Tracking Market Gamma

Modeling when and where market makers need to hedge their options positions can explain movements in the underlying stock and give profitable insight to traders

What Is It?

The options market, which allows traders to take bets on underlying stock direction and volatility, is facilitated by option market makers who generate their revenue by providing liquidity. When you buy or sell an option, you're directly transacting with an options market maker. If we think AAPL's earning report is going to be more explosive than everyone expects, we can express our view with options. Options market makers, the party we're transacting with, are in a completely different business. The last thing they want to do is bet on the underlying price movement or volatility- their goal is to provide liquidity and be compensated for it. Whether this be through the bid-ask spread or receiving rebates from the exchanges they make markets on, their business relies on providing liquidity- **not taking proprietary bets.**

Therefore, when we buy an AAPL call option, it should be of no surprise that market makers are interested in hedging this position's risk (they are *short* the call they sold to us). How? With the underlying stock[1]. How much? Pricing models tell them just how much underlying stock they need to buy. This is measured, initially, by the greek letter 'delta'. Delta is the change in option price given a 1 point change in the underlying price. Call option with a delta of 0.5? For every $1 AAPL goes up, the price of this option goes up by $0.50.

Example: If I buy a single call option with a delta of 0.25, I'm long 0.25 delta and the market maker is now short 0.25 delta (I bought what they sold to me). Listed options give the right to buy or sell 100 shares, so the market maker is essentially short 25 shares of stock (0.25*100). If the stock price goes up, they'll lose money on the short call option- and remember- they aren't in the business of taking directional bets like this. If they cannot cross this buy order with another customer's sell, *they will **buy** 25 shares of stock*. Now they are delta neutral. If the price goes up they'll lose money on the options position, but they'll make an equal amount of money on their 25 shares of long stock!

Simple enough? Unfortunately, delta is not a constant. That initial hedge is going to be insufficient as the price moves and the delta moves with it. Luckily, this change is measured by another greek: 'gamma'. Gamma is the change in *delta* given a 1 point change in the underlying.

Back to our example: The market maker is short 0.25 delta and long 25 shares of stock. The gamma of the option is 0.05 (the market maker is short 0.05 gamma). The stock price increases $1. Now, the call option's delta is 0.30 (0.25+0.05). Uh oh, they're short the equivalent of 30 shares via their short call but only holding 25 long shares. They need to go buy 5 more shares to get back to neutral.

If we can measure dealer gamma, we can predict how market makers are going to hedge, and we can determine how these flows might impact the underlying price[2].

Why does this matter? We are skipping ahead a bit, but imagine knowing that market makers are short $170,000,000 worth of AAPL gamma. This means that theoretically, for every $1 increase in AAPL, the 'house' will need to buy $170,000,000 worth of AAPL stock to get back to neutral. At a price of $130, that means market makers are going to need to buy up 1,307,692 shares of AAPL if its price goes up by a single dollar to maintain their hedge. We'll revisit this later in the paper, but hopefully it's becoming clear why knowing this information is valuable to traders.

The Early Days of Measuring Market Gamma (and why it doesn't work):

From page 1, we know that market makers are in the business of providing liquidity, not taking directional risk, and thus hedge out exposure we give them when we trade options. We can measure the initial hedge with the greek delta, but this initial hedge isn't going to cut it. The delta is going to change at a rate labeled 'gamma'. If you can get an accurate read on gamma, you'll have a numerical estimate as to when, where, and how much market makers will need to buy or sell in underlying stock.

Logically this leads to the question: *how do you know where dealers are long and short gamma?*

The answer? You have to measure where dealers are long options and where they are short options. This is our goal and the objective of numerous funds and firms. If you know where dealers are long and where they're short gamma, you know where they'll have to buy stock and when they'll have to sell it!

In the early days of gamma modeling, measures of gamma were solely for the SPX index. I.e. "SPX is long $1,680,142,000 worth of gamma'. These types of metrics were created by using option open interest (the first public model being *"GEX"* published by SqueezeMetrics[3]). The first published 'gamma' metric uses assumptions with option open interest numbers in order to get a 'SPX gamma exposure'. How? Assuming that **all** calls on the SPX are sold by traders and assuming that **all** puts on the SPX index are bought. *We know, it's far from true, but hang in there.*

Mini example: Consider this option chain:

Calls						Puts
Delta	Gamma	Open Interest	Strike	Delta	Gamma	Open Interest
0.43	0.1285	4,589	131	-0.57	0.1321	13,035
0.32	0.1143	9,120	132	-0.68	0.1150	16,240
0.21	0.0912	10,302	133	-0.79	0.0905	20,989
0.14	0.0675	25,285	134	-0.86	0.0659	12,545

So, going off our example, if we assume that all of the calls are sold by traders (dealers are long gamma in calls), we get a call gamma of:

(0.1285*4589)+(0.1143*9120)+(0.0912*10302)+(0.0675*25285) = + 4,278 shares per 1 point move

And a put gamma of:

(0.1321*13035)+(0.1150*16240)+(0.0905*20989)+(0.0659*12545) = - 6,315 shares per 1 point move

For a net 'gamma' of -2,037 shares per 1 point move. We iterate this process for **every** strike price, for **every** expiration, to get a net gamma reading. The basis for the early gamma metrics relies on the assumption of every call being sold (that is, the dealer is long gamma for ALL calls) and every put being bought (dealers being short gamma for every put option).

It turns out these assumptions worked relatively well on the SPX these past few years (data shown on the next page). But first, how do we *know* if it worked well (or didn't)?

Note on gamma measures and volatility: We would expect, if our model was accurate, negative gamma measures leading to more underlying volatility because dealers are buying into up markets and selling into down markets. Positive net gamma environments would result in the opposite- selling into up markets and buying dips, which keeps prices in place.

Here's a handy table to remember the difference:

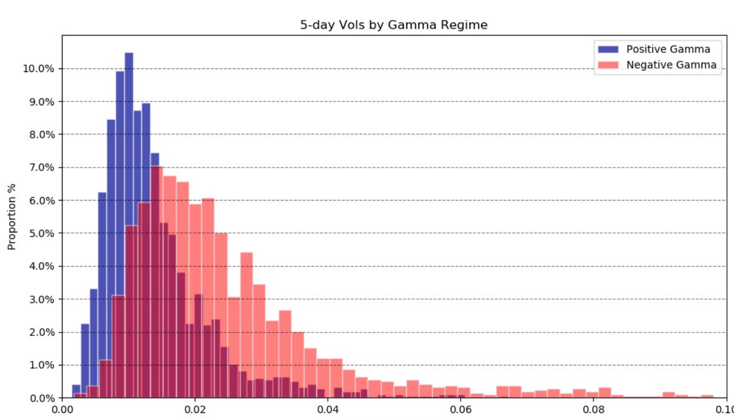
Table 4.1 Execution of Secondary Trades

	Long Gamma	Short Gamma
Buy Orders	Can buy when the market goes lower Posts a bid in the market that can guarantee a trade if the market goes through his level	Needs to buy when the market goes higher Leaves a stop order that becomes a market order triggered by the official printing of a given price
Sell Orders	Can sell when the market goes higher Posts an offer in the market that can guarantee a trade if the market goes through his level	Needs to sell when the market goes lower Leaves a stop order that becomes a market order triggered by the official printing of a given price

Taleb, Nassim

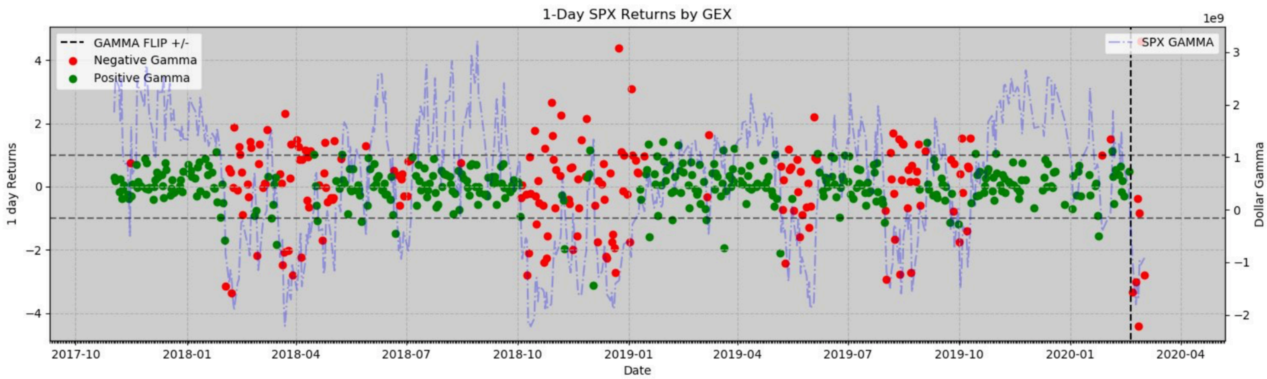
***The entire concept of a 'gamma squeeze' rests on this table.** You buy an OTM call, the dealer hedges and then places stop orders for the stock to BUY at a higher price. As stock price increases, the gamma of your OTM call increases (the closer to at-the-money the call gets, the larger the gamma gets). Thus, larger stop orders above the stock price are placed as more orders fill. Furthermore, gamma increases organically as time to expiration decreases- thus- not only are larger stop orders are submitted, but they are placed more *frequently*, causing a feedback loop as price increases!

The data appears to confirm the intuition. The histogram below shows distributions of 5-day S&P volatility in positive (blue) vs negative (red) gamma environments over the past decade. Notice the substantial "fattening" of the tail in the negative gamma regime (much higher volatility). Not so bad for simple assumptions!



Because a long-gamma market results in dealers selling rallies and buying dips, the realized volatility will be less than a short gamma market. If the model was accurately measuring gamma, we would see realized vol being smaller and tighter (vol of vol) for the long gamma market. At a first glance, this appears to be the case, but unfortunately it's misleading. More on this in a minute.

We also walked day by day from 2017 to the February 2020 sell off to illustrate the differences between positive and negative gamma environments in SPX. Notice the red dots (negative gamma day) resulting in much larger daily moves (Y-axis) than the green (positive gamma day):

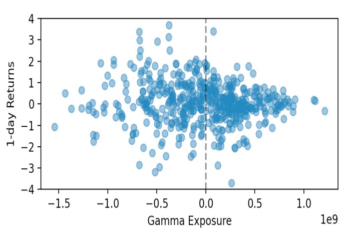


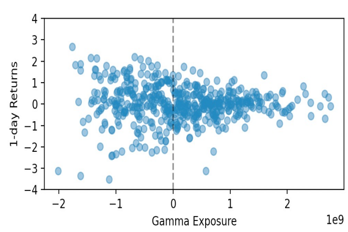
It *looks* compelling. **Great, we're done right? Just assume that all calls are sold and all puts are bought?.**

Unfortunately not- open interest models are inaccurate and unreliable when used as a trading indicator. Furthermore, they are simply useless when applied to individual equities.

Measuring Gamma Beyond Open Interest:

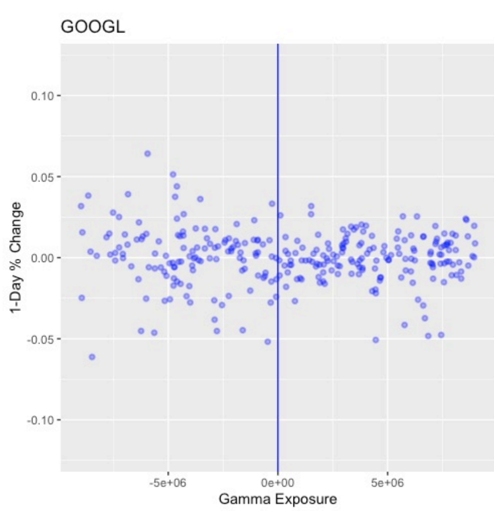
What happens when SPX sees sustained and consistent call buying? What happens when the entire SPX complex is dominated by premium selling? If the assumptions don't hold, the plots aren't as pretty. Check out an SPX plot of 2011-2012 and one of 2015-2016:





With gamma on the x-axis and returns on the y-axis, you'd expect a fan-like pattern: as gamma gets larger, the returns get tighter (long gamma stifles price action), and vice verse. While 2015-2016 looks somewhat convincing, using an open interest model in 2011-2012 would result in inaccurate gamma measures and trading losses.

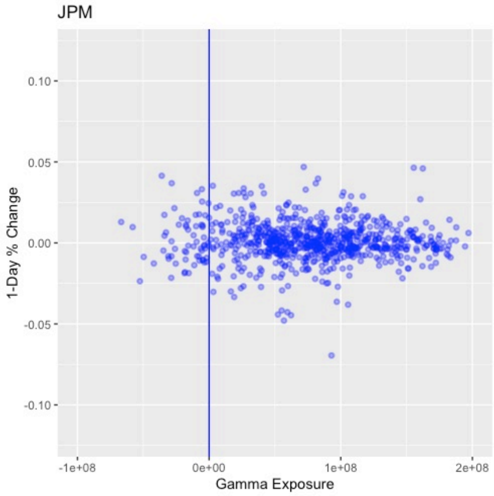
When you apply open interest assumptions to individual stocks, it only gets worse. Check out GOOGL and JPM below, with data from 2011 to today using open interest assumptions.

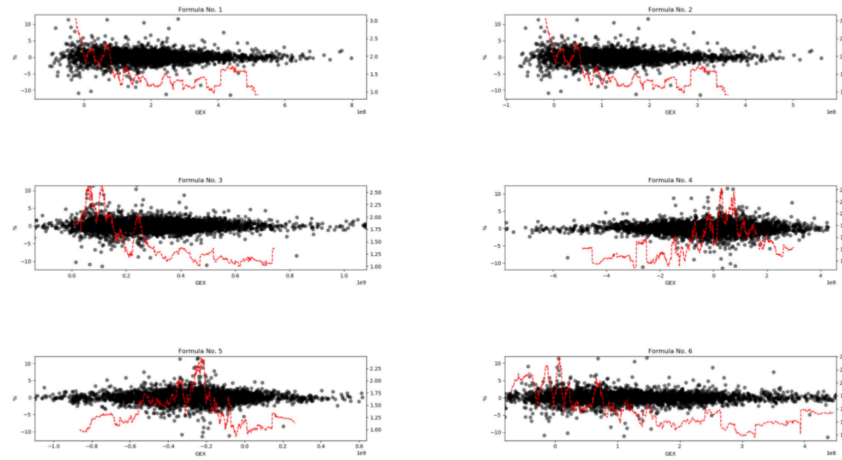


What happens if you tweak the assumptions? What if X% of calls ATM are sold and X% OTM are bought, etc? You could come up with infinite open interest assumption models. We came up with 6 and ran them on a handful of stocks, MSFT pictured below, over nearly 2 decades of data. Again, open interest just doesn't cut it.

X axis: Gamma exposure
Y axis: 1-day returns

An accurate model would produce a fan shape (positive gamma leading to lower volatility):





MSFT Returns with 6 Different Open Interest Models

How to Track Market Gamma Part 1- The Order Book:

Now we dive into the purpose of this paper and the challenge we set out to tackle.

First, a quick three-bullet recap:

- The options market is facilitated by option market makers who make their money by providing liquidity. These market makers facilitate transactions but don't want to take the 'opposite' side of every trade, so they hedge their position's delta exposure. Delta is not a constant, so they have to re-hedge. **This re-hedge is measured by the greek gamma.**

- **We know that dealer hedging flows are significant enough to influence underlying prices**. If we know how price will be influenced- i.e. there is going to be significant buying pressure as the underlying price increases towards some strike or there will be buying flows as it falls (large amounts of support in the market)- we can base trades around this.

- To measure gamma, you need to know where dealers are long options and where they are short. The early days of modeling gamma relied on assuming that groups of open interest were either *all* bought or *all* sold. This is a non-useful assumption as it changes with different 'regimes' (bull markets, bear, etc.). **Any model using open interest for gamma measures, especially on individual stocks, isn't practical or accurate.**

We know gamma is important, but how do we accurately model it?

Back to basics: We need to know where dealers are long options and where they're short options. In other words, we're interested in knowing where customers are short options and where they're long options.

That boils down to: when did someone buy an option, and when did someone sell one.

Open interest gives us no insight, so you'll need to get your hands dirty with transaction data.

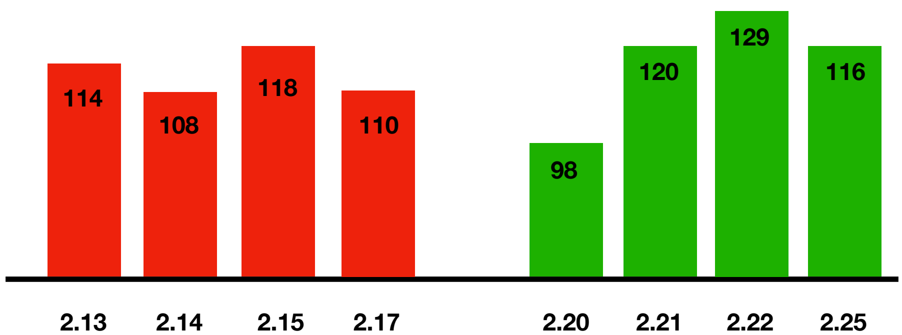
Time	Underlying	Option	Units	Price	Exchange Filled	Exchange A Best Bid	Exchange A Best Ask	Exchange A Best Bid Size	Exchange A Best Ask Size
12:51:01:01	AAPL	April 17 290 Call	300	2.19	A	2.16	2.20	1200	1680
12:51:01:49	AAPL	April 17 290 Call	100	2.20	A	2.17	2.21	1242	1560
12:51:02:00	AAPL	April 17 290 Call	14	2.20	C	2.18	2.22	1300	1710

Above is a small snippet of the option trading tape from an actual AAPL trading day. Of course, you'll want *all* exchange data, for every strike and every expiration, for the underlying.

This data is the foundation for being able to discern buys from sells in the options market.

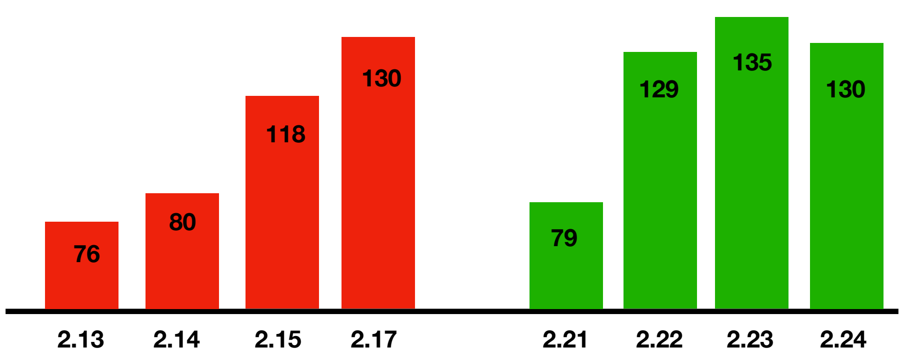
Let's start with our initial framework (one that we originally published in early 2020) of 'flagging'. We flag trades as either buys or sells depending on three things: quoting, quote adjustments, and book liquidity. It sounds complicated, but the underlying logic is plain and simple.

Let's say an order for the April 17 290 call is filled on Exchange A. At 12:51:01:05, the order book for the April 17 290 call looked like this (limit orders on the x-axis, size of the limits on the bars):



Imagine the order fills for 82 units at a price of 2.20. **The best bid** was 2.17 (where we sell and the market maker buys) and **the best ask was 2.20** (where we can buy and the market maker sells). Furthermore, we see how many units are currently bid or offered (we can buy up to 98 contracts at 2.20).

This order *crossed* the spread, and is likely a market buy order (if someone wanted to sell the contract, they'd have to find a buyer and the BEST available was at 2.17- this is where we'd expect the sell to fill). To confirm our suspicions, let's check the order book AFTER the order:

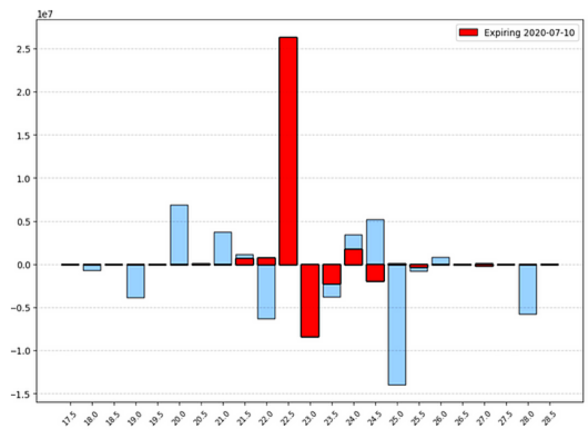


Notice how the new best ask is **lifted from 2.20 to 2.21**, while the best bid remains the same. Furthermore, the contracts offered at 2.21 drops from 120 to 79, indicating less liquidity being sold from market makers (liquidity was likely ‘eaten up’ by a buy order).

Probabilistically, someone bought 82 contracts of the April 17 290 call, and the market maker just sold short 82 contracts of the April 17 290 call. Think back to what we did for the open interest calculations to arrive at a gamma exposure metric:

We calculate the gamma of this contract at the time of the sale, then multiply that number by the 82 contracts and voila- we’ve arrived a measure of how much negative gamma (since this was a buy order) this trade added to a dealer’s book!

Repeat this process for every options trade, keeping a database of running gamma measures, and we can arrive at a net gamma exposure metric for a given security.



Hau Software

With astute data tracking, we’ll know *more* than just an aggregated metric- we know which **strikes** are long and short gamma. That information can be visualized like this. We know most long gamma is concentrated at that single 22.5 strike, resulting in selling anytime it goes above the strike and buying when it goes below it. Remember the table? Dealers have limit sells above 22.5 and limit buys below it! This results in a ‘pinning’, a gravitational pull towards the long gamma strike. This is valuable to traders (selling straddles?)…

Of course, not all trades fill at the best bid or the best ask so cleanly. Here’s a few examples:

- Trades that fill between the spread but take significant liquidity from one side of the book are likely buys or sells (we like to use ratios of the size on the best bid/ask pre-trade, post-trade, and compare them to the trade size).

Ex. Best bid is 1.20 with 1,000 units bid. Best Ask is 1.25 with 500 units offered. An order for 450 units fills at 1.24. The new best bid is 1.21 with the same 1,000 units and the new best ask is 1.25 for 70 units. This is flagged as a buy since significant liquidity is taken from the ask (and the best bid was lifted).

- Trades that fill between the spread but result in a quote adjustment on one side of the book are likely buys or sells.

Ex. Best bid is 1.20 with 1,000 units. Best Ask is 1.25 with 500 units. An order for 450 units fills at 1.24. The new best bid is 1.20, and the new best ask is 1.27. This is flagged as a buy.

We also watch things like lateral adjustments on both sides of the book, trades that look like spreads (call spreads, put spreads), and other parts of the ‘same’ book (22 strike calls affecting 23 strike calls etc). The process can get complicated, but the method is based on the tackling the simple question: is this trade more likely a buy or a sell?

Stepping through each millisecond of the trading day and applying this simple logic results in the foundation for building and tracking dealer gamma exposure.

Alright, but what about orders that are simply 'matched' between the spread? These orders are crossed and thus the dealer holds no risk on their book. And hey, what about those pesky 'flash' orders that appear just to throw off other market makers with no intention of actually trading at the price? What happens when trades happen 'upstairs' and are reported as trades on the tape later than when they actually filled?

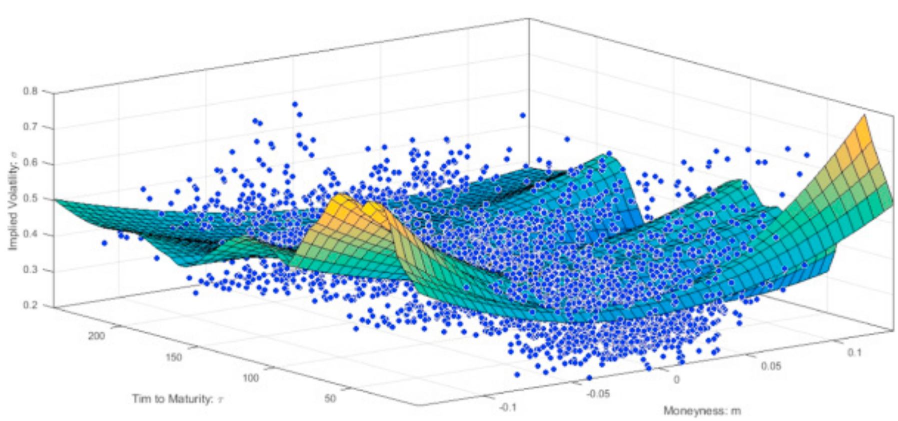
It's impossible to deny that simply looking at quotes to gain insight as to buy and sells may be misleading at times- especially in certain securities. How can we manage this?

How to Track Market Gamma Part 2- Beyond Quotes:

To solve this problem, we have to think about how market makers arrive at their quotes.

Options market makers are in a unique position to quote their option orders in volatility terms. That is, they may believe that AAPL's 290 April 17 Call should be priced at an implied volatility of 25. Thus, they may bid 23 IV and ask 27 IV. Using something like Black-Scholes, we can back out a price for an implied vol of 23 and an implied vol of 27, let's say 2.18 and 2.23. Great! We now have our quotes we can send to the market.

Market makers will fit an implied volatility surface, proprietarily, to model volatility in the underlying (say, AAPL) based on their assumptions and market fills. From this 3-dimensional surface (axis for strike price, axis for expiration, and axis for volatility), they get a bid ask spread for any listed option. Just like how we can see the price at which every trade filled, we can also compute the implied volatility for every trade fill (*think of Black-Scholes and other models as just a way to map implied volatility to price*):

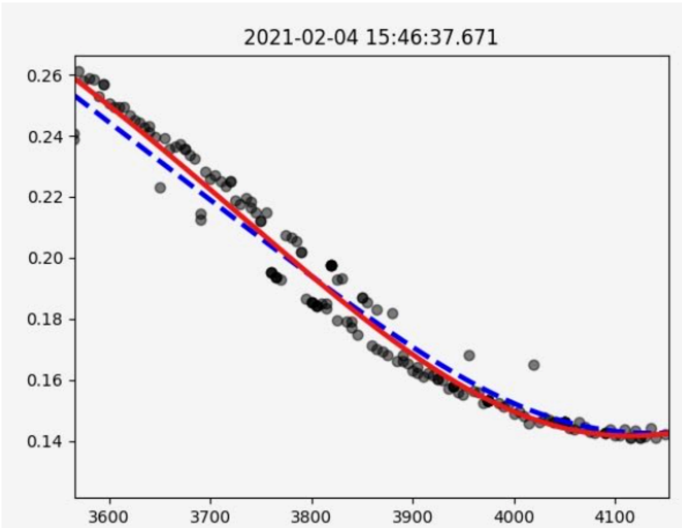


Jinzhong Wang, Shijiang Chen, Qizhi Tao, Ting Zhang

This may look complicated at first, but don't be alarmed. The market maker uses some model to estimate IV (Y-axis) for every time to maturity and every strike as orders come in through the day. Then, they offer to sell above it and buy below it. Imagine the plane to the left as their midpoint IV, and the blue dots as actual fills! Except these fills aren't as a fill price, they're at a fill "IV". Lucky for us, we can back out a price from an IV with a standard pricing formula.

This is a paper for the practical, non-academic trader, so let's stray away from 3-D models as they can cause unnecessary confusion. Let's break it down to 2-D by eliminating the time dimension by plotting the strikes on the x-axis and the IVs on the Y-axis (one 2-D plot for every expiration).

First, we have to model the *dealer's* implied volatility surface. To explain the details of the various models one could use would require an exponential increase this paper's page count. If you'd like



Hau Software

to venture into the world of stochastic volatility models and how to calibrate a volatility surface, feel free, but we'll keep it simple here. As trades come in, we 'fit' a market maker surface. This is an example from SPX call options expiring in two weeks (strikes on the x, implied vol on the Y). The surface is being modeled with the red and blue lines.

If the red and blue lines are the dealer's true surface (where they value IV at), they'll offer to sell above it and buy below it! Just like we flagged quotes, we can flag trades as buys, sells, or an order that is crossed (not adding gamma).

It's that intuitive.

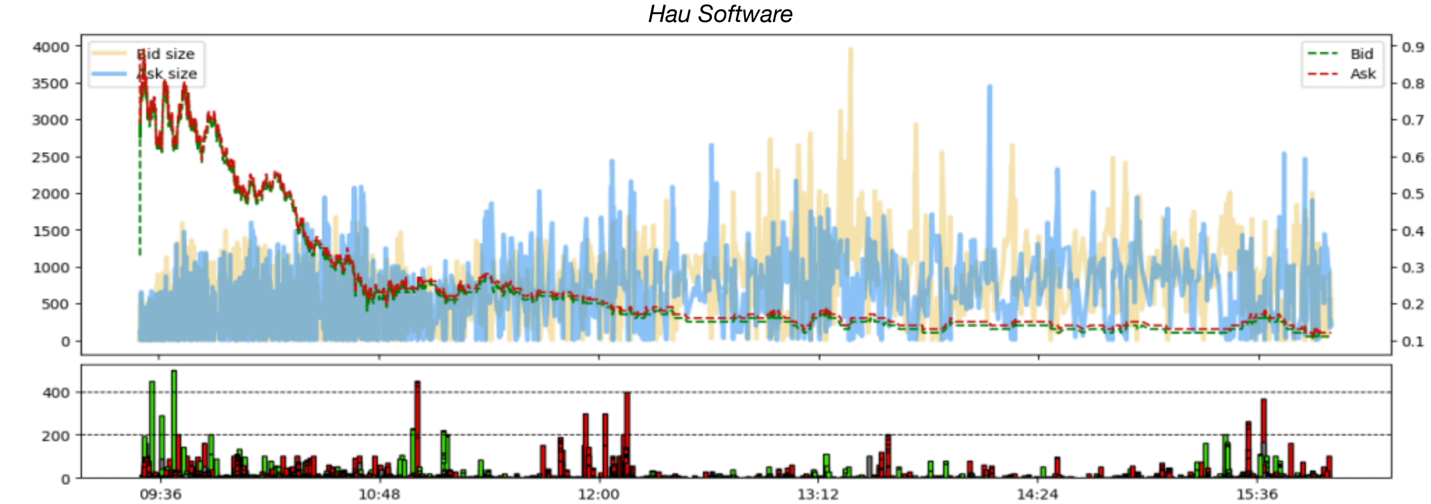
The best part? Where the quotes get misleading, the volatility surface comes in and sheds light. While market makers may flash orders and 'fake' liquidity on AAPL or SPX on the order book, they can't hide their volatility surface. Just like we step through the day with our order book simulator, we can step through the day with the implied volatility surface, calibrating after trades (information) come in to keep a live, up-to-the-millisecond model.

Once you get a good handle on when to flag using the surface and when to flag using the quotes and liquidity on the order book, you can run the process on every trading day.

Here is our program's output for the AAPL October 23, 2020 119 Call. Notice the bottom grid- those are all of this contract's orders for the trading day. Orders we flagged as buys are colored green and sells are red. The insights alone are interesting (large buy orders in the morning, followed by a random batch of selling around lunchtime, etc.). But really, we're interested in gamma- so sum up the gamma for those red orders and put that as long on a dealer's book (someone sold the call, the dealer bought the call), and do the opposite for the green orders.

23 OCT 2020 119.0 C

Hau Software

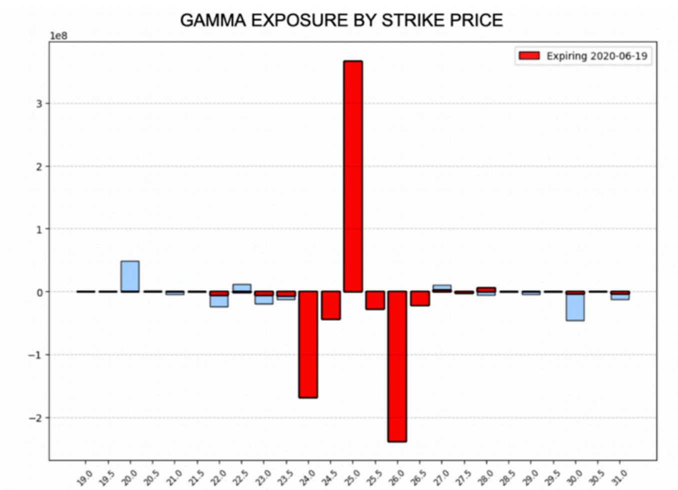


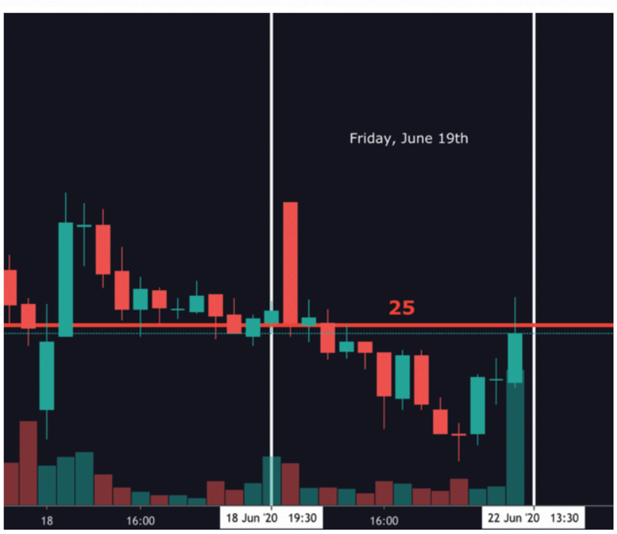
Application and Final Thoughts:

Having an accurate read on a gamma metric for the SPX or an individual stock has seemingly infinite applications- from simply backing out an estimate of volatility to knowing concentrated pin risk. Maybe you sell volatility around strikes with long gamma, or maybe you buy put spreads when you realize gamma has flipped to extremely low levels and price starts to drift lower.

Here's an example: You know that ABC has $100,000,000 worth of long gamma, and most of it is at the 30 strike or above. Furthermore, you know that ABC trades, on average, $1,000,000,000 in daily volume. Let's say that it's 3:00pm and ABC is approaching 30-31. About $500,000,000 worth of volume has traded today. Our gamma measure tells us there is somewhere in the neighborhood of $100,000,000 worth of selling pressure as ABC goes above 30. With oversimplification, we'd estimate that another $500,000,000 at MOST (conditional problems, yeah, we know) worth of trading volume for the day if we assume it'll trade around average volume. Thus, at a minimum, 20% of trading in the next hour will be there solely to sell as ABC's price drifts higher.

Take this actual example from BAC in June 2020: The week of June 19 started on Monday with a move up of 2.40%. Tuesday was up 4%. Wednesday? Down 2.65%. Thursday, as we approached expiration Friday, the price moved a mere 0.20% as BAC became in a long gamma regime.





Here's the gamma profile after the trading day Thursday- loads of gamma at 25 all expiring the next day. Market makers may hedge more aggressively on expiration day, especially close to market close if the underlying price is near their most significant exposure.

What happened on expiration day Friday, June 19th? Price initially fell from 25.50 to the coveted 25 with early morning hedging. The price drifted lower, down to 24.50 on lighter volume (volume bars on the bottom of the dark picture). Then, almost magically, in the last 30 minutes of trading, volume exploded and pushed the price back towards the magic 25, settling on the day at $24.97. This is exactly like our previous example.

Getting an accurate measure is a computationally intensive but not impossible task. Of course, there's never a 'perfect' measure, but probabilistically, if you know which orders are most likely buys, which are most likely sells, and which have been crossed- you'll have an accurate metric and an extremely useful tool to trade.

Questions, comments, concerns?
Email us at info@hauvolatility.com, visit us at hauvolatility.com, or follow us on Twitter: https://twitter.com/HauVolatility

References & Sources:

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

Bouchaud, J., Bonart, J., Donier, J., & Gould, M., Trades, quotes and prices (pp. 45-72) (2018). Cambridge University Press.

Jinzhong Wang, Shijiang Chen, Qizhi Tao, Ting Zhang, Modeling the implied volatility surface based on Shanghai 50ETF options, Economic Modeling, Volume 64, 2017, Pages 295-301, ISSN 0264-9993, https://doi.org/10.1016/j.econmod.2017.04.009.

Pearson, Neil D. and Poteshman, Allen M. and White, Joshua S., Does option trading have a pervasive impact on underlying stock prices? (February 23, 2007). AFA 2008 New Orleans Meetings Paper.

Prior Analytics, LLC (Squeeze Metrics), Gamma Exposure (GEX) [Pdf] (2017). Retrieved from https://squeezemetrics.com/download/white_paper.pdf

Sinclair, Euan, Option trading (pp. 139-147) (2013). Hoboken, N.J.: Wiley.

Taleb, Nassim, Dynamic hedging: managing vanilla and exotic options (pp. 89-94, 132-146) (1997). Wiley.

Notes:

1- Market makers are in the business of providing liquidity to the options market- not taking it. Therefore, hedging with other options is counterproductive to their business. If orders cannot be crossed or the risk cannot be transferred elsewhere, we assume they hedge with the liquid underlying stock.

2- See the above studies by Avellaneda et al. and by Pearson et al. for documentation and published research on how the options market can influence the underlying stock price and volatility.

3- The first public gamma model, relying on simple open interest assumptions for SPX, is detailed in the White Paper referenced above by Prior Analytics.