A Work Project presented as part of the requirements for the Award of a Master Degree in Finance
from the NOVA – School of Business and Economics

Measuring Dealer Inventories and Impact of Higher Order Greek Exposure: An Implied Volatility Modelling Approach

Sebastião Fernandes Vicente 44392

A Project carried out in the Master in Finance Program under the supervision of

Paulo M. M. Rodrigues

12.01.2022

Abstract

The purpose of this work is to theoretically model and empirically quantify the validity of four distinct trade classification methodologies for estimating market makers (MM) inventory and subsequent effect of higher order *greeks* exposure in explaining daily return variance. It begins by presenting a novel approach to trade classification based on modeling the implied volatility surface. This approach aims to be superior to current quote and open-interest based methods in that it does not require the use of quote data and by being a more dynamic way of classifying investor positioning.

To empirically test these models' accuracy, we first recreate and aggregate the MM greeks exposure across the proposed models and then regress the changes in inventory to changes in the underlying price. The empirical analysis performed over the period from 1^{st} January 2019 to 31^{st} December 2019 across five different stocks and one index reveal two important findings: (1) While the volatility model is preferable in terms of trade side classification, it falls short when used to estimate market maker inventory; and (2) The inclusion of *vega*, *vanna* and *volga* hedging effects turn out to be highly significant for explaining daily return variance. Quantitatively, accounting for these higher order greek effects results in a 66.04% increase in explanatory power when compared to merely accounting for gamma effects.

Keywords Volatility Surface Modeling; Market Maker Inventories; Options Greeks; Dynamic Hedging; Trade Classification;

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1 Introduction

In the colossal field of derivatives pricing, assumptions about the behavior of equities are made and underlying market dynamics are modeled in order to arrive at a price. While extensive research has been conducted, some critical assumptions that act as inputs to these models are still a stylized view of financial markets.

For this reason, a new field of research has emerged that concerns market micro-structure. This field focuses on the study of markets and how they operate, and is one of the fastest growing fields of financial research. More recently, due to the rising popularity of the options market, the analysis of flows and feedback effects regarding the hedging of these derivative instruments has caught the attention of the media, academics, and major market players.

These feedback effects have a deep impact on the underlying market parameters, such as spot prices and volatility. This is because the purpose of hedging is to replicate payoffs that are non-linear functions of the underlying price. To hedge, option holders have to buy / sell the underlying in accordance to their payoff structure in order to be neutral to any directional risk.

Since these effects caught the attention of academics, different bodies of research have emerged on the study of the impact of dynamic and the static hedging of derivatives. Some are focused on incorporating this effect into pricing models, while others are more concerned with describing the actual effect it has on the underlying market dynamics.

While important work has been done on the theoretical side, empirical work has been limited. This gap arises because the accurate study of the described effects requires the use of proprietary data (i.e data that allows for the identification of market maker inventory). To this day, to deal with the problem of lack of proprietary data, assumptions about options open interest are made. The main assumption is that all calls are sold (bought) by investors (market makers) and all puts are bought (sold) by investors (market makers).

While not ideal, and as curious as it may seem, the assumption established above has provided a useful framework for characterizing market maker exposure on the major stock indexes. So, what is the issue with using this assumption? The problem is twofold: (1) Index positioning changes with different market regimes; and (2) These assumptions of MM positioning may not fully apply to individual stocks.

As specified in (1), these assumptions are suspect at times because they fail to account for extended periods of out-the-money put premium harvesting by investors or extended periods of out-the-money call buying. The second issue (2) arises because individual stocks are subject to more variable patterns of trading than when contrasted with an index, where the effects of one single stock are dissolved. Board changes, product releases, acquisitions, and exposed fraud/negligence are all events intricate to that specific stock that can lead, at times, to a skew in investors' options positioning.

This work is of great importance because the biggest edges in trading are when players, that have a significant impact on the market due to their trading volume, are forced to act predictably. These flows that derive from the dynamic hedging activities conducted by market makers are a predictable part of the demand and supply that can be taken advantage of and is unlikely to disappear.

The objectives of this work are threefold: First, to synthesize the current literature on all the required concepts for comprehending the methodologies employed; second, to investigate how to classify trades by modeling the implied volatility surface and therefore constructing an inventory; and third, to use this inventory to quantify the higher greek exposure effects on the underlying market.

With this approach, it is intended to answer the two primary research questions of the paper:

1. Is volatility surface modeling a superior approach for estimating dealer inventory than current trade classification methods?
2. Besides *gamma*, what are the quantifiable effects of the other higher order *greeks* exposure?

2 Literature Review

This section provides an overview of the literature on three distinct topics that together make up the subject of the work carried out in this thesis.

2.1 Options hedging impact on the underlying market

Since options have been introduced, the question of how and to what extent the trading of these instruments affects the underlying equity dynamics has been a focal point of research among academics.

The current literature distinguishes between two channels (informational and non-informational) through which it is believed that options trading have an impact on the underlying. Notably, Hu (2014) presents strong evidence of an informational channel through which option trading impacts stock prices, by showing that market makers' hedging activities have strong predictive power of stock returns. Pan and Poteshman (2006) also provide evidence of a strong informational channel by taking advantage of put-call ratios. However, non-informational channels are also under current study. Avellaneda and Lipkin (2003), Ni et al. (2005) and Golez and Jackwerth (2012) document that the rebalancing and unwinding of market maker delta-hedges very close to the options expiration date impact the stock price by driving it towards the option strike.

More recently Ni et al. (2020) analyzed the effects of gamma imbalance on the returns and autocorrelation of returns. They find that there is a negative relation between stock volatility and gamma imbalance. While Ni et al. (2020) concentrate on daily data, Barbon and Buraschi (2020) resort to intraday data. Their research goes in line with the former, by finding that the gamma-imbalance is also negatively related to intraday volatility and autocorrelation of returns. However, they extend their research by documenting the impact of the same imbalance on the frequency and magnitude of flash crashes.

Finally, Baltussen et al. (2021) provide evidence that intraday momentum in futures contracts is strongly linked to the hedging demand coming from negative gamma exposure of market makers.

2.2 Trade side classification from intraday data

As intraday trade data becomes increasingly available from major exchanges, the improved ability to determine whether a trade was a buy or a sell order is of extreme importance. To solve this problem, researchers have been employing four competing algorithmic methods, such as the quote test, the tick rule (Finucane, 2000), the prevailing midpoint test - the LR algorithm (Lee and Ready, 1991) and the prevailing bid/ask test - the EMO algorithm (Michaely et al., 2000).

The quote rule classifies a transaction as a buy (sell) if the associated trade price is above (below) the midpoint of the bid and ask. The tick rule (Finucane, 2000), is based on price movements relative to previous trades. It compares if a trade price is above (below) to the previous price with upticks (downticks) being evidence of a buy (sell). The LR algorithm (Lee and Ready, 1991) is essentially a combination of the previous two rules: first, the trade is classified according to the quote rule, and then classifies the midpoint transaction with the tick rule. The EMO algorithm (Michaely et al., 2000), uses the quote rule to classify trades happening at the quoted bid/ask and the tick rule to classify all other trades.

Attempts at a modelling approach have been limited, with only Rosenthal (2012) providing significant improvements in this area. In his study, he incorporated all of the above tests into a linear modelling approach, accounting for autocorrelations and uncertainty about what quotes were at the time of the trade. This approach yielded the best results of all, with 1-2% more accuracy overall than currently used methods.

While these algorithms serve their main purpose on the cash market, our objective is to use them in the derivatives market, so we need to know what the performance is in that case. Using a proprietary dataset, Savickas and Wilson (2003) analyzed the performance rate of these classification methods for options trades and found that the correct classification rate for the quote rule is 83%, and for the LR, EMO and tick rules it is 80%, 77% and 59% respectively. It was also found that the major source of miss-classification came from outside-quote trades and reversed-quote trades (i.e., buying at the bid and selling at the ask).

2.3 Volatility surface modelling

Since the disclosure of the Black-Scholes formula, there has been extensive research conducted by both academics and practitioners on methods for modeling implied volatility and constructing volatility surfaces. Typical approaches used are based on (local) stochastic volatility models, Lévy processes, and models for the dynamics of implied volatility, based on parametric representations and non-parametric representations. A good overview of these approaches is given in Zhu (2014).

A parametric model comes with certain advantages. Observed implied volatilities can be inter- and extra-polated. Therefore, a parametric implied volatility model can be used to price new contracts for which there are no quotes on the market. The implied volatility in a parametric model is a function of strike and maturity with an explicit analytical expression.

The stochastic volatility inspired (SVI) model is part of the parametric representation family of models. It was developed at Merril Lynch in 1994 and was made public through a presentation by Gatheral in 2004 (Gatheral, 2004). It was chosen to be investigated further since it has the most potential outside of its already strong foundation and many advantages such as small computational time, relatively easy to calibrate to market data and a relatively good approximation of implied volatility for strikes deep in-the-money and out-of-the money for equities. Since its disclosure, a lot of investigation has been conducted regarding its peculiarities. One interesting advance was the Quasi-Explicit parameterization proposed by Marco and Martini (2009), which allowed for faster estimation of the parameters. However, the original SVI model, even though it gives great market fit, is missing two important features: it does not model the entire volatility surface, and there are no conditions imposed that grant the absence of arbitrage (Roper, 2010).

Ten years later, a new parameterization of the original SVI model was introduced for the entire volatility surface (Gatheral and Jacquier, 2014), called the surface SVI or SSVI which provided a good fit and tractable sufficient conditions that minimized the presence of arbitrage.

3 Methodology Implementation

3.1 Data & Pre-Processing

3.1.1 Data description

The dataset was constructed using four securities from distinct industries (Apple [Ticker: AAPL], Amazon [Ticker: AMZN], Boeing [Ticker: BA] and Exxon Mobil [Ticker: XOM]) and one index (S&P500 [Ticker: SPX]); see Table C.1 for the breakdown of the assets by industry and sector; see Table C.2 for the corresponding returns descriptive statistics. The equities used were selected based on their large daily option trading volume.

For this analysis, intraday option tick trade data from multiple exchanges (purchased from ivolatility.com) was used. The dataset comprises options transaction prices, quote and greeks data for the option and underlying securities across one unit of time in the whole sample period from January 2^{nd}, 2019 to December 31^{st}, 2019 (251 trading days).

3.1.2 Defining and removing inadequate data

Before creating data filters, it is first necessary to define what constitutes a “bad” input. The trade-off between data completeness and integrity must be balanced based on the sensitivity of the analysis to bad data. The emphasis of this study is on optimizing for integrity, which if not taken into consideration, would result in the most detrimental inaccuracy in the practical analysis.

The data is filtered by (i) removing all observations with non-positive prices/implied volatilities, (ii) removing all non-business days, (iii) removing all trades that are flagged as “late” or “canceled” by the exchange, (iv) removing duplicate trades. The cleaned dataset description with examples can be found in Table C.3.

3.2 Construction of the Implied Volatility Models (SVI&SSVI)

3.2.1 SVI Parameterization

The method that will be used for the construction of the implied volatility model is the stochastic volatility inspired parameterization or SVI parameterization. This method's popularity among practitioners is attributed to its two key properties (Gatheral and Jacquier, 2014):

(i) For a fixed time to maturity τ, the implied Black-Scholes variance $\sigma^2_{imp}(k,\tau)$ is linear in the moneyness k as $|k| \to \infty$;

(ii) It is relatively simple to fit to listed option prices.

The raw SVI parameterization of the total implied variance for a given parameter set $x = \{a, b, \rho, m, \sigma\}$ is given as:

$$w(k,x) = a + b\left\{\rho(k-m) + \sqrt{(k-m)^2 + \sigma^2}\right\} \tag{1}$$

where $a \in \mathbb{R}$, $b \geq 0$, $|\rho| < 1$, $m \in \mathbb{R}$, $\sigma > 0$ and $a + b\sigma\sqrt{1-\rho^2} \geq 0$ wich ensures that $w(k;x) \geq 0$ for all $k \in \mathbb{R}$.

3.2.2 SVI Calibration Process

The calibration of the SVI is based on using the raw parameterization in (1). The objective is to find the parameters that provide the best fit against the given market data. We can define the optimization problem as:

$$\min_{a,b,\sigma,\rho,m} \sum_{i=1}^{n} W_i (w_{raw}(k; a, b, \sigma, \rho, m) - \hat{w}_i)^2 \tag{2}$$

where w_{raw} is the raw parameterization depending on the parameter set (a, b, σ, ρ, m), $\hat{w}_i$ is the given market data defined in total implied variance and W_i are the weights for defining the goodness of fit of different data points.

One of the issues when solving this minimization problem is that it is non-linear in nature and generates many local minima points. The final solution is then, highly dependent on the starting parameters used as input to the optimization algorithm, and therefore the robustness of the fitted parameters is not guaranteed.

To solve this, the *Quasi-Explicit* parameterization proposed by Marco and Martini (2009) is applied so that it can be solved as a linear problem with the following variable change:

$$y(k) = \frac{k-m}{\sigma}.$$

Then, the *RAW* parameterization becomes:

$$\begin{aligned} w(k) &= a + b\sigma(\rho y(k) + \sqrt{y(k)^2+1}) \\ &= \hat{a} + dy(k) + cz(k) \end{aligned} \tag{3}$$

where $\hat{a} = a; \quad c = b\sigma; \quad d = \rho b\sigma; \quad z(k) = \sqrt{y(k)^2+1}.$

By picking an initial (σ, m) pair, the problem is transformed from a non-linear problem into:

$$P_{\sigma,m} := \min_{a,c,d \in D} \sum_{i=1}^{n} W_i(w(y_i) - \hat{w}_i)^2 \tag{4}$$

and the domain D of the parameter set *(a, c, d)* is given by

$$D = \begin{cases} 0 \leq c \leq 4\sigma \\ |d| \leq c \text{ and } |d| \leq 4\sigma - c \\ 0 \leq a \leq max(\hat{w}_i) \end{cases} \tag{5}$$

The problem in (4) is a typical least-squares problem with linear restrictions. This problem, due to its convexity, admits a single solution (if it exists), which will be the global minimum.

3.2.3 SSVI Parameterization

Gatheral and Jacquier (2014) introduced the Surface SVI (SSVI). This parameterization is a generalization of the original raw parametrization in (1) that takes the corresponding smiles into account in the estimation process. It aims to solve the problem of the restriction of the parameters to avoid the presence of butterfly and calendar arbitrage in a given smile and gives a single equation that fits the implied surface as a whole.

Specifically, let φ be a smooth function in $\mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{\tau \to 0} \theta_\tau \varphi(\theta_\tau)$ exists in $\mathbb{R}$. A SSVI surface is defined by:

$$w(k, \theta_\tau) = \frac{\theta_\tau}{2} \left\{ 1 + \rho \varphi(\theta_\tau) k + \sqrt{(\varphi(\theta_\tau) k + \rho)^2 + (1 - \rho^2)} \right\} \tag{6}$$

where $\theta_\tau := \sigma_{BS}^2(0, \tau)\tau$ and $\lim_{\tau \to 0} \theta_\tau = 0$.

It is now necessary to define the function φ, which is then going to be replaced in the SSVI's equation with its own parameters and we then have a function $w(k, \theta_\tau; \chi_{ssvi})$ that can be calibrated for a set of SSVI parameters χ_{ssvi} against the observed market values.

The function φ in (6) is parameterized as a power law type function, i.e.,

$$\varphi(\theta) = \frac{\eta}{\theta^\gamma (1 + \theta)^{1-\gamma}}. \tag{7}$$

3.2.4 No-Arbitrage Conditions

When we are presented with a situation in which we have a zero percent chance of losing and a non-negative chance of winning, we have an arbitrage opportunity. Market makers do not want this to happen since arbitrageurs could take advantage of this and cut into their profit margins, so their models often incorporate a set of rules for it not to happen when providing quotes. Carr and Madan (2005) state the following proposition and lemma concerning arbitrage on volatility surfaces:

Proposition 1. A volatility surface is free of static arbitrage if an only if the following conditions are satisfied:

(i) surface is free of calendar spread arbitrage;

(ii) each time slice is free of butterfly arbitrage.

Calendar spread arbitrage concerns arbitrage possibilities in the time to maturity direction of the surface whilst butterfly arbitrage concerns arbitrage possibilities in the moneyness direction.

Lemma 3.2.1. For a function φ of power law def.(7), the conditions on its parameters are

(i) $0 < \gamma < 1$ that optimize the conditions for no calendar arbitrage

(ii) $\eta(1+|\rho|) \leq 2$ that optimizes for the absence of butterfly arbitrage

3.2.5 SSVI Calibration Process

For the calibration of the SSVI parameters with φ as a type of power law function[1], we need the moneyness (k), time-to-maturity (τ) and implied volatility (σ) that will then be converted to total implicit variance (w).

The optimal choice of parameters can be found using least squares with an objective function set to the differences between the total implied variances given from the new parameterization and the original plus a large penalty P for static arbitrage. The general expression to minimize is then

$$\min_{\rho(\theta_\tau),\varphi(\theta_\tau)} \sum_{s=1}^{S} \sum_{i=1}^{n} W_{s,i}(\omega_{s,i}(k,\theta_\tau;\rho(\theta_\tau),\varphi(\theta_\tau)) - \hat{w}_{s,i})^2 + P \tag{8}$$

where S is the number of maturities in the data, n the number of data points per maturity, $\omega_{s,i}(k,\theta_\tau;\rho(\theta_\tau),\varphi(\theta_\tau))$ is the total implied variance given from the original parameterization and $\hat{\omega}_{s,i}$ is the total implied variance observed.

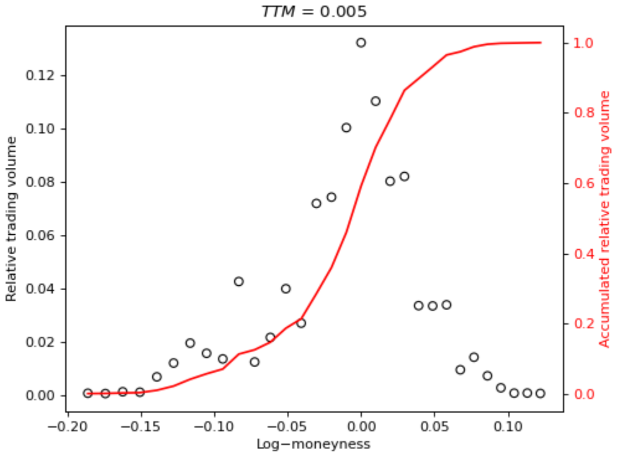
---

[1] The function defined for φ can be changed to another arbitrary function. The one used here (power-law) is one of the mostly used in the literature.

3.2.6 Weights vector

When calibrating a volatility model, the fits and subsequent errors are determined by the moneyness location of the greatest traded volume. Introducing weights in the objective function reflecting the significance of each moneyness value comes to mind as a natural solution to this problem. In this study's specific case, the most appropriate method is to give weights that gradually decrease from the at-the-money (ATM) point. This is because at any given time extant greeks exposure is going to be described by outstanding volume carrying large greek values - this happens at the ATM zone (See Figure 1).

Figure 1: **Concentration of trading volume**



Note: This Figure shows the concentration of trading volume around the at-the-money point of the nearest time-to-maturity. Approximately 80% of trade volume occurs within the *moneyness* interval [-0.05; 0.05].

For dealing with the concentration of trading volume, two function types come to mind - the absolute value (9) and exponential (10), i.e.,

$$W(k_i)_\tau = -|k_i| + max(k_\tau) \tag{9}$$

$$W(k_i)_\tau = e^{-|3k_i|} \tag{10}$$

Both functions will be used and the fit error for each will be compared in the results section. A graphical visualization of these functions is presented in fig.(B.1).

3.2.7 Fitting the data

Due to the sheer large number of trades that occur on any given trading day, the fitting of data for a time-to-maturity (τ) and moneyness (k) is comprised of the mean implied volatility of the midpoint of all trades happening exactly at the bid and ask. This value is then squared and multiplied by the corresponding time-to-maturity to obtain a total variance number used by the models,

$$w(k,\tau) = \left(\frac{\bar{IV}^{bid}_{k,\tau} + \bar{IV}^{ask}_{k,\tau}}{2} \right)^2 \tau. \tag{11}$$

With this procedure, the number of data points to fit and, subsequently, the computational power required are dramatically reduced while still feeding the model relevant input data for our objective. Table 1 gives an example of the contents in the file used for model fitting.

Table 1: **Example of a volatility model fitting file**

time-to-maturity (ttm)	moneyness (m)	variance (w)	theta (θ_{ATM})	weight (W)
0.019	-0.248	0.0079	0.0024	0.42
0.019	-0.223	0.0079	0.0024	0.51
...	...	...	...	...
0.537	0.104	0.238	0.055	0.44
0.537	0.190	0.437	0.055	0.39

Note: Example of the contents inside of a fitting file for a specific day (April 4^{th} 2019) and stock (AAPL).

3.2.8 Calculating the volatility's model calibration error

To measure the volatility models' fit against the market traded options, the root-mean-square error (RMSE) is used. For a certain options' time to maturity (τ) and log-moneyness (k) the error of a volatility surface's fit against the market is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{\tau}^{n} \sum_{k}^{m} (\hat{\sigma}_{imp} - \sigma_{imp})^2} \tag{12}$$

where $\hat{\sigma}_{imp}$ is the observed implied volatility and σ_{imp} is the predicted implied volatility.

3.3 Market Maker Inventory Estimation

3.3.1 Intraday Trade Cassification

The ability to classify trades is separated and evaluated across four distinct methodologies. Firstly, the BA (Bid & Ask) classification is based on the premise that trades occurring exactly at the quote (ask or bid) represent the true underlying trade direction (buy or sell). Secondly, the QR (Quote-Rule) is consistent with the findings in the literature on trade classification (Savickas and Wilson, 2003) which state that this technique has an overall classification accuracy of 83 percent for option transactions. Thirdly BSCP (all calls are sold (bought) by investors (MM's) and all puts are bought (sold) by investors (MM's)) is the generalization made across the literature related to market makers' gamma exposure studies. Finally, the proposed classification model of this work, the SVI/SSVI, builds on top of the BA approach by signing its unclassified trades (as buy or sell) based on their relative position to the surface. Table 2 below gives the exact rules for classification of each approach.

Table 2: **Trade classification algorithm rules**

Rule	Conditions	Classification
BA	$P_t = P_{Ask}$	Trade is classified as buyer-initiated
	$P_t \neq P_{Ask} \| P_{Bid}$	Trade is unclassified
	$P_t = P_{Bid}$	Trade is classified as seller-initiated
QR	$P_t > P_t^{mid}$	Trade is classified as buyer-initiated
	$P_t = P_t^{mid}$	Trade is unclassified
	$P_t < P_t^{mid}$	Trade is classified as seller-initiated
BCSP	$T_t = Put$	Trade is classified as buyer-initiated
	$T_t \neq Call \; \| \; Put$	Trade is unclassified
	$T_t = Call$	Trade is classified as seller-initiated
SVI & SSVI	$IV_t > \hat{IV}_t$	Trade is classified as buyer-initiated
	$IV_t = \hat{IV}_t$	Trade is unclassified.
	$IV_t < \hat{IV}_t$	Trade is classified as seller-initiated

Notes: Rules and description for each algorithms' trade direction classification. Each trade is signed as "1" for buy positions and as "-1" for sell positions. P_t denotes the transaction price at time t; $P_t^{mid} = \frac{P_t(ask)+P_t(bid)}{2}$; IV_t represents the trade implied volatility; $\hat{IV}_t$ is the implied volatility predicted by the model; and T_t refers to the transaction type.

3.3.2 Inventory Construction & Procedure

By aggregating the volume for each trade i and classification algorithm r on any given day, we may determine the market maker's higher greek exposure. For each trade occurring on the tape, the net dealer position is calculated as follows:

$$NetDealerPosition_i^r = Classification_i^r \times -1 \times Volume_i \times 100. \tag{13}$$

The notion of the -1 multiplication is that the market maker is on the opposing side of the underlying trade being made. Subsequently, the associated dollar nominated higher greeks exposure are calculated as in Carr and Wu (2020),

$$MM\_GEX(\$)_i^r = NetDealerPosition_i \times Gamma_i \times S_i^2. \tag{14}$$

This MM_{GEX} (Gamma Exposure) value measures the rate of change in the options delta and gives the dollar amount the MM has to invest in the market (for a one point move in the underlying) in order to be delta hedged. Additionally,

$$MM\_VEX(\$)_i^r = MM\_GEX(Dollar)_i^r \times \tau_i \times \sigma_i^2 \tag{15}$$

MM_{VEX} (Vega Exposure) is the first derivative of option prices with respect to change in underlying volatility. A positive (negative) VEX value, indicates the MM predominantly holds long (short) maturity ATM options. Moreover,

$$MM\_VAX(\$)_i^r = MM\_GEX(Dollar)_i^r \times [ln(\frac{K_i}{S_i}) + \frac{1}{2}\sigma_i^2 \tau_i] \tag{16}$$

MM_{VAX} (Vanna Exposure) is a cross-derivative with respect to delta and implied volatility. Positive (negative) VAX indicates that the inventory is comprised of long (short) calls and short (long) puts.

Finally,

$$MM\_VOX(\$)_i^r = MM\_VAX(Dollar)_i^r \times [-ln(\frac{S_i}{K_i}) - \frac{1}{2}\sigma_i^2\tau_i] \tag{17}$$

MM_{VOX} (Volga Exposure) is the rate of change of vega. A net positive (negative) MM portfolio VOX indicates a net long (short) positioning. For more intuition and explanation on these greeks see section A.2 in the Appendix.

The final inventory is calculated[2] end-of-day, and is comprised of a matrix of size $k \times n$ where k denotes the strike prices and n the expiration dates. This end-of-day inventory is carried forward to the next trading day until the current date matches a expiration date. When this happens, the corresponding column n is eliminated, since all remaining exposure at that time-to-maturity expires.

With this matrix calculated we have the market maker net volume exposure at each time-to-maturity and strike. Next, we have to determine the gamma values using (18) at each point in time for each $k \times n$ intersection point (since the parameters of the options traded on previous days will change).

$$\Gamma = \frac{N'(\frac{ln(S/K)+(\sigma^2/2)\tau}{\sigma\sqrt{\tau}})}{S\sigma\sqrt{\tau}} \tag{18}$$

where K is the strike price, S is the end of day spot price, σ is the interpolated implied volatility from the SVI model calibration for that day, τ is the time-to-maturity (annualized) and N' is the standard normal probability density function.

Finally, to arrive at the end-of-day total net greek exposure is a simple matter of adding the values across all k strikes and n maturities for each calculate higher greek i defined above.

$$MM_i^{Total} = \sum_k \sum_n MM_i$$

---

[2] Details on the Algorithm can be found in Section D of the Appendix

4 Results and Analysis

4.1 Volatility Models Calibration

An example of SVI & SSVI fitting to market data following sections 3.2.2 and 3.2.5 can be found in Figures B.2 & B.3, along with their $g(k)$[3] density and total variance plots in Figures B.4 & B.5.

Two tests will be used to evaluate the volatility model. First, the robustness of the methods will be tested. This means that we will test how often the models generate solutions that introduce static arbitrage. The second test will be in the view of quality. To test the quality, the fit's residuals against the market data will be analyzed.

Table 3: **Summary of number & percentage of cases where the models introduce arbitrage**

	SVI		SSVI	
Asset	Cal. Arb.	But. Arb	Cal. Arb.	But. Arb
AAPL				
EQU	2,271 (56.96%)	185 (4.62%)	968 (24.27%)	10 (0.25%)
ABS	2,228 (55.88%)	156 (3.91%)	981 (24.60%)	10 (0.25%)
EXP	2,223 (55.76%)	158 (3.96%)	941 (23.60%)	10 (0.25%)
AMZN				
EQU	2212 (58.14%)	237 (6.23%)	876 (24.02%)	56 (1.47%)
ABS	2189 (57.54%)	220 (5.78%)	646 (16.98%)	56 (1.47%)
EXP	2205 (57.97%)	236 (6.20%)	731 (19,21%)	56 (1.47%)
BA				
EQU	2248 (59.01%)	232 (6.06%)	945 (24.80%)	79 (2.07%)
ABS	2182 (57.28%)	176 (4.62%)	805 (21.13%)	78 (2.04%)
EXP	2222 (58.33%)	211 (5.53%)	861 (22.60%)	79 (2.07%)
XOM				
EQU	2376 (66.81%)	561 (15.77%)	745 (20.95%)	112 (3.14%)
ABS	2330 (65.52%)	415 (11.67%)	432 (12.14%)	111 (3.12%)
EXP	2366 (66.53%)	526 (14.79%)	587 (16.51%)	112 (3.14%)
SPX				
EQU	7087 (80.40%)	794 (9.00%)	4982 (56.52%)	517 (5.86%)
ABS	6980 (79.19%)	773 (8.77%)	4162 (47.22%)	513 (5.82%)
EXP	6988 (79.28%)	780 (8.84%)	4656 (52.82%)	511 (5.79%)

Notes: *But. Arb.* stands for butterfly arbitrage and show how many cases butterfly arbitrage was introduced. *Cal. Arb.* stands for calendar arbitrage and shows how many cases calendar arbitrage was introduced.

---

[3] $g(k) = (1 - \frac{kw'(k)}{2w(k)})^2 - \frac{w'(k)^2}{4}(\frac{1}{w(k)} + \frac{1}{4}) + \frac{w''(k)}{2}$, where $w(k)$ is defined in eq.(1)

Table 3 summarizes the robustness test's outcome. The type of arbitrage that is mostly present in the surfaces is calendar arbitrage. As expected, the SSVI method is introducing, on average, 55% and 67% less cases for calendar and butterfly arbitrage, respectively than the SVI. This is expected as the SSVI method is optimizing for reducing arbitrage conditions.

To assess the methods quality, the metric described in (12) will be used. Table 4 presents the overall result from the quality test and in Tables 5 & 6 we can see how the error is located between different maturity and moneyness intervals. To facilitate a visual comparison of the approaches, the progression of the total error for the weight technique with the lowest total RMSE is presented in Figures B.6a - B.6e.

Table 4: **Summary of quantified goodness of fit metrics**

	SVI			SSVI		
Asset	Tot.err	Mean.err	S.mean.err	Tot.err	Mean.err	S.mean.err
AAPL						
EQU	5.789	0.023	0.008	7.251	0.029	0.010
ABS	5.501	0.022	0.008	7.386	0.029	0.012
EXP	5.558	0.022	0.008	6.929	0.027	0.011
AMZN						
EQU	12.366	0.049	0.010	7.711	0.031	0.010
ABS	7.662	0.030	0.014	8.619	0.034	0.013
EXP	10.685	0.042	0.009	7.804	0.031	0.011
BA						
EQU	6.361	0.025	0.008	6.585	0.026	0.010
ABS	5.073	0.020	0.008	7.143	0.028	0.012
EXP	5.890	0.023	0.007	6.765	0.027	0.011
XOM						
EQU	2.693	0.011	0.005	5.012	0.020	0.011
ABS	4.884	0.019	0.024	5.599	0.022	0.014
EXP	2.648	0.011	0.005	5.279	0.021	0.012
SPX						
EQU	5.573	0.022	0.009	9.464	0.038	0.024
ABS	5.373	0.021	0.036	11.246	0.045	0.037
EXP	5.096	0.020	0.008	10.444	0.041	0.032

Notes: Tot.err stands for total error and gives the total error for all 251 trading days for each underlying. Mean.err stands for mean error and gives the mean error for each dataset. S.mean.err stands for smile mean error and gives the mean error for every smile in the data. EQU, ABS and EXP denote the weight functions defined in sub-section (3.2.6).

As per Table 4, both weight functions introduced manage to reduce the total error but SVI_{EXP} & SVI_{EQU} are almost always outperformed by SVI_{ABS} and $SSVI_{EXP}$ & $SSVI_{ABS}$ by $SSVI_{EQU}$.

It is self-evident that the SVI method generates significantly superior results to the SSVI approach in all circumstances. By and large, the former produces solutions that are roughly 20%-50% better than the latter. This is because it lacks arbitrage restrictions to minimize static arbitrage and naturally results in a more accurate fit.

Table 5: **Summary of quantified goodness of fit by maturity**

	SVI			SSVI		
Asset	Short	Medium	Long	Short	Medium	Long
AAPL						
EQU	0.041	0.010	0.005	0.033	0.013	0.006
ABS	0.027	0.008	0.005	0.034	0.013	0.007
EXP	0.036	0.008	0.004	0.031	0.012	0.007
AMZN						
EQU	0.053	0.011	0.004	0.033	0.012	0.006
ABS	0.033	0.010	0.008	0.037	0.014	0.008
EXP	0.046	0.009	0.004	0.033	0.012	0.008
BA						
EQU	0.029	0.007	0.004	0.030	0.008	0.006
ABS	0.022	0.008	0.010	0.033	0.010	0.008
EXP	0.027	0.007	0.004	0.031	0.009	0.007
XOM						
EQU	0.014	0.005	0.003	0.026	0.012	0.008
ABS	0.014	0.017	0.017	0.027	0.015	0.012
EXP	0.014	0.005	0.003	0.026	0.013	0.009
SPX						
EQU	0.024	0.016	0.008	0.042	0.025	0.018
ABS	0.018	0.016	0.042	0.049	0.034	0.026
EXP	0.022	0.013	0.008	0.046	0.030	0.023

Notes: Result is showing in what interval of maturity the mean smile error is located. *Short* is defined as the maturity interval (0, 0.05), *Medium* is defined as the maturity interval (0.05, 0.25) and *Long* is defined as the maturity interval (0.25, ∞).

As evident in Table 5, the shorter term maturities exhibit the greatest degree of inaccuracy. This can be attributed to the fact that put options of shorter time-to-maturities tend to have higher implied volatilities in the wings. This observed skew results in a steeper shape for the related volatility smile, which makes it more difficult for the model to calibrate, resulting in a greater inaccuracy.

Table 6: **Summary of quantified goodness of fit by moneyness**

	SVI			SSVI		
Asset	Short	Medium	Long	Short	Medium	Long
AAPL						
EQU	0.038	0.021	0.019	0.023	0.040	0.038
ABS	0.025	0.017	0.019	0.018	0.043	0.057
EXP	0.033	0.019	0.018	0.020	0.040	0.045
AMZN						
EQU	0.052	0.032	0.021	0.023	0.045	0.056
ABS	0.031	0.022	0.026	0.021	0.055	0.080
EXP	0.045	0.027	0.021	0.021	0.049	0.066
BA						
EQU	0.027	0.019	0.015	0.021	0.035	0.036
ABS	0.020	0.017	0.020	0.019	0.037	0.051
EXP	0.025	0.019	0.015	0.020	0.036	0.043
XOM						
EQU	0.011	0.009	0.008	0.014	0.026	0.034
ABS	0.017	0.014	0.021	0.013	0.027	0.045
EXP	0.011	0.009	0.008	0.013	0.026	0.040
SPX						
EQU	0.024	0.014	0.011	0.021	0.063	0.073
ABS	0.019	0.017	0.030	0.020	0.068	0.099
EXP	0.022	0.014	0.012	0.020	0.066	0.088

Notes: Result is showing in what interval of moneyness the mean smile error is located. *Short* is defined as the moneyness interval (0, 0.5), *Medium* is defined as the moneyness interval (0.5, 1) and *Long* is defined as the moneyness interval $(1, \infty)$.

Additionally, Table 6 demonstrates the success of introducing a weights function into the calibration process. By doing it, the mean error for *Short* moneyness intervals is reduced. This with the cost of increased inaccuracy in some circumstances for *Medium* and *Longer* moneyness intervals. However, as Table C.4 shows, this is a worthy tradeoff since around 50-75% of the traded volume is situated in *Short* moneyness intervals.

4.2 Classification Algorithms

The first objective of this work is to accurately flag intraday trading volume. The easiest and most accurate volume to flag are aggressive orders that cross the market at the quotes (i.e., trades occurring exactly at the bid and ask prices), because these orders are *most probably* filling on the quotes that encapsulate the underlying market maker quoting model. However, before employing a classification algorithm it is useful to get a clear picture of where the trades prices lie when compared to the prevailing quoted prices.

As the tables C.5 & C.6 show, a relatively large proportion of all trades (avg. 25.3% across the analyzed securities) occurs in- and out-side of the quotes. This is where the volatility models may bring added value by categorizing these trades more accurately than the current relative quote position approaches.

Table 7: **Algorithms direction classification bias**

Asset	BA	QR	BCSP	SVI	SSVI
AAPL					
Buyer-initiated	37.78	44.12	58.38	49.06	49.88
Seller-initiated	39.54	45.39	41.62	50.94	50.12
Unclassified	22.68	10.49	0.00	0.00	0.00
AMZN					
Buyer-initiated	33.42	48.91	61.93	49.98	52.70
Seller-initiated	33.10	47.90	38.07	50.02	47.30
Unclassified	33.48	3.19	0.00	0.00	0.00
BA					
Buyer-initiated	34.02	47.95	59.67	50.28	50.51
Seller-initiated	33.94	47.26	40.33	49.72	49.48
Unclassified	32.04	4.79	0.00	0.00	0.00
XOM					
Buyer-initiated	39.05	45.76	59.36	48.48	47.89
Seller-initiated	41.92	46.88	40.64	51.52	52.11
Unclassified	19.03	7.36	0.00	0.00	0.00
SPX					
Buyer-initiated	40.74	46.78	42.67	51.99	49.34
Seller-initiated	41.62	46.91	57.33	48.01	50.66
Unclassified	17.64	6.31	0.00	0.00	0.00

Note: All values should be read in percentages

As can be seen from Table 7, the studied approaches of this work, as well as the BCSP, leave no trade unclassified. On the contrary, the applicability and accuracy of the BA and QR procedures is rather low for the objective of having an accurate measure of MM inventory exposure, with a high percentage of unclassified trades (avg. 24.97% and 6.42% respectively), across the assets.

4.3 Market Maker Inventory Exposure

As outlined in the literature, market makers hedging activity as price changes in the underlying is assumed to generate feedback effects. Following the same line of argument, market makers must also be hedging in accordance to changes in volatility (vega, vanna, volga). In this sub-section, we will validate and quantify these feedback effects empirically with the data on the MM greeks exposure that was reconstructed from the intraday options trade data.

The analysis will be articulated around the following null hypothesis:

H_0 : *The inclusion of other greek factors does not help explain more market return variation.*

To this end, we first estimate the baseline univariate regression model in (19) that only accounts for gamma effects, i.e.,

$$U_{i,t}[\%] = GEX_t[\frac{Chng}{bn}] + U_{i,t-1}[\%] + \varepsilon_{i,t}. \tag{19}$$

Also the effects of the other higher order greeks are estimated using multivariate regression,

$$U_{i,t}[\%] = GEX_{i,t}[\frac{Chng}{bn}] + VEX_{i,t}[\frac{Chng}{bn}] + VAX_{i,t}[\frac{Chng}{bn}] + VOX_{i,t}[\frac{Chng}{bn}] + U_{i,t-1}[\%] + \varepsilon_{i,t} \tag{20}$$

where $U_{i,t}$ is the realized percent change for the underlying asset i in the time period between t and $t+1$, and $GREEK_{i,t}$ is the reconstructed daily dollar change in market maker exposure scaled by billions of dollars. Additionally, we control for the previous days underlying asset return. As the returns depend on the intraday hedging cycle, we validate the model with daily data. The calculation of the greeks exposures included in these regressions are described in sub-section 3.3.2.

The estimation results for (19) are summarized in Tables 8 & 9 and for (20) in Tables 10-12. Plots of the regressions are also shown in Figures B.7a - B.7e.

Table 8: **Univariate regression**

The table shows OLS regression results of eq.(19). The dependent variable is the percentage change in asset price. The independent variable is the absolute change in *MM* inventory *gamma* exposure defined in eq.(14) and scaled by billions of dolars. The sample covers 251 daily observations from January 2^{nd} through December 31^{st} 2019. Reported t-statistics are in parentheses. Key: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Asset	BA	QR	BCSP	SVI	SSVI
AAPL					
Constant	0.33***	0.33***	0.31***	0.33***	0.32***
	(3.39)	(3.40)	(3.52)	(3.50)	(3.40)
GEX Chng	-0.04**	-0.04***	0.01***	0.03***	-0.03***
	(-2.44)	(-2.63)	(7.57)	(4.70)	(-3.19)
AAPL Lag	-0.10*	-0.10*	-0.08	-0.09	-0.09
	(-1.71)	(-1.73)	(-1.44)	(-1.59)	(-1.63)
$R^2(\%)$	3.34	3.72	19.65	9.13	4.92
AMZN					
Constant	0.09	0.10	0.08	0.09	0.09
	(1.07)	(1.11)	(1.00)	(1.02)	(1.02)
GEX Chng	-0.07***	-0.06***	0.00***	0.00	-0.00
	(-4.66)	(-5.72)	(7.60)	(0.83)	(-0.08)
AMZN Lag	0.05	0.06	0.04	0.00	0.01
	(0.83)	(1.02)	(0.68)	(0.06)	(0.15)
$R^2(\%)$	8.10	11.70	18.96	0.28	0.01
BA					
Constant	0.03	0.03	0.03	0.04	0.04
	(0.30)	(0.30)	(0.35)	(0.31)	(0.33)
GEX Chng	-0.03	-0.04	0.04***	0.07***	-0.11***
	(-0.31)	(-0.51)	(8.87)	(2.74)	(-3.04)
BA Lag	0.01	0.01	0.04	0.02	0.01
	(0.12)	(0.14)	(0.66)	(0.32)	(0.18)
$R^2(\%)$	0.04	0.11	24.18	2.96	3.62

Table 9: **Univariate regression** (Continuation of Table 8)

Asset	BA	QR	BCSP	SVI	SSVI
XOM					
Constant	0.01	0.01	0.01	0.01	0.01
	(0.18)	(0.19)	(0.17)	(0.20)	(0.20)
GEX Chng	-0.11	0.12	0.31***	0.09	0.20
	(-0.79)	(0.92)	(10.86)	(0.63)	(1.39)
XOM Lag	-0.03	-0.02	0.03	-0.02	-0.02
	(-0.42)	(-0.39)	(0.63)	(-0.37)	(-0.33)
$R^2(\%)$	0.31	0.41	32.38	0.22	0.84
SPX					
Constant	0.12**	0.12**	0.12***	0.12**	0.12**
	(2.51)	(2.51)	(2.68)	(2.55)	(2.55)
GEX Chng	-0.00	-0.00	0.00***	0.00***	-0.00**
	(-0.24)	(-0.48)	(8.52)	(3.96)	(-2.44)
SPX Lag	-0.09	-0.09	-0.04	-0.08	-0.10
	(-1.47)	(-1.48)	(-0.75)	(-1.34)	(-1.63)
$R^2(\%)$	0.90	0.97	23.38	6.80	3.21

In the baseline univariate regression, *GEX Chng* shows highly significant parameter estimates (at the 1% level) which is indicative that changes in MM gamma inventory exposure have a high relevance for the daily variation of returns. I estimate a positive effect of change in MM_{GEX} on same day stock prices. From the R^2, we may infer that market maker gamma exposure changes explain on average 23.71% of the daily variation in returns.

Comparing the classification approaches, we constantly see that the BCSP mehod is the superior, followed by SVI, SSVI, QR and BA. The case is not the same for AMZN, in which the quote based methods have a higher explainable ability than the volatility models. This indicates to us, that in this market, the market makers are the primary liquidity providers near- and at- the quotes.

By examining the magnitude of the coefficients, we can deduce that the feedback effect is stronger (per billion dollar change in gamma exposure) in BA and XOM than in the rest, which can probably be explained by the lower liquidity of the spot market of these and thus the higher market share of the delta hedging strategy (and subsequently the higher impact of the traded volume).

Table 10: **Multivariate regression**

The table shows OLS regression results of eq.(20). The dependent variable is the percentage change in asset price. The independent variables are the absolute change in *MM* inventory *gamma* eq.(14), *vega* eq.(15), *vanna* eq.(16) and *volga* eq.(17) exposures scaled by billions of dolars. A control variable for the previous days asset return is also included. The sample covers 251 daily observations from January 2^{nd} through December 31^{st} 2019. Reported t-statistics are in parentheses. Key: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.015$

Asset	BA	QR	BCSP	SVI	SSVI
AAPL					
Constant	0.33***	0.33***	0.29***	0.32***	0.32***
	(3.40)	(3.44)	(3.92)	(4.04)	(3.52)
GEX Chng	-0.04**	-0.05***	0.01***	0.03***	-0.04***
	(-2.55)	(-2.84)	(6.08)	(5.37)	(-4.00)
VEX Chng	-10.67*	-11.21***	3.95***	12.11***	-9.18***
	(-1.69)	(-2.61)	(7.08)	(3.45)	(-4.25)
VAX Chng	0.82	-0.05	-0.49***	2.83***	-1.06***
	(0.84)	(-0.06)	(-9.46)	(8.53)	(-2.68)
VOX Chng	1.83	4.21	-1.93***	-11.20***	4.50**
	(0.44)	(1.50)	(-2.95)	(-5.50)	(2.31)
AAPL Lag	-0.10*	-0.10*	-0.06	-0.09*	-0.11*
	(-1.73)	(-1.75)	(-1.32)	(-1.93)	(-1.92)
Adj. $R^2(\%)$	3.19	4.51	42.94	35.44	11.60
AMZN					
Constant	0.10	0.09	0.08	0.08	0.10
	(1.11)	(1.17)	(1.21)	(1.10)	(1.22)
GEX Chng	-0.06***	-0.04***	0.00***	-0.00*	0.00
	(-3.89)	(-4.28)	(7.06)	(-1.71)	(0.48)
VEX Chng	-5.46	-5.33*	2.46***	4.33**	-5.24***
	(-1.15)	(-1.92)	(7.44)	(2.31)	(-3.26)
VAX Chng	-2.62**	-3.66***	-0.20***	1.37***	-1.19***
	(-2.29)	(-5.07)	(-7.04)	(12.29)	(-5.01)
VOX Chng	0.11	-2.54	-1.86***	-5.25***	5.01***
	(0.02)	(-0.60)	(-5.84)	(-3.27)	(2.91)
AMZN Lag	0.06	0.09	0.03	0.02	-0.03
	(1.01)	(1.49)	(0.56)	(0.48)	(-0.46)
Adj. $R^2(\%)$	9.04	19.79	43.06	38.26	14.55

Table 11: **Multivariate regression** (Continuation of Table 10)

Asset	BA	QR	BCSP	SVI	SSVI
BA					
Constant	0.03	0.03	0.04	0.04	0.04
	(0.30)	(0.30)	(0.42)	(0.38)	(0.40)
GEX Chng	-0.06	-0.06	0.03***	0.04*	-0.07**
	(-0.67)	(-0.77)	(5.35)	(1.77)	(-2.33)
VEX Chng	52.96*	19.32	13.76***	9.63	-48.11***
	(1.87)	(0.87)	(8.89)	(0.55)	(-4.40)
VAX Chng	3.02	-6.33**	-0.52***	9.20***	-6.03***
	(0.76)	(-2.00)	(-3.78)	(6.43)	(-3.72)
VOX Chng	-65.56**	-51.77**	-5.86***	-53.39***	85.88***
	(-2.58)	(-2.22)	(-4.05)	(-3.13)	(6.01)
BA Lag	-0.01	-0.00	0.05	0.03	0.01
	(-0.15)	(-0.02)	(1.13)	(0.43)	(0.19)
Adj. $R^2(\%)$	1.97	2.11	45.64	16.06	24.03
XOM					
Constant	0.01	0.02	0.01	0.02	0.01
	(0.20)	(0.22)	(0.20)	(0.24)	(0.21)
GEX Chng	-0.00	0.21	0.36***	0.22	0.33**
	(-0.03)	(1.57)	(11.49)	(1.57)	(2.21)
VEX Chng	-36.02**	-18.61	-5.83	-31.14*	-20.06
	(-1.98)	(-0.97)	(-1.60)	(-1.68)	(-1.13)
VAX Chng	23.53***	20.26***	-0.69	30.79***	16.39***
	(4.03)	(3.06)	(-1.31)	(5.62)	(3.04)
VOX Chng	-54.67**	-53.56**	-10.33***	-57.28**	-61.83***
	(-2.48)	(-2.20)	(-2.81)	(-2.35)	(-2.95)
XOM Lag	-0.01	-0.01	0.02	0.01	0.00
	(-0.16)	(-0.20)	(0.40)	(0.09)	(0.08)
Adj. $R^2(\%)$	5.01	3.34	35.19	10.96	4.36

Table 12: **Multivariate regression** (Continuation of Table 10)

Asset	BA	QR	BCSP	SVI	SSVI
SPX					
Constant	0.12**	0.12**	0.11***	0.12**	0.12**
	(2.56)	(2.51)	(2.74)	(2.58)	(2.54)
GEX Chng	-0.00	-0.00	0.00***	0.00***	-0.00**
	(-0.76)	(-0.87)	(8.30)	(4.14)	(-2.49)
VEX Chng	0.44*	0.27*	0.10***	0.20*	0.01
	(1.70)	(1.88)	(3.81)	(1.88)	(0.09)
VAX Chng	0.07	0.03	0.00*	-0.02	-0.02
	(1.45)	(0.71)	(1.70)	(-1.44)	(-0.77)
VOX Chng	0.29*	0.17	0.05***	0.06	0.07
	(1.74)	(1.36)	(3.78)	(1.04)	(0.91)
SPX Lag	-0.10*	-0.09	-0.01	-0.08	-0.10
	(-1.70)	(-1.44)	(-0.26)	(-1.29)	(-1.62)
Adj. $R^2(\%)$	1.84	0.97	30.04	9.15	2.06

In the multivariate regression, market makers inventory greek exposure metrics show strong significance. For AAPL, AMZN and BA, all of the greeks show significance at the 1% significance level. For XOM and SPX, both GEX and VOX changes show significance at the 1% level while VEX and VAX have no significance for XOM and are significant at the 1% and 10%, respectively for SPX.

Additionally, the coefficients for GEX and VEX are positive, whereas those for VAX and VOX are negative. The positive sign on VEX suggests that market makers as buyers of longer dated volatility are coincident with positive return days. The negative slope on VAX implies that market makers accumulate short put / long call inventory on down market days. The negative slope of VOX suggests short market maker inventory in the wings, which results in down asset return days.

The inclusion of the additional higher greek terms drastically increases the explanatory power of the model (Adj. R^2 of 39.37%). While it is expected for the R^2 to increase with the addition of more variables, the adjusted R^2 metric is utilized in this multivariate case because it has the potential to be more accurate in that it accounts for the number of independent variables in the model.

5 Discussion

5.1 Summary of Findings

The purpose of this work project was to contribute to the emerging literature documenting a non-informational channel through which option markets affect the underlying stock's price dynamics. The contribution is three-fold: first, it provides a novel method of identifying trade direction and subsequent market maker inventory exposure, and second, it examines the effect of other higher order greek exposure on the prevailing underlying asset returns.

The empirical experiments conducted reveal that the application of implied volatility modeling methods as a novel approach to classification of option transaction has yielded highly promising results. Contrary to quote-based methods, this leaves no trade unclassified. However, it was concluded that the task of classifying trade direction does not directly translate to the task of classifying dealer inventory. While the SVI and SSVI are better than BA and QR, BCSP is still the superior approach across the analyzed securities.

This research also indicates that gamma exposure is only the tip of the iceberg when it comes to explaining market returns variation. Three considerably less accounted for greeks (*vega*, *vanna* and *volga*) are responsible for a more significant portion of non-fundamental daily market behavior. It was demonstrated that by controlling for these other factors over the common gamma exposure, an increase of 66.04% explanatory power was obtained, therefore we reject the null hypothesis proposed.

This work highlights the informational power in modelling these often-overlooked variables, which are especially pertinent not only for option and stock market participants wanting to take advantage of these effects, but also for regulatory bodies, as extreme levels of these exposures can lead to feedback effects causing down- or up-ward price spirals.

5.2 Review of assumptions and approach

The main assumptions serving as the basis of this research are that all trades are facilitated by market makers, and that their method for quoting prices is based on the relative position to the modelled implied volatility surface.

The results show that the two assumptions mentioned above do indeed have explanatory power. However, the sensitivity of these research results might be application specific. Although the methodology presented seems more appropriate than current methods and gives a more complete view of the total exposure, the primary purpose of this paper is not to prove that this specific approach is the state of the art in classifying dealer inventory. Rather, it is my desire and intention that the examination of the assumptions and evidence presented here, will result in improved inferences in all subsequent applications.

5.3 Recommendations for Future Work

Potential investigations following the results from this work are:

- Using an alternative fitting methodology. Other than SVI, there are other potential candidates for volatility modeling (Heston, SABR, ...), and since the need for these flagging models is not to price contracts, but rather, to inductively model how someone else is pricing contracts, purely descriptive mechanisms like polynomial fits could also yield promising results.
- Further modelling of the market makers quoting policy. These players often have an optimal quoting policy given their current inventory and fundamental view. By incorporating more parameters to model the willingness of market maker to take on more risk, one can possibly achieve a realistic classification system and benefit from increased accuracy.
- Generating trading strategies. This work sheds light on other factors influencing the underlying price dynamics that can be taken advantage of. For a market participant wanting to obtain material rewards, this would be a worthwhile endeavor.

References

Avellaneda, M. and M. D. Lipkin (2003). A market-induced mechanism for stock pinning. *Quantitative Finance 3*(6), 417–425.

Baltussen, G., Z. Da, S. Lammers, and M. Martens (2021). Hedging demand and market intraday momentum. *Journal of Financial Economics 142*(1), 377–403.

Barbon, A. and A. Buraschi (2020). Gamma fragility. *Working paper, University of St.Gallen, School of Finance*.

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

Carr, P. and D. B. Madan (2005). A note on sufficient conditions for no arbitrage. *Finance Research Letters 2*(3), 125–130.

Carr, P. and L. Wu (2020). Option profit and loss attribution and pricing: A new framework. *Journal of Finance 75*(4), 2271–2316.

Finucane, T. J. (2000). A direct test of methods for inferring trade direction from intra-day data. *The Journal of Financial and Quantitative Analysis 35*(4), 553–576.

Gatheral, J. (2004). A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. *Presentation at Global Derivatives & Risk Management, Madrid*.

Gatheral, J. and A. Jacquier (2014). Arbitrage-free svi volatility surfaces. *Quantitative Finance 14*(1), 59–71.

Golez, B. and J. C. Jackwerth (2012). Pinning in the s&p 500 futures. *Journal of Financial Economics 106*(3), 566–585.

Hu, J. (2014). Does option trading convey stock price information. *Journal of Financial Economics 111*(3), 625–645.

Lee, C. M. C. and M. J. Ready (1991). Inferring trade direction from intraday data. *The Journal of Finance 46*(2), 733–746.

Marco, S. D. and C. Martini (2009). Quasi-explicit calibration of gatheral's svi model. *Zeliade White Paper*, 1–15.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science 4*(1), 141–183.

Michaely, R., K. Ellis, and M. O'Hara (2000). When the underwriter is the market marker: An examination of trading in the ipo aftermarket. *The Journal of Finance 55*, 1039–1074.

Ni, S., N. Pearson, and A. Poteshman (2005). Stock price clustering on option expiration dates. *Journal of Financial Economics 78*(1), 49–87.

Ni, S., N. Pearson, A. Poteshman, and J. White (2020). Does option trading have a pervasive impact on underlying stock prices? *The Review of Financial Studies 34*.

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

Roper, M. (2010). Arbitrage free implied volatility surfaces. *The University of Sydney*.

Rosenthal, D. W. R. (2012). Modeling trade direction. *Journal of Financial Econometrics 10*(2), 390–415.

Savickas, R. and A. J. Wilson (2003). On inferring the direction of option trades. *Journal of Financial and Quantitative Analysis 38*(4), 881–902.

Zhu, A. (2014). Implied volatility modelling. *University of Waterloo*.

A Theoretical Background

A.1 Options

At its core, this thesis is a study of options. An option is a financial derivative that gives the owner the right, but not the obligation, to buy or sell an underlying asset at an agreed-upon strike price K at a later point in the future T. Option contracts that give the right to buy the underlying asset are known as call options, whilst options that give the owner the right to sell an underlying asset are known as put options.

There are primarily two types of options classified by their time of exercise: European and American options. European options can be exercised only at the maturity date, while American options can be exercised at any point in time before expiration.

While European options have a closed formula for determining their price, American options do not. Due to the possibility to exercise the options at any time, the prices of American options can not be calculated explicitly but need to be approximated. The most popular numerical methods for this are binomial tree, lattice, partial differential equation and Monte Carlo.

Along this thesis, European options are assumed for every analysis and for this reason, whenever reference to options is made, the word 'European' is omitted for simplicity purposes.

A.1.1 Black-Scholes Formula

Fischer Black, Myron Scholes (Black and Scholes, 1973) and Robert Merton (Merton, 1973) provided in 1973 a breakthrough discovery in the options pricing world. Through their work a explicit formula for pricing European options was presented.

They started with the dynamics of price path for an asset S_t with W_t the Brownian motion under the risk-free measure, as the Geometric brownian motion.

$$dS_t = rS_t dt + \sigma dW_t$$

By assuming that the return of the riskless asset r and volatility of stock returns σ are constant and that the stock pays no dividend, a PDE was found for a payoff function $V(S_t,t)$

$$\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2 S_t^2\frac{\partial^2 V}{\partial S_t^2}+rS_t\frac{\partial V}{\partial S_t}-rV=0$$

With this in mind, the formulas for pricing of call and put options were found.

Definition A.1 (Black-Scholes Formula)**.** The price of an European call option for a non-dividend-paying underlying stock is given by the Black-Scholes Formula:

$$C(S_t,t,T,\sigma,K)=S_tN(d_1)-Ke^{-r(T-t)}N(d_2)]$$

While the price of the corresponding put option based on put-call parity with discount factor $e^{-r(T-t)}$:

$$P(S_t,t,T,\sigma,K)=Ke^{-r(T-t)}-S_t+C(S_t,t)$$

Where S_t is the price of the underlying asset, N($\cdot$) is the cumulative distribution function for the normal distribution, r is the risk free rate and

$$d1=\frac{ln(s/K)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}$$
$$d2=d1-\sigma\sqrt{T}$$

A.2 The Greeks

At financial institutions, risk is measured in a variety of ways. The most common way is to make sure individual traders exposure to certain markets is not too large. Only measuring individual risk is not sufficient, and risk management teams often aggregate the risk and assess it in terms of capital adequacy and compliance.

On a individual level, each trader needs to monitor how sensitive their portfolios are to changes in underlying market conditions. "The Greeks" are partial derivatives of the price and measure sensitivity of the options value to changes in parameter values, while holding the all other parameters fixed.

This knowledge of sensitivity is critical not only for financial institutions but also for market makers, because by monitoring the sensitivity of the Greeks to their portfolio, they are then able to make important decisions with respect to hedging.

A.2.1 Delta (Δ)

Delta (Δ) measures the (expected) change in the option's value (V) with respect to changes in the underlying security spot price (S_0)

$$\Delta = \frac{\partial V}{\partial S_0}$$

The calculation of Δ is different for call and put options

$$\Delta_{Call} = N(d1)$$

$$\Delta_{Put} = N(d1) - 1$$

Important observations about Δ:

1. Delta can be seen as the probability that an option will expire in-the-money.
2. Delta increases as an option gets further in-the-money.
3. As expiration approaches, changes in stock value will cause more dramatic changes in delta.

A.2.2 Gamma (Γ)

Gamma (Γ), also referred to as the ”Convexity Factor” because it measures the (expected) change in Delta Δ with respect to the underlying asset spot price S_t

$$\Gamma = \frac{\partial \Delta}{\partial S_t}$$

Calculation of Γ is identical for put and call options, *ceteris paribus*, and is given by

$$\Gamma = \frac{N'(d1)}{S\sigma\sqrt{T-t}}$$

Important observations about Γ:

1. Gamma is an indicator of how frequently a market maker needs to re-balance his hedging ratio. If Γ is large, Δ is highly sensitive to S_t and vice versa.
2. Gamma is positive for buy positions and negative for sell positions.
3. Gamma is highest for shorter dated options. (Gamma increases with a decrease in time-to-maturity.)

A.2.3 Vega (ν)

Vega (ν) measures the (expected) change in the options value (V) with respect to the change in volatility (σ) of the underlying security

$$\nu = \frac{\partial V}{\partial \sigma}$$

Calculation of ν goes as follows:

$$\nu = SN'(d_1)\sqrt{T-t}$$

Important observations about ν:

1. A rise in volatility increases the options value.
2. Vega is highest when the underlying price is near the option's strike price.
3. Vega declines as the option approaches expiration.

A.2.4 Vanna

Vanna measures the (expected) move of delta (Δ) with respect to changes in implied volatility (σ).

$$Vanna = \frac{\partial \Delta}{\partial \sigma}$$

The calculation of Vanna goes as follows:

$$Vanna = \sqrt{T-t}N(d_1)(1-d_1)$$

Important observations about Vanna:

1. Vanna is positive (negative) for OTM calls (puts).
2. Rising volatility increases deltas because it expands the underlying price distribution.
3. Is often used when testing sensitivity of hedge ratios to volatility.

A.2.5 Volga

Volga measures the rate of change in Vega (ν) with respect to a unit change in implied volatility (σ).

$$Volga = \frac{\partial \Delta}{\partial \sigma}$$

The calculation of Volga goes as follows:

$$Volga = \sqrt{T-t}N(d_1)(\frac{d_1 d_2}{\sigma})$$

Important observations about Volga:

1. Vega is maximum for ATM options, subsequently Volga is zero at this point.
2. Volga is positive for OTM calls and puts
3. For volatility movements, Volga is to Vega what Gamma is to Delta

A.3 Volatility

Volatility (usually denoted by σ) measures the dispersion around the mean return of a financial instrument. Statistically, volatility is often measured as the standard deviation of the prices.

Volatility comes in many different flavors, with the two main types being historical volatility and implied volatility. Historical volatility is obtained using past prices of the financial instruments. Implied volatility on the other hand, is a dynamic figure that represents the market's consensus on the forward-looking volatility of a security.

A.3.1 Implied Volatility

The famous market crash of 1987 sparked a shift in the options pricing field, as before this time, traders relied heavily on the use of the Black-Scholes formula with parameters obtained from historical data to price their options. The problem with using it, is that it assumes constant market volatility and log-normality of stock returns.

During this crash, the high levels of volatility in the market made traders vulnerable to considerable amount of risk. This made them rethink how they used this tool, and what parameters were used as input.

Since then, the formula was no longer used for the pricing process but instead for quoting option prices in terms of their implied volatility. This value was no longer seen as constant, meaning that one of the assumptions of the Black-Scholes formula was no longer true. Since then, volatility has been written as a function of the strike and time to maturity instead of a being a fixed parameter.

Definition A.2. Implied volatility is a function of the strike K and time to maturity T where C, s and r are observed in the market.

$$\sigma_{implied} = f(K, T | C, s, r)$$

Accurate measurement and modeling of implied volatility is of extreme importance and utility since the pricing of options contracts, risk management metrics and effectiveness of trading strategies greatly depend on it.

Moreover, when implied volatility is plotted as a function of the strike K, displays local maximum values for extreme strike prices far from the spot price S_t, which visually results in a skew or ”smile” instead of the expected flat surface as per Black-Scholes. That is the reason why it is often referred to as a volatility smile when one speaks of the implied volatility for a certain time to maturity T

A.3.2 Implied Volatility Surface

Given all options available on the market for a underlying security, the asset-specific implied volatility can be derived for each option. This set of discrete data points can then be interpolated to create a surface, known as the implied volatility surface. This surface is very useful because it can output a implied volatility figure for any combination of strike price and time to maturity.

It has been observed empirically that the surface has some general profile of characteristics:

1. The surface has a smile profile in the strike-price depended direction of the surface.
2. In the maturity direction, the surface has a linear leaning profile (term structure).
3. The curvature of the smile flattens with longer maturities (a.k.a deformation).

The surface also changes in time. The observed time dependent characteristics are mainly:

1. Implied volatility display high (positive) auto-correlation and mean-reversion. This is also known as the volatility clustering.
2. Returns of the underlying asset and return of implied volatility are negatively correlated. This is also known as the leverage effect.
3. Relative movements within the implied volatility surface have little correlation with the underlying.

B Figures

Figure B.1: **Visualization of weights functions**

Shows the graphical visualization of the absolute and exponential weights functions described in eq.(9) and eq.(10) for a sample moneyness k in the interval [-1.5; 1.5].

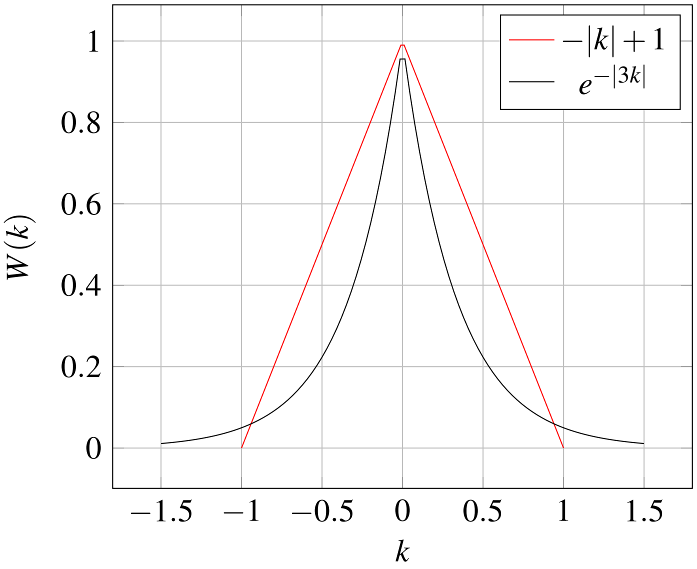


Figure B.2: **Plots of the SVI fits**

Shows the SVI fits of AAPL quote-midpoint implied volatilities for each time-to-maturity listed on May 1^{st}, 2019

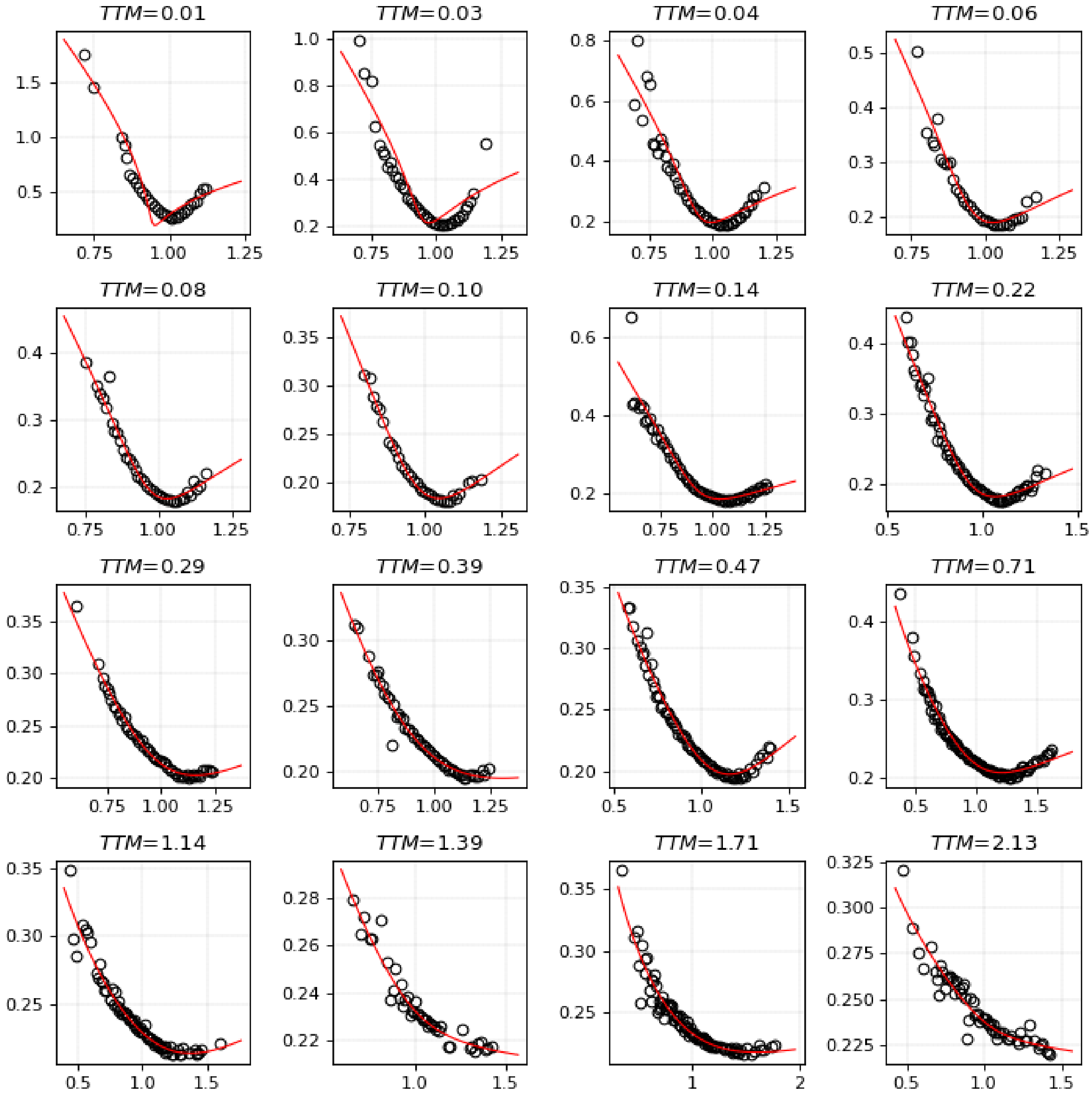


Figure B.3: **Plots of the SSVI fits**

Shows the SSVI fits of AAPL quote-midpoint implied volatilities for each time-to-maturity listed on May 1^{st}, 2019

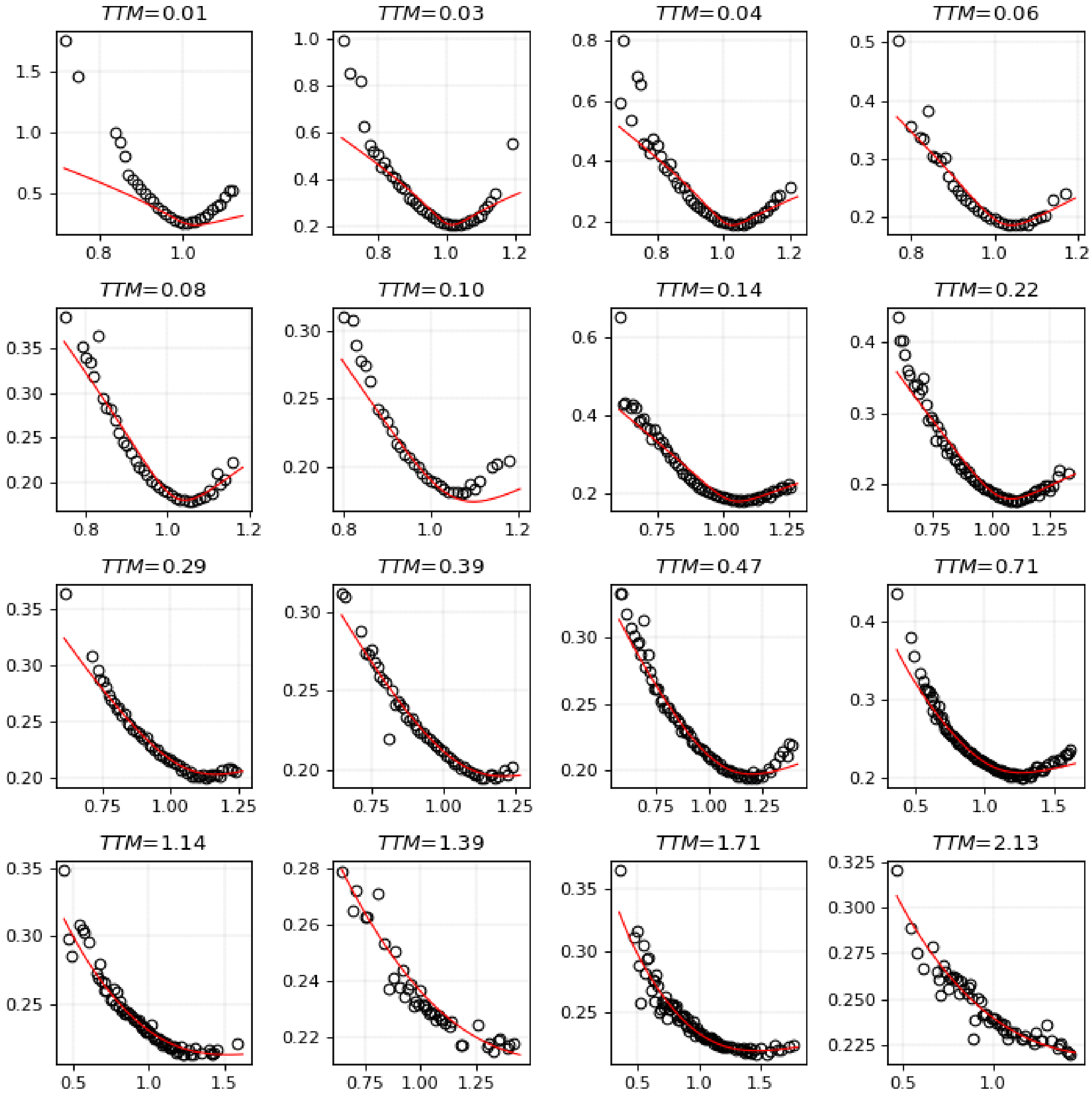
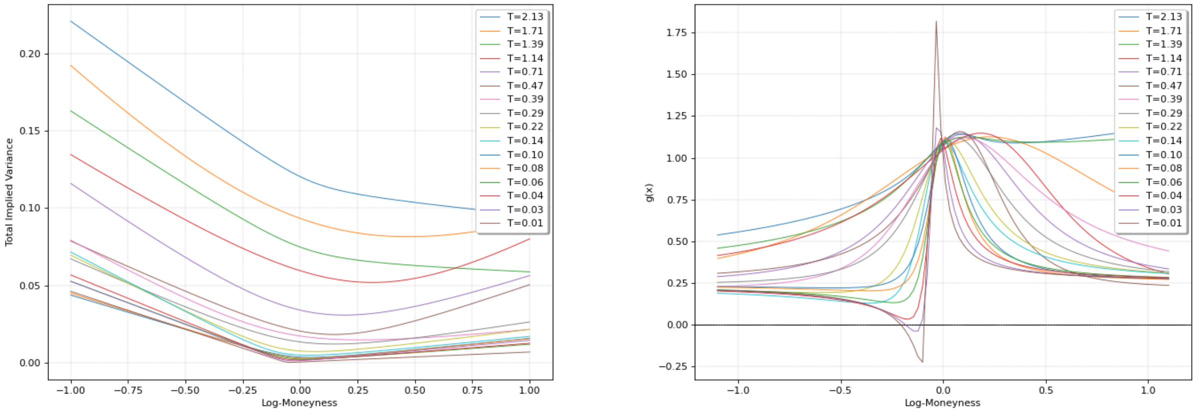


Figure B.4: **SVI static arbitrage tests**

Shows the SVI static arbitrage tests for AAPL on May 1^{st}, 2019. In (a), the total variance smiles are plotted against market data, if any lines intersect, there exists arbitrage. In (b) is the depiction of the *Durrleman's* condition, if the lines fall below 0 there exists arbitrage. Both calendar and butterfly arbitrage are introduced in this instance.

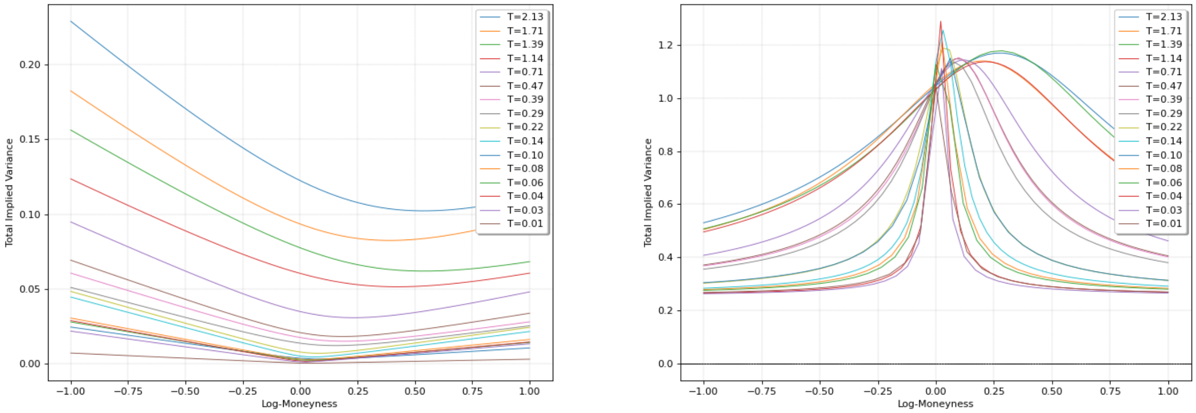


(a) Calendar Arbitrage

(b) Butterfly Arbitrage

Figure B.5: **SSVI static arbitrage tests**

Shows the SSVI static arbitrage tests for AAPL on May 1^{st}, 2019. In (a), the total variance smiles are plotted against market data, if any lines intersect, there exists arbitrage. In (b) is the depiction of the *Durrleman's* condition, if the lines fall below 0 there exists arbitrage. No arbitrage opportunities are introduced in this instance.

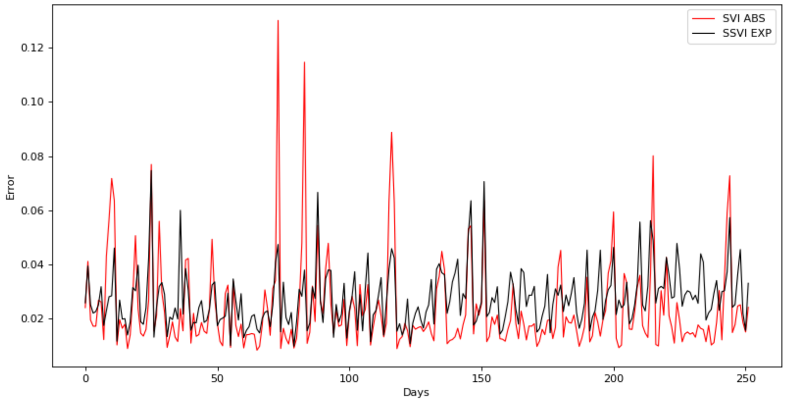


(a) Calendar Arbitrage

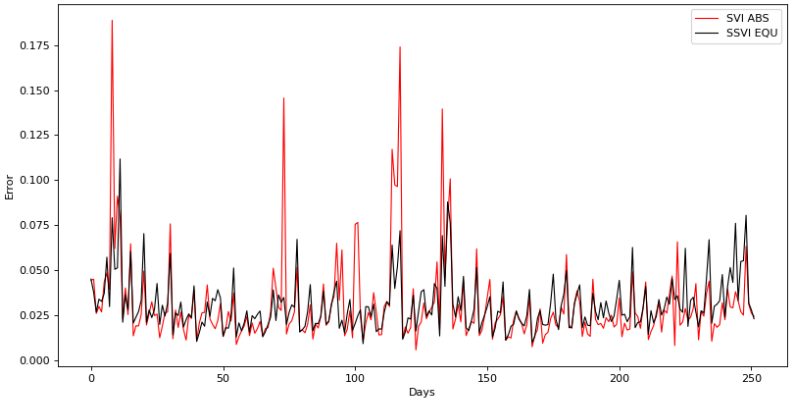
(b) Butterfly Arbitrage

Figure B.6: **Model root-mean-square errors**

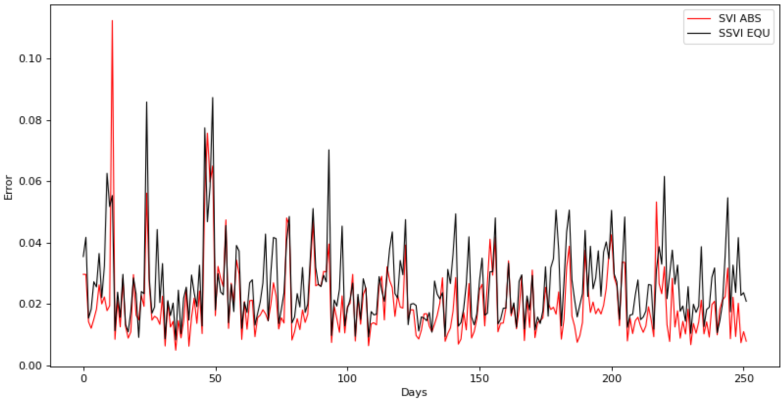
The following figures compare evolution of the optimal weight function error (defined as the function with the lowest total error), following eq.(12) for the two implied volatility models over the course of the 251 trading days of 2019.



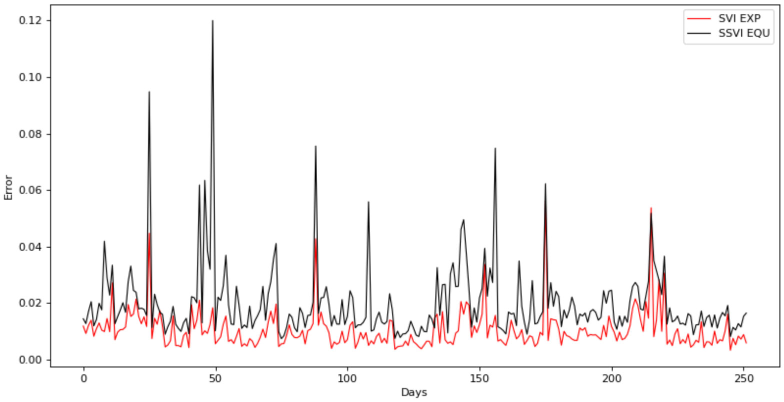
(a) AAPL



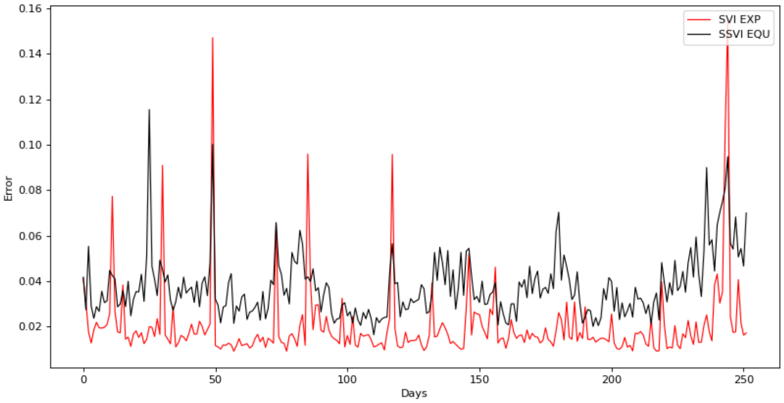
(b) AMZN



(c) BA



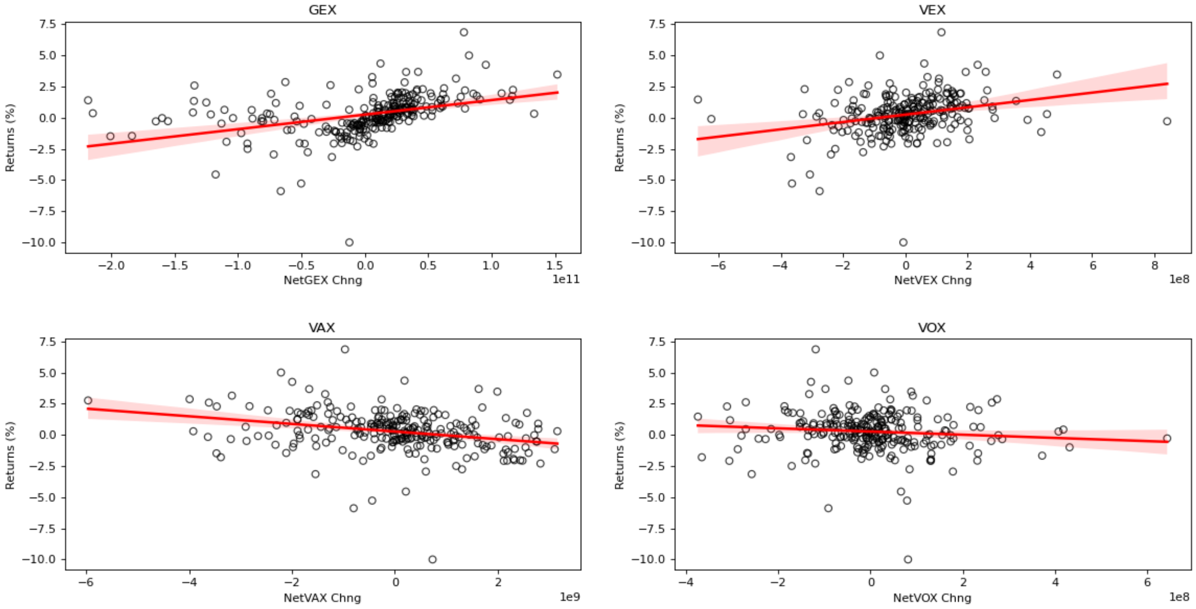
(d) XOM



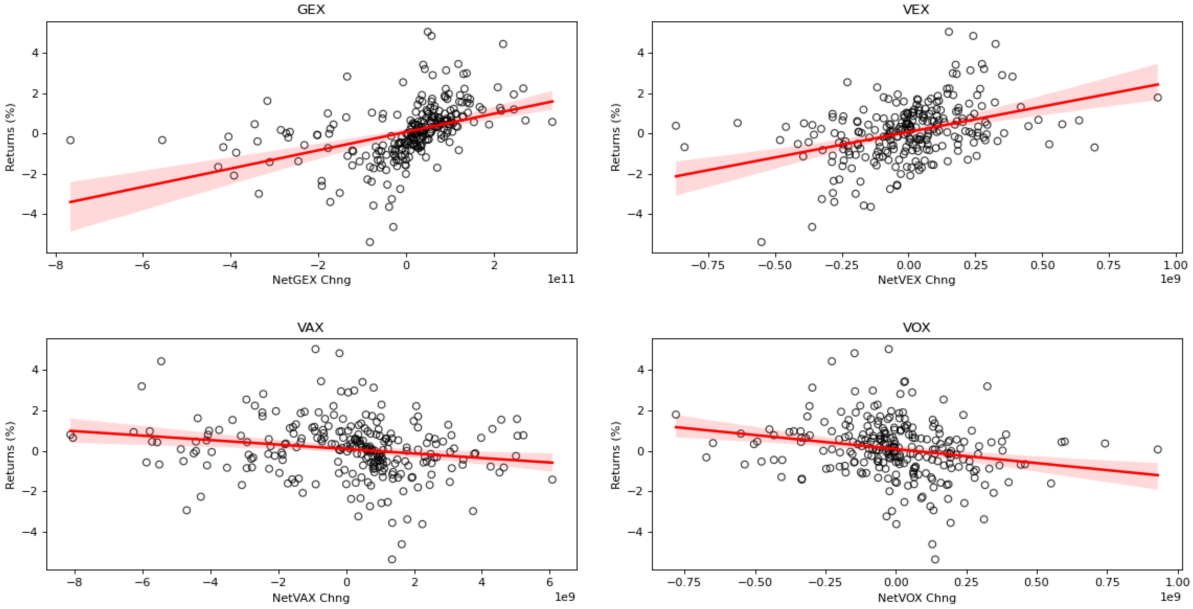
(e) SPX

Figure B.7: **Inventory greeks change vs. Underlying price change**

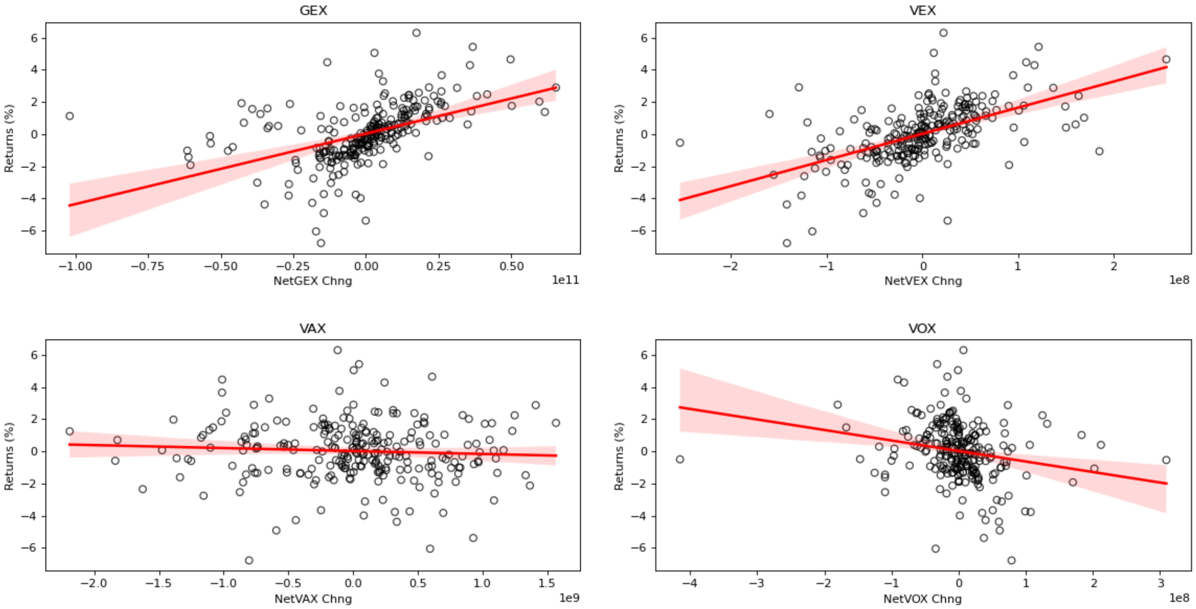
Shows the scatter plot of MM inventory *greeks* exposure absolute change against the underlying asset percent change. The red line represents the linear regression fit. The shaded red area denotes the estimations' 95% confidence interval.



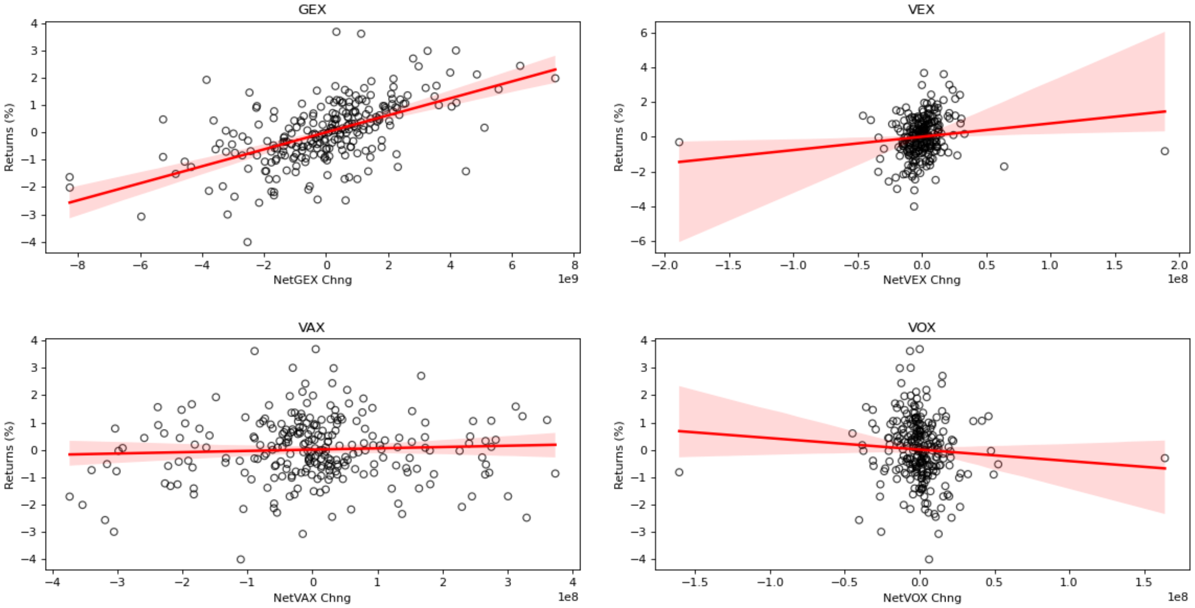
(a) AAPL



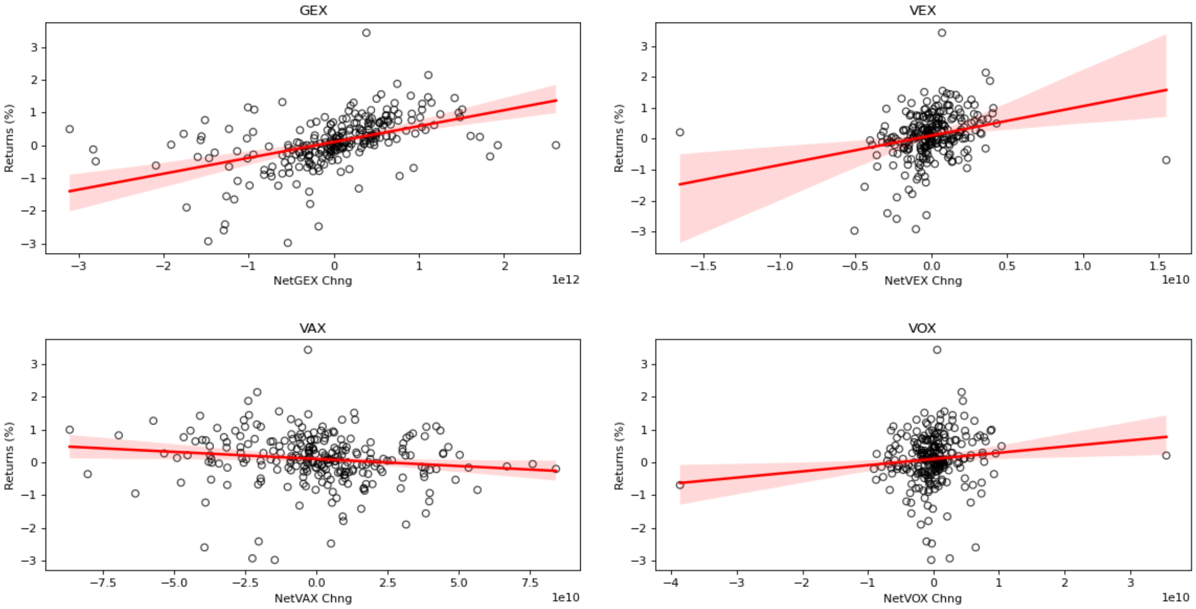
(b) AMZN


GEX
Returns (%)
6
4
2
0
−2
−4
−6
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
NetGEX Chng
1e11
VEX
Returns (%)
−2
−1
0
1
2
NetVEX Chng
1e8
VAX
Returns (%)
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
NetVAX Chng
1e9
VOX
Returns (%)
−4
−3
−2
−1
0
1
2
3
NetVOX Chng
1e8

(c) BA


GEX
Returns (%)
4
3
2
1
0
−1
−2
−3
−4
−8
−6
−4
−2
0
2
4
6
8
NetGEX Chng
1e9
VEX
Returns (%)
6
4
2
0
−2
−4
−6
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
NetVEX Chng
1e8
VAX
Returns (%)
−4
−3
−2
−1
0
1
2
3
4
NetVAX Chng
1e8
VOX
Returns (%)
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
NetVOX Chng
1e8

(d) XOM


GEX
Returns (%)
NetGEX Chng
1e12
VEX
NetVEX Chng
1e10
VAX
NetVAX Chng
1e10
VOX
NetVOX Chng
1e10

(e) SPX

C Tables

Table C.1: **Securities acquired for analysis**

The table shows the breakdown of each analyzed security by ticker, sector and industry.

Asset	Ticker	Sector	Industry
Stocks			
Apple	AAPL	Technology	Consumer Electronics
Amazon.com	AMZN	Consumer Cyclical	Internet Retail
The Boeing Company	BA	Industrials	Aerospace & Defense
Exxon Mobil	XOM	Energy	Oil & Gas Integrated
Index			
S&P 500	SPX	-	-

Table C.2: **Securities returns descriptive statistics**

The table shows the returns descriptive statistics. All securities had a positive cumulative return in the analysis period.

Statistic	AAPL	AMZN	BA	XOM	SPX
Mean	0.26	0.08	0.02	0.00	0.10
Standard Dev.	1.65	1.43	1.82	1.16	0.79
Minimum	-10.00	-5.38	-6.78	-4.01	-2.97
25th Percentile	-0.48	-0.68	-1.03	-0.69	-0.25
50th Percentile	0.30	0.08	0.05	0.00	0.10
75th Percentile	1.16	0.83	1.12	0.75	0.59
Maximum	6.85	5.04	6.29	3.68	4.43
Skewness	-0.91	-0.05	-0.17	-0.01	-0.58
Kurtosis	7.24	1.92	1.76	0.79	3.26

All values (except skewness and kurtosis) should be read in percentages.
Kurtosis is obtained using Fisher's definition.

Table C.3: **Dataset after applying the data processing step**

Description of the fields contained in the final dataset after the initial data processing, along with description and example set.

Variables	Description	Example Set
General		
Trade Date	Date of the transaction.	01/02/2019
Trade Time	Exact time of transaction.	09:30:01.102
Option		
Trade Price	Price at which the option was transacted.	3.45
Trade Size	Volume of the transaction.	12
Option Expiration	Date at which the option expires.	03/15/2019
Type	Call option C or put option P.	C
Strike K	The price paid for the option if exercised.	140.0
Bid price	The highest price a buyer is willing to pay.	3.00
Ask price	The lowest price a seller is willing to accept.	3.50
Log-Moneyness M *[1]	Relative position of strike price and underlying.	0.07
Time to maturity τ *[2]	Time remaining until expiration (Annualized).	0.162
ATM Variance θ_{ATM} *[3]	At-the-money variance.	0.021
IV & Greeks		
IV	Calculated option implied volatility.	0.32
Gamma	Change in delta for $1 increase in the underlying	0.01
Underlying Asset		
Last Price	The quoted market price at transaction time	150.29

*[1] Deducted by $M = K/S$

*[2] Deducted by $\tau = DTE/365$

*[3] Obtained via spline interpolation $\theta_{ATM} = \sigma^2(0, TTM) \times TTM$

Table C.4: **Trading volume by moneyness and maturity location**

Result is showing in what interval of moneyness and maturity the percentage of trading volume is located. *Short* moneyness is defined as the interval (0, 0.05), *Medium* moneyness is defined as the interval (0.05, 0.1) and *Long* moneyness is defined as the interval (0.1, ∞). *Short* maturity is defined as the interval (0, 0.05), *Medium* maturity is defined as the interval (0.05, 0.25) and *Long* maturity is defined as the interval (0.25, ∞).

	Moneyness			Maturity		
Asset	Short	Medium	Long	Short	Medium	Long
AAPL	69.68	15.28	15.04	72.42	19.93	7.65
AMZN	77.53	12.64	9.83	82.53	13.37	4.1
BA	67.81	17.6	14.59	71.34	20.84	7.82
XOM	72.91	16.42	10.66	47.16	35.22	17.62
SPX	50.31	16.83	32.86	58.48	35.28	6.24

All values should be read in percentages

Table C.5: **Comparison of trade prices to prevailing quotes**

This table categorizes all trades of the analyzed assets by comparing the price at which each trade was executed to the bid/ask price of the prevailing quote. The prevailing quote is defined as the market quote immediately before each trade.

Asset & Variable	Description	Count
AAPL		
$P_k > A_k$	Above ask.	23,278 (0.22%)
$P_k = A_k$	At ask.	4,093,314 (39.11%)
$(P_k < A_k)\&(P_k > M_k)$	Above midpoint and below ask.	638,820 (6.11%)
$P_k = M_k$	At midpoint.	1,086,177 (10.38%)
$(P_k > B_k)\&(P_k < M_k)$	Above bid and below midpoint.	633,655 (6.05%)
$P_k = B_k$	At bid.	3,966,308 (37.9%)
$P_k < B_k$	Below bid.	23,763 (0.23%)
Total		10,465,315 (100%)
AMZN		
$P_k > A_k$	Above ask.	36,628 (0.47%)
$P_k = A_k$	At ask.	2,529,962 (32.71%)
$(P_k < A_k)\&(P_k > M_k)$	Above midpoint and below ask.	1,140,027 (14.74%)
$P_k = M_k$	At midpoint.	244,192 (3.16%)
$(P_k > B_k)\&(P_k < M_k)$	Above bid and below midpoint.	1,152,841 (14.91%)
$P_k = B_k$	At bid.	2,592,705 (33.52%)
$P_k < B_k$	Below bid.	38,045 (0.49%)
Total		7,734,400 (100%)
BA		
$P_k > A_k$	Above ask.	8,689 (0.27%)
$P_k = A_k$	At ask.	1,077,832 (33.63%)
$(P_k < A_k)\&(P_k > M_k)$	Above midpoint and below ask.	429,176 (13.39%)
$P_k = M_k$	At midpoint.	152,005 (4.74%)
$(P_k > B_k)\&(P_k < M_k)$	Above bid and below midpoint.	434,680 (13.56%)
$P_k = B_k$	At bid.	1,094,498 (34.15%)
$P_k < B_k$	Below bid.	8,276 (0.26%)
Total		3,205,156 (100%)

Table C.6: **Comparison of trade prices to prevailing quotes** (Continuation of Table C.5)

This table categorizes all trades of the analyzed assets by comparing the price at which each trade was executed to the bid/ask price of the prevailing quote. The prevailing quote is defined as the market quote immediately before each trade.

Asset & Variable	Description	Count
XOM		
$P_k > A_k$	Above ask.	866 (0.15%)
$P_k = A_k$	At ask.	241,657 (41.38%)
$(P_k < A_k) \& (P_k > M_k)$	Above midpoint and below ask.	31,504 (5.39%)
$P_k = M_k$	At midpoint.	42,416 (7.26%)
$(P_k > B_k) \& (P_k < M_k)$	Above bid and below midpoint.	37,874 (6.49%)
$P_k = B_k$	At bid.	228,836 (39.18%)
$P_k < B_k$	Below bid.	857 (0.15%)
Total		584,010 (100%)
SPX		
$P_k > A_k$	Above ask.	3,636 (0.1%)
$P_k = A_k$	At ask.	1,492,900 (41.09%)
$(P_k < A_k) \& (P_k > M_k)$	Above midpoint and below ask.	209,195 (5.76%)
$P_k = M_k$	At midpoint.	226,399 (6.23%)
$(P_k > B_k) \& (P_k < M_k)$	Above bid and below midpoint.	212,576 (5.85%)
$P_k = B_k$	At bid.	1,484,426 (40.86%)
$P_k < B_k$	Below bid.	3,975 (0.11%)
Total		3,633,107 (100%)

D Algorithms

Algorithm 1 Inventory Construction Pseudo-Code

```
Require: $tradesFile$
  $r \leftarrow classificationAlgorithm$
  $inventory[k][e] \leftarrow empty$
  for $trade$ in $tradesFile$ do
      $MM_{position} \leftarrow Classification^{r}_{trade} \times -1 \times Volume_{trade} \times 100$
      if $inventory[k][n]$ is $empty$ then
          $inventory[trade_{strike}][trade_{expiration}] \leftarrow MM_{position}$
      else
          if $m < trade_{expiration}$ then
              DELETE $inventory[.][trade_{expiration}]$
          else
              $inventory[trade_{strike}][trade_{expiration}] += MM_{position}$
          end if
      end if
  end for
```