1-2011

Consolidating information in option transactions

Jianfeng HU
Singapore Management University, JIANFENGHU@smu.edu.sg

Richard HOLOWCZAK

Liuren WU

DOI: https://doi.org/10.2139/ssrn.1740046

Consolidating Information in Option Transactions*

Richard Holowczak

Zicklin School of Business, Baruch College

Jianfeng Hu

Zicklin School of Business, Baruch College

Liuren Wu

Zicklin School of Business, Baruch College

First draft: January 13, 2011

Abstract

Underlying each stock trades hundreds of options at different strike prices and maturities. The order flows from these option transactions reveal important information about the underlying stock price. How to aggregate the trade information of different option contracts underlying the same stock presents an interesting and important question for developing microstructure theories and price discovery mechanisms in the derivatives markets. This paper takes options on QQQQ, the Nasdaq 100 tracking stock, as an example and examines different order flow consolidation mechanisms in terms of their effectiveness in extracting information about the underlying stock price and volatility movements. The analysis leads us to propose an aggregation weighting scheme that depends both on the liquidity of each option contract and the contract's risk exposure, delta for stock price movement information and vega for volatility movement information. Based on this weighting scheme, we identify significantly positive correlations between the aggregate option order flows and the realized returns and volatilities. In particular, the delta buy pressure positively predicts the underlying return and the vega buy pressure positively predicts the change of volatilities.

Keywords: Options; QQQQ; Information trading; High frequency; OPRA; Option order flow.

*Jianfeng Hu thanks CUNY Graduate Center for graduate research grant. We welcome comments, including references we have inadvertently overlooked. Richard.holowczak@baruch.cuny.edu (Holowczak), jianfeng.hu@baruch.cuny.edu (Hu) and liuren.wu@baruch.cuny.edu (Wu). The reviews presented here are solely those of the authors.

1. Introduction

In the absence of market frictions and under the log-normal stock price dynamics assumed in Black and Scholes (1973) and Merton (1973), options can be perfectly replicated by a portfolio of risk free bond and the underlying instrument. Option trading is thus redundant in this idealistic world. In reality, however, the market shows a strong demand for options for two major reasons. First, the risks in the stock market cannot be completely spanned by the stock trading alone. For example, the presence of discontinuous stock price movements of random sizes necessitates the inclusion of options across a whole spectrum of strikes to span the jump risk (Carr and Wu (2004)). The presence of stochastic volatility in stock movement, on the other market, makes the options market the de facto market for trading volatility risks (Carr and Wu (2009)).

The second major reason for options trading is informational. Even in the absence of stock price jumps and stochastic volatility, investors may choose to trade options to gain exposure to the stock given the high leverage provided by options (Black (1975)). Easley, O'Hara, and Srinivas (1998) further argue that informed traders may prefer the options market because they can better hide themselves among the multiple option contracts available on one security. Trading options also allows the informed traders to take advantage of volatility information that is not profitable on the stock market alone according to the first argument above. The disadvantage of trading options for stock exposure is the much higher transaction cost associated with options trading. Thus, only when the perceived information advantage (and hence price movement) is large enough, do the benefits of high leverage and multiple contract availability overshadow the large transaction costs (Holowczak, Simaan, and Wu (2006)).

There is a long history and a long list of studies on the information flow between the options market and stock market. One remaining challenge is how to effectively aggregate the information in the multiple option contracts underlying the same stock. When the underlying stock price moves, no arbitrage dictates that the prices on all the option contracts underlying this stock will move accordingly. When an options market maker takes on a position in any of the options underlying the same stock, the market maker will use the same stock to perform delta hedging. Hence, it is important to aggregate the information from the diverse option transactions at different strikes and maturities before one links the option transactions to stock price movements.

So far, most studies either use only one pair of option contracts (e.g. see Chan, Chung, and Fong (2002) and Holowczak, Simaan, and Wu (2006)) or simply regard different contracts as equally informative (e.g. see Easley, O'Hara, and Srinivas (1998), Chakravarty, Gulen, and Mayhew (2004), Cao, Chen, and Griffin (2005), and Pan and Poteshman (2006)) to simplify the problem. Picking one pair of contracts while discarding all the others can potentially distort the estimated relations due to missing variable problems, because options underlying the same stock share tight linkages. One can think of the case where the chosen option contract has a small transaction while most other options experience large transactions pointing to an opposite direction for the stock price movement. In this case, the large transactions of the omitted option contracts, rather than 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. Two studies take different methods by assigning different weights to option contracts though. Bollen and Whaley (2004) examine the impact of absolute delta weighted option order flows on implied volatility functions. Ni, Pan, and Poteshman (2008) use price-scaled *vega* weighted option volumes to predict realized volatilities in the cross section.

Another important but rarely raised question in this strand of empirical research is how to aggregate the order flows across trades with different sizes. While most researchers focus on option volumes (Easley, O'Hara, and Srinivas (1998), Chan, Chung, and Fong (2002), Chakravarty, Gulen, and Mayhew (2004), Cao, Chen, and Griffin (2005), Pan and Poteshman (2006), Ni, Pan, and Poteshman (2008)), some scholars just count number of trades to construct ”option buy pressures” (Bollen and Whaley (2004) and Holowczak, Simaan, and Wu (2006)). By counting number of trades, they presume that trades are equally informative across sizes and this method is likely to overstate the impact of small trades given the fact that option trade size is skewed toward the right. If informed traders only trade these tiny size options, they must be very patient to build a desirable position and this significantly increases the risk of failure if the information can be revealed soon. Aggregating volumes gives more information weight to large trades with an underlying assumption that the information content is linear to trade size. However, informed traders may split their orders to disguise themselves among uninformed orders and they should avoid trading large size trades to stay away from public attention. Anand and Chakravarty (2003) find that medium (small) size option trades are used to achieve ”stealth trading” when trading volume is high (low). Another fact undermining the informativeness of large trades is that they can often be pre-negotiated. All researchers in this literature face a tradeoff of assigning weights to trade size when they construct their empirical measure of option order flows.

In this paper, to answer the first question we propose a mechanism to aggregate option transactions across all strikes and maturities on the same stock, and we test its effectiveness against four alternatives (one pair, equal weighting, Greek-weighting, and price-scaled-Greek-weighting) in terms of their effectiveness in extracting information about future stock price and volatility movements. To extract 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 exposure. Accordingly, aggregations of buy and sell orders on call and put options should take on opposite signs. A classic 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. The second consideration is leverage. Given limited capital and private information, an investor would want to maximize its delta exposure per dollar spent on the contract. The delta of an option scaled by the option's value represents the stock risk exposure per dollar spent. Finally, the investor must takes into account the different transaction costs on the options contract in terms of both bid-ask spreads and market impacts. The options market liquidity concentrates on short-term near-the-money options. Although the stock exposure per dollar spent is the highest for far out-of-the-money options, the high bid-ask spread relative to the option value makes these contracts prohibitively expensive to trade. We combine all three considerations to generate an aggregate net price buy pressure (PBP) for option contracts. We show that this PBP measure generates significant predictions on future stock price movements.

We also propose an aggregate order flow measure that reveals the information in the underlying volatility. In this case, we focus on the volatility risk exposure of each option contract instead of its delta exposure. We combine the *vega* exposure with the leverage and liquidity concerns to generate an aggregate net volatility buy pressure (VBP) for the option contracts. We find that this VBP measure predicts future stock volatility as measured by equity option implied volatilities.

To address the second question on how to aggregate across trade size, we directly test the effectiveness of four alternatives (number of trades, volume, log(volume), log(volume+1)) and find that log(volume) outperforms the rest regardless of how we aggregate order flows across strikes and maturities.

We also examine the effect of a volume filter on option contracts. Following Easley, O'Hara, and Srinivas (1998), many empirical works exclude inactive (usually deeply OTM) option contracts when measuring option order flows. This treatment creates a potential problem as informed traders might trade these OTM options for the great leverage they provide regardless of illiquidity and filtering may throw away valuable information. Comparing the effects of filtered and unfiltered option order flows, we find mixing results of the effect of filtering but generally the impact is not significant.

Many studies investigate the information flow between the options market and the stock market, often with conflicting findings. Early studies such as Manaster and Jr (1982) and Bhattacharya (1987) find that the options market reveals information about the underlying security prices. Easley, O'Hara, and Srinivas (1998) do not find the option prices are informative but they find option volumes are informative about future stock prices although the signs are not as expected. Using the information share approach developed by Hasbrouck (1995), Chan, Chung, and Fong (2002) and Chakravarty, Gulen, and Mayhew (2004) find that the stock market leads the option market in price discovery. Holowczak, Simaan, and Wu (2006) find that the statistical significance of price discovery varies with option trading intensity and sidedness. Using a unique dataset, Pan and Poteshman (2006) find option call-put volume ratios predict future stock returns; and Ni, Pan, and Poteshman (2008) find that daily dollar-*vega* weighted order flow predicts future realized volatility.

Our work contributes to the literature by providing a systematic analysis on the aggregation of option transactions across different strikes, maturities and sizes, an issue that has been largely ignored or avoided in the literature. As we have argued earlier, one cannot possibly obtain robust results on the information flow between the options market and the stock market without first resolving the aggregation issue. We focus our analysis on the aggregation of option transactions, but the same mechanism can also be applied to aggregations of option quotes. Our finding of significant relationship between option order flows and stock market movement in high frequency public data also provides direct empirical evidence for information trading on options market.

The rest of the paper is organized as follows. Section 2 gives details about our empirical measures of option order flows. Section 3 describes the data we use. Section 4 gives the results of our comparison of buy pressures. And section 5 concludes.

2. Empirical specifications of option order flows

In this section, we describe the different measures of option order flow we construct. There are three main dimensions in our considerations. First, we want to aggregate across strikes and maturities to address the leverage and liquidity effects and we propose five alternatives in this dimension. Second, we want to aggregate across sizes and we propose four alternatives. Third, we want to extract information in option trading of underlying price and volatility so we construct option price buy pressures and volatility buy pressures respectively. Interacting all three dimensions, we have 40 different measures of option order flows. Further implementing a volume filter, we double the measures except for those including one_pair because this method nests the volume filter as will be explained in details later. We end up with 72 measures described in the sections that follow.

2.1. Aggregate across strikes and maturities

We use five alternatives to aggregate information in option trading across strikes and maturities:

1. *one_pair*: Following Chan, Chung, and Fong (2002) and Holowczak, Simaan, and Wu (2006), our first option order flow measure picks the most active calls and puts to represent all options out of liquidity concerns. On each trading day in the sample, we form pairs of call and put at the same strike and maturity and sort these pairs by the total number of trades. We then choose the most traded pair to construct option order flows. These options are always close-to-money and near-maturity options and constitute 24% of the total trades in our sample. We have also tested using the most traded calls and puts separately instead of sorting the volume by pairs and our results still hold.

To extract the price information from option trading, we construct a net price buy pressure, defined as

$$PBP\_one\_pair = Call\_BP - Put\_BP, \tag{1}$$

where *Call_BP* and *Put_BP* denote the net buy pressures (buy-sell) of the most traded calls and puts, respectively. As Easley, O'Hara, and Srinivas (1998) argue, buying a call and selling a put has positive price pressure on the underlying security while selling a call and buying a put has negative price pressure. Thus our *PBP_one_pair* is expected to be positively correlated to the underlying price if option trading is informative.

To extract the volatility information, we construct a net volatility buy pressure, defined as

$$VBP\_one\_pair = Call\_BP + Put\_BP. \tag{2}$$

Since the volatility change has the same pricing impact on both calls and puts, we sum the net buy pressures of calls and puts to predict underlying volatility. If the volatility increases, options generally become more valuable. Thus we expect our *VBP_one_pair* to be positively correlated to underlying volatility if option trading has volatility information.

2. *equal_weighting*: This is the most commonly used aggregation method in the literature (e.g. see Easley, O'Hara, and Srinivas (1998)). Equal weighting assumes the informed trader has no preference over any particular contract thus options across different strikes and maturities are equally informative. We define the equal weighted price buy pressure as

$$PBP\_equal\_weighting = \sum_{K,T} Call\_BP_{K,T} - Put\_BP_{K,T}, \tag{3}$$

where $Call\_BP_{K,T}$ and $Put\_BP_{K,T}$ are the net buy pressures of call and put options with strike price at K and time to maturity at T. Similarly we define the equal weighted volatility buy pressure as

$$VBP\_equal\_weighting = \sum_{K,T} Call\_BP_{K,T} + Put\_BP_{K,T}, \tag{4}$$

3. *Greek_per_share*: The derivatives of option prices to the underlying security price and volatility are well known as *delta* and *vega*. Using the Greeks as weights in aggregation takes into account the option's exposure to underlying price and volatility. An option with a larger delta is more sensitive to the price movement in the underlying security and would result in greater profit per share for an informed trader. Consequently, one can expect options with large delta to be more informative. So we define delta weighted price buy pressure as

$$PBP\_delta\_per\_share = \sum_{i} delta_i, \tag{5}$$

where $delta_i$ is price implied Black and Scholes (1973) *delta* for each trade *i*. Calls and puts with the same strike and maturity have opposite signs of delta. This mechanism gives more information weight to ITM options which is consistent with the experimental finding that informed traders may favor ITM options by Jong, Koedijk, and Schnitzlein (2001). Similarly, we define *vega* weighted price buy pressure as

$$VBP\_vega\_per\_share = \sum_{i} vega_i, \tag{6}$$

where $vega_i$ is price implied Black-Scholes (1973) *vega* for each trade. Calls and puts with the same strike and maturity have the same *vega*. And the *vega* weighting gives more weight to ATM options for large volatility exposure. We label these two order flows as *Greek_per_share* weighted because the Greeks used here are those of one option contract. Through out the paper we shall always refer to delta and *vega* as defined here unless otherwise specified. Bollen and Whaley (2004) use absolute delta weighted option volumes to predict the shape of implied volatility. But in this paper we test the predictability of these weighted order flows on the underlying market movement.

4. *Greek_per_dollar*: Ni, Pan, and Poteshman (2008) use price-scaled *vega*-weighted option volume to predict the realized volatility because the price-scaled *vega* can be thought as the return to the option position. Following their method, we also construct a price scaled delta buy pressure as

$$PBP\_delta\_per\_dollar = \sum_i \frac{delta_i}{p_i}, \tag{7}$$

and a price scaled *vega* buy pressure as

$$VBP\_vega\_per\_dollar = \sum_i \frac{vega_i}{p_i}, \tag{8}$$

where p_i is the price of option trade i. This weighting method considers the leverage effect of different options and assigns more weight to OTM options because OTM options are usually cheap. Chakravarty, Gulen, and Mayhew (2004) find that OTM options have higher information share in price discovery than ITM and ATM options. If it is true that informed traders prefer OTM options, this weighting method should outperform the rest.

5. *HHW*: While the Greek-weighting methods discussed above take into account the option exposure and leverage effect, they do not include the liquidity effect. If an option contract is highly illiquid or has thin market depth, it is difficult to realize a desirable profit from trading that contract alone and we would expect such contracts to be less informative about the underlying market. Choosing the most traded options as representatives goes to the other extreme as it ignores the option exposure and leverage effect and focuses only on liquidity. As an effort to combine all these considerations, we propose two aggregated measures of the option order flow as an information index. To predict the underlying stock price movement, we propose an aggregate net price buy pressure measure, defined as

$$PBP\_HHW = \sum_{K,T} \frac{Call\_BP_{K,T} - Put\_BP_{K,T}}{M} \cdot n(d_1), \tag{9}$$

where the summation is over all available strikes (K) and maturities (T), $n(d_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{d_1^2}{2}}$ denotes the standard normal probability density functions with $d_1 = \frac{ln(F/K)+\frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}$ denoting the standardized variable in the Black-Scholes formula that underlies the delta calculation, and $M = max\{1, T*12\}$ denotes the maturity in months with a minimum truncated at one. The aggregate measure takes into consideration the delta exposure, leverage, and liquidity. First, at the same strike and maturity, the weight is the same in magnitude but opposite in sign for the call and put option contracts. One can regard this weighting as an analog to the put-call parity condition. Since the difference between the forward values of the call and put contract at the same strike and maturity is equal to the forward value of the underlying stock, absent from the effects of volatility, we use this weighting on order flows to reflect the order flow on the underlying stock price movement while minimizing the impact from volatility trades. Second, across different strikes, the aggregation puts more weight on near the money options through the $n(d_1)$ weighting because near the money options are usually more liquid. Third, across different maturities, the option value scales approximately in the order of $\sqrt{T}$ We divide the weight by maturity to further punish longer-term contracts for liquidity concerns. We convert the maturity scaling in months and set the minimum to one month to avoid extreme weighting for options at very short maturities. One month or shorter represents the nearest month option.

To predict stock volatility movements, we also propose an aggregate net *vega* buy pressure (VBP) measure,

$$VBP\_HHW = \sum_{K,T} \frac{Call\_BP_{K,T} + Put\_BP_{K,T}}{\sqrt{M}} \cdot n(d_1). \tag{10}$$

Here, since both call and put options have positive *vega* exposures, we use the same weight (both magnitude and sign) for the call and put options at the same strike and maturity. Furthermore, since the *vega* of the option is proportional to the square root of maturity $\sqrt{T}$, it cancels with the $\sqrt{T}$ scaling in the option value. Hence, the denominator is $\sqrt{M}$ instead of M to punish long-dated options for liquidity concerns.

2.2. Aggregate across sizes

We consider four different ways to aggregate the information in trades of the same option contract with different sizes.

1. *number_of_trades*: Assuming option trades are equally informative regardless of their sizes, we can simply count number of trades as Bollen and Whayley (2004) and Holoczak, Simaan, and Wu (2006) and the option order flows are exactly as defined in the previous subsection. This aggregation will usually be

biased towards small trades in the sample.

2. *volume*: Assuming the information content in options trades is linear to the trade size, we can sum volumes as Easley, O'Hara, and Srinivas (1998) and others. For each option order flow defined above, we multiply each trade implied buy pressure with its size and sum up the products within a period of time to obtain an observation. This aggregation will usually be biased towards large trades in the sample.

3. *log(volume)*: The third method we use is multiplying natural logarithm of each trade size instead of trade size. Using logarithm smoothes the distribution of size and there is a more important reason. By using logarithm, we assume that the information content is concave to size and both volume and number of trades will matter. For example, two trades of the same size N are as informative as one trade of size 2N under volume summation. However, they will result in $2logN$ under this new aggregation method, which is greater than the information content of a single trade, $log2N$, if N is greater than or equal to 3. As a result, the larger number of trades to accumulate a bulk volume, the more informative the aggregated order flow will become. We propose this logarithm weight because we think informed traders tend to avoid large trades and would rather split their tentative order into small sizes to hide their intention. This method is supposed to combine the effects of both number of trades and trade volume.

4. *log(volume+1)*: A potential problem with log(volume) aggregation is that it completely ignores those trades of only one contract. To address this concern, we use log(volume+1) to aggregate across size. Consequently this method will give more weight to small trades than log(volume) aggregation.

2.3. Volume filter

So far we have five ways of aggregate across strikes and maturities, four ways to aggregate across sizes, and two types of order flows to predict underlying return and volatility separately. Interacting all three dimensions gives us 40 different option order flows. Now we turn to the last consideration, volume filter. Most empirical works adopt a volume filter to exclude inactive option contracts. For example, Easley, O'Hara, and Srinivas (1998) exclude those contracts with less than 50 trades a day; Bollen and Whaley (2004) filter out deeply OTM and deeply ITM option contracts with extreme delta values; using the most active options, Chan, Chung, and Fong (2002) and Holowczak, Simaan, and Wu (2006) explicitly apply an extreme volume filter and only consider the highest volume put/call pairs.

There exist both theoretical (Easley, O'Hara, and Srinivas (1998)) and empirical (Chakravarty, Gulen, and Mayhew (2004) evidence that OTM options might be preferred by informed traders. The volume filter may discard valuable information in these infrequently traded OTM options with the benefit of less noise. The overall effect is still unclear though. We thus construct option order flows with volume filter on and off to see if it has any impact. In this paper, we set the volume filter to 50 trades a day. This filter has no impact on those order flows constructed with *one_pair* because *one_pair* itself is an extreme filter. As a result, with consideration of the volume filter we expand our option order flow set to 40 unfiltered and 32 filtered. Using alternative filtering methods does not change our results qualitatively.

3. Data

3.1. Data sources and sample selection

Our option data comes from the Option Price Reporting Authority (OPRA), which records every option quote and trade message across all option exchanges in the United States. The underlying security price data comes from NYSE Trade and Quote (TAQ) database.

We pick the option ticker "QQQQ", an ETF of NASDAQ 100 index traded in all major exchanges in the United States, because QQQQ is the most actively traded ticker in options market (Holowczak, Simaan, and Wu (2006)). The high trading volume facilitates our high frequency research so that we can examine the effect of option buy pressures in shorter windows than the traditional 5 minute interval. Our sample covers 231 trading days from February 1st to December 29th in 2006. There are 1,572,865 trades in our raw sample. From this raw sample, we exclude data errors and options that expire within 10 calendar days because these close-to-maturity options can exhibit abnormal trading behaviors. Options quotes depend on underlying spot prices. In absence of continuous trading of the underlying security, option quotes become less reliable. Therefore, we exclude all off-hour trades. To control the options opening rotation effect, we also exclude trades recorded by OPRA within fifteen minutes after market opening and five minutes before market closing. Our final sample contains trades recorded by OPRA between 9:45:00 am and 3:54:59 pm EST only. These treatments reduce the sample size to 1,107,061 trades in all, and 4,792 trades daily on average. The average trade size is 70 lots but the median is only 5 lots. This clearly shows that in this sample number of trades is dominated by small trades while volume is dominated by large trades.

According to Table 1, unlike equity options which usually have most trades on at-the-money (ATM) options, QQQQ has 41.26% of trades on out-of-the-money (OTM) options and a slightly lower proportion of 36.69% on ATM options. The proportions in Table 1 are computed with number of trades only. Table 1 also shows that the majority (79.91%) of trades are short term options expiring in two months. Breaking down the sample into call and put option groups, we find that for QQQQ, puts are traded more often than calls and also at larger sizes. OTM options account for 42.88% in all puts traded, larger than that of calls (39.40%) and ATM options account for 34.02% in puts, less than that of calls (39.76%). The difference may reflect the use of QQQQ options as hedging devices. Both call and put groups constitute most of near-maturity trades. One concern regarding our choice of the underlying security is that information trading is less likely to occur for an ETF and related options because of less information asymmetry. While it is true that the macro variables driving the ETF's price and volatility are more transparent than single stock fundamentals, there can be as much if not more liquidity information in trading the ETF which dictates the movement of the underlying price and volatility in the near future ranging from a few seconds to a few hours. Also using QQQQ should make it more difficult for us to detect any effect of option order flows if information trading is less possible. Any significant result from our study will thus not be shadowed by the choice of this underlying security.

[Table 1 about here.]

3.2. Trade signing

OPRA does not have a flag on whether the non-market maker party in a trade is the buyer or the seller. In order to determine the option order flows, we follow Lee and Ready (1991) to classify trades into three categories: buyer-initiated, seller-initiated and unclassified. The signing algorithm is as follows: if a trade price is above the last effective mid quote, it is classified as buyer-initiated; if a trade price is below the mid quote, it is classified as seller-initiated; if a trade price falls exactly on the mid quote and 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; everything else is unclassified. With this signing algorithm, we are able to classify most trades, leaving only 0.86% in the unclassified category. These trades normally occur in market opens when there are no valid quotes or last different prices. We discard these unclassified trades because there is no reason to expect that our results are driven by this treatment.

Unlike Lee and Ready (1991), however, we examine the last (t-0) effective quotes on the same exchange rather than the five-second proceeding (t-5) quotes. The reporting lag is unnecessary for OPRA data because we find that the proportion of trade-through trades (with the price outside the quote bounds) increases in the time lag of quotes used and using t-0 has the largest proportion of trades right on the bid or the ask. We have also checked OPRA National Best Bid and Offer (NBBO) and find NBBO underperforms t-0 quotes from the same exchange. This data feature shows that OPRA is efficient in recording option trades and quotes.

We acknowledge that our signing algorithm is not perfect. But the errors in trade classification will bias against any significant findings. Thus we are comfortable with the current signing algorithm. Table 1 shows that the overall sample is slightly imbalanced with 51.81% of trades are initiated by buyers. Calls are more balanced with 49.38% buy trades. Puts are more imbalanced with 54.51% buy trades.

3.3. Frequency

We conduct our research at different frequencies but mainly report results from a relatively short time window of one minute. Information spreads fast between integrated markets (underlying market and option market in this case). Therefore, the impact of option buy pressure is not likely to last long and a long observation window may not be sensitive enough to capture the impact of information trading. A short observation window can increase the underlying price sensitivity to the option buy pressure and potentially the statistical significance of the correlations. However, it may reduce the economic significance on the other hand as short term price and volatility changes can be tiny. Also it tends to creates more uninformative observations with 0 trades because the arrival of trades is not continuous. There is clearly a tradeoff between a long observation window and a short one. Given the ETF nature of our underlying security, we are more concerned about liquidity based information trading in options market. We choose relatively short windows because it is more risky to trade on liquidity information in long horizons. Also under normal market conditions, liquidity shocks can be easily absorbed in price discovery. Therefore, we construct one minute non-overlapping option buy pressures as our main sample rather than the traditional five minute interval observations. We have tried alternative frequencies ranging from five seconds to half an hour and found that these alternatives do not alter our conclusion about the effectiveness of different aggregation methods tested.

3.4. Main variables

We calculate returns and volatilities of the underlying security QQQQ as our dependent variables. Prewhitening returns are calculated as the difference between log mid quotes of the National Best Bid and Ask (NBBO) at the beginning and the end of each one minute interval during trading hours. Table 2 reports that the mean of this raw return is -0.0063 basis points (bp). Bid ask bounce of the underlying price can cloud our findings. Following Easley, O'Hara, and Srinivas (1998), we use MA(1) process to remove the autocorrelation of the returns for each trading day and we call the residuals from this model excess returns. The mean of one minute excess return is -0.0022 bp in our sample with standard deviation of 4.5157 bp. As a proxy for market volatility, we calculate the standard deviation of second by second returns within each one-minute observation window and normalize it to annual volatility assuming price follows a log normal process. Table 2 reports that the mean of this annualized volatility is only 3.03%. We find using other volatility measures such as (high-low)/average and option implied volatility does not change our results qualitatively.

[Table 2 about here.]

Table 2 also reports the mean and standard deviation of each option order flow we construct in section 2. We use three month implied volatility from Bloomberg database and assume 0 interest rate and dividend rate to compute the B-S Greeks. The mean price buy pressures are generally negative and the mean volatility buy pressures are all positive although none is statistically different from 0. One thing to notice is that unfiltered option order flows are always stronger than the filtered except for *HHW_volume*. We apply Augmented Dickey-Fuller (1979) test for all variables with lags of 20. Not surprisingly, the p-values are all less than 0.001, strongly rejecting the null hypothesis of non-stationary time series.

4. Static comparison

4.1. Price buy pressures

An informed trader can trade options to profit from directional movement of the underlying security price. If our price buy pressures are correctly constructed, they are supposed to be positively correlated with the excess returns. Moreover, we want to test the effectiveness among the different price buy pressures. Table 3 reports our main results.

Easley, O'Hara, and Srinivas (1998) find option volumes are informative about stock prices. However, they find a significantly negative coefficient of the contemporaneous five-minute option volume instead of the hypothesized positive one. Contradicting their result, we find that all contemporaneous correlations are significantly positive as expected in our sample as shown in Panel A of table 3, supporting the hypothesis of information trading in options market.

[Table 3 about here.]

The second result emerging from Panel A of table 3 is that using log(volume) aggregation always outperforms the rest regardless of what method is used to aggregate across strikes and maturities. Log(volume+1) always comes next. Between number of trades and volume, number of trades has twice to three times larger coefficients than volume except for *one_pair*. The coefficients are very close for number of trades and volume aggregation with this extreme volume filter active. The result does suggest that empirical studies of option order flows need a good balance between number of trades and volume and log(volume) works well.

Among the aggregation methods across strikes and maturities, *delta_per_share* always outperforms the rest with *equal_weighting* and our *HHW* measure are closely tied in the second place. The differences between these three are small. *Delta_per_dollar* comes after with a larger gap and *one_pair* always has the smallest coefficient. A surprising result is the underperformance of *delta_per_dollar* to *delta_per_share* because the former is expected to capture the leverage effect which should be one of the main reasons for information trading in options market. One possible explanation is that *delta_per_dollar* aggregation overstates the effect of penny trades of deeply OTM options and ignores the liquidity effect which is equally important.

Finally we examine the filtering effect in Panel A of table 3. Apparently filtering volumes does not improve the contemporaneous correlations and actually in most cases it weakens the correlations instead. The single largest contemporaneous correlation belongs to unfiltered log(volume) and *delta_per_share* weighted price buy pressure (0.4112). Also using unfiltered data and log(volume) weighting across sizes, *equal_weighting* and *HHW* also generate strong contemporaneous correlations of 0.3962 and 0.3883, respectively.

To show the results are not driven by our choice of one minute observation window, we do the same analysis for different observation lengths ranging from 5 seconds to 15 minutes. The contemporaneous correlations are plotted in figure 1. Generally we see that the contemporaneous correlations increase in the observation length and the main results from Panel A of table 3 holds in different observation lengths, i.e. *delta_per_share* and log(volume) always outperform the other aggregation methods in their categories. Figure 1 also shows that different aggregation methods across strikes and maturities converge to a great extent under log(volume) aggregation except for *one_pair*. It seems the contemporaneous correlation is more sensitive to the aggregation method across sizes than the aggregation method across strikes and maturities.

[Figure 1 about here.]

Now let us turn to the predictability of our price buy pressures. Panel B of table 3 reports the correlations of the excess underlying returns and one period lagged price buy pressures. Comparing the aggregation methods across sizes first, we can see that number of trades is not informative about future returns. Within the rest three methods, volume seems to lead the race. Among the aggregation methods across strikes and maturities, *HHW* always comes in the first place with *equal_weighting* and *delta_per_share* following up closely. One surprising result is filtering slightly improves predictability of price buy pressures and *delta_per_share* does not well predict future excess return although it has strong contemporaneous correlation with excess returns. Figure 2 plots the correlations in different observation lengths.

[Figure 2 about here.]

What can we say about the effectiveness of these price buy pressures then? Despite some contradicting results of contemporaneous and forward looking correlations, some results are pretty informative. First of all, *equal_weighting*, *delta_per_share*, and *HHW* are better aggregation methods across strikes and maturities than *one_pair* and *delta_per_dollar*, suggesting the liquidity effect needs to be considered together with leverage and price exposure of the options. Within these three methods, however, it is hard to say which one always outperforms. Second, volume filter does not have much impact on the effectiveness of price buy pressures. It weakens the contemporaneous signals but improves the forecasting ability slightly. Third, among aggregation methods across sizes, log(volume) works best in contemporaneous correlations and volume weighting excels in forecasting.

4.2. Volatility buy pressures

Private information about market volatility is also profitable by trading options and we expect the volatility buy pressures constructed in section 2 to be positively correlated to underlying volatilities. Table 4 reports the correlations.

[Table 4 about here.]

Panel A of table 4 reports the contemporaneous correlations between volatility buy pressures and underlying volatilities. Similarly to the results of price buy pressures, log(volume) outperforms the other aggregation methods across sizes. Unlike for price buy pressures, however, the volume filter improves the contemporaneous correlations a lot. Unfiltered *HHW* outperforms the rest of unfiltered volatility buy pressures among aggregations across strikes and maturities. But once filtered, *vega_per_share* catches up quickly. Figure 3 plots the correlations in different observation lengths and it is clear that the order of effectiveness is persistent.

[Figure 3 about here.]

We examine the predictability of volatility buy pressures in Panel B of table 4. Volatility buy pressures show a consistent pattern of predictability with contemporaneous relationships. Log(volume) and *HHW* outperform the other aggregation methods in each category and the single biggest coefficient belongs to unfiltered *HHW* log(volume). Figure 4 plots the predictability of volatility buy pressures in different observation lengths and there is a clear pattern that log(volume) with either *vega_per_share* or *HHW* should be favored.

[Figure 4 about here.]

5. Conclusion

Our analysis above generally shows that the aggregation method of option order flows is not a trivial question in order to extract the information about the underlying market. We summarize our findings from the static comparison as follows: 1) aggregate option order flows are informative about both underlying returns and volatilities, i.e. price buy pressures positively predict underlying returns and volatility buy pressures positively predict underlying volatilities; 2) log(volume) weighting generally outperforms the other aggregation methods across sizes and only loses to volume weighting in predicting future returns; 3) *Greek_per_share*, *equal_weighting* and *HHW* generate more informative option buy pressures than *one_pair* and *Greek_per_dollar* ; 4) trading activity filter has less impact on the effectiveness of price buy pressures than of volatility buy pressures. Which aggregation method to choose depends on the specific empirical question.

References

Anand, A., and S. Chakravarty, 2003, "Stealth Trading in Options Market," working paper, Syracuse University.

Bhattacharya, M., 1987, "Price Changes of Related Securities: The Case of Call Options and Stocks," *Journal of Financial and Quantitative Analysis*, 22, 1–15.

Black, F., 1975, "Fact and Fantasy in the Use of Options," *Financial Analysts Journal*, 31, 36–72.

Black, F., and M. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81(3), 637–654.

Bollen, N. P., and R. E. Whaley, 2004, "Does Net Buying Pressure affect the Shape of Implied Volatility Functions?," *Journal of Finance*, 59(2), 711–753.

Cao, C., Z. Chen, and J. M. Griffin, 2005, "Information Content of Option Volume Prior to Takeovers," *Journal of Business*, 78, 1073–1109.

Carr, P., and L. Wu, 2004, "Static Hedging of Standard Options," working paper, Bloomberg and Baruch College.

Carr, P., and L. Wu, 2009, "Variance Risk Premiums," *Review of Financial Studies*, 22(3), 1311–1341.

Chakravarty, S., H. Gulen, and S. Mayhew, 2004, "Informed Trading in Stock and Option Markets," *Journal of Finance*, 59, 1235–1258.

Chan, K., Y. P. Chung, and W. M. Fong, 2002, "The Informational Role of Stock and Option Volume," *Review of Financial Studies*, 14, 1049–1075.

Easley, D., M. O'Hara, and P. S. Srinivas, 1998, "Option Volume and Stock Prices: Evidence on Where Informed Traders Trade," *Journal of Finance*, 53, 431–465.

Hasbrouck, J., 1995, "One Security, many Markets: Determining the Contributions to Price Discovery," *Journal of Finance*, 50(4), 1175–1199.

Holowczak, R., Y. Simaan, and L. Wu, 2006, "Price Discovery in the U.S. Stock and Stock Options Markets: A Portfolio Approach," *Review of Derivatives Research*, 9, 37–65.

Jong, C. D., K. C. G. Koedijk, and C. R. Schnitzlein, 2001, "Stock Market Quality in the Presence of a Traded Option," working paper, Erasmus University Rotterdam.

Lee, C. M. C., and M. J. Ready, 1991, "Inferring Trade Direction from Intraday Data," *Journal of Finance*, 46, 733–746.

Manaster, S., and R. J. R. Jr, 1982, "Option prices as predictors of equilibrium stock prices," *Journal of Finance*, 37, 1043–1058.

Merton, R. C., 1973, "An Intertemporal Asset Pricing Model," *Econometrica*, 41, 867–887.

Ni, S. X., J. Pan, and A. M. Poteshman, 2008, "Volatility Information Trading in Option Market," *Journal of Finance*, 63, 1059–1091.

Pan, J., and A. M. Poteshman, 2006, "The Information in Option Volume for Future Stock Prices," *Review of Financial Studies*, 19, 871–908.

Table 1

Data Description

The sample includes all option trades with underlying ticker ’QQQQ’ between 9:30:00 am or after 4:00:00 pm EST from 02/01/2006 to 12/29/2006 recorded by the Option Price Reporting Authority (OPRA) with maturity over 10 calendar days. Percentages are computed with number of trades. abs(delta) is the absolute value of Black-Scholes (1973) model implied *delta* computed with 0 interest rate and dividend rate. Near-maturity options are those expiring within 60 calendar days. The trades are classified into buy, sell, and unclassified categories with Lee and Ready (1991) method.

Statistics	All options	Calls	Puts
Number of trades	1,107,061	515,359	591,702
Mean daily number of trades	4,792	2,231	2,561
Mean trade size	70	56	73
Std trade size	690	724	658
Median trade size	5	5	6
Mean daily volume	311,246	124,147	187,100
Percentage of near-the-money $0.375 <= abs(delta) <= 0.625$	36.69	39.76	34.02
Percentage of out-of-the-money $abs(delta) < 0.375$	41.26	39.40	42.88
Percentage of near-maturity	79.91	80.22	79.64
Percentage of buy	52.12	49.38	54.51
Percentage of unclassified	0.86	1.02	0.73

Table 2

Underlying returns, volatility, and option buy pressures

This table reports the main variables in the analysis. Panel A reports the mean and standard deviation of dependent variables. Prewhitening returns are calculated as the difference between log prices (mid quote NBBO) at the beginning and the end of each one minute observation. Excess returns are the residual of an MA(1) model of the returns. Volatilities are calculated as the annualized standard deviation of second by second NBBO returns within each one minute observation. Panel B and C report the means and standard deviations of the net price buy pressures and volatility buy pressures as detailed in section 2, respectively. The column heads represent the aggregation methods across option strikes and maturities and the row names are the aggregation methods across trade sizes. We report both filtered (without contracts that have less than 50 trades on each day) and unfiltered buy pressures where applicable. Standard deviations are in parentheses.

Panel A: Dependent Variables

Prewhitening return (bp)	excess return (bp)	volatility
-0.0063	-0.0022	0.0303
(4.5374)	(4.5157)	(0.0190)

Panel B: Net Price Buy Pressure

	one_pair	equal_weighting		delta_per_share		delta_per_dollar		HHW	
		all	filtered	all	filtered	all	filtered	all	filtered
number_of_trades	-0.26	-0.68	-0.53	-0.27	-0.19	-0.71	-0.70	-0.18	-0.17
	(11.43)	(17.55)	(17.00)	(6.01)	(5.54)	(28.15)	(28.05)	(5.05)	(4.98)
volume	0.75	-10.98	-6.31	-7.45	-5.37	10.47	7.13	-2.38	-2.83
	(922.50)	(2826.03)	(2397.36)	(1101.64)	(959.55)	(2514.73)	(2412.66)	(812.85)	(792.19)
log(volume)	-0.13	-0.95	-0.62	-0.48	-0.30	-0.44	-0.40	-0.25	-0.21
	(6.34)	(19.76)	(16.66)	(10.24)	(8.46)	(14.09)	(12.94)	(6.17)	(5.80)
log(volume+1)	-0.28	-1.22	-0.85	-0.56	-0.37	-0.84	-0.80	-0.33	-0.28
	(10.30)	(23.91)	(20.98)	(11.34)	(9.50)	24.65	(23.89)	(7.29)	(6.93)

Panel C: Net Volatility Buy Pressure

	one_pair	equal_weighting		delta_per_share		delta_per_dollar		HHW	
		all	filtered	all	filtered	all	filtered	all	filtered
number_of_trades	0.07	0.64	0.55	3.50	2.79	3.53	2.11	0.19	0.18
	(11.36)	(16.30)	(16.18)	(71.64)	(68.90)	(468.22)	(466.99)	(4.72)	(4.71)
volume	4.32	55.83	39.65	334.42	221.02	421.50	190.11	15.91	12.87
	(916.33)	(2768.59)	(2317.24)	(18168.46)	(10738.59)	(38446.62)	(29358.83)	(816.65)	(772.81)
log(volume)	0.16	0.90	0.70	5.42	4.15	5.90	2.12	0.27	0.24
	(5.71)	(13.07)	(11.68)	(87.27)	(64.60)	(170.95)	(137.37)	(4.29)	(4.13)
log(volume+1)	0.18	1.17	0.95	6.82	5.31	7.15	3.12	0.36	0.32
	(9.83)	(17.42)	(16.34)	(100.68)	(80.71)	(371.69)	(356.56)	(5.44)	(5.31)

Table 3

Price buy pressures and returns

This table reports the correlations between underlying excess returns and the net price buy pressures. Excess returns are the residual of an MA(1) model of the returns. The column heads represent the aggregation methods across option strikes and maturities and the row names are the aggregation methods across trade sizes. We report both filtered (without contracts that have less than 50 trades on each day) and unfiltered buy pressures where applicable. Superscription a, b, and c represent significance level at 1%, 5%, and 10%, respectively.

Panel A: Contemporaneous correlation

	one_pair	equal_weighting		delta_per_share		delta_per_dollar		HHW	
		all	filtered	all	filtered	all	filtered	all	filtered
number_of_trades	0.0502[a]	0.1869[a]	0.1515[a]	0.2855[a]	0.2344[a]	0.0746[a]	0.0664[a]	0.1923[a]	0.1768[a]
volume	0.0629[a]	0.0808[a]	0.0807[a]	0.1052[a]	0.1010[a]	0.0619[a]	0.0590[a]	0.0933[a]	0.0899[a]
log(volume)	0.2208[a]	0.3962[a]	0.3746[a]	0.4112[a]	0.3887[a]	0.3358[a]	0.3213[a]	0.3883[a]	0.3748[a]
log(volume+1)	0.1500[a]	0.3593[a]	0.3265[a]	0.4035[a]	0.3745[a]	0.2153[a]	0.1963[a]	0.3587[a]	0.3430[a]

Panel B: Forecasting t+1 returns

	one_pair	equal_weighting		delta_per_share		delta_per_dollar		HHW	
		all	filtered	all	filtered	all	filtered	all	filtered
number_of_trades	-0.0035	0.0001	0.0008	-0.0002	0.0010	-0.0015	-0.0011	0.0013	0.0017
volume	0.0107[a]	0.0084[b]	0.0089[a]	0.0103[a]	0.0099[a]	0.0040	0.0033	0.0110[a]	0.0100[a]
log(volume)	0.0089[a]	0.0035	0.0054	0.0027	0.0051	0.0018	0.0028	0.0061[c]	0.0069[b]
log(volume+1)	0.0022	0.0026	0.0041	0.0020	0.0043	-0.0003	0.0003	0.0050	0.0058[c]

Table 4

Volatility buy pressures and volatilities

This table reports the correlations between underlying volatilities and the net volatility buy pressures. Volatilities are calculated as the annualized standard deviation of second by second NBBO returns within each one minute observation. The column heads represent the aggregation methods across option strikes and maturities and the row names are the aggregation methods across trade sizes. We report both filtered (without contracts that have less than 50 trades on each day) and unfiltered buy pressures where applicable. Superscription a, b, and c represent significance level at 1%, 5%, and 10%, respectively.

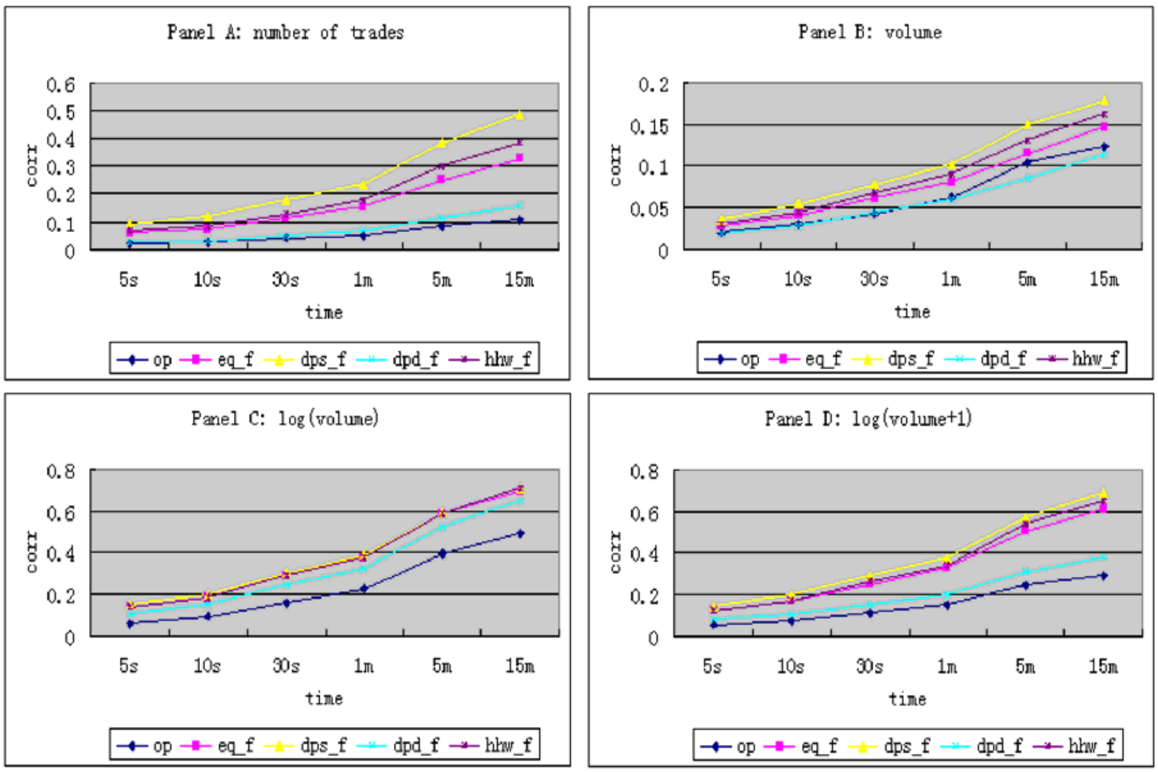
Panel A: Contemporaneous correlation

	one_pair	equal_weighting		delta_per_share		delta_per_dollar		HHW	
		all	filtered	all	filtered	all	filtered	all	filtered
number_of_trades	-0.0030	0.0044	0.0040	0.0094^a	0.0093^a	-0.0001	-0.0002	0.0074^b	0.0075^b
volume	0.0063^c	0.0091^a	0.0092^a	0.0090^a	0.0135^a	-0.0004	0.0020	0.0101^a	0.0098^a
log(volume)	0.0109^a	0.0136^a	0.0205^a	0.0049	0.0198^a	-0.0097^a	-0.0041	0.0178^a	0.0205^a
log(volume+1)	0.0028	0.0116^a	0.0151^a	0.0085^b	0.0188^a	-0.0038	-0.0014	0.0165^a	0.0182^a

Panel B: Forecasting t+1 returns

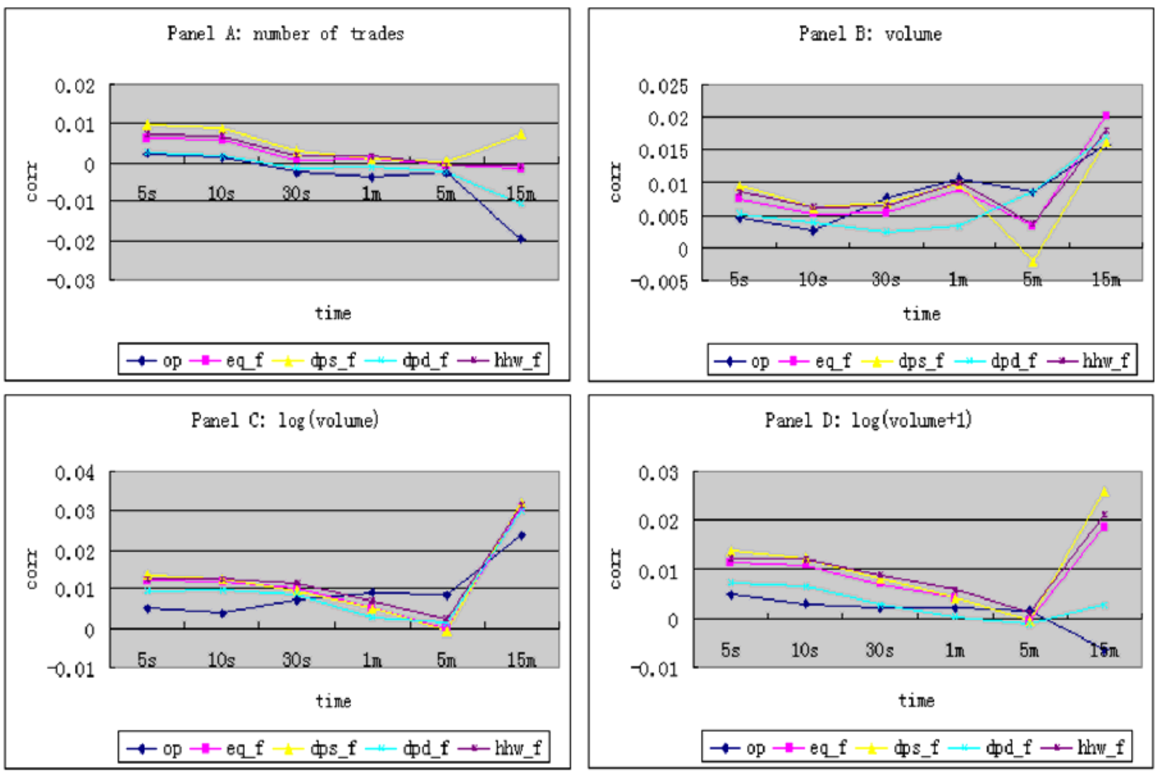
	one_pair	equal_weighting		delta_per_share		delta_per_dollar		HHW	
		all	filtered	all	filtered	all	filtered	all	filtered
number_of_trades	-0.0024	0.0049	0.0038	0.0056	0.0046	0.0012	0.0010	0.0044	0.0038
volume	0.0062^c	0.0074^b	0.0073^b	0.0056	0.0111^a	-0.0002	0.0021	0.0086^b	0.0079^b
log(volume)	0.0099^a	0.0190^a	0.0198^a	0.0120^a	0.0200^a	-0.0037	-0.0016	0.0203^a	0.0199^a
log(volume+1)	0.0029	0.0151^a	0.0144^a	0.0116^a	0.016^a	-0.0003	0.0005	0.0161^a	0.0153^a

Figure 1
Contemporaneous correlations between PBP and excess returns over different observation lengths



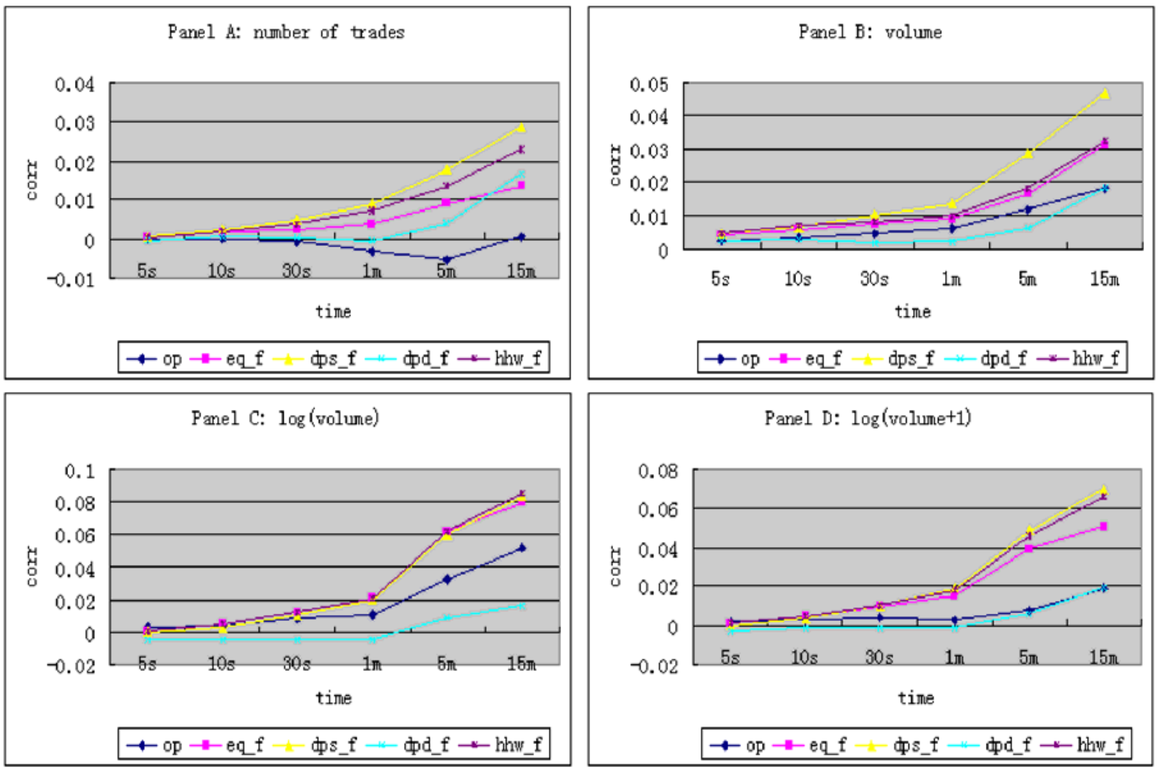
This figure plots the contemporaneous correlations between net option price buy pressures and underlying excess returns. The observation window ranges from 5 seconds to 15 minutes. The price buy pressures are calculated using filtered data (exclude options with less than 50 contracts on each day).

Figure 2
Forecasting excess returns over different observation length



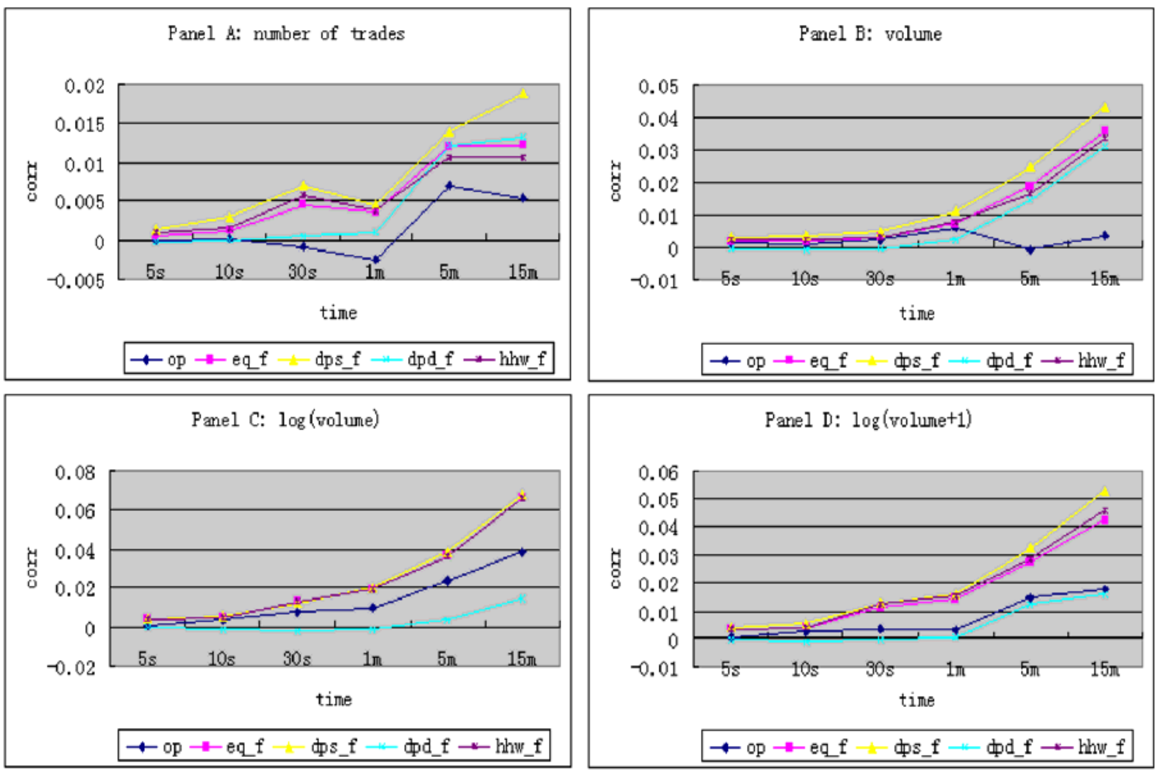
This figure plots the correlations between underlying excess returns and one-period lagged net option price buy pressures. The observation window ranges from 5 seconds to 15 minutes. The price buy pressures are calculated using filtered data (exclude options with less than 50 contracts on each day).

Figure 3
Contemporaneous correlations between VBP and volatilities over different observation lengths



This figure plots the contemporaneous correlations between net option volatility buy pressures and underlying volatilities. The observation window ranges from 5 seconds to 15 minutes. The price buy pressures are calculated using filtered data (exclude options with less than 50 contracts on each day).

Figure 4
Forecasting volatilities over different observation length



This figure plots the correlations between underlying volatilities and one-period lagged net option price buy pressures. The observation window ranges from 5 seconds to 15 minutes. The price buy pressures are calculated using filtered data (exclude options with less than 50 contracts on each day).