations. However, for some days, the option prices move too much to use the lagged prices. If the price of the identical option contract yesterday was less than 0.5x, or more than 2.0x, of the price of the option at today's close, we have chosen to use only today's price to approximate the volume. Additionally, some options do not have data from yesterday. This can be due to several factors, such as the options contract only entering the market today, or that the option was too far away from the money yesterday, and as such not included in our data set, but that the underlying price moved closer to that strike price and was included in the data set the day after. In any such cases, only today's price has been used.

The earnings announcements data contains data on the dates of earnings announcements. For the companies where earnings are released before the market open, the date of the earnings announcements has been kept unchanged. For the companies which report after the market has closed, the earnings announcement date has been changed to the next following (working) day. This is because we do not care about the date of the announcement, but rather the effect that earnings have on share prices. If the company reports after the market close, the impact will not be seen until the day after. The same methodology has been followed for the CPI data, although this is identical for all stocks, and accounting for the fact that CPI data is always released pre-market. The dates in both data sets have been assigned dummy values of "1" or "0", where "1" indicates that there is an earnings announcement or CPI release that day.

Data Credibility and Accuracy

Data on share prices and volumes on individual shares, as well as for SPY, is from Bloomberg. Earnings announcements are also from Bloomberg, and we trust this data to be accurate and reliable. The CPI release dates are from the official Bureau of Labor Statistics website[3], and we also believe this data to be accurate and reliable.

Our options data is from Optionistics, whose market data is "provided by a third party and believed to be accurate and reliable". Optionistics states that the third party is a service called HistoricalOptionData.com[4], but they do not independently verify the data themselves. HistoricalOptionData's parent company is DeltaNeutral[5], which is a certified vendor for the Option Price Reporting Authority (OPRA)[6]. OPRA collects "last sale and quotation information from the national securities exchanges that have been approved by the Securities and Exchange Commission to provide markets for the listing and trading of exchange-traded securities options" [7]. As such, we have confidence in that the data is accurate and reliable.

3 https://www.bls.gov/cpi/
4 https://www.optionistics.com/
5 https://historicaloptiondata.com/
6 https://www.opraplan.com/
7 https://www.opraplan.com/

To ensure the accuracy of this data, we have compared two sets of options data from our Optionistics source with the equivalent Bloomberg data. We examined Apple option prices, volumes and implied volatility for options traded on 01.11.2023 and expiring on 17.11.2023. Additionally, we have compared MSFT option prices, volumes, and implied volatility for 01.11.2023, on options expiring on 03.11.2023. When comparing the two data sources, as seen in Table 1 and 2, both the price as well as the volume data was identical. Given the test discussed here, as well as the data vendors and parent companies discussed earlier, we believe that the options data we use is accurate and reliable.

Ethical and Legal

We do not believe our data sourcing and collection are at risk of any legal or ethical violations. The data collected has been paid for through a subscription, and we do not publish, transmit, reproduce, or distribute the underlying data. It is our opinion that we do not breach Optionistics Terms and Conditions. The data from Bloomberg is retrieved from a license and has not been distributed. The data from the Bureau of Labor Statistics is publicly available, and to the best of our knowledge, does not pose risks or challenges from a non-commercial view.

As we are collecting purely financial, non-personal, data, from a third-party vendor, we do not believe that there are ethical violations or considerations in the sourcing of the data.

7. Results and Analysis

In this section we will present and discuss our findings from running the regressions discussed in Chapter 5 "Research Methodology". We will also touch upon potential limitations in the regression and the data we have, as well as assumptions that we have taken.

Clarification

The call-to-stock volume and put-to-stock volume ratios range from 0 to 100, while the parameter estimates are given in percentages, and a parameter estimate of 0.1 translates to a 0.1% effect on returns. The way to interpret this is that an increase of 1 percentage point in the option-to-stock ratio, for example from 1% to 2%, leads to next-day returns being 1 times the parameter estimate higher or lower. If the parameter estimate is 0.1, an increase of 1 percentage point in the option-to-stock ratio leads to 0.1% higher next-day returns.

Pooled Regression

We begin by examining a pooled OLS regressions on logarithmic returns without controlling for time- and entity-fixed effects. We test our first hypothesis, that put volume is more statistically significant in determining next-day returns than call volume. The call-to-stock volume ratio exhibits a negative coefficient of -0.0457, statistically significant at the 5% level, as can be inferred by Table 3 in the first column. The put-to-stock volume ratio coefficient is positive at 0.0478, fairly similar to the call-to-stock volume ratio in absolute terms but is not statistically significant at the 5% level.

The first results indicate that heightened call-to-stock volume is a negative predictor of returns, while heightened put-to-stock volume is a positive predictor of returns, indicating that market participants are, on average, wrong about the direction they believe the market will take. From here on, the call-to-stock volume ratio and put-to-stock volume ratio will only be referred to as call ratio and put ratio respectively. These are the two parameters of interest, while all the other parameters are control variables.

As we believe earnings announcements of the individual stocks as well as the release dates for the CPI can have a significant impact on the returns, we include these control variables in our next regression. As can be inferred from Table 3, in the second column, there are minor changes in the parameter estimates of the call and put ratios from the introduction of these terms. The parameter estimates change from -0.0457 to -0.0477 for the call ratio, and from 0.0478 to 0.0475 for the put ratio, and the call ratio is still statistically significantly different from zero while the put ratio is not.

In the regression a one standard deviation shock to the call ratio and put ratio respectively, leads to -0.0537% and 0.0405% lower/higher returns, respectively, the next day. On an annualized basis, in arithmetic terms, this computes to -13.52% and 10.21% respectively. This is a clear indication that heightened call volume is bearish for next-day returns, and that heightened put volume is bullish for next-day returns, and that these parameter estimates are economically significant. A potential reason for this is that options traders are overly optimistic or pessimistic ahead of mean reversions in the market. It should be noted that this is only for next-day returns, and they might have different predictive implications looking several days or weeks out.

Pooled Regression – Year for Year

We also run regression on the individual years, as can be inferred by columns "2021", "2022 & "2023" in Table 3, and all of these include the control variables "earnings announcement" and "cpi". We perform these regressions to examine our second hypothesis that the coefficients for call- and put-ratios differs from each other for the three years due to different market structures. In all three years, the coefficient for the call ratio is negative, and statistically significantly different from zero in 2022 and 2023 at the 1% and 5% level respectively. The put ratio is positive for all years, but only statistically significantly different from zero in 2023, with a significance level of 1%.

When it comes to the values of the parameter estimates, these also change from the full sample regression and across the individual years. We can see that in 2021 the call ratio coefficient was -0.0233, approximately half of the parameter estimate for the full sample, while 2022 and 2023 had parameter estimates of -0.1769 and -0.1181. This suggests that in 2021, the effect of call volume was less negatively significant on next-day returns, which may be explained by the fact that it was a strong bull market. However, 2023 was a strong bull market as well, but here we notice a large negative coefficient, which suggests that this might not be the reason. 2022, which was a year with large negative moves, has the largest coefficient in absolute value, and may be explained by the equities market trending downwards that year. While we can see that the effect does differ across the years, confirming our second hypothesis, it is difficult to see a very clear pattern in the driving factors for these differences. We should also note that the values for the call ratios differed in 2021, 2022 and 2023, with mean values of 1.4757, 0.9939 and 0.9629 respectively. Still, the difference in the parameter estimates differ enough to weigh up for these differences.

Similarly to the call ratios, the put ratios also vary across the years. The parameter estimates change from 0.0475 in the full sample, to 0.0724, 0.0556 and 0.3349 in 2021, 2022 and 2023 respectively. The strong outlier here is the 2023 coefficient, which also happens to be the only coefficient that is statistically significantly different from zero, even at the 1% level. It is difficult to interpret what the driver behind such a large impact would be. This impact could be driven by individual stock effects or by the high market volatility experienced in 2023. In times of large market downturns, options traders could have been too pessimistic right before direction changes in the market. However, 2022 was also a very volatile year, so it is uncertain whether this could be the reason.

Fixed Effects Regression – Control for Security Specific Effects

Our data consists of 15 different stocks which means we are working with firms which have unique attributes in terms of sectors, business models, products and cyclicality. They are also likely to exhibit different patterns in the returns of the underlying shares, the volume of those shares, and the size and characteristics of the traded options. We therefore introduce entity fixed effects regressions on logarithmic returns where the additional intercept μ_i represents the security-specific variation in logarithmic returns for security i. The security-specific intercepts can be seen in Figure 6. It should be noted that AMZN is not in this figure, as AMZN is the intercept, in order to avoid the dummy variable trap.

The results from the full sample regression are largely similar with our first set of pooled regression on logarithmic returns. Call volume has a negative effect on next-day returns, and put volume has a positive effect on next-day returns, as seen in Table 4, on all samples and time frames. We can see that the call ratio coefficient, on the full sample regression including the two dummy variables "earnings announcement" and "cpi", is slightly higher than before. It increases from -0.0477 in the pooled regression to -0.0526 in the fixed effects regression, but is no longer statistically significantly different from zero. The loss of statistical significance, with minor changes in the coefficient value, may be indicative of reduced statistical power rather than an omitted variable. The put ratio is practically unchanged, from 0.0475 in the pooled regression to 0.0473 in the fixed effects regression. A one standard deviation shock to the call ratio and put ratio respectively, leads to -0.0618% and 0.0401% lower/higher returns, respectively, next period. On an annualized basis this equates to -15.56% and 10.10% respectively, similar to our findings from the pooled regression.

When looking at the individual years, there are larger changes in the size and statistical significance in the parameter estimates. In 2021, 2022 and 2023 the call ratio coefficients change from -0.0233, -0.1769 and -0.1181 in the pooled regression to -0.0186, -0.1232 and -0.2457 in the fixed effects regression. Only the last coefficient, -0.2457, is statistically significantly different from zero. While the introduction of entity specific intercepts alters the parameter estimates it is difficult to see an obvious pattern. We observe that while the values of some parameter estimates increase, others decrease. Similarly, some estimates experience improved statistical significance, while others become less statistically significant.

There is however an interesting finding when examining the put ratios. While the put ratio coefficients in the pooled regressions for the individual years were quite different from each other, the put ratio coefficients in the fixed effects regressions are quite similar. The values are 0.0978, 0.1083 and 0.1053. As discussed earlier in the pooled regression, the 2023 regression had an abnormally high put ratio coefficient of 0.3349, vs. 0.0724 and 0.0556 for 2021 and 2022 respectively. The relatively similar values in the fixed effects regression may suggest that the results gained from the pooled regression were due to one, or several, individual securities skewing the results. As can be seen in Figure 6, this might be due to single name stocks like NVDA, LLY, or XOM, which have large positive security specific intercepts. The effect from put volume on next-day returns may therefore be quite similar for all three years. Still, as we see large differences between the coefficients for the call ratio, we still conclude that the effect of options volume, and especially from call options, do differ across the individual years.

Another difference between the pooled regression and the fixed effects model is the change in the p-values of the models. For the pooled regression, the p-value was below 0.05 for the full sample regression including "earnings announcement" and "cpi" dummy variables, 2022, and 2023. For the fixed effects regression, only 2023 has a p-value below 0.05. This indicates that there might not be any statistically appealing reason to why we should use a fixed effects model, meaning that introducing security specific intercepts are of little use. We should also note that in the five regressions shown in Table 4 fixed effects regression where we include the two dummy variables "earnings announcement" and "cpi", none of the individual security specific intercepts are statistically significantly different from zero.

Pooled and Fixed Effects Regressions on Excess Returns – Robustness Check

These regressions are similar to the pooled regression run on the logarithmic returns, but instead run on excess adjusted returns as seen in Table 5. As previously discussed, the excess adjusted return is the return of the underlying shares in excess of what the CAPM would predict, on a daily basis, and using 6-month rolling betas. The motivation behind this is that we want to adjust the model for each share's different betas and volatility and is in some sense a robustness check. Previous papers mentioned in the literature review have used different methods (Johnson & So, 2012), but with similar motivations. On the other hand, there is noise in the estimation of betas used in the CAPM, which reduce the informativeness of excess adjusted returns, and logarithmic returns may therefore be better to work with.

Doing this, the parameter estimates change, and in some cases the sign changes from positive to negative, or inverse. For instance, this is shown in Table 5, where all regressions, except the one run on the 2023 data, have positive coefficients for the call ratio. This implies, contrary to our previous findings, that heightened call volume predicts positive excess returns. Put ratios still have a positive coefficient, except for in 2022, and as such mostly finds the same findings as previously.

However, we are cautious about taking these findings at face value. The statistical significance for the individual coefficients, as well as the p-values for the models, are not there anymore. None of the parameters which we aim to discuss, the call and put ratios, are statistically significantly different from zero. The p-values for the models themselves are also never statistically significant at the 5% level, implying that none of the pooled regressions involving excess returns are predictive enough in explaining excess returns to be significant.

For consistency, we have also performed fixed effects regressions on excess returns following the previous methodology, as can be seen in Table 6. This, in a sense, accounts for the security specific effects twice, once through the beta calculations in the excess returns and another time through entity specific intercepts. There are variations and inconsistencies in the parameters of interest here too, and the standard errors are very large. The statistical significance of the models themselves, interpreted through their respective p-values, is never different from zero at the 5% significance level.

As none of the individual coefficients of estimates, neither the models themselves, are statistically significant at the 5% level, we determine that excess returns are not something we will consider in our analysis. This may arrive from several reasons; the stock betas are difficult to calculate accurately and is prone to noise and short-term variations, as can be inferred from Figure 4. The noise in the estimation of the betas may cause noise in the excess returns, which could be what is causing the loss in statistical significance. Furthermore, the way we calculate excess returns are perhaps not the best way to do it to capture the effects we want to account for. It might be that the best way to account for these effects is through the fixed effects regression, although as previously discussed this is also not the most accurate model. Additionally, the results may tell us that there is no relation between the options volume and the broader market (S&P). In previous literature, authors have worked with both logarithmic returns (Roll et al., 2010) as well as some form of excess or abnormal returns (Johnson & So, 2012). Going forward, our regressions will only include logarithmic returns, as we see these being the preferred type of returns to work with.

Examining different options maturities – Pooled Regression

Additionally, we run regression on a sample containing volumes on options trading up to and including 7 days to expiry (7D), as well as volume on options with more than 7 days to expiry (7D+). The intuition behind this, as previously discussed, is the fact that retail traders are proportionally more present in short-dated options than what they are in long-dated options.

Starting with our pooled logarithmic regressions, we find that there are, in general, minimal changes to the parameter estimates. As can be seen in Table 8, the parameter estimate for the call ratio full sample regression, using 7D options data, is -0.0672, an increase from -0.0477 in Table 3 which contains options data with all maturities (hereby referred to as AllD). However, it does not differ substantially from the estimate we retrieve from the 7D+ data, with a parameter estimate of -0.0653, shown in Table 9. The equivalent parameter estimates for the put ratio exhibits larger differences, taking on values of 0.0527 and 0.0665 in the 7D and 7D+ sample respectively. Neither the put- or call-ratio coefficients in the 7D and 7D+ regressions are statistically significant, although the 7D+ regression has lower standard errors. The higher absolute value of the coefficients relative to the AllD regressions is because the AllD contains more observations and higher volumes. The mean values for both call and put ratios are therefore naturally higher, making the parameter estimates look comparatively smaller, as can be seen in Table 7 and in Tables 12 and 13.

Looking at how a one standard deviation shock to the call- and put-ratio affects next-day returns, we get relatively similar results for the 7D and 7D+ regression. A one standard deviation shock to the call-ratio results in -0.0431% and -0.0432% lower next-day returns for the 7D and 7D+ sample respectively, corresponding to annual values of -10.85% and -10.89% respectively. A similar shock in the put ratio leads to higher next-day returns of 0.0252% and 0.0312% respectively for 7D and 7D+, corresponding to 6.35% and 7.87% respectively. It is clear that traders in the short-dated market do not perform worse than traders in the longer dated market, and if anything may actually perform better.

The lower standard errors in the estimates for the 7D+ regression, compared to the 7D regression, might be because the informativeness of options with more than seven days to maturity is marginally better, and less clouded, but it could also be due to the larger sample size. The 7D options data contains 45 706 individual options, while the 7D+ options data contains 266 994 individual options, which can reduce the effect of large outliers and therefore lead to more accurate values. Additionally, by splitting the maturities and running separate regressions, we should be able to examine the findings that informed traders use options trading to take a view on future returns (Easley, O'hara and Srinivias, 1998). The 7D+ volume should have a higher proportion of professional and informed traders, but the signs are still negative for call options and positive for put options. This indicates that traders participating in the longer-dated options market are not more informed than those participating in the short-dated options market, and as such do not appear to have more valuable information which can be used to profitably trade markets. It may also be that the introduction and growth of shorter dated options, alongside increased retail participation, have changed the market dynamics in both the shorter and longer dated option markets.

Examining the individual years we find larger differences. The effect of the call-ratio on next-day returns in the 2021 7D regression almost disappears, with a very small negative coefficient of -0.0012, less than a tenth of the size of the 7D+ regression. The put ratio also differs a lot, with values of 0.0380 and 0.1735 for the 7D, 7D+ and AllD regressions respectively. We also see deviations in 2022 and 2023, but overall there is no clear difference in either direction.

In both the 7D and 7D+ data, call ratios have negative coefficients and put ratios have positive coefficients, equivalent results to what is seen from the options data with all of the maturities. This is to say that even when controlling for maturities, which we believe to better capture retail and professional trading, there is no change in the pattern seen previously. An increase in call options volume affects next-day's returns negatively, while an increase in the put options volume affects next-day's returns positively.

Examining different options maturities – Fixed Effects Regression

We now proceed with the same data as before, but now using a fixed effects regression as previously discussed. Entity fixed effects run on options with different maturities are generally consistent with the findings in the AllD regression, as call ratios have negative coefficients for both the 7D and 7D+ regressions. However, when examining the put ratio we find that in 2021, for options with up to and including 7 days to expiry, the put coefficient is negative at -0.0306, as can be seen in Table 10. This result is inconsistent with previous regressions on logarithmic returns, where we have had consistently positive put ratio coefficients. Additionally, this result is in the 7D sample, the one which contains proportionally more retail traders, and proportionally less informed traders. When looking at the 7D+ sample we find a large positive number of 0.2281, displayed in Table 11. However, given that the put coefficient of -0.306 is small and not statistically significantly different from zero, while the 0.2281 is strongly positive and statistically different from zero at the 5% level, we believe that we should be cautious in drawing any conclusions based on these findings.

Overall, we find that there is no significant difference between the two regressions. The reason for these inconclusive results might be because our 7D and 7D+ regressions are not a useful proxy for retail and professional traders respectively, and that we therefore are not able to test what we want to test. The other explanation might be that we have in fact used suitable proxies, but that retail trading, which is present in both short and longer dated options market, make the results noisy and hard to interpret. A third explanation is that professional traders are not better at predicting future returns and do not possess valuable private information. We believe that the most likely explanation is the latter, as we have shown that retail traders clearly are more skewed towards short-dated options than towards long-dated options (Poser, 2023b).

Testing our model on the market (S&P 500)

Our final set of regressions aims to test our model on the broader market, where the SPY ETF is proxied for the S&P500 Index. In this regression the call and put volumes are, respectively, summed together cross-sectionally, giving us a single observation for call and put volumes at each date. The same procedure is done on stock volume, while returns are retrieved from the SPY. These summed values are regressed on the logarithmic returns. We aim to know; are our 15 stocks a representative sample that can be used to predict S&P 500 returns? Our preferred method here is a standard OLS regression using logarithmic returns and the call and put ratios. There is not a cross sectional sample as we only work with the SPY data, and therefore there is no possible fixed effects regressions.

All call ratios have a negative sign, both for the full sample including and excluding the two dummy variables "earnings announcement" and "cpi", as well as for the individual years, as can be seen in Table 14 . The put ratios, however, have a positive sign for all five regressions, except for 2023 where there is a negative sign. Additionally, the p-values of the models themselves are not statistically significant, with the lowest p-value being only 0.2256. We find, in general, the same results which we found in the pooled regression, that the call ratios and put ratios have negative and positive coefficients respectively. However, the large standard errors, and thereby lack of statistical significance, leads us to be cautious in interpreting the results.

We believe that this lack of significance is due to a potential combination of the following reasons. The first reason can be that the 15 stocks are not a good enough proxy for the S&P, and that they therefore hold little explanatory power. Potential ways to solve this would be to use a broader sample of stocks, more closely resembling the S&P, using correct weighting corresponding to the S&P weighting, or just using options data on the SPY itself. The other reasons which could have caused our insignificant results, are that SPY logarithmic returns may be seen in the same light as the excess adjusted returns regression that have been run. These are far more statistically insignificant when compared to their respective logarithmic counterparts, and it may be this effect that we are seeing. Lastly, it is possible that the broader market behaves differently towards options volume and with different dynamics than what is seen when compared to individual stocks.

Limitations and assumptions

We believe it is important to highlight the limitations and assumptions which are inherent in the analysis. First of all, there are other variables which could explain the returns, other than those that we have included. Some examples of these are announcement dates on other macro numbers like CPI, such as the dates where the Fed Funds Rate is determined, monthly releases of unemployment numbers, GDP releases, and many other examples. It has been documented that option volume tends to heighten ahead of earnings announcements (Roll et al., 2010 & Truong, 2012). It is therefore not unreasonable that options trading would heighten ahead of other known releases of news and numbers, such as macro releases, announcements which market participants know will cause volatility in the markets.

NYSE statistics show that over 80% of trading in mega-cap stocks occurs on non-Alternative Trading System (ATS) venues, compared to 60% for other stocks (Poser, 2023a). According to NYSE, a large part of this non-ATS off-exchange activity is reported by the wholesalers, who handle the order flow from retail brokers. As our data sample consists of the 15 stocks with the largest market caps we include the stocks with the highest shares of retail traders. This might cause our data sample to be biased towards more speculation driven movements, thus making our results less consistent. Consequently, the predictability of options trading on the underlying stocks could be obscured by the retail-dominated trading in these stocks.

The entire analysis is conducted solely on in-sample data, which is not unusual considering that this thesis does not aim to create a profitable trading strategy or create a portfolio with a better risk-adjusted return than what the market can provide. Still, we exercise caution in extrapolating the findings to out-of-sample future data. When markets, and market dynamics, change, as we saw in 2021, relationships can turn, or otherwise increase or decrease. Using our estimated regressions on out-of-sample data could be helpful in ensuring the robustness of our results, but our goal is mainly to explain the dynamics of options volume on stock volume.

8. Conclusion

We find that, in almost all cases, next-day returns are more negative given a high degree of call options volume and are more positive given a high degree of put options volume. When regressing on logarithmic returns, excluding the summed regression, the call-ratios are in many cases statistically significant, and more so than the put ratios. However, the put ratios tend to have a bigger absolute value than the call-ratios, although only 1/10 are statistically significantly different from zero. This is in contrary to our belief that the increased presence of retail trading would make options markets, and especially call options, less informational and less significant. Fixed effects regressions generally do not change the findings, except for the significance level of some variables and models. This implies that the results are not driven by any individual stock contributes, but that the results are broad and driven by the entire sample.

Additionally, we show that although the parameter estimates and statistical significance levels vary year by year, they have the same sign (+ or -). This suggests that, even across different market structures, call volume predicts negative returns, and put volume predicts positive returns. However, the explanatory power that they have may change.

When performing regressions on different maturities to closer examine retail options trading, we are not able to find significant differences in options volume where retail is proportionally more present, volume on options with 7 or less days to expiry, versus volume on options with more than 7 days to expiry. Therefore, we conclude that professional traders are not better at positioning themselves in the options market when compared to retail traders, when judging by next-day returns.

We also find that our sample of options on the 15 largest stocks in the US is not able to help in explaining future returns for the market. This is possibly due to a sample which is too small and not representative of the broader market.

In general, our results tend to go in the direction of mean reversion. This could be due to liquidity effects and option sellers. Option sellers may want to hedge their short call options by buying the underlying stock, which could cause the stock to increase. The next day however the stock may revert back, because the move was not fundamentally driven, and this effect would be stronger within low liquidity stocks. However, as can be seen in Figure 7, there is no clear relationship between the spread on traded shares and the coefficient values for call- and put-ratios seen in Figure 8 and 9. The parameter estimates for individual stocks can be, in absolute values, very strong, regardless of whether the spread is low or high. Additionally, as can be seen in Figure 10 and 11, the majority of the coefficient values are not statistically significantly different from zero at the 5% significance level.

We conclude with finding that market participants are not rational. They tend to buy more call options than usual ahead of next-day negative returns and similarly have heightened put options volume ahead of next-day positive returns. The market also seems to be dominated by traders which do not possess private information, as they are on average wrong about the direction.

We believe these conclusions are important in understanding how options markets behave in today's world. Compared to previous papers using options data from between the 1990's and early 2010's, we use newer data from the period 2021-2023. The options market faced significant structural changes in the previous 10 years, both in terms of its size, dynamics and market participants. Our findings and conclusions shed light on this. This thesis also differentiates itself from previous papers in that it examines the difference between shorter and longer dated options market, motivated by the increase in shorter dated options trading (see Figure 1).

9. Further Research

We believe that there are many areas for further research on this topic. Creating a portfolio trading strategy which trades on the information presented, would be an interesting case study in order to examine whether it is possible to trade based upon these relationships. This could also examine whether there are arbitrage opportunities in the market, or if trading costs outweigh the returns. A trading strategy could be based upon deciles of options volume, and creating portfolios that go short or long certain deciles. We believe that if this were to be done, daily rebalancing may be too impractical due to liquidity issues and trading fees, and expanding the horizon from next-day returns to next-week returns could be an alternative. Expanding the time horizon is also interesting, because while we see a certain degree of mean reversion, it is very short-term. Weekly data may lead to different results and conclusions.

We also believe further research into investigating if retail traders and professional traders have different abilities in predicting future returns is of great interest. We performed a regression analysis in this manner, using option maturities as an attempt to determine this, but we acknowledge that this method is far from perfect. If someone were able to attain more granular data and be better able to sort professional from non-professional trades from each other, they would be able to investigate this issue further. This, together with gathering a larger dataset that include less popular stocks among retail traders could add further depth to the findings presented in this paper. This is to a large extent why we find the options trading so interesting, as the large influx of non-professional traders have likely had an effect on the market.

10. Bibliography

Bergsma, K., Fodor, A., Singal, V., & Tayal, J. (2020). Option Trading After the Opening Bell and Intraday Stock Return Predictability. *Financial Managemen*t, *49*(3), 769-804. https://doi.org/10.1111/fima.12284

Breusch, T.S. (1978). Testing for Autocorrelation in Dynamic Linear Models. *Australian Economic Papers, 17*(31), 334-355. https://doi.org/10.1111/j.1467-8454.1978.tb00635.x

Brooks, C. (2019). *Introductory Econometrics for Finance (4th edition)*. Glasgow: Bell & Bain.

Bryzgalova, S., Pavlova, A., & Sikorskaya, T. (2023). Retail Trading in Options and the Rise of the Big Three Wholesalers. *Journal of Finance, 78*(6), 3465-3514. https://doi.org/10.1111/jofi.13285

Easley, D., O'hara, M. & Srinivas, P.S. (1998). Option Volume and Stock Prices: Evidence on Where Informed Traders Trade. *The Journal of Finance, 53(*2), 431-465. https://doi.org/10.1111/0022-1082.194060

Ge, L., Lin, T.C. & Pearson, N.D. (2016). Why Does the Option to Stock Volume Ratio Predict Stock Returns? *Journal of Financial Economics, 120*(3), 601-622. https://doi.org/10.1016/j.jfineco.2015.08.019

Godfrey, L.G. (1978). Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables. *Econometrica, 46*(6), 1293-1301. https://doi.org/10.2307/1913829

Johnson, T.L., & So, E.C. (2012). The Option to Stock Volume Ratio and Future Returns. *Journal of Financial Economics, 106*(2), 262-286. https://doi.org/10.1016/j.jfineco.2012.05.008

Macrotrends (n.d.). https://www.macrotrends.net/2526/sp-500-historical-annual-returns

Mackintosh, P. (2023, April 13). *Figure 1*, What's Driving the Growth in Options Trading. *Nasdaq*. https://www.nasdaq.com/articles/whats-driving-the-growth-in-options-trading

Mark, S. (2021, March 11). How Robinhood and Reddit Have Changed Options Trading -- and How You Can Profit. *The Street*. https://www.thestreet.com/investing/how-robinhood-and-reddit-have-changed-options-trading

Ni, S.X., & Pan, J. (2020). Trading Options and CDS on Stocks under the Short Sale Ban. https://doi.org/10.2139/ssrn.1572462

Ni, S.X., Pearson, N.D., Poteshman, A.M., & White, J. (2021). Does Option Trading Have a Pervasive Impact on Underlying Stock Prices? *Review of Financial Studies*, *34*(4), 1952-1986. https://doi.org/10.1093/rfs/hhaa082

Poser, S.W. (2023a, June 20). A breadth of Fresh Air. *The New York Stock Exchange*. https://www.nyse.com/data-insights/a-breadth-of-fresh-air

Poser, S.W. (2023b, December 4). *Figure 2*, *Figure 3*, Trends in Options Trading. *The New York Stock Exchange*. https://www.nyse.com/data-insights/trends-in-options-trading

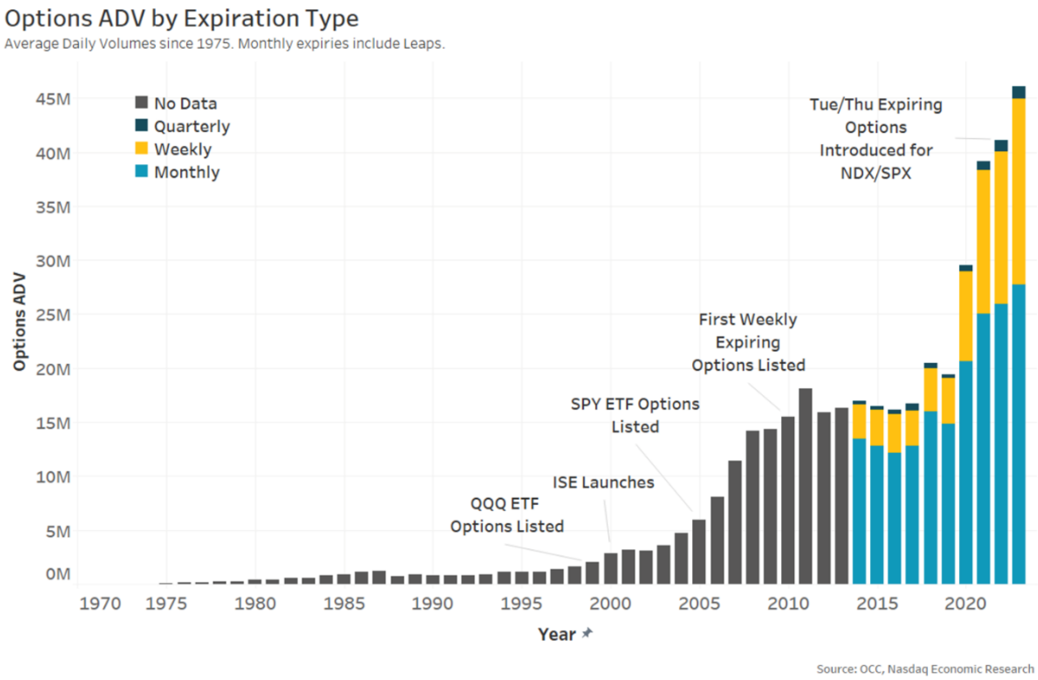
Roll, R., Schwartz, E., & Subrahmanyam, A. (2010). O/S: The Relative Trading Activity in Options and Stock. *Journal of Financial Economics*, *96*(1), 1-17. https://doi.org/10.1016/j.jfineco.2009.11.004

Truong, C. (2012). Options Trading and the Extent That Stock Prices Lead Future Earnings Information. *Journal of Business Finance & Accounting*, *39*(7-8), 960-996. https://doi.org/10.1111/j.1468-5957.2012.02293.x

11. Appendix

Figure 1: Average Daily Option Volumes since 1975, from Nasdaq[8].

This figure shows the daily options volume since 1975 and up to early 2023 and is retrieved from Nasdaq. The figure shows a clear increase in daily options volume from the 1990's and up to the early 2010's, before it temporarily plateaued. In 2020 options volume increased 50% year-on-year, and it continued to increase in the following years. The shorter-dated options market, here illustrated by weekly expirations, grew faster than the options market as a whole in 2020-2023.



[8] https://www.nasdaq.com/articles/whats-driving-the-growth-in-options-trading

Figure 2: Share of Total Options Volume Trading by Time to Expiration, adapted from NYSE.

This figure shows the share of options volume trading by time to expiration with data collected from Nasdaq. The figure the time period between November 2019 – September 2023, and shows an increase in the proportional volume of options trading with time to expiration of less than 30 days. In the same time, options with between 1-3 months' time to expiry have decreased proportional to overall options trading volume, and so has options with more than 3 months to expiry.

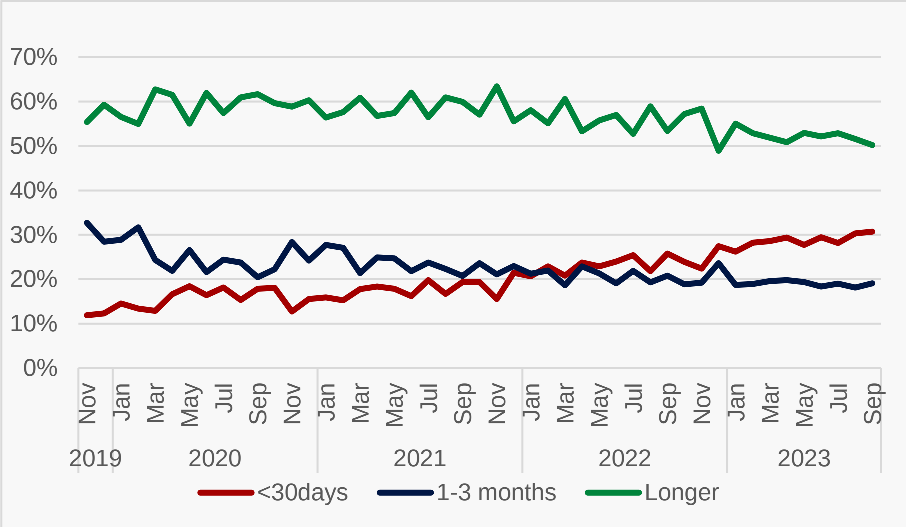


Figure 3: Short-Dated Volume Shares (Retail vs All Share of Trading Volume Short-dated, overrepresentation) adapted from NYSE.

This figure shows the share of options volume which is short-dated (less than 30 days to expiry) with data collected from Nasdaq. The figure shows the share of short-dated options trading for all volume ("All"), consisting of both retail as well as professionals and institutions traders, and for retail traders only ("Retail"). Short-dated options trading, here defined as less than 30 days to expiry, is clearly more popular amongst retail traders than for other traders.

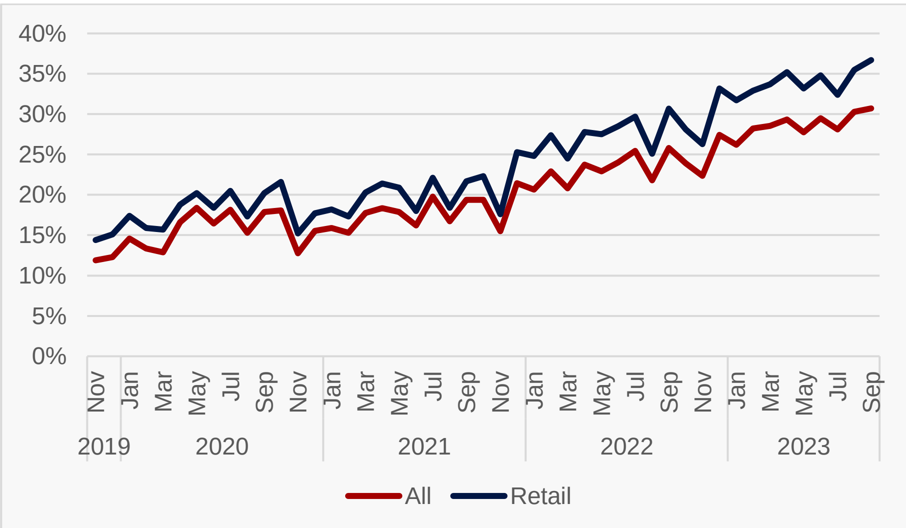


Figure 4: Overview of Rolling 6-month Beta for All 15 Stocks.

This figure illustrates the betas we have computed for each individual stock. The beta is calculated on a rolling 6-month basis, using the return of the individual security, the S&P 500 index, and the risk-free rate (US 10Y).

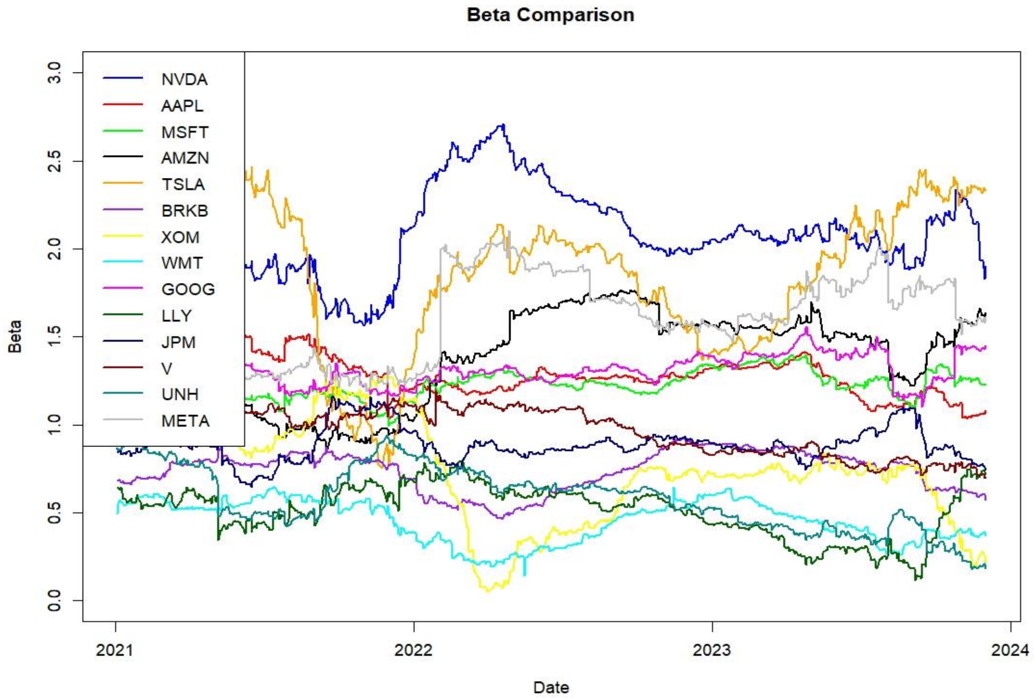


Figure 5: Options Volume Relative to Stock Volume for All Stocks

This figure shows the options volume relative to the stock volume, where the options and stock volume for all individual stocks, in US dollars, have been summed together for each period. The figure shows that options volume relative to stock volume peaked in late 2021, at a value of between 7-8% and has since declined towards 2-3%.

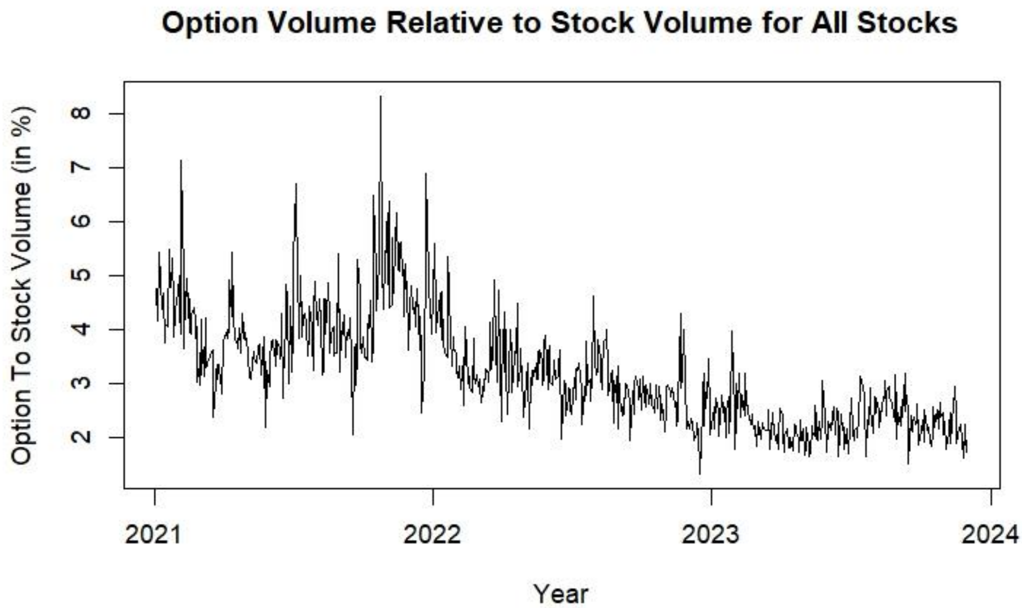


Figure 6: Security Specific Intercept for Logarithmic Returns

This bar figure shows the security specific intercept for each individual stock in the fixed-effects regression on logarithmic returns. Note that to exclude the dummy variable trap an entity-specific intercept must be dropped, which in this case has been AMZN. The figure illustrates that different securities have different intercepts, with the intercept representing logarithmic returns. A higher security specific intercept represents higher logarithmic returns, all else being equal.

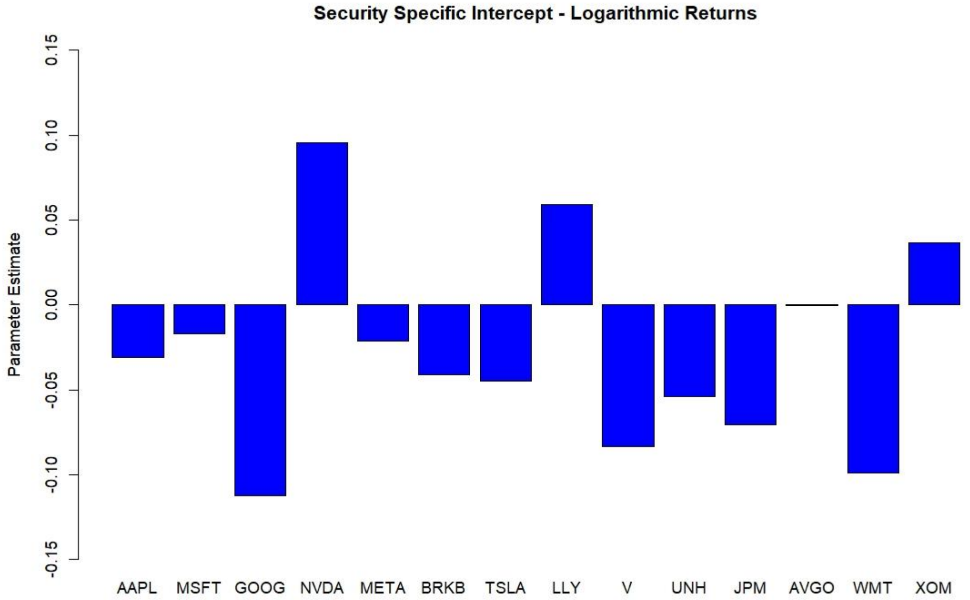


Figure 7: Average Spread on Shares Traded Between 2021 and 2023 (in basis points)

This figure shows the average spread, in basis points, on shares traded between 2021 and 2023, for each of the 15 stocks in our data. The figure shows that liquidity across individual stocks is not constant, and that there are large differences between shares such as AAPL and AMZN.

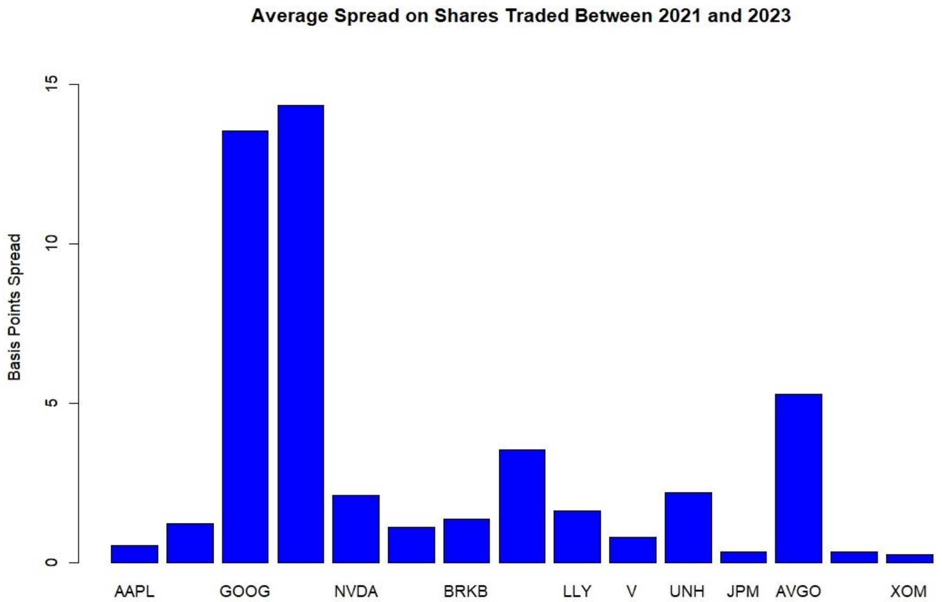


Figure 8: Parameter Estimate Call Options Logarithmic Returns – Full Sample

This figure shows the parameter estimates for the call ratio, defined as the call options volume relative to the underlying stock's volume, on next period logarithmic returns, when running individual regressions using data from each individual stock. For example, "aapl" shows the parameter estimate of the call ratio using only return data from Apple shares and options. The figure shows large differences, but in general the parameter estimates are negative, indicating that higher call ratios predict lower next period returns.

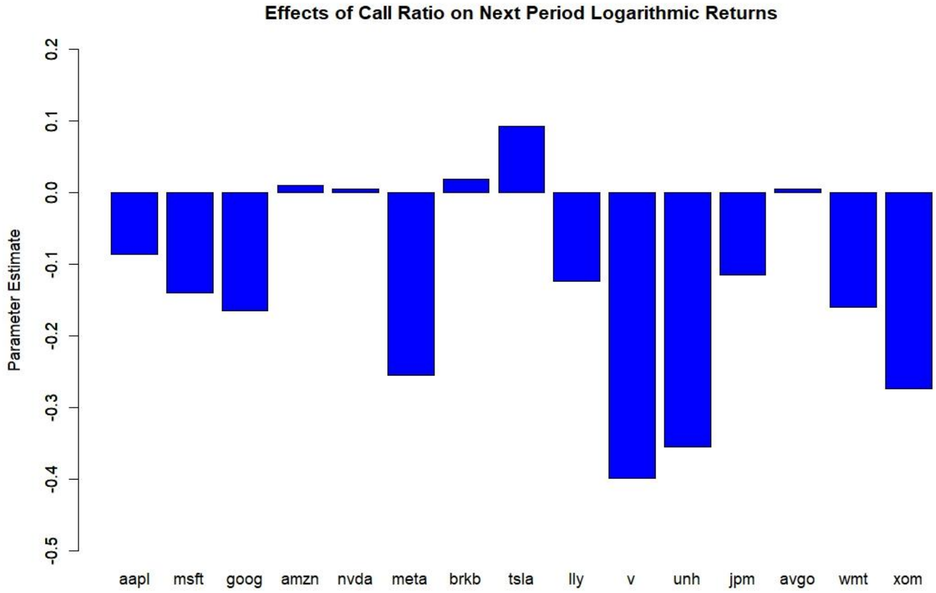


Figure 9: Parameter Estimate Put Options Logarithmic Returns – Full Sample

This figure shows the parameter estimates for the put ratio, defined as the put options volume relative to the underlying stock's volume, on next period logarithmic returns, when running individual regressions using data from each individual stock. The figure shows large differences, but in general the parameter estimates are positive, indicating that higher put ratios predict higher next period returns.

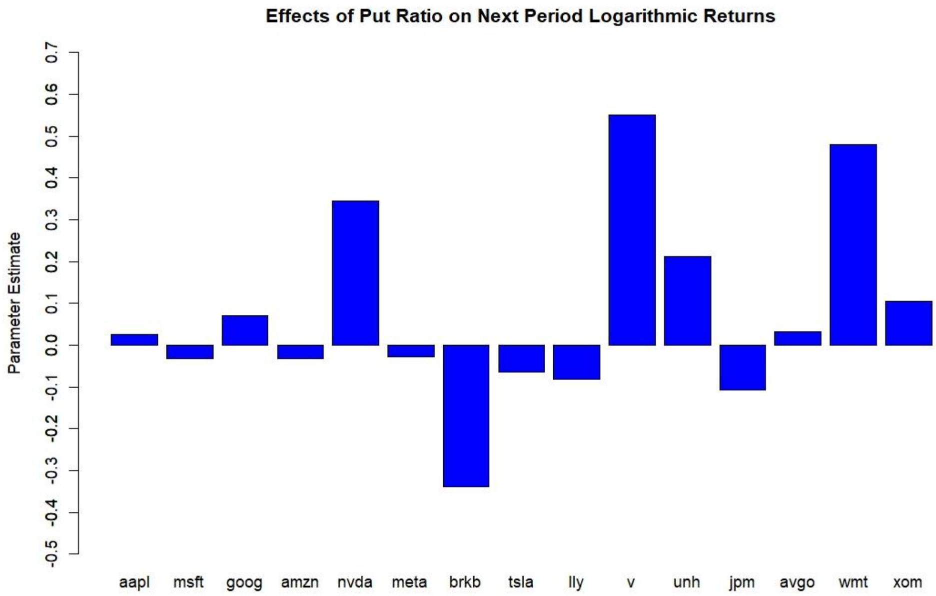


Figure 10: Confidence Interval Call Ratio Individual Stock Regressions on Logarithmic Returns

This figure shows the 95% confidence interval for the parameter estimates of the call ratios (see Figure 8). The interval shows the range of values which we can be 95% confident in contains the true value of the parameter. Most of the call ratios are not statistically significantly different from zero at the 5% significance level, except for Visa, United Health Group and Exxon Mobil.

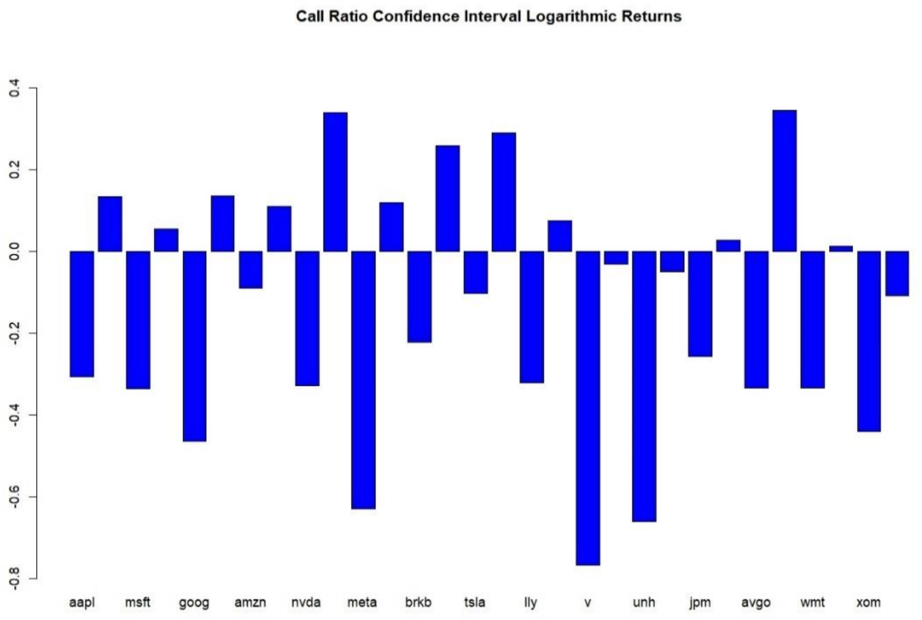


Figure 11: Confidence Interval Put Ratio Individual Stock Regressions on Logarithmic Returns

This figure shows the 95% confidence interval for the parameter estimates of the put ratios (see Figure 9). All parameter estimates, except for Walmart, are not statistically significantly different from zero at the 5% significance level.

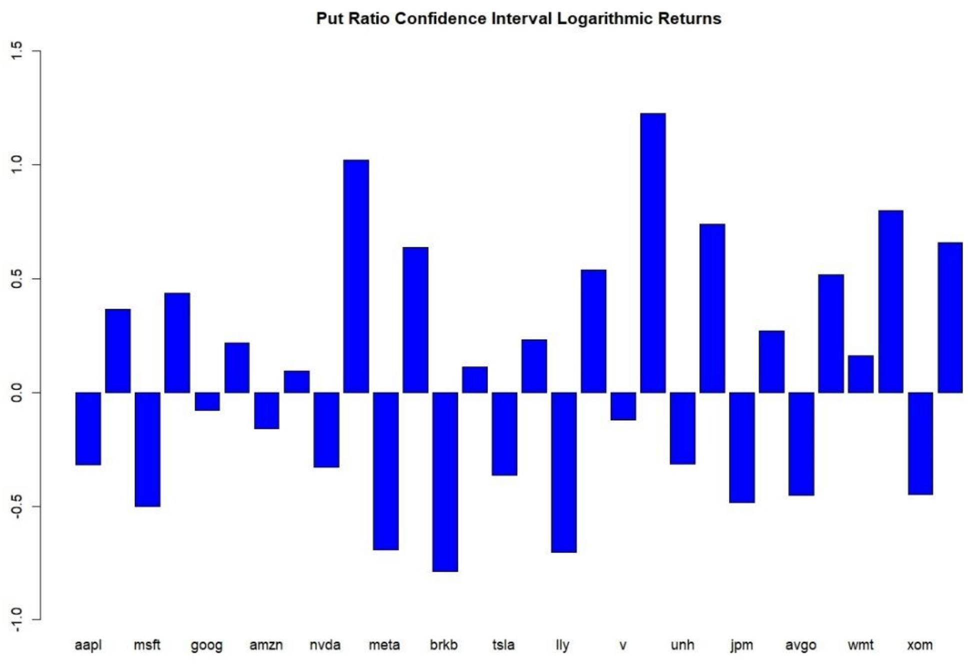


Table 1: Difference Between Bloomberg and Optionistics – Apple (AAPL)

This table shows the difference between the data collected from Bloomberg and Optionistics, when looking at "Last Price" and "Volume". This table examines options where Apple is the underlying security, for both call and put options. The data is from 01.11.2023, for options maturing 17.11.2023. This table was constructed to check if the Optionistics data was accurate when compared to the data from Bloomberg. There is no difference in Last Price and Volume for all observations, as they are identical.

Difference Bloomberg - Optionistics										
Underlying	Call/Put	Date	Maturity	Strike	Last Price Optionistics	Last Price Bloomberg	Last Price Difference	Volume Optionistics	Volume Bloomberg	Volume Difference
AAPL	C	01.11.2023	17.11.2023	155,0	18,40	18,40	0,00	1	1	0
AAPL	P	01.11.2023	17.11.2023	155,0	0,37	0,37	0,00	3 976	3 976	0
AAPL	C	01.11.2023	17.11.2023	157,5	16,80	16,80	0,00	44	44	0
AAPL	P	01.11.2023	17.11.2023	157,5	0,51	0,51	0,00	1 020	1 020	0
AAPL	C	01.11.2023	17.11.2023	160,0	14,75	14,75	0,00	146	146	0
AAPL	P	01.11.2023	17.11.2023	160,0	0,74	0,74	0,00	11 102	11 102	0
AAPL	C	01.11.2023	17.11.2023	162,5	12,75	12,75	0,00	119	119	0
AAPL	P	01.11.2023	17.11.2023	162,5	1,03	1,03	0,00	1 046	1 046	0
AAPL	C	01.11.2023	17.11.2023	165,0	10,65	10,65	0,00	387	387	0
AAPL	P	01.11.2023	17.11.2023	165,0	1,45	1,45	0,00	7 857	7 857	0
AAPL	C	01.11.2023	17.11.2023	167,5	8,77	8,77	0,00	483	483	0
AAPL	P	01.11.2023	17.11.2023	167,5	2,04	2,04	0,00	1 469	1 469	0
AAPL	C	01.11.2023	17.11.2023	170,0	6,94	6,94	0,00	2 794	2 794	0
AAPL	P	01.11.2023	17.11.2023	170,0	2,76	2,76	0,00	2 401	2 401	0
AAPL	C	01.11.2023	17.11.2023	172,5	5,31	5,31	0,00	3 511	3 511	0
AAPL	P	01.11.2023	17.11.2023	172,5	3,65	3,65	0,00	3 322	3 322	0
AAPL	C	01.11.2023	17.11.2023	175,0	3,87	3,87	0,00	4 227	4 227	0
AAPL	P	01.11.2023	17.11.2023	175,0	4,75	4,75	0,00	3 958	3 958	0
AAPL	C	01.11.2023	17.11.2023	177,5	2,70	2,70	0,00	1 575	1 575	0
AAPL	P	01.11.2023	17.11.2023	177,5	6,02	6,02	0,00	142	142	0
AAPL	C	01.11.2023	17.11.2023	180,0	1,75	1,75	0,00	4 095	4 095	0
AAPL	P	01.11.2023	17.11.2023	180,0	7,58	7,58	0,00	1 336	1 336	0
AAPL	C	01.11.2023	17.11.2023	182,5	1,05	1,05	0,00	2 466	2 466	0
AAPL	P	01.11.2023	17.11.2023	182,5	9,65	9,65	0,00	58	58	0
AAPL	C	01.11.2023	17.11.2023	185,0	0,60	0,60	0,00	3 124	3 124	0
AAPL	P	01.11.2023	17.11.2023	185,0	11,46	11,46	0,00	105	105	0

Table 2: Difference Between Bloomberg and Optionistics – Microsoft (MSFT)

This table shows the difference between the data collected from Bloomberg and Optionistics, when looking at "Last Price" and "Volume". This table examines options where Microsoft is the underlying security, for both call and put options. The data is from 01.11.2023, for options maturing 03.11.2023. This table was constructed to check if the Optionistics data was accurate when compared to the data from Bloomberg. There is no difference in Last Price and Volume for all observations, as they are identical.

Difference Bloomberg - Optionistics										
Underlying	Call/Put	Date	Maturity	Strike	Last Price Optionistics	Last Price Bloomberg	Last Price Difference	Volume Optionistics	Volume Bloomberg	Volume Difference
MSFT	C	01.11.2023	03.11.2023	330,0	16,60	16,60	0,00	686	686	0
MSFT	P	01.11.2023	03.11.2023	330,0	0,13	0,13	0,00	10 486	10 486	0
MSFT	C	01.11.2023	03.11.2023	332,5	14,15	14,15	0,00	199	199	0
MSFT	P	01.11.2023	03.11.2023	332,5	0,19	0,19	0,00	3 324	3 324	0
MSFT	C	01.11.2023	03.11.2023	335,0	11,68	11,68	0,00	3 863	3 863	0
MSFT	P	01.11.2023	03.11.2023	335,0	0,28	0,28	0,00	17 530	17 530	0
MSFT	C	01.11.2023	03.11.2023	337,5	9,45	9,45	0,00	1 332	1 332	0
MSFT	P	01.11.2023	03.11.2023	337,5	0,45	0,45	0,00	6 725	6 725	0
MSFT	C	01.11.2023	03.11.2023	340,0	7,20	7,20	0,00	6 805	6 805	0
MSFT	P	01.11.2023	03.11.2023	340,0	0,74	0,74	0,00	17 752	17 752	0
MSFT	C	01.11.2023	03.11.2023	342,5	5,07	5,07	0,00	7 293	7 293	0
MSFT	P	01.11.2023	03.11.2023	342,5	1,21	1,21	0,00	12 079	12 079	0
MSFT	C	01.11.2023	03.11.2023	345,0	3,33	3,33	0,00	25 319	25 319	0
MSFT	P	01.11.2023	03.11.2023	345,0	1,99	1,99	0,00	17 940	17 940	0
MSFT	C	01.11.2023	03.11.2023	347,5	2,02	2,02	0,00	12 135	12 135	0
MSFT	P	01.11.2023	03.11.2023	347,5	3,15	3,15	0,00	7 183	7 183	0
MSFT	C	01.11.2023	03.11.2023	350,0	1,10	1,10	0,00	31 496	31 496	0
MSFT	P	01.11.2023	03.11.2023	350,0	4,60	4,60	0,00	892	892	0
MSFT	C	01.11.2023	03.11.2023	352,5	0,53	0,53	0,00	6 367	6 367	0
MSFT	P	01.11.2023	03.11.2023	352,5	6,65	6,65	0,00	265	265	0
MSFT	C	01.11.2023	03.11.2023	355,0	0,25	0,25	0,00	7 027	7 027	0
MSFT	P	01.11.2023	03.11.2023	355,0	9,00	9,00	0,00	476	476	0
MSFT	C	01.11.2023	03.11.2023	357,5	0,14	0,14	0,00	4 548	4 548	0
MSFT	P	01.11.2023	03.11.2023	357,5	12,40	12,40	0,00	30	30	0
MSFT	C	01.11.2023	03.11.2023	360,0	0,06	0,06	0,00	3 041	3 041	0
MSFT	P	01.11.2023	03.11.2023	360,0	16,21	16,21	0,00	6	6	0

Table 3: Pooled Logarithmic Returns Regression Results

*This table shows the results from the pooled regression using logarithmic returns. The table contains information on individual parameter estimates, their standard errors in parentheses, as well as the p-value for the F-test where the null hypothesis is that all of the slope coefficients are jointly equal to zero. One star, *, indicates statistical significance at the 5% level, while two stars, **, indicates statistical significance at the 1% level. The intercept of 0.0845 indicates a logarithmic return of 0.0845%. The parameters of interest are the call and put ratios, as well as the p-value.*

As can be seen in the table, all parameter estimates for the call to stock volume are positive, while those for the put to stock volume are negative. Additionally, most of the call to stock volume parameter estimates are statistically significant at the 5% level, and far more so than the put to stock volume parameter estimates.

Variable	**Full Sample**	**Full Sample**	**2021**	**2022**	**2023**
Intercept	0.0845	0.0722	0.1226	0.0244	0.0335
	(0.0264)**	(0.0261)**	(0.0373)**	(0.0555)	(0.0551)
$logret_{t-1}$	0.0040	0.0042	-0.0238	0.0146	0.0285
	(0.0127)	(0.0127)	(0.0211)	(0.0200)	(0.0199)
call $ratio_{t-1}$	-0.0457	-0.0477	-0.0233	-0.1769	-0.1181
	(0.0232)*	(0.0233)*	(0.0309)	(0.0592)**	(0.0508)*
put $ratio_{t-1}$	0.0478	0.0475	0.0724	0.0556	0.3349
	(0.0316)	(0.0314)	(0.0439)	(0.0548)	(0.1001)**
earnings announcement	-	0.5521	-0.3789	0.7807	1.2600
	(-)	(0.4595)	(0.4665)	(0.9602)	(0.8589)
cpi	-	0.1259	-0.1576	0.0370	0.5887
	(-)	(0.1116)	(0.1332)	(0.2633)	(0.1342)**
p-value (F-test)	0.2236	0.0041**	0.0510	0.0084**	>0.0000**

Table 4: Entity Fixed Effects Logarithmic Returns Regression Results

This table shows the results from the entity fixed effects regression using logarithmic returns. The results do not largely differ from the results gathered from the pooled regression in terms of parameter estimates and signs, but there is loss of statistical significance both in the individual parameter estimates and for the F-test (indicated by the p-value).

Variable	**Full Sample**	**Full Sample**	**2021**	**2022**	**2023**
Intercept	0.1042	0.1035	-0.0046	-0.1343	0.3881
	(0.1158)	(0.1156)	(0.3161)	(0.2337)	(0.1741)*
$logret_{t-1}$	0.0040	0.0041	-0.0244	0.0115	0.0264
	(0.0126)	(0.0127)	(0.0231)	(0.0201)	(0.0201)
call $ratio_{t-1}$	-0.0493	-0.0526	-0.0186	-0.1232	-0.2457
	(0.0278)	(0.0280)	(0.0377)	(0.0689)	(0.0687)**
put $ratio_{t-1}$	0.0496	0.0473	0.0978	0.1083	0.1053
	(0.0359)	(0.0357)	(0.0676)	(0.0671)	(0.1242)
earnings announcement	-	0.5541	-0.3803	0.7502	1.3265
	(-)	(0.4600)	(0.4646)	(0.9608)	(0.8583)
cpi	-	0.1262	-0.1582	0.0343	0.5962
	(-)	(0.1117)	(0.1345)	(0.2632)	(0.1297)**
p-value (F-test)	0.8281	0.1793	0.4204	0.1749	>0.0000**

Table 5: Pooled Excess Adjusted Returns Regression Results

This table shows the results from the pooled regression using excess adjusted returns. Excess adjusted returns are the returns of a stock in excess of what the CAPM predicts. This regression is performed as a form of robustness check. There is a large loss of significance in the parameters of interest, and the results are mixed.

Variable	**Full Sample**	**Full Sample**	**2021**	**2022**	**2023**
Intercept	0.0227	0.0234	0.0082	0.0245	-0.0090
	(0.0216)	(0.215)	(0.0356)	(0.0352)	(0.0524)
exc_adj_{t-1}	-0.0079	-0.0082	-0.0489	-0.0038	0.0321
	(0.0128)	(0.0129)	(0.0240)*	(0.0182)	(0.0246)
call $ratio_{t-1}$	0.0009	0.0006	0.0029	0.0037	-0.0302
	(0.0229)	(0.0229)	(0.0351)	(0.0360)	(0.0552)
put $ratio_{t-1}$	0.0088	0.0087	0.0100	-0.0124	0.1551
	(0.0282)	(0.0282)	(0.0527)	(0.0381)	(0.0967)
earnings announcement	-	0.1302	0.1441	0.2825	-0.0939
	(-)	(0.1898)	(0.2645)	(0.2829)	(0.4163)
cpi	-	-0.0507	-0.0463	-0.0466	-0.0780
	(-)	(0.0748)	(0.1357)	(0.1180)	(0.1360)
p-value (F-test)	0.8227	0.8067	0.0845	0.8641	0.186

Table 6: Entity Fixed Effects Excess Adjusted Returns Regression Results

This table shows the results from the entity fixed effects regression using excess adjusted returns. This regression is performed as a form of robustness check. There is a large loss of significance in the parameters of interest, and the results are mixed.

Variable	**Full Sample**	**Full Sample**	**2021**	**2022**	**2023**
Intercept	-0.0538	-0.0512	0.0943	0.0351	-0.0894
	(0.1050)	(0.1050)	(0.2683)	(0.1635)	(0.1573)
exc_adj_{t-1}	-0.0086	-0.0089	-0.0497	-0.0071	0.0294
	(0.0127)	(0.0129)	(0.0241)*	(0.0176)	(0.0249)
call $ratio_{t-1}$	0.0080	0.0075	-0.0037	0.0007	-0.0264
	(0.0259)	(0.0258)	(0.0354)	(0.0529)	(0.0532)
put $ratio_{t-1}$	0.0225	0.0221	-0.0076	-0.0126	0.1566
	(0.0339)	(0.0339)	(0.0812)	(0.0381)	(0.1163)
earnings announcement	-	0.1271	0.1469	0.2835	-0.0910
	(-)	(0.1891)	(0.2628)	(0.2817)	(0.4190)
cpi	-	-0.0516	-0.0456	-0.0472	-0.0775
	(-)	(0.0748)	(0.1362)	(0.1178)	(0.1357)
p-value (F-test)	0.8694	0.8812	0.8393	0.7284	0.6402

Table 7: Average Values of Select Parameters

This table shows the average values of some select parameters, such as the logarithmic returns, the excess adjusted returns, the call ratio, and the put ratio. The table also shows the one standard deviation of the mean in parentheses. For the "Full Sample", the logarithmic return value of 0.0682 represents 0.0682%, the excess adjusted return value of 0.0310 represents 0.0310%. The call- and put-ratio values of 1.1495 and 0.7665 represents 1.1495% and 0.7665% respectively. This indicates that the average value of the call options volume relative to the underlying stock volume in this sample is 1.1495%, and equivalently for put options volume.

Variable	Full Sample	2021	2022	2023
$logret_{t-1}$	0.0682	0.1288	-0.0933	0.1796
	(2.1564)	(1.8227)	(2.6041)	(1.9278)
exc_adj_{t-1}	0.0310	0.0212	0.0172	0.0568
	(1.7528)	(1.7881)	(1.6903)	(1.7814)
call $ratio_{t-1}$	1.1495	1.4757	0.9939	0.9629
	(1.1740)	(1.5797)	(0.9510)	(0.7049)
put $ratio_{t-1}$	0.7665	0.8076	0.8653	0.6125
	(0.8470)	(0.9775)	(0.9730)	(0.4094)

Table 8: Pooled Logarithmic Returns Regressions Results – Options With 7 or Less Days to Maturity

This table shows the results from the pooled regression using logarithmic returns, but only includes options data where the time-to-expiry is 7 days or less. This is done to differentiate between retail and professional traders, as retail traders are proportionally more present in the shorter dated market than in the longer dated market. Compared to the regression with data on all options regardless of expiration (see Table 3), there are few major changes when looking at the "Full Sample" containing the entire period, but larger changes on the individual years.

Variable	Full Sample	Full Sample	2021	2022	2023
Intercept	0.0796	0.0665	0.1376	-0.0436	0.0744
	(0.0213)**	(0.0204)**	(0.0310)**	(0.0436)	(0.0394)
$logret_{t-1}$	0.0033	0.0033	-0.0304	0.0106	0.0285
	(0.0129)	(0.0129)	(0.0204)	(0.0207)	(0.0208)
call $ratio_{t-1}$	-0.0646	-0.0672	-0.0012	-0.2000	-0.1431
	(0.0458)	(0.0459)	(0.0534)	(0.1083)	(0.1265)
put $ratio_{t-1}$	0.0551	0.0527	0.0380	0.0343	0.5008
	(0.0570)	(0.0564)	(0.0718)	(0.1020)	(0.1735)**
earnings announcement	-	0.5508	-0.3876	0.7690	1.2665
	(-)	(0.4595)	(0.4660)	(0.9617)	(0.8615)
cpi	-	0.1238	-0.1608	0.0365	0.5840
	(-)	(0.1116)	(0.1325)	(0.2627)	(0.1326)**
p-value (F-test)	0.4179	0.0081**	0.1371	0.0685	>0.0000**

Table 9: Pooled Logarithmic Returns Regressions Results – Options With More Than 7 Days to Maturity

This table shows the results from the pooled regression using logarithmic returns, but only includes options data where the time-to-expiry is more than 7 days. Compared to the regression with data on all options regardless of expiration (see Table 3), there are few major changes when looking at the "Full Sample" containing the entire period, but larger changes on the individual years.

Variable	Full Sample	Full Sample	2021	2022	2023
Intercept	0.0824	0.0699	0.1033	0.0596	0.0486
	(0.0286)**	(0.0282)*	(0.0417)*	(0.0606)	(0.0614)
$logret_{t-1}$	-0.0005	-0.0004	-0.0276	0.0047	0.0151
	(0.0120)	(0.0121)	(0.0198)	(0.0185)	(0.0198)
call $ratio_{t-1}$	-0.0620	-0.0653	-0.0495	-0.3195	-0.1487
	(0.0334)	(0.0336)	(0.0464)	(0.0919)**	(0.0686)*
put $ratio_{t-1}$	0.0661	0.0665	0.1735	0.0727	0.4168
	(0.0456)	(0.0456)	(0.0735)*	(0.0661)	(0.1581)**
earnings announcement	-	0.5507	-0.3729	0.7794	1.2670
	(-)	(0.4600)	(0.4660)	(0.9600)	(0.8583)
cpi	-	0.1271	-0.1548	0.0424	0.5940
	(-)	(0.1116)	(0.1320)	(0.2630)	(0.1354)**
p-value (F-test)	0.3519	0.0066**	0.0162**	0.0063**	>0.0000**

Table 10: Fixed Effects Logarithmic Returns Regressions Results – Options With 7 or Less Days to Maturity

This table shows the results from the fixed effects regression using logarithmic returns, but only includes options data where the time-to-expiry is 7 days or less. Compared to the regression with data on all options regardless of expiration (see Table 4), there are few major changes when looking at the "Full Sample" containing the entire period, but larger changes on the individual years.

Variable	Full Sample	Full Sample	2021	2022	2023
Intercept	0.1048	0.1027	0.3775	-0.3053	0.2276
	(0.1014)	(0.1006)	(0.2676)	(0.2279)	(0.1384)
$logret_{t-1}$	0.0030	0.0031	-0.0310	0.0048	0.0254
	(0.0129)	(0.0129)	(0.0230)	(0.0212)	(0.0205)
call $ratio_{t-1}$	-0.0714	-0.0769	-0.0473	-0.0462	-0.3612
	(0.0513)	(0.0514)	(0.0654)	(0.1154)	(0.1402)*
put $ratio_{t-1}$	0.0499	0.0434	-0.0306	0.1684	0.1530
	(0.0685)	(0.0678)	(0.1158)	(0.1550)	(0.1954)
earnings announcement	-	0.5538	-0.3744	0.7187	1.3156
	(-)	(0.4601)	(0.4650)	(0.9602)	(0.8595)
cpi	-	0.1240	-0.1618	0.0361	0.6000
	(-)	(0.1116)	(0.1347)	(0.2629)	(0.1301)**
p-value (F-test)	0.9142	0.2556	0.5606	0.2991	>0.0000**

Table 11: Fixed Effects Logarithmic Returns Regressions Results – Options With More Than 7 Days to Maturity

This table shows the results from the pooled regression using logarithmic returns, but only includes options data where the time-to-expiry is more than 7 days. Compared to the regression with data on all options regardless of expiration (see Table 4), there are few major changes when looking at the "Full Sample" containing the entire time period, but larger changes on the individual years.

Variable	**Full Sample**	**Full Sample**	**2021**	**2022**	**2023**
Intercept	0.0859	0.0795	-0.0550	-0.0103	0.3686
	(0.0993)	(0.0990)	(0.2081)	(0.2067)	(0.1762)*
$logret_{t-1}$	-0.0007	-0.0006	-0.0290	0.0025	0.0134
	(0.0120)	(0.0121)	(0.0200)	(0.0187)	(0.0193)
call $ratio_{t-1}$	-0.0651	-0.0695	-0.0390	-0.2621	-0.3156
	(0.0403)	(0.0406)	(0.0546)	(0.1038)*	(0.0909)**
put $ratio_{t-1}$	0.0753	0.0742	0.2281	0.1432	0.1405
	(0.0470)	(0.0470)	(0.0942)*	(0.0692)*	(0.1929)
earnings announcement	-	0.5515	-0.3711	0.7657	1.3124
	(-)	(0.4600)	(0.4636)	(0.9614)	(0.8565)
cpi	-	0.1272	-0.1567	0.0394	0.5908
	(-)	(0.1118)	(0.1334)	(0.2633)	(0.1301)**
p-value (F-test)	0.8846	0.2247	0.2382	0.1376	>0.0000**

Table 12: Average Value of Select Parameters – Options With 7 or Less Days to Maturity

This table shows the average values of some select parameters, such as the logarithmic returns, the call ratio, and the put ratio, on the options data with 7 or less days to maturity. The table also shows the one standard deviation of the mean in parentheses.

Variable	**Full Sample**	**2021**	**2022**	**2023**
$logret_{t-1}$	0.0682	0.1288	-0.0933	0.1796
	(2.1564)	(1.8227)	(2.6041)	(1.9278)
call $ratio_{t-1}$	0.3861	0.5179	0.3504	0.2809
	(0.6407)	(0.8565)	(0.5654)	(0.3530)
put $ratio_{t-1}$	0.2547	0.5179	0.2933	0.1812
	(0.4782)	(0.5486)	(0.5545)	(0.2342)

Table 13: Average Value of Select Parameters – Options With More Than 7 Days to Maturity

This table shows the average values of some select parameters, such as the logarithmic returns, the call ratio, and the put ratio, on the options data with more than 7 days to maturity. The table also shows the one standard deviation of the mean in parentheses. Compared to Table 12, we can see that the values here are larger as the volume of options with more than 7 days to maturity is usually higher than the volume of options with 7 or less days to maturity.

Variable	**Full Sample**	**2021**	**2022**	**2023**
$logret_{t-1}$	0.0682	0.1288	-0.0933	0.1796
	(2.1564)	(1.8227)	(2.6041)	(1.9278)
call $ratio_{t-1}$	0.7633	0.9578	0.6435	0.6819
	(0.6616)	(0.8832)	(0.5031)	(0.4494)
put $ratio_{t-1}$	0.5117	0.5244	0.5720	0.4314
	(0.4699)	(0.5191)	(0.5564)	(0.2434)

Table 14: Regression of Summed Options Volume and Summed Stock Volume on SPX Returns

This table shows the results from the regression where all options volume is summed together and divided by the summed volume of stock volume and is then run on SPX returns. This regression is inconsistent in its results, and there is no statistical significance in any of the individual parameter estimates. Additionally, in all cases, we can't reject the null hypothesis that the value of all slope coefficients are jointly equal to zero at the 5% significance level.

Variable	**Full Sample**	**Full Sample**	**2021**	**2022**	**2023**
Intercept	-0.0248	-0.0272	-0.0096	-0.2167	0.5166
	(0.1400)	(0.1398)	(0.3115)	(0.4963)	(0.3238)
$logret_{t-1}$	0.0311	0.0308	-0.0418	0.0961	0.0615
	(0.0469)	(0.0472)	(0.0681)	(0.0723)	(0.0803)
call $ratio_{t-1}$	-0.0328	-0.0324	-0.0121	-0.1648	-0.2777
	(0.0434)	(0.0433)	(0.0582)	(0.1378)	(0.1644)
put $ratio_{t-1}$	0.0929	0.0914	0.1048	0.2833	-0.0788
	(0.1057)	(0.1053)	(0.1499)	(0.2552)	(0.2512)
cpi	-	0.0716	-0.1808	-0.0586	0.4724
	(-)	(0.2831)	(0.2320)	(0.7308)	(0.2518)
p-value (F-test)	0.7602	0.8613	0.7236	0.6139	0.2256