Aggregating Information in Option Transactions

RICHARD HOLOWCZAK, JIANFENG HU, AND LIUREN WU

In the absence of market frictions and under the geometric Brownian motion stock price dynamics assumed in Black and Scholes [1973] and Merton [1973], options can be perfectly replicated by a portfolio of a risk-free bond and the underlying stock. Options contracts are redundant. In reality, however, the market shows a strong demand for options for several reasons. First, the risks in the stock market cannot be completely spanned by stock trading alone. For example, the presence of discontinuous stock price movements of random size necessitates the inclusion of options across a whole spectrum of strikes to span the jump risk (Carr and Wu [2014]). The presence of stochastic volatility (Engle [2004]), on the other hand, makes the options market the de facto market for trading volatility risk (Carr and Wu [2009]).

Another reason for options trading is informational. Investors may choose to trade options to gain exposure to the stock, because of the high leverage provided by options (Black [1975]). Informed traders may also prefer the options market because they can better hide themselves among the multiple option contracts available on one security (Easley et al. [1998]).

A large number of studies have investigated the information flow between the options market and stock market. One of the key challenges in such studies is how to effectively aggregate the information in the multiple option contracts underlying the same stock. When an investor has private information on a stock, the investor can trade on this information through many combinations of calls and puts across a wide range of strikes.
and maturities. The exposure for market makers is not just the quote size they offer on one particular option contract, but rather the aggregation of the quote size they provide over the whole range of option contracts on the security. Further complicating the issue is the fact that investors can use options to trade on different types of information. For example, investors who predict rising stock prices can take long positions in call options and short positions in put options simultaneously, whereas investors who predict rising volatility can take long positions in both call and put options while minimizing their exposure to the directional movement of the stock.\(^2\) Therefore, finding the appropriate way to aggregate the particular type of information from the many option contracts is critical for market makers, who must determine their quote-setting and updating strategies, and for researchers to develop market microstructure theories in the derivatives market.

Most existing studies either choose one pair of option contracts (e.g., Chan et al. [2002] and Holowczak et al. [2006]) or regard different contracts as equally informative (e.g., Easley et al. [1998], Cao et al. [2005], and Pan and Poteshman [2006]) in inferring the directional movement of the stock price. Picking one pair of contracts while discarding all the others amounts to throwing away a large amount of information, and can potentially distort the estimated relations. One can imagine a case where the chosen option contract has a small transaction, while most other options experience large transactions, pointing to the opposite direction for the stock price movement. In this case, the large transactions of the omitted option contracts, rather than the small transaction of the chosen contract, are likely to dictate the direction of the stock price movement. Equal weighting can be equally problematic, as informed traders do not randomly pick an option contract to trade. Instead, they will consider market depth, liquidity, and leverage to optimize their contract allocation. More recently, Bollen and Whaley [2004] examine the impact of absolute delta-weighted option order flows on implied volatility functions. Ni et al. [2008] use price-scaled, vega-weighted option volumes to predict realized volatilities in the cross section. The rationales for these particular weighting choices are, however, not well understood.

This article proposes a mechanism to align the aggregation of option transactions with the particular type of information that one intends to extract. In extracting the information on stock price movement, the first consideration is the stock price exposure. A call option has positive stock price exposure, and a put option has negative stock price exposure. Accordingly, aggregations of buy and sell orders on call and put options should take on opposite signs. A standard measure for the stock price exposure is the delta of the option, which measures how much the option price moves when the underlying stock price moves by one dollar. Furthermore, when aggregating information, one must be mindful of interference from transactions with other purposes. For example, when an investor buys far out-of-the-money put options, the purpose is more likely to be buying protection against corporate default, in the case of an individual stock, and protection against market crash, in the case of a stock index. Similar considerations apply when extracting information on volatility movement. In this case, the risk exposure can be measured by the vega of the option contract, which measures the option’s price sensitivity to the underlying return volatility movement. Meanwhile, one must also be mindful that short-term and long-term volatilities can be driven by different factors. Vega exposures from short-term and long-term option transactions can be employed for different purposes.

We combine these considerations to generate both an aggregate delta order imbalance (ADOI) and an aggregate vega order imbalance (AVOI) from option transactions. We analyze how the two aggregate order imbalance measures relate to stock return and volatility movements, respectively. We find that the aggregate delta order imbalance is positively correlated with both contemporaneous and future stock returns, but the predictive power declines quickly as the prediction horizon increases. Little predictive power is left after one minute. The aggregate vega order imbalance is also positively correlated with both contemporaneous and future realized return volatilities. Furthermore, the predictive power of the aggregate vega order imbalance lasts much longer (as long as 15 minutes), suggesting that information on stock return dissipates much faster than information on volatility.

Our work contributes to the literature by providing a systematic analysis on the aggregation of option transactions across different strikes and maturities. One cannot possibly obtain robust results on the information flow between the options market and the stock market without first resolving the aggregation issue.
In what follows, after a description of the data and sample properties, we propose delta and vega order imbalance measures to capture the information on the stock price movement and its volatility variation, respectively. Then we report the empirical results regarding the relation between the aggregate delta order imbalance measures and stock returns, followed by the results on the relation between aggregate vega order imbalance measures and realized volatilities.

**DATA**

In the United States, listed options are traded on several exchanges simultaneously. A national market system is established among these exchanges so that option transactions always happen at the best bid and offer across these exchanges. The participant exchanges form a policy committee with representatives from each exchange in the form of the Option Price Reporting Authority (OPRA). The exchanges implement policies and procedures set forth in the OPRA plan. Trade and quote information from the participating exchanges is disseminated to the public through OPRA. In our analysis, the options quote and transaction data are from OPRA. The corresponding trade and quote data on the underlying stock are obtained from the New York Stock Exchange Trade and Quote database.

**Sample Selection and Summary Statistics**

The analysis in this article is based on one underlying security, the NASDAQ 100 index tracking stock (QQQQ), over a span of 231 trading days, from February 1, 2006, to December 29, 2006. Options on QQQQ are listed on all options exchanges in the United States and are among the most actively traded stock options.

The sample contains 1,572,865 trade records on QQQQ options. We filter the trading records by excluding 1) before- and after-market trades, 2) trades that happen within the first 15 minutes of the market open and the last five minutes before the market close, 3) trades flagged as “late” or “cancel” by OPRA, and 4) trades on options that expire within 10 calendar days. The filtering reduces the sample to 1,087,778 trade records. Exhibit 1 reports the summary statistics on the filtered sample. Among the 1,087,778 trades, 506,948 of them (46.6%) are call options, and 580,830 (53.4%) are put options. Trade sizes for put options average higher, at 73 lots per trade, than for call options, at 56 lots per trade. As a result, the average daily trading volume for the put options, at 184,494 lots, represents a larger percentage (60%) of the total trading volume.

When classifying the transactions in terms of the moneyness of the options, Exhibit 1 shows that 44.19% of the transactions have strikes close to the spot, with the absolute delta of options between 37.5% and 62.5%.

**Exhibit 1**

Summary Statistics of the QQQQ Options Trade Sample

<table>
<thead>
<tr>
<th>Statistics</th>
<th>All</th>
<th>Calls</th>
<th>Puts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trade and Trade Size Statistics:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of trades</td>
<td>1,087,778</td>
<td>506,948</td>
<td>580,830</td>
</tr>
<tr>
<td>Average number of trades per day</td>
<td>4,709</td>
<td>2,195</td>
<td>2,514</td>
</tr>
<tr>
<td>Mean trade size</td>
<td>65</td>
<td>56</td>
<td>73</td>
</tr>
<tr>
<td>Median trade size</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Std of trade size</td>
<td>694</td>
<td>729</td>
<td>662</td>
</tr>
<tr>
<td>Mean daily volume</td>
<td>307,170</td>
<td>122,676</td>
<td>184,494</td>
</tr>
<tr>
<td><strong>Percentages of Trade Records Classified into Different Categories:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$37.5% \leq</td>
<td>\delta</td>
<td>\leq 62.5%</td>
<td>44.19</td>
</tr>
<tr>
<td>$</td>
<td>\delta</td>
<td>&lt; 37.5%</td>
<td>37.23</td>
</tr>
<tr>
<td>$</td>
<td>\delta</td>
<td>&gt; 62.5%</td>
<td>18.58</td>
</tr>
<tr>
<td>$T \leq 60$ days</td>
<td>79.91</td>
<td>80.25</td>
<td>79.62</td>
</tr>
<tr>
<td>$T &gt; 60$ days</td>
<td>20.09</td>
<td>19.75</td>
<td>20.38</td>
</tr>
</tbody>
</table>

Notes: Statistics are computed on the filtered sample of trade records on QQQQ options from February 1, 2006, to December 29, 2006. The filtering excludes after-hours transactions, transactions at the first 15 minutes of the market open and last five minutes before market close, trades flagged as “late” or “cancel,” and trades on options with fewer than 10 calendar days before expiry. Delta is computed using the Black-Scholes [1973] formula.
A total of 41.26% of the transactions are out-of-the-money options, with the absolute delta below 37.5%, and only 14.55% are in-the-money options, with the absolute delta above 62.5%. A higher percentage of put options (39.5%) are out-of-the-money, compared to 34.81% for call options.

Across option maturities, 79.91% of the transactions are on short-term options expiring in less than 60 days. The actual percentage is even higher, since we have filtered out options expiring in less than 10 days.

Trade size averages at 65 contracts per trade, with a median of five lots, a minimum of one lot, and a maximum of 275,000 lots. Transactions are heavily concentrated on small sizes. In particular, 394,555 trades (36.27% of the total number) contain only one lot of each trade. We refer to these trades as “odd-lot” trades. Exhibit 2 plots the histogram of the trade size represented in natural logarithms of the number of lots. The highest bar on the left represents the odd-lot trades. Excluding the odd-lot trades, the remaining trade records have an average of 102 lots, a median of 15 lots, and a mode of ten lots.

**Stock Returns and Return Volatilities**

To gauge the effectiveness of the different order imbalance aggregation methods, we measure the correlation between the aggregate order imbalance measures and the stock returns and return volatilities at different leading and lagging horizons.

Stock returns are defined as the difference between the log mid quotes of the National Best Bid and Offer (NBBO):

$$R_{-h,t} = \ln \left( \frac{P_t}{P_{t-h}} \right)$$  \hspace{1cm} (1)

where $P_t$ denotes the midpoint of the NBBO price at time $t$ and $h = 5, 10, 30, 60, 300, 900$ seconds. Realized return volatilities over the same time spans are computed using second-by-second NBBO mid-quote returns:

$$V_{-h,t} = \sqrt{\frac{1}{h} \sum_{i=h+1}^{t-1} R_{t,i}^2}$$  \hspace{1cm} (2)

The second-by-second returns are zero when the NBBOs do not update within the second.

Exhibit 3 reports the summary statistics of the return and volatility estimates over different horizons. The statistics are computed on estimates over non-overlapping samples. Thus, the number of observations, as shown in the second column, decline as the horizon increases. Panel A reports the statistics on the stock returns. The returns average close to zero during the sample period. The standard deviation estimates increase with the return horizon, and translate into about half a basis point per second. The minimum and maximum returns show the possibility of large movements in short sample periods. The last column in Panel B shows that the stock returns computed over the mid quotes exhibit little serial correlation.

---

**EXHIBIT 2**

Histogram of QQQQ Options Transaction Size

Notes: The histogram is on 1,087,778 filtered trade records for QQQQ options from February 1, 2006, to December 29, 2006. Number of trades in millions is plotted against the natural log of the number of lots per trade.
Panel B reports the statistics on the volatility estimates. The mean volatility estimates show an upward-sloping term structure as the estimates become larger over longer horizons. The standard deviations of the volatility estimates become smaller at longer horizons, as a larger number of second-by-second returns are included in computing the return volatilities. The minimum volatility estimates are zero across all horizons from five seconds to 15 minutes, suggesting that there are time periods when the NBBO midpoints are not updated over horizons as long as 15 minutes. On the other hand, the maximum volatility estimate is much higher when the horizon is shorter, suggesting that the volatility estimates over shorter horizons can be much noisier and are more sensitive to a single large price movement.

### AGGREGATING OPTION ORDER IMBALANCE

Aggregating option order imbalance involves four major steps. First, the OPRA database shows only the transaction price and size, but not the information on who initiated the transaction. The first step is to determine the direction of each transaction, i.e., whether the transaction is initiated by a buyer or by a seller. Each transaction is between two counterparties. We regard the initiator as the party who pays the bid–ask spread to enter the transaction. The other party (either a market maker or a limit order provider) receives the spread by providing liquidity to the initiator. Thus, a transaction that happens at the ask is regarded as initiated by the buyer, as the buyer pays more than the mid-quote to enter the transaction, whereas the seller receives more than the mid-quote in entering the transaction. On the other hand, when a transaction happens at the bid, the seller is regarded as the initiator, because the seller receives less than the mid-quote and thus pays a spread over the mid to enter the transaction.

The second step involves aggregating the buy and sell transactions over a certain horizon to come up with an aggregated order imbalance estimate. The imbalance estimate depends on both the aggregation horizon and the weight assigned to each transaction, of possibly different sizes.

Order imbalance calculation for the stock market involves only the two steps just discussed. For the options market, however, one also needs to aggregate the order imbalance across the hundreds of different option contracts on the same underlying stock. We classify this aggregation into two categories: The first is to aggregate the call and the put order imbalance at the
same strike and maturity, and the second is to aggregate order imbalances across different strikes and maturities. We discuss each step in detail in the following subsections.

### Determining the Direction of Each Transaction

We adopt the procedure proposed by Lee and Ready [1991] to determine the direction of each transaction for each option. If the trade price is above the last effective mid quote from the same exchange, it is classified as a buyer-initiated transaction. If the trade price is below the mid quote, it is classified as a seller-initiated transaction. If the trade price falls exactly on the mid quote, but it is higher than the last different trade price, it is classified as buyer-initiated. If a trade price falls exactly on the mid quote and is lower than the last different trade price, it is classified as seller-initiated.

This procedure is able to classify most transactions, leaving only 0.86% of the transactions unclassified. The unclassified transactions normally occur in market opens when there are no valid quotes. We discard these unclassified transactions from our analysis. Lee and Ready [1991] compare the trade price at time \(t\) with the quotes five seconds ago \((t-5)\) to determine the trade direction. The time shift is applied to accommodate potential reporting delays in the transactions. In our application, we compare the trade price at time \(t\) with the most recent quote at time \(t\), thus without applying any time shift. We have experimented with different degrees of time shifts and find that, during our sample period, matching trades with quotes with zero time delay generates the largest proportion of trades at exactly the bid or the ask. Our finding suggests that there are no systematic reporting delays in the stock options market during our sample period.

We compare the price of each transaction with the quotes from the same exchange to determine the direction of the transaction. An alternative is to compare the transaction price with the NBBO to determine the trade direction. However, we find that 85.64% of the transaction prices fall on the bid or the ask of the same exchange, but only 65.51% of the transaction prices fall on the national best bid or offer. The reason for this is that the SEC makes exceptions for trades executed at prices inferior to the NBBO from the order protection rule if that execution trading center displayed an NBBO quote within the previous second.\(^4\) In other words, the benchmark for determining trade-through is the NBBO at the previous second. In a market with fast-moving quotes and insignificant reporting lag, the trade direction can be more effectively identified by matching the transaction to the most recent quote on the same exchange rather than the NBBO price.

The trade direction is slightly imbalanced, with 52.06% of trades initiated by buyers in the entire sample. Call options are more balanced, with 49.32% trades being buyer-initiated, and put options are more imbalanced, with 54.46% trades being buyer-initiated.

### Aggregating Information on Each Contract

For each option contract, we measure the trade imbalance over a fixed time horizon using the concept of a net order imbalance, defined as the buy transactions minus the sell transactions over this time period. We use \(\text{COI}(K,T)\) to denote the net order imbalance from a call option at strike \(K\) and expiry \(T\), and use \(\text{POI}(K,T)\) to denote the net order imbalance of a put option at strike \(K\) and expiry \(T\). Two decisions must be made in computing the net order imbalance: 1) Over how long a time period do we aggregate the transactions? and 2) What is the weight assigned to each transaction of possibly different sizes?

The aggregation horizon choice represents a balance between reducing random noise and catching the information dissemination cycle. On one hand, for sparsely traded contracts, aggregating over a short horizon can generate very noisy estimates, as very few transactions happen within the short time interval. Increasing the aggregation interval will include more transactions from both sides into the calculation and can thus result in a smoother estimate for the trade imbalance. On the other hand, when trading frequency is high and information disseminates very quickly, the order imbalance can disappear quickly. Therefore, one will need a shorter horizon to capture the order imbalance. Taking both sides into consideration, the aggregation horizon should be long enough to include a reasonable number of transactions, but short enough to reveal the dissemination process of an information event.

This article considers the aggregation of net order imbalances over different horizons from five seconds to 15 minutes. By investigating the information flow at
different horizons, we can infer the speed of information dissemination on the security. Nevertheless, our horizon choice dictates that our analysis focuses on the short-term (intraday) information flow on returns and volatilities, rather than information flow over longer horizons (such as months).

To aggregate different transactions over the chosen horizon, the literature mostly considers two weighting schemes. One is to weigh all transactions equally regardless of trade size, with the order imbalance representing the net number of buy transactions over the chosen horizon. The other is to weigh each transaction by the size of the transaction, with the order imbalance essentially capturing the net buying volume over the chosen horizon.

Several empirical studies on stock market microstructure have found that the number of trades can be more informative than trade volume, e.g., Jones et al. [1994], Ané and Geman [2000], and Izzeldin [2007]. However, using number of trades can overstate the importance of very small trades. In the stock options market, odd-lot trades are generally considered as uninformative retail trades. Indeed, on some options exchanges such as the International Securities Exchange, a large proportion of the odd-lot trades are rewarded to the primary market makers as a compensation for their extra responsibilities (Simaan and Wu [2007]). On the other hand, using volume weighting may overstate the importance of very small trades. In the stock options market, very large trades are often negotiated in the upstairs market and are put into print at a later time. As a result, the reported large trade tends to be considered a lagged report and is not as informative about the current market. The same practice also happens on the options exchanges.

In this article, we balance the different considerations by applying a concave function on the trade size. Specifically, to aggregate the order imbalance, we use the natural logarithm of the trade size as the weighting so that 1) we assign zero weight to the odd-lot transactions, since they are commonly regarded as uninformative, and 2) the weight increases with the trade size, but only concavely, so that we do not overweight very large transactions. Our empirical experiments show that the two extreme aggregating methods using number of trades and volume always underperform log volume weighting. Therefore, in this article we report results only for log volume weighting.5

There are two option contracts at each strike-maturity point: one call option and one put option. The two contracts have the opposite stock price exposure but the same volatility exposure. Therefore, if the aggregation is meant to extract information about the underlying stock price movement, one should assign positive weights to the net order imbalance of the call option and negative weights to the net order imbalance of the put option, so that the aggregated imbalance reflects the risk exposure on the underlying stock price. On the other hand, if the aggregation is meant to extract information about the underlying volatility variation, one should assign positive weights to the net order imbalance of both the call option and the put option, as both contracts have positive exposures to volatility.

In the option-pricing literature, the stock price sensitivity of an option is measured by delta, the partial derivative of the option value with respect to the underlying stock price. The sensitivity of an option to volatility is measured by vega, the partial derivative of the option value with respect to return volatility. In aggregating information from the call and the put options at the same strike and maturity, we propose to use the two sensitivity measures as the weight in constructing two types of order imbalance, the delta order imbalance (DOI) and the vega order imbalance (VOI), formally defined as

\[
DOI(K,T) = N(d)COI(K,T) - (1 - N(d))POI(K,T)
\]

\[
VOI(K,T) = n(d)\sqrt{T}[COI(K,T) + POI(K,T)]
\]

where \(N(d)\) and \(n(d)\) denote the cumulative density and probability density of a standard normal variable, respectively, and

\[
d = \frac{\ln(F/K) + \frac{1}{2} \sigma^2 T}{\sigma \sqrt{T}}
\]

with \(F\) denoting the forward price and \(\sigma\) being some volatility estimate for the underlying stock. In implementation, we use an average implied volatility estimate from the previous day to proxy \(\sigma\). We set the interest rates and dividend yields to zero and normalize the stock price level to one in computing the delta and vega weights. By aggregating the order imbalance based on the delta exposure, Equation (3) measures the aggregate information regarding the stock price movement. Thus, we also
call the delta order imbalance the stock order imbalance. On the other hand, through vega weighting, Equation (4) measures the aggregate information regarding the stock return’s volatility. As such, the vega order imbalance is essentially the volatility order imbalance. We use the terms interchangeably, with delta and vega highlighting the sensitivity measures, whereas stock and volatility highlight the information target.

A stock transaction can express a view only on the directional movement of the stock, with a buy to express the view of future stock price going up and a sell to express the view of future stock price going down. By contrast, an option transaction can express views not only on the directional movement of the stock price, but also on the variation of volatility. The aggregation method applied to the option order imbalance depends crucially on what information one strives to extract from the option transactions. This article represents the first effort to formalize this idea through the proposed risk exposure weighting in aggregating call and put order imbalances in Equations (3) and (4).

Aggregating Information Across Strikes and Maturities

To aggregate the stock and volatility information across different strikes and maturities, one must also consider the potential interference from other risk dimensions.

Most options transactions are concentrated at short maturities and at strike prices close to the spot level. These options contracts tend to have narrower bid–ask spreads. Thus, everything else being equal, informed traders are likely to allocate more of their capital to the most actively traded options to mitigate market impact and to reduce transaction cost. On the other hand, when transactions happen at deep out-of-the-money regions (strikes far away from the spot) or at very long maturities, where the transaction costs are much higher, the motivations can be different from gaining short-term stock or volatility exposures. For example, deep out-of-the-money index puts are often bought for protection against market crashes rather than for short-term market movements, and long-term contracts are often traded for hedging even longer-term corporate structural deals, which may have little bearing on the short-term stock volatility movements.

We propose four weighting schemes across strikes and maturities that put different emphasis on these considerations:

1. **Greek weighting (GK)**, which ignores the other considerations and regards the delta and vega exposure as the only relevant consideration in aggregating the order flows. The exposure-based stock and volatility imbalances defined in Equations (3) and (4) are simply aggregated across different strikes and maturities without other weighting.

   \[
   ADOI_{GK} = \sum_{j=1}^{N} DOI(K_j, T_j),
   \]

   \[
   AVOI_{GK} = \sum_{j=1}^{N} VOI(K_j, T_j)
   \]

   where \( N \) denotes the number of strike-maturity points.

2. **Maturity discount (MD)**, which assigns less weight to order imbalances computed from long-term options.

   \[
   ADOI_{MD} = \sum_{j=1}^{N} e^{-M_j T_j} DOI(K_j, T_j),
   \]

   \[
   AVOI_{MD} = \sum_{j=1}^{N} e^{-M_j T_j} VOI(K_j, T_j)
   \]

   where \( M_j = \max(1, T_j \times 12) \) measures the maturity in months and floors the minimum maturity to one month. According to this weighting, order imbalances from options with maturities of one month or shorter have a weight of one, but order imbalances from longer-term options are discounted with a smaller weight, with the weight declining exponentially with increasing maturity.

3. **Strike discount (KD)**, which assigns less weight to order imbalances computed from options with strikes far away from the spot.

   \[
   ADOI_{KD} = \sum_{j=1}^{N} e^{-d_j T_j} DOI(K_j, T_j),
   \]

   \[
   AVOI_{KD} = \sum_{j=1}^{N} e^{-d_j T_j} VOI(K_j, T_j)
   \]
where the weight is one for at-the-money options 
\(d = 0\) and declines as the standardized moneyness 
\(d\) increases in absolute magnitude.

4. **Maturity and strike discount (MK),** which discounts across both the strike and the maturity dimension.

\[
ADOI_{MK} = \sum_{j=1}^{N} e^{-\frac{r}{2} \left[ \frac{M_j}{M} \right]^2} DOI(K_j, T_j),
\]

\[
AVOI_{MK} = \sum_{j=1}^{N} e^{-\frac{r}{2} \left[ \frac{M_j}{M} \right]^2} VOI(K_j, T_j)
\]

where the weight is one for at-the-money options 
\(d = 0\) at one month or shorter maturities, and declines as the standardized moneyness 
\(d\) increases in absolute magnitude and as the option maturity increases.

For comparison, we also document the effectiveness of two aggregating methods employed in the literature:

1. **One pair (OP),** which considers only the order imbalance from one strike-maturity point,

\[
ADOI_{OP} = \sum_{j=1}^{N} 1_{t_t, t_{M_j}, M_j} \left[ DOI(K_j, T_j) - POI(K_j, T_j) \right],
\]

\[
AVOI_{OP} = \sum_{j=1}^{N} 1_{t_t, t_{M_j}, M_j} \left[ DOI(K_j, T_j) + POI(K_j, T_j) \right]
\]

where \(1_{t_t, t_{M_j}, M_j}\) denotes a weight of one for the strike-maturity point closest to one month at-the-money and zero for all other strikes and maturities.

Instead of weighting the order imbalances using option sensitivities, this method treats calls and puts at the same strike and maturity as having opposite implications on the underlying stock price, and the same implication on the underlying return volatility. This extreme weighting of picking only one put-call pair is used in several studies, e.g., Chan et al. [2002].

2. **Equal weighting (EQ),** which considers all option contracts but regards them as equally informative.

\[
ADOI_{EQ} = \sum_{j=1}^{N} \left[ DOI(K_j, T_j) - POI(K_j, T_j) \right],
\]

\[
AVOI_{EQ} = \sum_{j=1}^{N} \left[ DOI(K_j, T_j) + POI(K_j, T_j) \right]
\]

This method does not consider sensitivity weighting either, and has been used in many studies, including Easley et al. [1998] and Cao et al. [2005].

**DELTA ORDER IMBALANCE AND STOCK RETURNS**

For both stock order imbalance and volatility order imbalance, we consider six weighting methods across different strikes and maturities (OP, EQ, GK, MD, KD, MK). For each measure, we consider six horizons \((h)\) at 5, 10, 30, 60, 300, and 900 seconds. We gauge the effectiveness of the various stock order imbalance measures in terms of both their contemporaneous impacts on the stock returns and their predictive power on future stock price movement. The contemporaneous impact analysis reveals the depth of the stock market as it measures how much the trade imbalance can move the stock market. By contrast, the predictive analysis explores whether there is extra information in the aggregated stock order imbalance measures that has not yet been fully revealed in the current stock price, but will show up in subsequent stock price movements.

**Contemporaneous Impacts**

Panel A of Exhibit 4 measures the contemporaneous correlation between the six measures of aggregate stock order imbalance across six different horizons and the stock returns over the same horizon. Under each measure, the contemporaneous correlation estimates increase with the aggregation horizon, showing the fact that aggregating over longer horizons removes more of the measurement noise due to discreteness in trading.

Among the six measures, one pair generates the weakest correlation estimates, highlighting the importance of aggregating option transactions across all strikes and maturities. Compared to equal weighting of the net volume across all strikes and maturities, aggregating delta exposure always increases the correlation estimates. This effect is stronger in short horizons, and
the difference is reduced when the observation length increases. Further discounting delta exposure along the maturity dimension reduces the correlation estimates, but applying discounting along the strike dimension can increase the correlation estimates when the aggregation horizon is longer than 60 seconds.

Taken together, the correlation estimates suggest that to capture the strongest contemporaneous impact, the aggregate stock price information should be measured in delta exposure across all contracts and discounted along the strike dimension but not along the maturity dimension. At the 15-minute horizon, the correlation estimate between stock returns and the aggregate delta order imbalance with strike discounting reaches 72.8%.

**Forecasting Relation**

Panel B of Exhibit 4 measures the forecasting correlation between the aggregate delta order imbalance measures and future stock returns with matching future horizons. As expected, the forecasting correlations are much lower than the contemporaneous correlation estimates. Furthermore, different from the contemporaneous correlation, the forecasting correlation estimates are the most positive at the shortest horizons. The estimates decline as the horizon increases, and most estimates become statistically insignificant when the horizon is longer than one minute.

Comparing the forecasting correlation estimates across different measures, we find again that the aggregate delta order imbalance measure with strike discounting generates the highest forecasting correlation, which starts at 2.7% at a 5-second horizon, and declines to 2.1%, 1.6%, and then 0.9% at 10-, 30-, and 60-second horizons, respectively. The forecasting correlation estimates at even longer horizons are no longer significant statistically.

Similar to the results for contemporaneous correlations, picking one pair generates the weakest forecasting power. Equal weighting of net volume across all contracts generates weaker correlation estimates than delta weighting. Strike discounting helps improve the predictability of the measure slightly, whereas maturity discounting reduces the performance.

Exhibit 4 matches the return forecasting horizon with the aggregation horizon for the order imbalance. This matching is not particularly necessary. Exhibit 5 focuses on the best-performing aggregate delta order imbalance with strike discounting as an example, and reports the forecasting correlation estimates at different aggregation horizons (p) and forecasting horizons (h). The estimates show that the forecasting power is strongest when both the aggregation horizon and the forecasting horizon are short. The predictability declines quickly as either the aggregation horizon or the forecasting horizon increases. This finding suggests that information about the stock price movements dissipates very fast.
VEGA ORDER IMBALANCE AND RETURN VOLATILITIES

Similar to the analysis on stock order imbalance measures, we analyze the effectiveness of vega order imbalance measures in terms of their contemporaneous and forecasting relations with realized return volatilities.

Contemporaneous Impacts

Panel A of Exhibit 6 reports the contemporaneous correlation estimates between the aggregate vega order imbalance measures and the stock return volatilities over the same horizon. The contemporaneous correlation estimates are much smaller than those between stock returns and stock order imbalances. The estimates are virtually zero at short horizons. Only at horizons longer than one minute do we observe significant and consistent correlation estimates across the different vega order imbalance measures.

The insignificant correlation estimates at shorter horizons are a combined result of several considerations. First, as in the case of stock order imbalance, the vega order imbalance estimates are noisy at short horizons. Second, the volatility estimator is also noisier and less accurate at a shorter horizon because fewer second-by-second returns are included in its calculation. Third, given the larger transaction cost on options compared to the transaction cost on the underlying stock, information in return volatility disseminates slower than that in

E X H I B I T  6
Linking Aggregate Volatility Order Imbalance to Contemporaneous and Future Stock Volatilities

| Panel A: Contemporaneous Correlations, Corr($ADOI_{\Delta t}$, $R_{\Delta t}$) |
|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | OP              | EQ              | GK              | KD              | MD              | MK              |
| 5                | 0.001           | -0.002***       | -0.005***       | -0.005***       | 0.000           | 0.000           |
| 10               | 0.001           | 0.001           | -0.004***       | -0.004***       | 0.003**         | 0.003**         |
| 30               | 0.066***        | 0.006**         | -0.002          | -0.002          | 0.011***        | 0.011***        |
| 60               | 0.011***        | 0.013***        | 0.005           | 0.005           | 0.019***        | 0.019***        |
| 200              | 0.030***        | 0.054***        | 0.038***        | 0.039***        | 0.062***        | 0.062***        |
| 900              | 0.049***        | 0.079***        | 0.060***        | 0.062***        | 0.091***        | 0.090***        |

| Panel B: Forecasting Correlation, Corr($ADOI_{\Delta t}$, $R_{\Delta t}$) |
|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | OP              | EQ              | GK              | KD              | MD              | MK              |
| 5                | 0.000           | 0.003***        | 0.002*          | 0.002*          | 0.004***        | 0.004***        |
| 10               | 0.001           | 0.003**         | 0.001           | 0.002           | 0.005***        | 0.005***        |
| 30               | 0.066***        | 0.011***        | 0.005**         | 0.005**         | 0.012***        | 0.012***        |
| 60               | 0.008**         | 0.019***        | 0.012***        | 0.012***        | 0.021***        | 0.021***        |
| 200              | 0.018**         | 0.035***        | 0.027***        | 0.028***        | 0.037***        | 0.037***        |
| 900              | 0.037***        | 0.060***        | 0.041***        | 0.042***        | 0.066***        | 0.066***        |

Notes: Entries report the contemporaneous correlation in Panel A and predictive correlation in Panel B between stock return volatilities and aggregate volatility order imbalance measures. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.
AGGREGATING INFORMATION IN OPTION TRANSACTIONS

S PRING 2014

stock price. Only over a longer horizon does one observe contemporaneous co-movements between vega order imbalance and realized volatility.

At longer horizons, the contemporaneous correlation estimates become significantly positive. Among the six different measures, one pair (OP) generates the lowest correlation estimate, again showing the value of incorporating all option contracts. Compared to equal weighting of volume across contracts, vega weighting reduces the correlation estimates. However, maturity discounting on vega weighting significantly increases the correlation estimates and generates the largest correlation estimates at 5-minute and 15-minute horizons. Strike discounting has a negligible effect.

Interestingly, maturity discounting reduces the correlation between delta order imbalance and stock returns, but increases the correlation between vega order imbalance and volatility. On the other hand, strike discounting has a more beneficial effect on stock order imbalance than on volatility order imbalance.

Intuitively, one can buy options at different strikes to target stock price movements of different magnitudes. Options with strikes close to the spot have concentrated exposures to short-term, small market movements, whereas options with extremely low strikes serve more as an insurance against rare, but large, market disruptions. The weight discounting along the strike dimension makes the stock order imbalance more focused on strikes close to the spot and thus generates more concentrated exposures to short-term stock price movements. As a result, its correlation estimates with short-term stock returns are the highest.

By contrast, return volatility movements can show distinct short-term and long-term volatility movements, with transactions on short-term options revealing more of short-term volatility movements and transactions on long-term options revealing more of long-term volatility movements. Therefore, discounting along the maturity dimension focuses the volatility order imbalance more on short-term volatility movements, and accordingly enhances the correlation with short-term realized volatilities.

Forecasting Relation

Panel B of Exhibit 6 reports the forecasting correlation between aggregate vega order imbalance measures and future return volatilities of the same horizon. The estimates show that vega order imbalance with maturity discounting not only generates the highest contemporaneous correlation with return volatility, but also shows the strongest predictability about future volatility movements.

Exhibit 7 focuses on the best-performing aggregate vega order imbalance with maturity discounting as an example, and reports the forecasting correlation estimates at different aggregation horizons (p) and forecasting horizons (h). The estimates show that aggregating the vega order imbalance over a short time interval (such as five seconds) is not going to generate meaningful order imbalances. Although the correlation can be statistically significant, it is close to zero. By contrast, if we aggregate the vega order imbalance over a longer period, such as 15 minutes, the order imbalance is informative in pre-

**EXHIBIT 7**

Linking Aggregate Volatility Order Imbalance to Return Volatilities over Different Horizons

<table>
<thead>
<tr>
<th>Forecasting Horizon h:</th>
<th>5</th>
<th>10</th>
<th>30</th>
<th>60</th>
<th>300</th>
<th>900</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregation Horizon p:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.003***</td>
<td>0.004***</td>
<td>0.005***</td>
<td>0.006***</td>
<td>0.007***</td>
<td>0.007***</td>
</tr>
<tr>
<td>10</td>
<td>0.004***</td>
<td>0.005***</td>
<td>0.007***</td>
<td>0.008***</td>
<td>0.010***</td>
<td>0.009***</td>
</tr>
<tr>
<td>30</td>
<td>0.008***</td>
<td>0.009***</td>
<td>0.012***</td>
<td>0.014***</td>
<td>0.016***</td>
<td>0.016***</td>
</tr>
<tr>
<td>60</td>
<td>0.010***</td>
<td>0.013***</td>
<td>0.017***</td>
<td>0.021***</td>
<td>0.022***</td>
<td>0.022***</td>
</tr>
<tr>
<td>300</td>
<td>0.022***</td>
<td>0.027***</td>
<td>0.034***</td>
<td>0.038***</td>
<td>0.042***</td>
<td>0.045***</td>
</tr>
<tr>
<td>900</td>
<td>0.033***</td>
<td>0.040***</td>
<td>0.052***</td>
<td>0.058***</td>
<td>0.071***</td>
<td>0.081***</td>
</tr>
</tbody>
</table>

Notes: Entries report the forecasting correlation between the aggregate volatility order imbalance with maturity discounting over different aggregation periods (p) and stock return realized volatilities over different future horizons (h), Corr(AVOI<sub>p,h</sub>, \( V_{t+h} \)). *** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.
dicting return volatilities over all considered horizons from five seconds to 15 minutes.

CONCLUSION

In the stock market, researchers have found that stock order imbalance, defined as the difference between buy and sell transactions over a certain period of time, contains important information about current and future stock price movements. In this article, we explore the information content of order imbalances on stock options. The analogy between the two markets, however, stops here, because of the hundreds of options series at different strike prices and maturities that trade on each stock. To extract information on the underlying stock price, we must find an appropriate way to aggregate the transactions on the hundreds of option contracts based on the same stock.

Furthermore, investors can gain stock exposure only by trading a particular stock, but they can gain many different types of exposure by trading options. Through options trading, investors can gain exposures not only on the stock price movement, but also on the stock volatility. By appropriately positioning the option strikes and maturities, they can also gain other types of exposures, such as credit risk on a company or crash risk on the market, and they can further distinguish between exposures on short-term versus long-term volatilities. These different types of possible exposures present both a challenge and an opportunity in the order aggregation. The challenge is that when one tries to extract information about a particular exposure (say, stock price movement), one must be mindful of the potential interferences from trades for other purposes. Only after one controls for the trades for other purposes can one effectively extract the information on one particular dimension. On the other hand, the options market also provides unique opportunities, as one can extract certain types of information, such as volatilities, from options transactions that are not as easily extractable from stock market transactions alone.

This article focuses on information in option transactions related to the stock price movement and the volatility movement and systematically investigates different considerations in the order aggregation. The analysis shows that an effective aggregation method must account for each contract’s different exposure to the stock price and volatility movements, and accommodate concerns on interference from other potential risk dimensions, such as market crashes and long-term versus short-term volatility factors. The article identifies significant relations, both contemporaneous and predictive, between the appropriately aggregated options order flow and the stock return and the return volatility.

ENDNOTES

The authors thank Stephen Figlewski (the editor), Turan Bali, Armen Hovakimian, Neil Pearson, Henry Schwartz, Robert Schwartz, and seminar participants at Baruch College, the 2010 Northeast Business and Economics Association(NBEA) annual conference, the 2011 Financial Management Association (FMA) annual conference, the 21st FDIC-Cornell-University of Houston Derivative Securities and Risk Management Conference, the SMU-ESSEC Conference on Empirical Finance and Financial Econometrics, and the Fifth Annual Risk Management Conference (Singapore) for comments. Jianfeng Hu thanks the CUNY Graduate Center for graduate research grant. All remaining errors are ours.

1See, for example, Manaster and Rendleman Jr. [1982], Bhattacharya [1987], Anthony [1988], Stephan and Whaley [1990], Finucane [1991], Chan et al. [1993], Easley et al. [1998], Jarnecic [1999], Chan et al. [2002], Chakravarty et al. [2004], and Holowczak et al. [2006].

2Furthermore, investors can use deep out-of-the-money puts to position themselves against market crashes (Carr and Wu [2008]) or corporate defaults (Carr and Wu [2010, 2011], and they can use calendar spreads to sharpen their bets on volatility term structures (Egloff et al. [2010]).

3During our sample period, options on QQQQ have the highest trading volume, at 107,218,968 lots, followed by SPX, at 101,227,467 lots, IWM at 84,787,711 lots, and SPY at 63,416,811 lots. The highest trading volume on single-name stocks is on AAPL, at 32,983,347 lots.

4This rule takes into consideration that it is practically difficult for exchanges to disseminate quote messages to the whole market in real time. See Paragraph (b)(8) of Rule 611 in SEC release 34-51808, page 152. A detailed discussion of this exception can be found in McInish and Upson [2013].

5In principle, one can also experiment with other concave weighting functions. Our own preliminary analysis shows that the additional gains from such experiments are small.

REFERENCES


*To order reprints of this article, please contact Dewey Palmieri at dpalmieri@iijournals.com or 212-224-3675.*