On the Informational Role of US Equity Options

by

Jianfeng Hu

A dissertation submitted to the Graduate Faculty in Business in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2013

This manuscript has been read and accepted for the Graduate Faculty in Business in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Committee:

Date

Professor Liuren Wu
Chair of Examining Committee

Date

Professor Joseph B. Weintrop
Executive Officer

Professor Turan Bali

Professor Armen Hovakimian

Professor Lin Peng
Supervision Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

On the Informational Role of US Equity Options

by

Jianfeng Hu

Adviser: Professor Liuren Wu

This dissertation consists of three parts of related research on the information flow across the stock market and the options market. In the first part, I investigate the optimal method of aggregating stock price and volatility information in option transactions across different option strike prices and maturities. I propose weighting methods based on option risk exposures (*Greeks*) and find that these aggregation methods empirically outperform the traditional methods for the NASDAQ 100 tracking stock (QQQQ). In the second part, I focus on the stock price information from option trading in the stock cross section. I decompose total stock order imbalance into an option market induced component and a stock market induced component independent of options market activity. I find that the option order imbalance has significant predictive ability in the cross section but the independent stock order imbalance has only transitory price impact. The return predictability of option order imbalance is greater for firms with higher levels of information asymmetry, larger transaction costs, and when option trading is active. I also find that the return predictability mainly comes from negative option order imbalance and the imbalance also predicts cumulative abnormal returns five days before earnings announcements. In the third part, I examine the impact of option listing on the underlying stock market quality. I find empirical evidence that option introduction reduces the probability of informed trading by attracting liquidity based traders more than information based traders. The robust findings of significant amount of information content in the options market reject the notion of options as redundant assets and raise more questions on how derivative securities should be priced and regulated.

Acknowledgements

I sincerely thank my advisor Dr. Liuren Wu for guiding me into the fruitful world of academic research on finance and economics. I am deeply indebted to him for his immense help in every stage of my research development. I would like to thank Dr. Turan Bali, Dr. Armen Hovakimian, and Dr. Lin Peng for their critical comments and mentorship, and Dr. Richard Holowczak and Henry Schwartz for help with data. The last part of the dissertation was finished during my visit to the Risk Management Institute at National University of Singapore. For helpful comments, I also thank Sumit Agarwal, Linda Allen, Gurdip Bakshi, Alexander Barinov, Utpal Bhattacharya, Jin-Chuan Duan, Fangjian Fu, Harmeet Goindi, Richard Herron, Craig Holden, Jian Hua, Robert Jarrow, Robert Jennings, Mehdi Karoui (discussant), Robert Kimmel, Roger Loh, Sophie Ni, Neil Pearson, Robert Schwartz, Zhaogang Song (discussant), Anand Srinivasan, Viktoriya Staneva, Debby Su, Charles Trzcinka, Jun Tu, Zhenyu Wang, Mitch Warachka (discussant), Jason Wei, Joseph Weintrop, Yangru Wu, Rui Yao, Joe Zhang, Xiaofei Zhao (discussant), Hao Zhou, and the seminar participants at Baruch College, Central University of Finance and Economics, City University of Hong Kong, DePaul University, Drexel University, Fordham University, Hong Kong University of Science and Technology, Indiana University, Rutgers University Camden, National University of Singapore, Singapore Management University, Sun Yet-San University, Sungkyunkwan University, Trade Alert LLC, University of Georgia, University of Massachusetts Amherst, University of Missouri Columbia, the Fifth Annual Risk Management Conference (Singapore), the Sixth Annual Risk Management Conference (Singapore), 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, the 2012 European Finance Association Doctoral Tutorial in Copenhagen, the 2012 Northern Finance Association Conference, the 22nd Annual Derivatives Securities and Risk Management Conference at the FDIC, and the Sixth Annual Conference on Asia-Pacific Financial Markets. I also want to thank Qianqian Wang for excellent research assistance at the Risk Management Institute of National University of Singapore. The generous research grants and travel grants from the Graduate Center and the Business Ph.D. program are greatly appreciated. Finally, I would like to thank my family for their unconditional and unlimited support.

Contents

List of Tables

List of Figures

Chapter 1

Aggregating Information in Option Transactions

Summary

Underlying each stock trades hundreds of options at different strike prices and maturities. The order flow from these option transactions reveals important information about the underlying stock price movement and its volatility variation. How to aggregate the trade information of different option contracts underlying the same stock presents an important challenge for developing microstructure theories and understanding price discovery mechanisms in the derivatives market. This paper takes options on QQQQ, the Nasdaq 100 tracking stock, as an example and examines different order flow aggregation methods in terms of their effectiveness in extracting information about the underlying stock price movement and its volatility variation. The analysis shows that an effective aggregation method must account for each contract's different exposure to the stock price and volatility movements, and also accommodate concerns on liquidity and interference from other potential risk dimensions, such as market crashes and long-term versus short-term volatility factors. The paper identifies significant relations, both contemporaneous and predictive, between the appropriately aggregated options order flow and the stock return and the return volatility, and finds that information on stock returns dissipates much faster than information on return volatility.

1. Introduction

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 riskfree 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 2004). 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 can 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, O'Hara, and Srinivas 1998).

A long list of studies have investigated the information flow between the options market and stock market.[1] 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 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

[1]See, for example, Manaster and Rendleman (1982), Bhattacharya (1987), Anthony (1988), Stephan and Whaley (1990), Finucane (1991), Chan, Chung, and Johnson (1993), Easley, O'Hara, and Srinivas (1998), Jarnecic (1999), Chan, Chung, and Fong (2002), Chakravarty, Gulen, and Mayhew (2004), and Holowczak, Simaan, and Wu (2006).

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, Chung, and Fong 2002, Holowczak, Simaan, and Wu 2006) or regard different contracts as equally informative (e.g., Easley, O'Hara, and Srinivas 1998, Cao, Chen, and Griffin 2005, 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 think of the 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, Pan, and Poteshman (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 paper proposes a mechanism to align the aggregation of option transactions with the particular type of information 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

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

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.

2. 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 New York Stock Exchange Trade and Quote database.

2.1. Sample Selection and Summary Statistics

The analysis in this paper is based on one underlying security, the NASDAQ 100 index tracking stock (QQQQ), over a span of 231 trading days from February 1 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.[3]

The sample contains 1,572,865 trade records on QQQQ options. We filter the trading records by excluding (i) before- and after-market trades, (ii) trades that happen within the first 15 minutes of the market open and the last five minutes before the market close, (iii) trades flagged as "late"

[3]During 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.

Table 1.1
Summary Statistics of the QQQQ Options Trade Sample
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-hour 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 less than 10 calendar days before expiry. Delta is computed using the Black-Scholes (1973) formula.

Statistics	All	Calls	Puts
Trade and trade size statistics:			
Total number of trades	1,087,778	506,948	580,830
Average number of trades per day	4,709	2,195	2,514
Mean trade size	65	56	73
Median trade size	5	5	6
Std of trade size	694	729	662
Mean daily volume	307,170	122,676	184,494
Percentages of trade records classified into different catetories:			
$37.5\% \leq \lvert\text{delta}\rvert \leq 62.5\%$	44.19	44.92	43.55
$\lvert\text{delta}\rvert < 37.5\%$	37.23	34.81	39.34
$\lvert\text{delta}\rvert > 62.5\%$	18.58	20.27	17.11
$T \leq 60$ days	79.91	80.25	79.62
$T > 60$ days	20.09	19.75	20.38

or “cancel” by OPRA, and (iv) trades on options that expire within 10 calendar days. The filtering reduces the sample to 1,087,778 trade records. Table 1.1 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, table1.1 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%. 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.

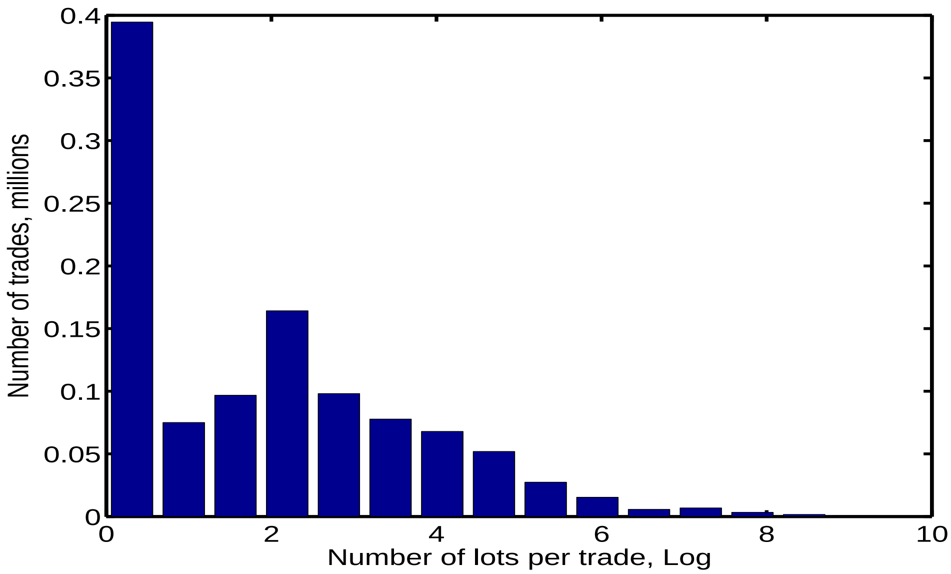
Across option maturities, 79.91% of the transactions are on short-term options expiring in less than 60 days. The actual percentage is even higher since we have filtered out options expiring in less than 10 days.

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 for each trade. We refer to these trades as “odd-lot” trades. Figure 1.1 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.

2.2. 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

Figure 1.1
Histogram of QQQQ Options Transaction Size
The histogram is on 1,087,778 filtered trade records for QQQQ options from February 1st to December 29th, 2006. Number of trades in millions is plotted against the natural log of the number of lots per trade.



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):

$$R_{t-h,t} = \ln(P_t/P_{t-h}), \tag{1.1}$$

where P_t denotes the mid point 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:

$$V_{t-h,t} = \sqrt{\frac{1}{h}\sum_{i=t-h}^{t-1} R^2_{i,i+1}}. \tag{1.2}$$

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

Table 1.2 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 A shows that the stock returns computed over the mid quotes exhibit little serial correlation.

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 mid-points 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.

3. AGGREGATING OPTION ORDER IMBALANCE

Aggregating option order imbalance involves four major steps. First, the OPRA database only shows 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 a seller. Each transaction is between two counterparties. We regard the side who pays the bid-ask spread to enter the transaction as the initiator. The other side (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 paying a

Table 1.2
Summary Statistics of Stock Returns and Return Volatilities
Entries report the summary statistics, including number of observations (Nobs), mean, standard deviation, minimum, maximum, and first-order autocorrelation, on the stock returns and return volatilities computed over different horizons from five seconds to 15 minutes. Returns are calculated as log difference of the midpoint of the National Best Bid and Ask (NBBO) prices, $R_{t-h,t} = \ln(P_t/P_{t-h})$. Return volatilities are computed using second-by-second NBBO midpoint returns over the shown horizons, $V_{t-h,t} = \sqrt{\frac{1}{h}\sum_{i=t-h}^{t-1} R_{i,i+1}^2}$. The second-by-second returns are zero if the NBBO mid-points are not updated during the second. Both returns and return volatilities are in basis points.

Horizon, h	Nobs	Mean	Std	Minimum	Maximum	Auto
Panel A. Stock returns, $R_{t-h,t}$						
5	1025640	-0.001	1.246	-47.005	131.354	0.065
10	512820	-0.001	1.819	-48.266	133.921	0.024
30	170940	-0.003	3.200	-85.255	133.921	0.002
60	85470	-0.006	4.537	-86.513	130.069	-0.013
300	17094	-0.032	9.860	-122.331	63.975	-0.002
900	5544	-0.023	16.947	-105.886	82.772	0.004
Panel B. Return volatilities, $V_{t-h,t}$						
5	1025640	0.241	0.433	0.000	58.743	0.210
10	512820	0.298	0.396	0.000	41.557	0.265
30	170940	0.383	0.314	0.000	23.993	0.382
60	85470	0.420	0.263	0.000	16.977	0.485
300	17094	0.456	0.193	0.000	9.098	0.615
900	5544	0.466	0.175	0.000	5.261	0.663

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 only involves the above two steps. 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.

3.1. 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 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.

3.2. 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 $COI(K,T)$ to denote the net order imbalance from a call option at strike K and expiry T, and use $POI(K,T)$ to denote the net order imbalance of a put option at strike K and

[4]This 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 (2012).

expiry T. Two decisions must be made in computing the net order imbalance: (1) How long a time period do we aggregate the transactions over? (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. One thus 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 paper 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, Kaul, and Lipson (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 informational importance of the very large trades. In the stock market, very large trades are often negotiated in the upstairs market and are put into the 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 paper, we balance the different considerations by applying a concave function on the trade size. Specifically, we use the natural logarithm of the trade size as the weighting to aggregate the order imbalance so that (i) we assign zero weight to the odd-lot transactions since they are commonly regarded as uninformative, and (ii) the weight increases with the trade size but only concavely so that we do not overweigh 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, we report only results for log volume weighting in the paper.[5]

3.3. Aggregating Information at Each Strike-Maturity Combination

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.

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

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 option 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

$$
\begin{aligned}
DOI(K,T) &= N(d)COI(K,T) - (1-N(d))POI(K,T), & (1.3)\\
VOI(K,T) &= n(d)\sqrt{T}\left[COI(K,T) + POI(K,T)\right], & (1.4)
\end{aligned}
$$

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

$$
d = \frac{\ln(F/K) + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}, \tag{1.5}
$$

with F denoting the forward price and σ being some volatility estimate for the underlying stock. In implementation, we use an average implied volatility estimate from the previous day to proxy σ. 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 (1.3) 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 (1.4) 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 highlighting the information target.

A stock transaction can only express a view 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 paper represents the first effort to formalize this idea through the proposed risk exposure weighting in aggregating call and put order imbalances in equations (1.3) and (1.4).

3.4. 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 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 that 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 (1.3) and (1.4) are simply aggregated across different strikes and maturities without other weighting.

$$ADOI_{GK} = \sum_{j=1}^{N} DOI(K_j, T_j), \quad AVOI_{GK} = \sum_{j=1}^{N} VOI(K_j, T_j), \tag{1.6}$$

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.

$$ADOI_{MD} = \sum_{j=1}^{N} e^{-(M_j-1)^2} DOI(K_j, T_j), \quad AVOI_{MD} = \sum_{j=1}^{N} e^{-(M_j-1)^2} VOI(K_j, T_j), \tag{1.7}$$

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.

$$ADOI_{KD} = \sum_{j=1}^{N} e^{-d_j^2/2} DOI(K_j, T_j), \quad AVOI_{KD} = \sum_{j=1}^{N} e^{-d_j^2/2} VOI(K_j, T_j), \tag{1.8}$$

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.

$$ADOI_{MK} = \sum_{j=1}^{N} e^{-d_j^2/2-(M_j-1)^2} DOI(K_j, T_j), \quad AVOI_{MK} = \sum_{j=1}^{N} e^{-d_j^2/2-(M_j-1)^2} VOI(K_j, T_j), \tag{1.9}$$

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 only considers the order imbalance from one strike-maturity point,

$$ADOI_{OP} = \sum_{j=1}^{N} 1_{d_j=0,M_j=1}(COI(K_j,T_j) - POI(K_j,T_j)),$$
$$AVOI_{OP} = \sum_{j=1}^{N} 1_{d_j=0,M_j=1}(COI(K_j,T_j) + POI(K_j,T_j)), \tag{1.10}$$

   where $1_{d_j=0,M_j=1}$ 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 to have 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, Chung, and Fong (2002).

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

$$ADOI_{EQ} = \sum_{j=1}^{N}(COI(K_j,T_j) - POI(K_j,T_j)), \quad AVOI_{EQ} = \sum_{j=1}^{N}(COI(K_j,T_j) + POI(K_j,T_j)). \tag{1.11}$$

   This method does not consider sensitivity weighting either, and has been used in many studies, including Easley, O'Hara, and Srinivas (1998) and Cao, Chen, and Griffin (2005).

4. 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

Table 1.3
Linking ASOI to Contemporaneous and Future Stock Returns
Entries report the contemporaneous correlation in Panel A and predictive correlation in Panel B between stock returns and aggregate delta order imbalance measures over different horizons. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

h	OP	EQ	GK	KD	MD	KM
Panel A. Contemporaneous correlations, $\text{Corr}(ADOI_{t-h,t}, R_{t-h,t})$						
5	0.071***	0.166***	0.182***	0.178***	0.165***	0.161***
10	0.100***	0.215***	0.232***	0.229***	0.213***	0.210***
30	0.170***	0.316***	0.332***	0.330***	0.316***	0.314***
60	0.238***	0.400***	0.415***	0.415***	0.400***	0.400***
300	0.437***	0.603***	0.610***	0.616***	0.599***	0.604***
900	0.547***	0.717***	0.718***	0.728***	0.707***	0.715***
Panel B. Forecasting correlation, $\text{Corr}(ADOI_{t-h,t}, R_{t,t+h})$						
5	0.011***	0.024***	0.027***	0.027***	0.025***	0.025***
10	0.009***	0.019***	0.020***	0.021***	0.019***	0.020***
30	0.010***	0.015***	0.015***	0.016***	0.014***	0.016***
60	0.013***	0.008**	0.007**	0.009***	0.008**	0.010***
300	0.006	-0.001	-0.001	0.001	0.000	0.003
900	0.025*	0.018	0.018	0.019	0.023*	0.024*

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 fully revealed in the current stock price, but will show up in subsequent stock price movements.

4.1. Contemporaneous Impacts

Panel A of table 1.3 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 reduces 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%.

4.2. Forecasting Relation

Panel B of table 1.3 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 5-second horizon, and declines to 2.1%, 1.6%, and then 0.9% at 10-, 30-, and 60-second horizon, respectively. The forecasting correlation estimates at even longer horizons are no longer significant statistically.

Table 1.4
Linking ASOI to Future Stock Returns Over Different Horizons
Entries report the forecasting correlation between the aggregate delta order imbalance with strike discounting over different aggregation periods (p) and stock returns over different future horizons (h), $\mathsf{Corr}(ADOI_{t-p,t}, R_{t,t+h})$. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Forecasting horizon h:	5	10	30	60	300	900
Aggregation horizon p:						
5	0.028***	0.024***	0.020***	0.015***	0.004***	0.002***
10	0.023***	0.021***	0.020***	0.015***	0.004***	0.001***
30	0.016***	0.017***	0.016***	0.012***	0.001**	-0.001*
60	0.011***	0.011***	0.010***	0.007***	-0.001***	-0.003***
300	0.002***	0.001***	0.000	-0.002***	-0.010***	-0.005***
900	0.000	-0.000	-0.002***	-0.004***	-0.008***	0.000

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.

Table 1.3 matches the return forecasting horizon with the aggregation horizon for the order imbalance. This matching is not particularly necessary. Table 1.4 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.

5. 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.

5.1. Contemporaneous Impacts

Panel A of table 1.5 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 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 five-minute and fifteen-minute horizons. Strike discounting has a negligible

Table 1.5

Linking AVOI to Contemporaneous and Future Stock Volatilities

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.

h	OP	EQ	GK	KD	MD	MK
Panel A. Contemporaneous correlations, $\text{Corr}(AVOI_{t-h,t}, R_{t-h,t})$						
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	0.006***	0.006**	-0.002	-0.002	0.011***	0.011***
60	0.011***	0.013***	0.005	0.005	0.019***	0.019***
300	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, $\text{Corr}(AVOI_{t-h,t}, R_{t,t+h})$						
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	0.006***	0.011***	0.005**	0.005**	0.012***	0.012***
60	0.008**	0.019***	0.012***	0.012***	0.021***	0.021***
300	0.018**	0.035***	0.027***	0.028***	0.037***	0.036***
900	0.037***	0.060***	0.041***	0.042***	0.068***	0.066***

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 a 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 reveal 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 enhance the correlation with short-term realized volatilities.

5.2. Forecasting Relation

Panel B of table 1.5 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.

Table 1.6 focuses on the best-performing aggregate vega order imbalance with maturity discounting as an example, and reports the forecasting correlation estimates at different aggregation

Table 1.6
Linking AVOI to Return Volatilities Over Different Horizons
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), $\text{Corr}(AVOI_{t-p,t}, V_{t,t+h})$. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Forecasting horizon h:	5	10	30	60	300	900
Aggregation horizon p:						
5	0.003***	0.004***	0.005***	0.006***	0.007***	0.007***
10	0.004***	0.005***	0.007***	0.008***	0.010***	0.009***
30	0.008***	0.009***	0.012***	0.014***	0.016***	0.016***
60	0.010***	0.013***	0.017***	0.021***	0.022***	0.022***
300	0.022***	0.027***	0.034***	0.038***	0.042***	0.045***
900	0.033***	0.040***	0.052***	0.058***	0.071***	0.081***

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 predicting return volatilities over all considered horizons from 5 seconds to 15 minutes.

6. 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 the current and future stock price movements. In this paper, 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 only gain stock exposure 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 the stock market transactions alone.

This paper focuses on the information in option transactions related to the stock price movement and 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 paper identifies significant relations, both contemporaneous and predictive, between the appropriately aggregated options order flow and the stock return and the return volatility.

Chapter 2

Does Option Trading Convey Stock Price Information?

Summary

An option market maker who executes an option order turns to the stock market to hedge away the underlying stock exposure. Thus, the *stock exposure* imbalance in option transactions translates directly into *order* imbalance in stock trades. In this paper, I decompose total order imbalance in stock transactions into the imbalance induced by option transactions and the imbalance induced by stock transactions independent of option trading activities. I find that the stock exposure imbalance induced by option transactions has strong predictive power of stock returns that does not reverse at long horizons. The independent stock order imbalance has a transitory price impact. The option-based stock return prediction increases with the level of information asymmetry, short-sale costs, and options market activity. The option-induced imbalance also predicts cumulative abnormal returns five days before earnings announcements.

1. Introduction

Many microstructure theories, such as those of Ho and Stoll (1981), Glosten and Milgrom (1985), Kyle (1985), and Easley and O'Hara (1987), all suggest that stock order flow can affect stock prices. Empirically, Chordia, Roll, and Subrahmanyam (2002) provide evidence that the stock order flow predicts future stock returns at the market level, and Chordia and Subrahmanyam (2004) provide evidence of the return predictability in the stock cross section. The stock options market provides an alternative for gaining stock exposures. Several studies, such as Easley, O'Hara, and Srinivas (1998) and Pan and Poteshman (2006), show that options order flow can also predict the underlying stock returns. In this paper, I examine how options order flow interacts with stock order flow to generate the stock return predictability.

When a customer executes an option order, the option market maker takes the opposite position by earning the bid-ask spread. Given relative scarcity of option transactions, it is difficult for the market maker to immediately unload the position via trades in opposite directions. Market makers often need to hold option positions for a long time, frequently until option expiry. To reduce risk exposure, it is now standard practice for market makers to perform delta hedging by trading on the underlying stocks. Therefore, if option transactions generate an imbalance in stock exposure, that stock exposure imbalance can be transferred to the stock market as a stock order imbalance through the delta hedging practice used by option market makers. As a result, some of the order imbalance in the stock market can be induced by option transactions.

To understand the interaction between the two markets, I decompose the aggregate stock order imbalance into two components: (i) an imbalance induced by option transactions and (ii) the remaining imbalance induced by stock market transactions that are unrelated to options market activities. To compute the option-induced order imbalance, I compute the hedging ratio, delta, of each option transaction using the real-time spot price and implied volatility. I use the delta of the option to capture the stock exposure of each option transaction, and I aggregate the delta of all option transactions within a certain period as the option-induced stock order imbalance, assuming that the market makers fully delta-hedge their option transactions and that the customers intentionally gain stock exposure from the option transactions and, hence, do not hedge their stock exposure.

I subtract that option-induced stock order imbalance from the total order imbalance to arrive at the residual imbalance that is induced purely by stock market investors and is unrelated to the option transactions. The decomposition enables me to separate the two sources of order imbalance and to investigate the role of each source of order imbalance in the stock return predictability.

I compute the daily stock order imbalance on a large cross section of stocks with options from April 2008 to August 2010. On average, there are 2,207 stocks each day in the sample. I analyze the return predictability of the order imbalance in the cross section. Several interesting results emerge: First, only option-induced order imbalance positively predicts the next day's stock returns. An investment analysis shows that firms in the highest quintile of option-induced order imbalance outperform those in the lowest quintile by 8.736 basis points on the next day (22% annualized, t-statistic = 6.03). The independent stock order imbalance has large contemporaneous price impact but shows no significant predictive ability for stock returns on the next day.

Second, the return predictability from the option-induced order imbalance does not reverse direction at longer horizons, suggesting that such predictability is more likely to be driven by permanent information flow than by temporary price pressure. Finally, in an intraday analysis at half-hour intervals, I also find that the option-induced imbalance has permanent price impact, whereas the independent stock imbalance generates transitory price impact. Those findings highlight the information content in option transactions and pinpoint the importance of separating it from other stock market transactions.

I then investigate how the information content of option transactions varies with different types of option contracts. I calculate option-induced order imbalances for at-the-money (ATM), in-the-money (ITM), and out-of-the-money (OTM) options separately. OTM options provide the highest leverage to investors but the transaction costs in terms of percentage bid-ask spread are also the largest for OTM options. I find that the predictive power of the option-induced order imbalance comes mainly from ATM and ITM options but not OTM options.

I also investigate how the return predictability of the option-induced stock imbalance varies with different types of firms. I classify firms based on three types of characteristics: level of information asymmetry, short-sale costs, and market trading activities. I apply five measures of information asymmetry: the probability of informed trading (PIN), the number of stock analysts following, the bid-ask spread on the stock market, the adverse-selection component of the bid-ask spread, and the firm size. When I organize firms into three groups based on each of the asymmetry variables, I find that the option-induced stock order imbalance always has the largest predictive coefficient and t-statistic in the group of informationally opaque firms that is, firms with high PIN, low analyst coverage, large spread, large adverse-selection component of the spread, and small market capitalization. My findings support the hypothesis that the return predictability of options order flow is driven by informed trading in the options market.

When I classify firms into three groups based on institutional holding, the group with low institutional ownership generates both the largest predictive coefficient estimate and the highest statistical significance. That finding supports the hypothesis that option trading conveys more information for firms with lower institutional ownership because it is more costly to borrow shares to meet short interests on such stocks. As a result, informed traders with negative news may have to place more orders in the options market as an alternative to short selling (Figlewski and Webb 1993, Johnson and So 2012).

I also find that option-induced order imbalance becomes more informative when option trading is active and total option trading volume is high. Easley, O'Hara, and Srinivas (1998) predict in their theoretical work that informed traders prefer to trade in deep markets. My finding supports that theoretical prediction. In particular, consistent with the short-sale-constraint story, I find that predictive price response to option-induced order imbalance is asymmetric: negative order imbalance has stronger predictive power than does positive order imbalance.

In an event study on earnings announcements, I find that option-induced order imbalance predicts cumulative abnormal returns five days before the announcement. The predictability is robust to alternative event windows and is greater when the earnings surprise is large and the dispersion of analyst forecasts is high. The results suggest that option trading is more informative when there exists substantial asymmetry of information and the profit from informed trading is high.

The findings in this paper have at least two important implications: First, the analysis underlines the importance of separating option-induced stock order imbalance from order imbalance generated by pure stock market investors. The existing literature often compares order flows and trading volumes from the two markets directly. Through careful separation, I show that options order flow contains an important informative component – about future stock price movement – that is not in stock order flow. Therefore, this decomposition is critical in measuring information content in order flows.

Second, my findings also elucidate the ongoing debate regarding where informed traders place their orders. A long list of empirical studies spotlights investigations of the cross-market information flow between the stock market and the options market.[1] In a closely related study, Pan and Poteshman (2006) show that put-call option volume ratio predicts future stock prices in the cross section. This paper takes a step further to investigate the interaction of options and stock order flows at both interday and intraday frequencies. To the best of my knowledge, this is the first paper to show that permanent price information in stock order flow is induced mostly by option transactions, suggesting that informed traders do use options.

The rest of the paper is organized as follows. Section 2 reviews the literature on the price impact of option trading. Section 3 proposes a decomposition method for total stock order imbalance and develops the hypotheses for empirical tests. Section 4 describes the data used in this study. In section 5, I report the empirical results of the hypothesis tests. Section 6 provides additional analysis on the predictive power resulting from option transactions. Section 7 concludes.

[1] See, for example, Manaster and Rendleman (1982), Bhattacharya (1987), Anthony (1988), Stephan and Whaley (1990), Chan, Chung, and Johnson (1993), Easley, O'Hara, and Srinivas (1998), Chan, Chung, and Fong (2002), Chakravarty, Gulen, and Mayhew (2004), Cao, Chen, and Griffin (2005), Holowczak, Simaan, and Wu (2006), Pan and Poteshman (2006), Ni, Pan, and Poteshman (2008), and Muravyev, Pearson, and Broussard (2013).

2. Background and motivation

The actual trading of derivative securities can convey important information in a market with information asymmetry. Black (1975) first notes the possibility of the options market as an alternative trading venue for informed traders because option contracts provide higher leverage. Biais and Hillion (1994) examine the impact on informed trading from option introduction and find that informed traders' profit can either increase or decrease depending on type of liquidity orders. Focusing on how private information gets incorporated into security prices, Easley, O'Hara, and Srinivas (1998) formalize a two-market microstructure model whereby informed traders can choose to trade either stock or options. The authors show that there exists a pooling equilibrium whereby informed traders trade both stock and options if both the leverage of the options and the options market liquidity are sufficiently high. They further argue that the availability of multiple contracts in the options market can present difficult learning problems for uninformed traders, thus making option contracts more attractive to informed traders. If informed traders trade in the options market, options order flow can contain permanent price information about underlying stocks.

Options order flow can gain return predictability in another important but less discussed way. Starting with Black and Scholes (1973) and Merton (1973), modern derivative pricing theory is built on the no-arbitrage argument. Other than the price relationship between options and the underlying security, the no-arbitrage argument also explains how to hedge option positions by using the underlying security, which implies reverse dependence on trading volume and order flow. After an option trader executes a transaction, the option market maker takes the opposite position and gains risk exposure to the underlying price movement as well as other factors such as return volatility. Because options market liquidity is usually not high enough for market makers to unload inventory immediately (liquidity is further diluted to hundreds of different option contracts on the same underlying security), market makers can hedge the underlying price risk by transacting in the underlying market. The hedge ratio is captured by option price sensitivity to the underlying price movement, delta. Meanwhile, non-market makers can also perform delta hedging if they intend to gain risk exposures other than delta. In so doing, they not only trade stocks to hedge delta by themselves, but also induce option maker makers to perform delta hedging too. Therefore, the options order flow can lead to subsequent stock order flow through delta hedging transactions; and information in the options order flow is also passed to the stock market. However, even without private information, hedging transactions can cause transitory price pressure on the underlying market in the same way as do stock transactions that are unrelated to the options market. Empirically, Ni, Pearson, and Poteshman (2005) show that the practice of delta hedging causes underlying stock prices to cluster at option strike prices. [2]

The theoretical discussion has motivated many empirical studies on the informational linkage across the two markets. A group of researchers focus on the relationship between actual stock prices and option implied stock prices. Early studies such as Manaster and Rendleman (1982) and Bhattacharya (1987) find that the options market leads the stock market in price discovery. However, more evidence points to the opposite direction. For example, Chakravarty, Gulen, and Mayhew (2004) find the information share of options quotes is less than 20% on average. Holowczak, Simaan, and Wu (2006) show that the information share of options quotes is even decreasing over time, and they argue that it is due to the prevailing use of computers for automatic updating of option quotes. In a recent study, Muravyev, Pearson, and Broussard (2013) find it is option quotes, but not stock quotes, that adjust to eliminate arbitrage opportunities across the two markets. Those authors conclude that the options market does not play a role in price discovery. Another strand of empirical research directly investigates options order flow. Many studies – including Easley, O'Hara, and Srinivas (1998), Poteshman (2006), Holowczak, Hu, and Wu (2011) and Dong and Sinha (2011) – find that options order flow is able to predict future stock returns in time series regressions for a small group of stocks.

In a closely related study, Pan and Poteshman (2006) investigate information content of options order flow in the cross section of stocks. Using a unique data set, those authors construct put-call volume ratios from option buyers to open new positions. They find that the open-buy put-call ratio negatively predicts returns and the return predictability lasts over three weeks. However, most

[2] In the structured equity product market, Henderson and Pearson (2010) find that hedging transactions raise the prices of underlying stocks by almost 100 basis points on the pricing date.

studies on options order flow do not control stock order flow. It is still unclear how options and stock order flows interact to generate return predictability. Since informed traders can trade in both markets, it is important to understand whether the options market makes a marginal contribution to embedding private information into stock prices.

Chan, Chung, and Fong (2002) investigate the informational role of order flows and quote revisions on the two markets. They find that options order flow does not have a pricing effect but the stock order flow does. Cao, Chen, and Griffin (2005) examine order flow information around mergers and acquisitions events. Their results suggest that options volume imbalance becomes significantly informative about future stock prices right before merger announcements but remains silent during normal periods. Overall, no significant return predictability is found for options order flow once stock order flow is controlled for. However, the stock order flow analyzed in these two studies are total stock order flow, which nests the option-induced stock transactions. The mixed stock order flow can absorb both information content and return predictability from options order flow. Therefore, information content in different markets must be separated carefully before any lead-lag analysis can be conducted.

One challenge of measuring information content in options order flow is that option trading can be motivated by risk factors other than the underlying stock price. For example, Ni, Pan, and Poteshman (2008) show that net demand of volatility risk in the options market positively predicts the next day's realized volatility in the cross section. It is therefore critical to specify the type of information to extract from options order flow and design an effective measure for that purpose. The same problem occurs when measuring price pressure from delta hedging transactions. If volatility traders use options-only strategies such as straddles and strangles, they need not hedge in the stock market. Such trades have neutral impact on option market makers' delta positions and hedging demand. However, volatility traders can also use a combination of options and stock to construct delta-neutral positions, and it is possible that both volatility traders and option market makers will hedge delta positions in the stock market. With the presence of volatility traders in the options market, measuring underlying stock price information and price pressure from options order flow is no longer straight forward. I propose a solution to the problem in next section.

This paper is also motivated by some recent development in the options market. Most studies use data from before 2000, when the options market in the United States is not yet a consolidated national market (see Battalio, Hatch, and Jennings 2004). Multiple listings on different exchanges and price competitions are not common at that time, and transaction costs in the options market are very high. After 2000, as a result of the SEC's efforts and technology development in quote updating and order routing, the options market has experienced tremendous growth and consolidation. Options market quality and liquidity have improved significantly. For informed traders, those changes lead to lower transaction costs and higher trading profits and can affect their order allocation decisions. Therefore, it is worthwhile to reexamine the information content of options order flow in a more recent sample period. Also, lack of liquidity in the past significantly limits the sample size of empirical analysis. For example, Chan, Chung, and Fong (2002) have only 14 stocks in their sample, out of which, the most active option ticker, General Electric (GE), has an average daily trading volume of 2,595 option lots. By comparison, GE's option trading volume reaches 134,874 lots on April 1, 2008, the first day in my sample; and the average option trading volume of all common stocks is 3,255 lots on that day. The growth of the options market facilitates the expansion of the sample coverage. In this study, I include all common stocks with options, and I focus on the cross-sectional relationship between order flows and future returns.

3. Methodology

This section begins by explaining how to measure price information in options and stock order flows. It then describes the primary hypotheses and the empirical testing methods.

3.1. Option order imbalance

A remarkable feature of the options market is that hundreds of option contracts can underlie the same security. For example, on January 2, 2009, there are 478 option contracts underlying the common stock of Apple Inc., spanning 61 strike prices from \$12.5 to \$400 [3] and 6 expiry dates from 15 days to 750 days. Because each contract has its unique combination of put/call type, strike price, and maturity, the trading process for each option contract does not necessarily reveal the underlying price information in the same way. Measuring the price information in the options order flow is challenging. In a recent study, Holowczak, Hu, and Wu (2011) show that risk exposure to the underlying stock price (delta) is an important consideration when one is aggregating stock price information in option transactions. Following those authors' intuition, I propose a standardized measure of option order imbalance (OOI):

$$OOI_{it} = \frac{\sum_{j=1}^{N} 100 Dir_{itj} \cdot delta_{itj} \cdot size_{itj}}{Num\_shares\_outstanding_i}. \tag{2.1}$$

The option order imbalance, OOI_{it}, is measured for stock i on day t. Dir_{itj} is a dummy variable equal to 1 if the j^{th} option trade on stock i is initiated by the buyer, and -1 if the trade is initiated by the seller, according to certain trade signing algorithms. $delta_{itj}$ is the option price sensitivity to the underlying stock price, and $size_{itj}$ denotes the trade size in option lots (100 shares of the underlying stock). The numerator is thus the net delta position of non-market makers in the options market. The order flow is more likely to be imbalanced when trading is inactive, and that imbalance can reflect mere noise caused by uninformed trades. To address this issue, I scale net delta volume with the number of common shares outstanding for cross-sectional standardization. [4]

OOI measures the directional trading intention of non-market makers in the options market. To link OOI to information content and hedging demand in the stock market, I make the following assumption.

Assumption: On average, option trading on non-delta risk exposures has neutral impact on delta imbalance in the options market, and uninformed option trading on delta risk exposure is also balanced.

[3] The closing price of Apple stock is \$90.75 on that day.

[4] I have also experimented with alternative scalers such as total delta volume and total stock volume for robustness checks. My primary results hold qualitatively, but these alternative scalers are more likely to generate outliers during inactive trading periods. Hence, I do not report those results in the paper.

Under the assumption, OOI is driven by informed traders with advanced stock price information, and the magnitude of the imbalance then measures those informed traders' aggressiveness and amount of information regarding the underlying stock price. Moreover, OOI also determines net delta hedging demand and price pressure caused in the stock market. This is true because informed traders intentionally seek delta exposures in the options and do not hedge, and only option market makers perform delta hedging. Market makers take the opposite delta position to OOI, and they will then have to transact the same number of stocks in the stock market to offset the delta position. Although it is unknown when option market makers will perform delta hedging after receiving option orders, one can still infer option market makers' net stock demand at least on a daily basis because option market makers tend to go home flat–meaning, they normally keep delta-neutral positions overnight. Note that option market makers do not necessarily wait until the end of the day to hedge when the net delta position is known. It is difficult to track all of the stock orders submitted by option market makers, and total transaction volume can be much higher than the numerator in equation (2.1). Nonetheless, one can learn option market makers' net demand for number of shares to hedge the newly established option positions if the hedge ratio does not fluctuate too much intraday.[5] One should be reminded that OOI does not include stock demand from dynamic hedging on past option positions. Those hedging transactions cannot be identified without knowing the market makers' net positions, and they do not contain any new information. Therefore, in this study, they are treated as liquidity trades in the stock market.

3.2. Stock order imbalance

Defining option order imbalance in terms of delta imbalance has a great advantage because now I can calculate the stock order imbalance (SOI) that is not related to options market activity by removing the delta imbalance from the total stock order imbalance. Subsequently, the stock order

[5]Although option order imbalance can also be measured by number of trades and dollar volume, it is more difficult to relate such imbalance to hedging demand because delta hedging should be based on number of shares.

imbalance induced by non-options market activity is defined as

$$SOI_{it} = TOI_{it} - OOI_{it} = \frac{\sum_{j=1}^{N} Dir_{itj} \cdot size_{itj}}{Num\_shares\_outstanding_i} - OOI_{it}, \tag{2.2}$$

where TOI_{it} is the total stock order imbalance in terms of trading volume, and Dir_{itj} and $size_{itj}$ are the direction dummy and the size of the j^{th} trade of firm i on day t in the stock market, respectively. I also scale that order imbalance with total number of shares outstanding to keep it consistent with OOI. After subtraction of OOI_{it}, the remainder is simply the stock order imbalance unrelated to options order flow.

3.3. Main test

To gauge return predictability from order flows in the two markets, I estimate the following equation in the cross section:

$$Ret_{i,t} = \alpha + \sum_{k=1}^{5} \beta_1^k SOI_{i,t-k} + \sum_{k=1}^{5} \beta_2^k OOI_{i,t-k} + \theta X_{i,t-1} + \varepsilon_{i,t}, \tag{2.3}$$

where for stock i, $Ret_{i,t}$ is the stock return on day t, calculated using the midpoint of the last valid National Best Bid and Offer (NBBO) spread and where $X_{i,t-1}$ is a set of control variables, including the closing bid-ask spread of the stock, the stock turnover ratio, log trading volumes in the two markets, and the stock returns and the equally weighted option returns from day $t-5$ to $t-1$.[6]

Hypothesis 1: SOI positively predicts future stock returns–at least for some k, $\beta_1^k > 0$. *If the predictability comes from the price impact of liquidity trades, the predictive relationship reverses its sign at longer horizons. If the predictability comes from informed trading in the stock market,*

[6]In theory, the private information in order flows is about specific firms rather than the entire market and the imbalance should have greater predictive ability about idiosyncratic returns. In unreported tests, I use risk-adjusted returns as the dependent variable and find stronger results. I have also experimented with open-to-close NBBO midpoint returns, returns calculated using transaction prices and value-weighted option returns and found largely the same results

the predictive relationship does not reverse its sign.

With asymmetric information in the market, stock order flows can reveal private information; and SOI is able to predict future permanent price changes (see, e.g., Glosten and Milgrom 1985, Kyle 1985, Easley and O'Hara 1987). Positive (Negative) SOI should reflect the buying (selling) of informed traders with good (bad) news. Alternatively, even in the absence of information asymmetry, market makers can lift the quoted price on positive order imbalance and reduce the quoted price on negative order imbalance to entice offsetting orders and return to their optimal portfolios. As a result, SOI also positively predicts the future price changes. However, that price impact is short-lived. Once the price pressure from the order imbalance disappears, the stock price will return to its fundamental value. Thus, a reverse predictive relationship exists at longer horizons.

Hypothesis 2: OOI positively predicts future stock returns controlling for SOI–that is., at least for some k, $\beta_2^k > 0$*. If the predictability comes from price pressure of the option market makers' delta hedging, the predictive relationship reverses its sign at longer horizons. If the predictability comes from informed trading in the options market, the predictive relationship does not reverse.*

Similar to SOI, OOI can predict future returns due to either price pressure of delta hedging trades or informed trading in the options market. Controlling SOI, OOI then captures the marginal contribution of options order flow in price discovery.

3.4. Option leverage

Hypothesis 3: If informed traders use options for reasons of leverage, they should prefer options that give them the highest leverage. Therefore, the order imbalance of these options will be more informative about future stock returns.

To test that hypothesis, I divide all option transactions into three groups based on their moneyness, delta: (i) out-of-the-money (OTM, $|delta| < 0.375$), (ii) at-the-money (ATM, $0.375 \leq |delta| \leq 0.625$), and (iii) in-the-money (ITM, $|delta| > 0.625$). Although OTM options have the smallest delta, they provide the highest leverage for every dollar investment because of their low prices. ITM options have the lowest dollar leverage among the three groups. I calculate OOI separately for the three option groups and estimate the following equation:

$$Ret_{i,t} = \alpha + \beta_1 SOI_{i,t-1} + \beta_2 OTM\_OOI_{i,t-1} + \beta_3 ATM\_OOI_{i,t-1} + \beta_4 ITM\_OOI_{i,t-1} + \theta X_{i,t-1} + \varepsilon_{i,t}, \quad (2.4)$$

where $Ret_{i,t}$, $SOI_{i,t-1}$, and $X_{i,t-1}$ are the same as defined before, and $OTM\_OOI_{i,t-1}$, $ATM\_OOI_{i,t-1}$, and $ITM\_OOI_{i,t-1}$ are delta order imbalances calculated using OTM, ATM, and ITM options respectively.

3.5. Level of information asymmetry

If the price impact of OOI comes from informed trading, I expect the predictive ability of OOI to be greater for firms that are more likely to have information asymmetry. I use five proxies for the level of information asymmetry: PIN of Easley and O'Hara (1992), the number of analysts following, the percentage bid-ask spread, the adverse-selection component of the bid-ask spread, and the firm size. To calculate the adverse selection component of the spread, I use both the Glosten and Harris (1988) and Lin, Sanger, and Booth (1995) models.

Hypothesis 4: If options order flow reveals private information, the predictive ability of OOI should be greater for firms with more information asymmetry–that is, high PIN, low analyst coverage, large bid-ask spread, and large adverse selection component of the bid-ask spread.

Based on each proxy of information asymmetry, I first sort all firms by ascending order every day, and divide the full sample into three subgroups–low, medium, and high–with cutoff points at the 30th and 70th percentiles. I then reestimate equation (2.3) in each subgroup and compare the estimated coefficients and t-statistics of OOI across subgroups.

3.6. Institutional ownership

Other than level of information asymmetry, short-sale costs will also have an impact on the predictive ability of OOI. It is well-known that options are used as devices to circumvent the short-sale constraint. For stocks that are more difficult to short, option trading can have greater informational benefits because informed traders with negative news can be forced to trade options only. Institutional ownership as a proxy for the market supply of short interests is known to be negatively related with the difficulty of short selling (Asquith, Pathak, and Ritter 2005). On the other hand, institutions are also more likely to have an information advantage over retail investors. Thus, I propose the following hypothesis.

Hypothesis 5: Return predictability from OOI can decrease in a stock's institutional ownership because institutional ownership reduces the short-sale costs. Alternatively, return predictability from OOI can increase in a stock's institutional ownership because institutional investors can be better informed.

To test that hypothesis, I reestimate equation (2.3) in three subgroups based on institutional ownership: low ($< 30\%$), medium ($30 \sim 70\%$), and high ($\geq 70\%$) and compare both the statistical and economic significances of the estimated OOI coefficients in the subgroups.

3.7. Market activeness

Liquidity can also affect the return predictability of OOI if it is due to informed trading.

Hypothesis 6: The price impact of OOI increases when market liquidity improves because informed traders can conceal their trades better in high option volume stocks.

To test that hypothesis, I divide the full sample into three option volume groups and reestimate equation (2.3) in the subgroups.

4. Data

Analysis of the order flows in both markets requires the merging of several databases. This section describes the details of the data sources, the sample selection, and the variable construction.

4.1. Options market activity

I obtain option transaction data on 611 trading days from April 2008 to August 2010 from Trade Alert LLC, a specialized option market data vendor. The data include all trade messages recorded by the Options Price Reporting Authority (OPRA), a national information processor that consolidates market information generated by option trading on all US option exchanges. Trade Alert matches option transaction data with underlying spot price and computes option implied volatility for each transaction from a binomial tree in real time. Following the quote rule, Trade Alert also classifies each trade as either buyer-initiated or seller-initiated. A trade is classified as buyer-initiated if the price is above the last effective mid quote price, and as seller-initiated if the price is below the mid quote price. Such comprehensive trade data facilitate a cross-sectional study over a relatively long period.

I exclude all indexes, units, ADRs, REITs, closed end funds, ETFs, and foreign firms and focus on common stocks only (CRSP share codes 10 and 11). For computing OOI, I exclude options expiring within 10 calendar days and the following trades: (i) off-hour trades, (ii) trades at market open (the first 15 minutes) and market close (the last five minutes), (iii) trades that are reported to the OPRA as "late" or "cancel", (iv) trades flagged as part of structured trades, and (v) data errors such as 0 strike price or 0 trade price. I filter out (i), (ii), and (iii) because the trade classification is less reliable for those trades; (iv) because those trades are less likely to be about stock price information than about volatility information, etc.; and (v) for obvious reasons. I also exclude stock-split days and dividend days because of their complex implications for options pricing and trading.

In table 2.1, I present the primary statistics for the full options sample as well as for four transaction-type groups: buy call, sell call, buy put, and sell put. Because the quote rule is unable to classify trade directions when the trade price falls exactly on the mid quote or when there is no valid quote at all, some trades in the sample cannot be classified into any of the four transaction-type groups. I do not report the statistics for that group because those trades are not of interest to the study.[7] On average, 1,670 stocks per day have at least one valid option transaction, with the maximum at 1,909 and the minimum at 1,395. Call options are traded more often than put options because the average daily number of firms with call transactions exceeds the average daily number of firms with put transactions in both the buy and sell categories in panel A. Trade level statistics are reported in panel B. The average trade size is 17.45 lots in the full sample. The call option trades are, on average, smaller than the put option trades. However, call options have a much larger average daily number of trades than put options do. Therefore, the average daily volume of call options also exceeds that of put options by approximately 0.77 million lots. The average premium of all single-name options traded on a day exceeds $1 billion. Lakonishok, Lee, Pearson, and Poteshman (2007) find that non-market maker participants take net long positions in daily option volume from 1990 to 2001.Panel B confirms their finding by showing that for both call and put options, there are more buy trades than sell trades in terms of both volume and premium. Panel C presents the distribution of trading volume across option moneyness and time to expiration. Across the moneyness, OTM options are the most frequently traded, accounting for over 50% of the entire market volume; ATM options account for 33.91% of the total volume; and ITM options have the smallest share, with only 14.14%. The same pattern can be found in the four transaction-type groups. Additionally, across the three moneyness regions, non-market maker participants take net long positions except for ITM puts. Order imbalance is largest for OTM options and smallest for ITM options. Across maturities, approximately 37% of the transactions are for options expiring in 30 calendar days. Options expiring in 31 to 60 days account for 27.91% of the total volume. Given the fact that I exclude all option trades expiring in less than 10 days, most trades are on short term options. The same pattern exists in the four transaction-type groups, and the buy volume always

[7] The trade direction is assigned to 0 for these trades. Therefore, unclassified trades have no impact on OOI in empirical analysis. Omitting that group, however, can cause the full-sample statistic to be different from the sum or weighted average of the four transaction-type groups.

exceeds the sell volume.

At the beginning of the sample period, there are seven option exchanges in the OPRA plan: the American Stock Exchange (AMEX), the Boston Options Exchange (BOX), the Chicago Board Options Exchange (CBOE), the International Securities Exchange (ISE), NASDAQ, the New York Stock Exchange ARCA (NYSE), and the Philadelphia Stock Exchange (PHLX). During the sample period, another participant exchange, the Better Trading System (BATS), joins the OPRA on February 26, 2010. The last panel of table 2.1 presents the volume share by exchanges. The CBOE and the ISE lead in market share as each of them attracts more than one quarter of the total volume: 29.66% and 27.43%, respectively. The PHLX is also competitive, attracting 16.92% of the total volume, followed by the NYSE at 11.59% and the AMEX at 7.15%. The BOX and NASDAQ are relatively small markets; neither has more than 5% of market share. The BATS does not have much market share because it is newly established. The Herfindahl index of the market share is 0.213, indicating that the US options market is moderately concentrated during the sample period.

4.2. Other data sources

I obtain stock transaction data from the NYSE Trade and Quote (TAQ) database. I follow Lee and Ready (1991) in signing trade directions. After applying the quote rule, if the trade is still unclassified, I compare the trade price with the last different trade price. If the current trade price is above (below) the last different price, it is classified as buyer-initiated (seller-initiated). Unlike Lee and Ready (1991), however, I collapse trades made at the same second into one record weighted by dollar volume and use the most updated NBBO prices one second before the trade time ($t-1$) for trade signing. In addition to excluding canceled trades and data errors, I also exclude trades made within the first 15 minutes and the last five minutes of trading to increase the accuracy of trade signing and to match the observation period of OOI.

Finally, stock returns are calculated using the midpoints of the closing bid-ask spreads collected from the CRSP. The number of analysts following is extracted from I/B/E/S, and the institutional ownership data from Form 13-F filings in the Thomson Reuters database.

Table 2.1

Summary statistics of the options market

This table describes options market activity from April 2008 to August 2010. Only single-name equity options with maturities of more than 10 days are included in the sample. The following trade records are excluded: (i) off-hour trades, (ii) trades at market open (the first 15 minutes) and market close (the last five minutes), (iii) trades that are reported to the OPRA as "late" or "cancel", (iv) trades flagged as part of structured trades, and (v) data errors such as 0 strike price or 0 trade price. The trade direction is based on the quote rule. Panel A provides descriptive statistics of number of firms per day. Panel B reports trade level statistics. Trading volume is in option lots (equivalent to 100 shares of the underlying stocks). Panel C shows trading volume breakdown (as percentages of the total volume) based on option moneyness, $delta$, and maturity, T. Panel D gives the volume breakdown by option exchanges.

Statistics	All	Buy Call	Sell Call	Buy Put	Sell Put
Panel A: Number of firms per day					
Mean	1,670	1,390	1,400	1,179	1,164
Max	1,909	1,709	1,738	1,634	1,503
Min	1,395	1,091	1,086	839	874
Std	98	120	115	120	102
Panel B: Transaction statistics					
Mean trade size	17.45	17.74	16.16	18.80	17.67
Mean daily number of trades	273,102	74,114	75,321	49,145	47,147
Mean daily trade volume	4,765,903	1,314,526	1,217,082	924,072	832,880
Mean daily premium (in millions)	1,031.36	254.07	246.89	223.15	206.45
Panel C: Percentages of trading volume by moneyness and maturity					
$OTM : \|delta\| < 37.5\%$	51.94	13.21	11.77	11.62	10.21
$ATM : 37.5\% \leq \|delta\| \leq 62.5\%$	33.91	10.11	9.52	5.75	5.21
$ITM : \|delta\| > 62.5\%$	14.14	4.26	4.25	2.03	2.06
$10 < T \leq 30$ days	36.64	10.25	9.22	7.12	6.46
$31 < T \leq 60$ days	27.91	7.78	7.05	5.46	4.82
$T > 60$ days	35.45	9.55	9.27	6.81	6.20
Panel D: Percentages of trading volume by exchange					
American Stock Exchange	7.15	1.93	1.82	1.32	1.28
BATS	0.05	0.01	0.01	0.01	0.01
Boston Options Exchange	4.70	1.24	1.17	1.01	0.97
Chicago Board Options Exchange	29.66	8.76	8.05	5.79	5.06
International Securities Exchange	27.43	7.06	6.88	5.58	4.97
Nasdaq	2.50	0.64	0.65	0.45	0.44
New York Stock Exchange ARCA	11.59	2.69	2.62	2.08	1.90
Philadelphia Stock Exchange	16.92	4.79	4.40	3.22	2.91

To be included in the final sample, a stock-day observation must have valid stock price information from the CRSP with share code 10 or 11; and the stock must have options listed on a US exchange. On average, the final sample includes 2,207 stocks a day.

4.3. Statistics of the main variables

I construct SOI and OOI as discussed earlier. If there is no option (stock) transaction on a particular day, OOI (SOI) is set to zero. I use the Black and Scholes (1973) model to calculate option delta and for simplicity assume a 0% interest rate and a 0% dividend rate. Panel A of table 2.2 presents the time series averages of the cross-sectional statistics for the main variables. For ease of reporting, I scale stock returns and all order imbalance variables to basis points (bp). The average total order imbalance (TOI) is close to zero (-0.782 bp), with the median even smaller (-0.348 bp). The standard deviation of TOI reaches 21 bp, approximately 27 times its mean. As a result, TOI is not significantly different from zero, indicating that the stock market is well balanced overall. In extreme cases, the maximum imbalance goes beyond 2.7% of all shares outstanding and the minimum is below -3.98%. SOI has statistics very similar to those of TOI with a mean of -0.773 bp and a standard deviation of 21.71 bp. OOI is even smaller on average, with a mean of -0.016 bp. Compared with the stock market, the options market is a more balanced market because the skewness of OOI (0.272) is much smaller than SOI (-3.058). Although the tails of OOI are not as long as SOI, OOI has larger excess kurtosis (327) than SOI (182), indicating that both SOI and OOI have fat tails. The average size of the bid-ask spread is 0.28% of the midpoint price on the stock market. It is interesting to compare the imbalance variables with the turnover ratio because the turnover ratio is also standardized using number of shares outstanding. The average turnover ratio is approximately 1.38%. At its maximum, over 46% of the shares outstanding change hands on one day. Stock and options volumes are reported as log total volumes traded. Average log stock volume is around 13.2 and average log options volume is 7.9. It is obvious that the options market is still not the same size as the stock market.

In panel B, I report the time series averages of the daily cross-sectional correlations. The stock

Table 2.2

Descriptive statistics of main variables

Panel A reports the time series averages of the cross-sectional statistics. *Ret* is the daily return calculated using the midpoint of the last National Best Bid and Offer (NBBO) quote before the close of the stock market. *TOI* is the total stock order imbalance. *SOI* is the order imbalance induced by stock market investors other than option market makers. *OOI* is the option order imbalance measured as *delta* imbalance. *spread* is the percentage bid-ask spread of the mid quote in the stock market. *turnover* is the total stock trading volume scaled by number of shares outstanding, reported as percentages. Vol_{stock} and Vol_{option} are log total stock and option trading volumes from 9:45 to 15:55, respectively. *Ret*, *TOI*, *SOI*, and *OOI* are reported as basis points. Panel B presents the time series averages of the contemporaneous cross-sectional correlations. Panel C gives the cross-sectional averages of autocorrelations up to five lags for each variable.

Panel A: Descriptive statistics

Variables	N	Mean	Std	Median	Min	Max	Skew	Kurt
Ret	1,348,555	4.602	407.019	-6.005	-3269.897	7193.655	2.698	111.010
TOI	1,348,555	-0.782	21.178	-0.348	-397.915	269.771	-3.276	181.018
SOI	1,340,407	-0.773	21.710	-0.343	-403.796	282.378	-3.058	182.394
OOI	1,340,407	-0.016	5.517	-0.001	-106.223	109.196	0.272	327.125
spread	1,348,555	0.279	0.636	0.131	-0.550	14.153	9.929	180.013
turnover	1,348,552	1.375	1.923	0.945	0.018	46.133	10.558	238.320
Vol_{option}	1,348,555	7.900	4.593	9.003	0.000	17.667	-0.658	-0.663
Vol_{stock}	1,346,313	13.195	1.558	13.083	7.623	19.464	0.249	0.149

Panel B: Correlations

	Ret	TOI	SOI	OOI	spread	turnover	Vol_{option}	Vol_{stock}
Ret	1							
TOI	0.211	1						
SOI	0.188	0.962	1					
OOI	0.082	0.067	-0.188	1				
spread	-0.007	-0.040	-0.040	0.000	1			
turnover	0.068	-0.072	-0.071	0.001	-0.034	1		
Vol_{option}	0.027	0.019	0.019	-0.000	-0.258	0.325	1	
Vol_{stock}	0.025	0.008	0.006	0.005	-0.233	0.434	0.674	1

Panel C: Autocorrelations

lag	Ret	TOI	SOI	OOI	spread	turnover	Vol_{option}	Vol_{stock}
1	-0.021	0.105	0.101	0.028	0.170	0.436	0.322	0.358
2	-0.031	0.053	0.051	0.009	0.157	0.309	0.265	0.295
3	0.016	0.035	0.034	0.006	0.152	0.253	0.233	0.256
4	-0.018	0.028	0.028	0.004	0.153	0.226	0.215	0.239
5	-0.028	0.026	0.026	0.003	0.152	0.207	0.202	0.219

returns have large and positive contemporaneous correlations with all order imbalance variables, and the correlation with TOI reaches 0.211, suggesting the order imbalances have strong impact on contemporaneous stock prices. The correlation between TOI and SOI is 0.962, but the correlation between TOI and OOI is only 0.067, suggesting that the majority of TOI is determined by stock market investors. SOI has a larger contemporaneous correlation with returns (0.188) than does OOI (0.082). The correlation between SOI and OOI is negative (-0.188). 50.63% of the observations have SOI and OOI with different signs. A natural explanation of that finding is that although I exclude reported structured trades, investors may have used options to hedge their newly established stock positions–for example, covered calls and protected puts. Such trades are less likely to be based on private information and should weaken the return predictability by adding noise to OOI. The imbalance variables have small correlations with the spread, turnover, and trading volumes. The stock market and the options market tend to be active at the same time because the correlation between the stock volume and the options volume is 0.674.

Panel C presents the cross-sectional averages of the autocorrelations for each variable up to five lags. The autocorrelations of TOI and SOI are strong and positive. For example, the autocorrelation of TOI is 0.105 at the first lag; and it gradually decays to 0.026 at the fifth lag. By comparison, autocorrelation is much smaller for OOI. It is only 0.028 at the first lag and almost dies out at the fifth lag. The bid-ask spread, the turnover ratio, and the stock and options volumes all have highly positive and persistent autocorrelations.

The relationship between OOI and SOI is further examined in table 2.3, which shows the statistics of two OOI-to-SOI ratios: one original and one using the absolute values. To address the ratio explosion when SOI is close to zero, I winsorize both ratios at the 0.5% and 99.5% levels. The original OOI-to-SOI ratio has a mean of -0.028. The standard deviation of the ratio is 0.986. The average absolute ratio is 0.341, but the median is only 0.033. Even the 90th percentile absolute ratio is smaller than 1 (0.629). It is clear that OOI is relatively small compared with SOI and that TOI is dominated by SOI.

Table 2.3
Option order imbalance relative to stock order imbalance
This table compares the magnitude of option order imbalance (*OOI*) and stock order imbalance (*SOI*). Two ratios–one original and one using absolute values–are calculated in the full sample, and detailed statistics are presented. Both ratios are winsorized at 0.5% and 99.5%.

	OOI/SOI	$\lvert OOI/SOI \rvert$
MIN	-19.689	0
1st percentile	-3.354	0
10th percentile	-0.309	0
25th percentile	-0.040	0.002
MEDIAN	-0.000	0.033
75th percentile	0.028	0.174
90th percentile	0.202	0.629
99th percentile	3.088	6.388
MAX	25.094	35.898
MEAN	-0.028	0.341
STD	0.986	1.360

5. Empirical results

5.1. Predicting stock returns using SOI and OOI

Table 2.4 contains the main results using Fama and MacBeth (1973) regressions. Reported are average slope coefficients estimated from the cross-sectional regressions and the Newey and West (1987) adjusted t-statistics. The coefficient estimates show little autocorrelation. For many series, the autocorrelation estimates are insignificant even at the first lag. The maximum length of significant lag for all series is three. To be conservative in the reported t-statistics, I choose to use the maximum length of three lags for all Newey-West standard error calculation. I first test the predictive ability of TOI in the first column by estimating the following equation:

$$Ret_{i,t} = \alpha + \beta TOI_{i,t-1} + \varepsilon_{i,t}. \quad (2.5)$$

TOI has an average slope coefficient of -0.002, and the t-statistics is -0.05, suggesting TOI has no significant predictive power. Column (2) presents regression results from using decomposed order imbalances. The OOI coefficient 0.59 is statistically significant at the 1% level (t-statistic = 5.59). The SOI coefficient is positive but insignificant (t-statistic = 0.54). Only OOI positively and significantly predicts the next day's returns. I then add lagged order imbalance variables to the model to investigate whether the predictive relations would reverse at longer horizons. Column (3) shows that SOI on day $t-2$ has a significant and negative coefficient of -0.106 (t-statistic = -2.66) but the lagged OOIs have only insignificant coefficients. The full model in equation (2.3) is examined in column (4). With microstructure controls, OOI still positively predicts the next day's returns. The reversal effect of SOI becomes less significant as the $t-2$ SOI coefficient drops to -0.065 (t-statistic = -1.9). I also find that both stock and option trading volumes significantly predict returns, but in opposite directions. Large stock volume is associated with positive future returns, but large option volume is associated with negative returns. That finding is consistent with a recent paper by Johnson and So (2012), which shows that option-to-stock volume ratio negatively predicts stock returns.

There can be concerns that during the sample period, the market experiences a major financial crisis in 2008, when unusual market volatility and frequent regulation intervention could have distorted the generality of the results. For a robustness check, I reestimate the full model using data from 2009 and 2010 only. The results are reported in column (5). Comparing columns (4) and (5), I find the results to be quite similar, except that in the after-crisis period, the first lag SOI becomes marginally significant. For robustness check, I also measure SOI and OOI at one hour before market close and repeat the full regression in column (6). The results are largely the same.

The results in this table have important implications for the first two hypotheses. On one hand, I find SOI has weak predictive power of future returns. The reversal effect on day $t+2$ indicates transitory price pressure from SOI rather than information. On the other hand, OOI positively predicts future returns and the predictability does not reverse at longer horizons, thereby providing unambiguous evidence that options order flow contains a significant amount of private information about the underlying stock's price movement.

Table 2.4

Daily regressions of stock returns on lagged stock and option order imbalances

The first column reports the Fama-MacBeth regression results of the following equation:

$$Ret_{i,t} = \alpha + \beta TOI_{i,t-1} + \varepsilon_{i,t}.$$

$Ret_{i,t}$ is stock i's return calculated using the midpoint of the last NBBO quote before market close on day t. $TOI_{i,t-1}$ is stock i's total order imbalance on day $t-1$. The rest of the columns present the Fama-MacBeth regression results based on the following equation:

$$Ret_{i,t} = \alpha + \sum_{k=1}^{5} \beta_1^k SOI_{i,t-k} + \sum_{k=1}^{5} \beta_2^k OOI_{i,t-k} + \theta X_{i,t-1} + \varepsilon_{i,t}.$$

$SOI_{i,t-k}$ is stock i's order imbalance induced by stock market investors other than option market makers on day $t-k$. $OOI_{i,t-k}$ is the option order imbalance measured as *delta* imbalance. $X_{i,t-1}$ is a set of control variables on day $t-1$ including $Ret_{i,t-k}$, stock returns for the previous five days; $OpRet_{i,t-k}$, equally-weighted option returns across all option contracts on stock i for the previous five days; *spread*, the percentage stock bid-ask spread; *turnover*, the ratio of total stock trading volume to the number of shares outstanding; Vol_{stock}, log total stock volume from 9:45 to 15:55; and Vol_{option}, log total option volume from 9:45 to 15:55. Column (5) gives results in the subsample period from 2009 to 2010. Column (6) reports regression results using order imbalances measured at 15:00 every day. Standard errors are calculated with Newey-West adjustment to three lags. T-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
intercept	4.957	3.978	3.899	-35.116***	-26.165**	-29.801**
	(0.51)	(0.41)	(0.40)	(-2.90)	(-2.09)	(-2.53)
TOI_{t-1}	-0.002					
	(-0.05)					
SOI_{t-1}		0.023	0.037	0.019	0.073*	0.120***
		(0.54)	(0.85)	(0.48)	(1.65)	(2.73)
SOI_{t-2}			-0.106***	-0.065*	-0.054*	-0.064*
			(-2.66)	(-1.90)	(-1.67)	(-1.72)
SOI_{t-3}			0.046	0.034	0.027	0.045
			(1.02)	(1.12)	(0.89)	(1.17)
SOI_{t-4}			0.057	-0.008	-0.041	0.002
			(0.60)	(-0.27)	(-1.11)	(0.06)
SOI_{t-5}			-0.059	-0.028	-0.087	0.042
			(-0.65)	(-0.38)	(-0.84)	(1.15)
OOI_{t-1}		0.590***	0.531***	0.402***	0.336***	0.641***
		(5.59)	(5.12)	(4.17)	(2.96)	(3.57)
OOI_{t-2}			0.036	0.077	0.088	0.145
			(0.27)	(0.62)	(0.72)	(0.59)
OOI_{t-3}			-0.070	-0.075	-0.090	-0.043
			(-0.71)	(-0.79)	(-0.83)	(-0.30)
OOI_{t-4}			-0.058	-0.016	-0.094	-0.157
			(-0.49)	(-0.16)	(-0.89)	(-0.94)
OOI_{t-5}			-0.018	-0.029	-0.088	-0.169
			(-0.16)	(-0.26)	(-0.72)	(-1.06)

Table 2.4 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)
Ret_{t-1}				0.003	0.006	0.002
				(0.71)	(1.21)	(0.54)
Ret_{t-2}				-0.004	-0.006	-0.004
				(-1.15)	(-1.37)	(-1.15)
Ret_{t-3}				-0.005	-0.005	-0.005
				(-1.39)	(-1.19)	(-1.49)
Ret_{t-4}				-0.001	-0.001	-0.001
				(-0.30)	(-0.37)	(-0.34)
Ret_{t-5}				-0.003	0.000	-0.002
				(-0.62)	(-0.04)	(-0.54)
$OpRet_{t-1}$				-1.468	-1.581	-1.689
				(-0.87)	(-0.96)	(-0.99)
$OpRet_{t-2}$				1.059	0.517	1.243
				(0.64)	(0.31)	(0.75)
$OpRet_{t-3}$				0.686	1.816	0.833
				(0.45)	(1.09)	(0.54)
$OpRet_{t-4}$				1.092	1.098	1.096
				(0.80)	(0.91)	(0.81)
$OpRet_{t-5}$				-0.787	-0.528	-0.829
				(-0.55)	(-0.34)	(-0.58)
spread				4.258	8.342**	4.524
				(1.32)	(1.96)	(1.40)
turnover				0.008	0.000	0.011
				(0.88)	(-0.01)	(1.23)
Vol_{stock}				2.773***	2.791***	2.341***
				(3.52)	(3.18)	(3.01)
Vol_{option}				-0.763***	-0.671**	-0.679***
				(-3.41)	(-2.44)	(-3.06)
N of obs	1348290	1342322	1297690	1294304	892120	1294304
Adj R-sq	0.008	0.009	0.021	0.067	0.063	0.067

Table 2.5
Alphas of order imbalance strategies
This table reports average daily returns on equally-weighted quintile portfolios based on total order imbalance (TOI), stock order imbalance (SOI), and option order imbalance (OOI) as well as abnormal returns (alpha) from long-short portfolios. The portfolios are rebalanced every day at market close based on order imbalance signals from 9:45 to 15:55. Strategy alphas are reported in three forms: the original, the Fama-French three-factor, and four-factor adjusted alphas. All returns are reported as basis points. *Sharpe* is the annualized Sharpe ratio. T-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	TOI	SOI	OOI
low - 1	8.720	8.929	2.095
2	3.511	2.739	4.405
3	0.510	1.804	7.018
4	0.856	3.411	4.908
high - 5	11.333	10.396	10.831
5-1	2.613	1.467	8.736***
t-stat	(1.08)	(0.81)	(6.03)
FF3 alpha	2.552	1.486	8.787***
t-stat	(1.07)	(0.85)	(6.28)
FF4 alpha	2.268	1.228	8.600***
t-stat	(0.95)	(0.71)	(6.16)
Sharpe	0.696	0.426	3.754

I then perform a simple investment analysis to gauge the economic significance of the return predictability. On each day, quintile portfolios are formed based on each order imbalance variable. The top quintile portfolio with the most negative order imbalance is defined as the “sell” portfolio, and the bottom quintile portfolio with the most positive order imbalance is defined as the “buy” portfolio. A zero-investment portfolio is constructed by buying all of the stocks in the buy portfolio and selling all of the stocks in the sell portfolio with equal weight at market close. All of the portfolios are re-balanced the next day. Because I measure all order imbalance variables at 15:55 every day, there is a five-minute window for those trading strategies to be executed.

Table 2.5 gives the portfolio returns. TOI strategy generates a V-shape pattern in quintile portfolio returns, and the long-short portfolio does not generate a significant return (t-statistic = 1.08). The average returns on the quintile portfolios based on SOI have the same V-shape pattern, and there is no significant return on the long-short portfolio either. The returns on OOI portfolios increase almost monotonically across the quintile portfolios. The "sell" portfolio has an average return of 2.095 bp per day, and the "buy" portfolio has an average return of 10.831 bp. The daily abnormal return reaches 8.736 bp (t-statistic = 6.03), annualized to 22%. The abnormal return is significant at the 1% level, after controlling for Fama-French risk factors and the momentum factor. The annual Sharpe ratio of the OOI strategy reaches 3.754.

5.2. An intraday analysis

In this subsection, I investigate high-frequency return predictability from SOI and OOI at half-hour intervals. The empirical test is based on the following equation:

$$Ret_{t,t+1} = \alpha + \sum_{k=1}^{2} \beta_1^k SOI_{t-k,t-k+1} + \sum_{k=1}^{2} \beta_2^k OOI_{t-k,t-k+1} + \theta X_t + \varepsilon_{t+1}, \tag{2.6}$$

where $Ret_{t,t+1}$ is the stock return calculated using the midpoints of the first and the last NBBO quoted prices from time t to $t+1$, with the time unit being half an hour; and X_t includes the NBBO returns for the last two periods, the percentage bid-ask spread at time t, log total stock volume, and log total option volume in the last half hour. Firm subscription is omitted. When constructing order imbalances, I exclude the first half hour and the last hour of each trading day and I do not use order imbalances from the previous day in the regressions. I choose two lags of the imbalance variables in equation 2.6 as a result of balancing the goals of detecting potential reversal effects and preserving the sample size from missing values. Therefore, the analysis starts in the fourth half hour of a day. Each stock then has nine half-hour observations that can be used for estimating equation (2.6) on each day.

Table 2.6

Predicting stock returns using order imbalances at half-hour intervals

This table investigates intraday relations between stock returns and order imbalances. Each trading day is divided into 13 half-hour intervals, and the first three intervals and the last interval are excluded. Fama-MacBeth regression results are presented for the full sample as well as for the nine half-hour intervals separately from the following equation:

$$Ret_{t,t+1} = \alpha + \sum_{k=1}^{2} \beta_1^k SOI_{t-k,t-k+1} + \sum_{k=1}^{2} \beta_2^k OOI_{t-k,t-k+1} + \theta X_t + \varepsilon_{t+1}.$$

The dependent variable $Ret_{t,t+1}$ is the stock return calculated using the midpoint of NBBO quotes from time t to $t+1$, and the time unit is half an hour. $SOI_{t-1,t}$ is the stock order imbalance induced by stock market investors other than option market makers. $OOI_{t-1,t}$ is the option order imbalance measured as *delta* imbalance. X_t is a set of control variables at time t, including $Ret_{t-2,t-1}$ and $Ret_{t-1,t}$, the NBBO returns for the last two periods; *spread*, the percentage stock bid-ask spread; $StVol_{t-1,t}$, log total stock volume in the last half hour; and $OpVol_{t-1,t}$, log total option volume in the last half hour. The firm subscription is omitted for all variables. T-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1) all	(2) t=11:00	(3) t=11:30	(4) t=12:00	(5) t=12:30	(6) t=13:00	(7) t=13:30	(8) t=14:00	(9) t=14:30	(10) t=15:00
intercept	-2.488***	-5.249***	-2.448	-4.354**	-2.789**	0.396	-1.004	-3.382**	-1.678	-1.885
	(-3.78)	(-2.77)	(-1.34)	(-2.58)	(-2.00)	(0.13)	(-0.69)	(-2.14)	(-0.96)	(-0.88)
$SOI_{t-1,t}$	0.082**	0.093**	0.117***	0.090*	0.015	0.175***	0.062	0.067	-0.197	0.318
	(2.12)	(2.04)	(2.69)	(1.82)	(0.33)	(2.74)	(1.45)	(0.53)	(-1.37)	(1.02)
$SOI_{t-2,t-1}$	-0.075***	-0.106***	-0.056	-0.070**	-0.068*	-0.017	-0.104***	-0.003	-0.070	-0.176
	(-3.86)	(-2.93)	(-1.50)	(-2.02)	(-1.76)	(-0.32)	(-2.73)	(-0.07)	(-0.77)	(-1.55)
$OOI_{t-1,t}$	0.353***	0.098	0.540***	0.483***	0.314**	0.446***	0.166	0.509***	0.271*	0.351**
	(3.48)	(0.47)	(3.16)	(3.02)	(2.01)	(2.79)	(0.22)	(3.05)	(1.72)	(2.25)
$OOI_{t-2,t-1}$	-0.017	-0.019	0.067	0.071	0.091	-0.168	-0.010	-0.402	0.309*	-0.092
	(-0.18)	(-0.14)	(0.43)	(0.49)	(0.65)	(-0.89)	(-0.05)	(-0.54)	(1.94)	(-0.59)
$Ret_{t-1,t}$	-0.02***	-0.023***	-0.011***	-0.028***	-0.017***	-0.034***	-0.013***	-0.034***	-0.006	-0.014***
	(-14.54)	(-6.54)	(-3.00)	(-7.46)	(-4.54)	(-7.30)	(-3.35)	(-7.64)	(-1.54)	(-3.06)
$Ret_{t-2,t-1}$	-0.005***	-0.011***	-0.002	-0.006**	-0.001	-0.012*	-0.002	-0.008*	-0.001	-0.004
	(-3.94)	(-4.34)	(-0.68)	(-2.32)	(-0.19)	(-1.92)	(-0.56)	(-1.93)	(-0.13)	(-0.93)
$spread_t$	-1.773***	-2.678***	-1.017	-0.620	-0.806	-1.546**	-1.801***	-2.815***	-1.584**	-3.088***
	(-6.82)	(-3.70)	(-0.92)	(-1.04)	(-1.39)	(-2.00)	(-2.78)	(-4.34)	(-2.35)	(-4.04)
$StVol_{t-1,t}$	0.242***	0.234	0.190	0.383***	0.189	0.078	0.032	0.211	0.466***	0.392**
	(4.12)	(1.34)	(1.13)	(2.73)	(1.47)	(0.29)	(0.23)	(1.46)	(3.26)	(2.21)
$OpVol_{t-1,t}$	-0.058***	-0.093**	-0.009	-0.114***	-0.033	-0.086**	-0.027	-0.027	-0.039	-0.097**
	(-4.25)	(-2.27)	(-0.16)	(-3.61)	(-1.11)	(-2.41)	(-0.55)	(-0.79)	(-1.17)	(-2.65)

Column (1) of table 2.6 gives Fama-MacBeth regression results in the full sample. SOI positively predicts the next half-hour returns with an estimated coefficient of 0.082 (t-statistic = 2.12). However, the predictive relation reverses to significantly negative (t-statistic = -3.86) in the following half hour. The $SOI_{t-2,t-1}$ coefficient is -0.075–almost offsetting the $SOI_{t-1,t}$ impact completely. The $OOI_{t-1,t}$ coefficient is 0.353 (t-statistic = 3.48), and the second lag OOI has an insignificant coefficient (t-statistic = -0.18). The rest of the columns show regression results for each half-hour interval separately. For example, column (2) gives the time series averages of the coefficients estimated from cross-sectional regressions using only the first observation of the day. Out of the nine half-hour trading periods, both $SOI_{t-1,t}$ and $SOI_{t-2,t-1}$ have 5% significant coefficients in three periods; and $OOI_{t-1,t}$ has 5% significant coefficients in six periods. The high frequency results confirm the permanent price impact of OOI and the transitory price pressure of SOI.

5.3. Option leverage

Having established the link between options order flow and future stock prices, I then test hypothesis 3, that informed traders prefer to use options with higher leverage. Table 2.7 contains the regression results of equation (2.4). I first show univariate regression results. Column (1) demonstrates that OOI constructed using only OTM options does not predict returns. This OOI measure is not informative although OTM options provide option investors the largest leverage. Columns (2) and (3) show that OOI constructed using either ATM options or ITM options significantly predicts future return. The multivariate regression result in column (4) confirms the finding. The result is also robust after adding option volumes in column (5) and the full microstructure controls in column (6). Therefore, I find no empirical support for hypothesis 3. A possible explanation is that the transaction costs associated with OTM options could be too high. I find that during the sample period, the average percentage bid-ask spread of OTM options reaches 11.4%. For ATM and ITM options, the percentage bid-ask spreads are only 3.46% and 3%, respectively. This finding is consistent with the theoretical prediction of Johnson and So (2012). If informed traders buy OTM options to gain high leverage, it is likely that they would need to reverse trade positions rather than

Table 2.7
Return predictability from option order imbalance by moneyness groups
This table presents the Fama-MacBeth regression results of the following equation:

$$Ret_{i,t} = \alpha + \beta_1 OTM\_OOI_{i,t-1} + \beta_2 ATM\_OOI_{i,t-1} + \beta_3 ITM\_OOI_{i,t-1} + \theta X_{i,t-1} + \varepsilon_{i,t}.$$

$Ret_{i,t}$ is stock i's return calculated using the midpoint of the last NBBO quote before market close on day t. $OTM\_OOI_{i,t-1}$ is the option order imbalance measured as *delta* imbalance calculated using out-of-the-money (OTM) contracts. $ATM\_OOI_{i,t-k}$ is the option order imbalance calculated using at-the-money (ATM) contracts. $ITM\_OOI_{i,t-k}$ is the option order imbalance calculated using in-the-money (ITM) contracts. $X_{i,t-1}$ is a set of control variables on day $t-1$, including $SOI_{i,t-1}$, the stock order imbalance induced by stock market investors other than option market makers on day $t-1$; $Ret_{i,t-k}$, stock returns for the previous five days; *spread*, the percentage stock bid-ask spread; *turnover*, the ratio of total stock trading volume to number of shares outstanding; Vol_{stock}, log total stock volume from 9:45 to 15:55; Vol_{OTM}, Vol_{ATM}, and Vol_{ITM}, the log trading volumes for OTM, ATM, and ITM options, separately. Coefficients and t-statistics (in parentheses) are reported for main variables only. Standard errors are calculated with Newey-West adjustment to three lags. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
OTM_OOI_{t-1}	0.651			0.435	0.549
	(0.98)			(0.64)	(0.94)
ATM_OOI_{t-1}		0.778***		0.740***	0.608***
		(3.03)		(2.81)	(2.66)
ITM_OOI_{t-1}			1.068***	0.977***	0.599*
			(2.85)	(2.64)	(1.77)
control	no	no	no	no	yes

exercise the options to take profit. In so doing, they pay the full amount of the bid-ask spread, thereby sharply increasing their transaction costs. The results in table 2.7 suggest that consideration of transaction costs can outweigh the desire for leverage when informed traders allocate their orders across options with different moneyness.

5.4. Level of information asymmetry

This subsection tests hypothesis 4 that OOI has greater return predictability for firms with more information asymmetry. Based on each of the five proxies for information asymmetry, I divide the full sample into three subgroups: low, medium, and high. I then run full specification regressions of equation (2.3) in each subgroup. Because the main interest of the test is the predictive ability of OOI in subgroups, I report in table 2.8 only the estimated coefficients and *t*-statistics of OOI_{t-1}. Panel A shows that the OOI coefficient is significant at the 1% level (*t*-statistic = 2.85) in the high PIN group, marginally significant in the medium PIN group (*t*-statistic = 1.65), but not significant in the low PIN group. Panel B shows that the OOI coefficient is significant at the 1% level (*t*-statistic = 2.93) in the low analyst coverage group, significant at the 5% level in the medium coverage group (*t*-statistic = 2.03), but not significant in the high coverage group. Panel C shows that OOI has significant coefficients in all subgroups based on the bid-ask spread. However, the coefficient is 0.901 (*t*-statistic = 3.03) in the high spread group, which is much greater than the coefficient of 0.399 (*t*-statistic = 1.66) in the low spread group. Similar patterns of the OOI coefficients are found in panels D and E based on the adverse-selection component of the spread, and in panel F based on the firm size. Collectively, those results suggest that consistent with hypothesis 4, OOI is more informative for firms with higher levels of information asymmetry.

5.5. Institutional ownership

This subsection describes the testing of hypothesis 5 that information content of OOI increases in short-sale costs of a stock measured by institutional ownership. Panel G in table 2.8 gives the results. OOI is statistically significant at the 10% level in all ownership subgroups. However, the OOI coefficient is 0.777 (*t*-statistic = 3.64) in the low ownership group and 0.237 (*t*-statistic = 1.74) in the high ownership group. The differences of the magnitude of the coefficients and *t*-statistics demonstrate that OOI has a larger amount of information for firms with low institutional ownership and high short-sale costs, supporting hypothesis 5.

Table 2.8

Predictive power of order imbalances and firm characteristics

This table presents the subgroup results based on firm characteristics. For each firm characteristic variable, the full sample is divided into three groups: low (< 30th percentile), medium (30th $\sim$ 70th percentile), and high (> 70th percentile). Within each group, I estimate the following equation using Fama-MacBeth regressions:

$$Ret_{t,t+1} = \alpha + \sum_{k=1}^{2} \beta_1^k SOI_{t-k,t-k+1} + \sum_{k=1}^{2} \beta_2^k OOI_{t-k,t-k+1} + \theta X_t + \varepsilon_{t+1}.$$

Average slope coefficients and t-statistics (in parentheses) are reported for option order imbalance on day $t-1$ (OOI_{t-1}) only. *PIN* is the probability of informed trading from Easley and O'Hara (1992). *Analyst coverage* is the number of analysts following. *Spread* is the percentage stock bid-ask spread. *GH adverse spread* is the adverse selection component of the bid-ask spread based on the model of Glosten and Harris (1988). *LSB adverse spread* is the adverse-selection component of the bid-ask spread based on the model of Lin, Sanger, and Booth (1995). *Size* is the market capitalization. *Ownership* is the percentage of shares held by institutional investors. *Option volume* is the total option volume traded. Standard errors are calculated with Newey-West adjustment to three lags. T-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: PIN				*Panel B: Analyst coverage*			
	low	*medium*	*high*		*low*	*medium*	*high*
OOI_{t-1}	0.219	0.253*	0.505***	OOI_{t-1}	0.965***	0.455**	0.416
	(0.74)	(1.65)	(2.85)		(2.93)	(2.03)	(1.57)
Panel C: Spread				*Panel D: GH adverse spread*			
	low	*medium*	*high*		*low*	*medium*	*high*
OOI_{t-1}	0.399*	0.429**	0.901***	OOI_{t-1}	0.220	0.358*	0.532***
	(1.66)	(2.12)	(3.03)		(0.78)	(1.87)	(2.66)
Panel E: LSB adverse spread				*Panel F: Size*			
	low	*medium*	*high*		*small*	*medium*	*large*
OOI_{t-1}	0.278	0.513**	0.516***	OOI_{t-1}	1.216***	0.276	0.240*
	(1.44)	(2.09)	(3.28)		(2.82)	(1.24)	(1.83)
Panel G: Ownership				*Panel H: Option volume*			
	low	*medium*	*high*		*low*	*medium*	*high*
OOI_{t-1}	0.777***	0.688***	0.237*	OOI_{t-1}	62.376	1.699***	0.474***
	(3.64)	(2.69)	(1.74)		(1.46)	(2.66)	(4.18)

5.6. Market activeness

Hypothesis 6 is tested in this subsection. Panel G in table 2.8 shows that OOI does not significantly predict returns when option trading is inactive and the options trading volume is low. The OOI coefficients are statistically significant at the 1% level in both the medium volume group (t-statistic = 2.66) and the high volume group (t-statistic = 4.18), suggesting that the predictive ability of OOI comes from periods when the options market is active. The results are consistent with the information explanation for the return predictability of OOI.

6. Further analysis

The previous section shows that options order flow contains important price information about the underlying stocks. In this section, I perform additional analysis to better understand that information linkage.

6.1. Asymmetric price response

I first investigate whether there exist asymmetric price responses to the imbalance variables by estimating the following equation:

$$Ret_{i,t} = \alpha + \beta_1 SOI^{+}_{i,t-1} + \beta_2 SOI^{-}_{i,t-1} + \beta_3 OOI^{+}_{i,t-1} + \beta_4 OOI^{-}_{i,t-1} + \theta X_{i,t-1} + \varepsilon_{i,t}, \qquad (2.7)$$

where $SOI^{+}_{i,t-1} = max(SOI_{i,t-1}, 0)$, $SOI^{-}_{i,t-1} = min(SOI_{i,t-1}, 0)$, $OOI^{+}_{i,t-1} = max(OOI_{i,t-1}, 0)$, and $OOI^{-}_{i,t-1} = min(OOI_{i,t-1}, 0)$. The results are reported in the first column of table 2.9. The β_1 estimate is 0.286, statistically significant at the 1% level (t-statistic = 4.24). The β_2 estimate is -0.238, also significant at the 1% level (t-statistic = -3.63). The results suggest that both positive and negative SOI significantly predict stock returns, but in the opposite directions. Therefore, the overall SOI does not predict returns. The β_3 estimate is insignificant and the β_4 estimate is significant at the 1% level (t-statistic = 4.63), suggesting that the information content in OOI mainly comes from negative OOI, supporting the hypothesis that informed traders use options to get around the short-sale constraints when they acquire negative information.

6.2. Nonlinear price impact

This subsection investigates nonlinearity in the predictive relationship. Both SOI and OOI have fat tails. It is possible that the stock price response will be different in the tails of order imbalances. I then estimate the following specification:

$$Ret_{i,t} = \alpha + \beta_1 SOI_{i,t-1} + \beta_2 OOI_{i,t-1} + \beta_3 SOI^2_{i,t-1} \cdot S1_{i,t-1} + \beta_4 OOI^2_{i,t-1} \cdot S2_{i,t-1} + \theta X_{i,t-1} + \varepsilon_{i,t}, \tag{2.8}$$

where S_1 and S_2 are dummy variables that equal 1 when SOI and OOI are positive, respectively, and -1 otherwise. The interaction terms quadratically amplify the impact of the two imbalance variables and will parse out the tail effect. The second column of table 2.9 gives regression results. Both SOI and OOI coefficients are positive and statistically significant at the 5% level. No significant nonlinear effect is found on SOI because β_3 is insignificant. However, β_4 is negative and significant at the 1% level (t-statistic = -3.5), indicating that the predictive ability of OOI reduces in the tails.

6.3. Moving average and shocks

Both SOI and OOI have positive autocorrelations. In this subsection, I decompose each imbalance variable into a moving average (MA) component and a shock component to further investigate the sources of the price impact. For example,

$$OOI = OOI_{MA_k} + OOI_{shock_k}, \tag{2.9}$$

where OOI_{MA_k} is the average OOI from the previous k trading days. For simplicity, I omit the firm and time subscription in equation (2.9). The same decomposition is also performed for SOI. Then,

Table 2.9
Nonlinear price impact from option and stock order imbalances
The first column presents the Fama-MacBeth regression results from the equation:

$$Ret_{i,t} = \alpha + \beta_1 SOI^{+}_{i,t-1} + \beta_2 SOI^{-}_{i,t-1} + \beta_3 OOI^{+}_{i,t-1} + \beta_4 OOI^{-}_{i,t-1} + \theta X_{i,t-1} + \varepsilon_{i,t},$$

where $Ret_{i,t}$ is stock i's return calculated using the midpoint of the last NBBO quote before market close on day t, $SOI^{+}_{i,t-1} = max(SOI_{i,t-1}, 0)$, $SOI^{-}_{i,t-1} = min(SOI_{i,t-1}, 0)$, $OOI^{+}_{i,t-1} = max(OOI_{i,t-1}, 0)$, and $OOI^{-}_{i,t-1} = min(OOI_{i,t-1}, 0)$. $Xi,t-1$ includes the stock returns and equally-weighted option returns in the previous five days, the percentage bid-ask spread, the ratio of total stock volume to number of shares outstanding, log total stock and option volumes. The second column presents the regression results from the following equation:

$$Ret_{i,t} = \alpha + \beta_1 SOI_{i,t-1} + \beta_2 OOI_{i,t-1} + \beta_3 SOI^{2}_{i,t-1} \cdot S1_{i,t-1} + \beta_4 OOI^{2}_{i,t-1} \cdot S2_{i,t-1} + \theta X_{i,t-1} + \varepsilon_{i,t},$$

where $S1_{i,t-1}$ and $S2_{i,t-1}$ are dummy variables that equal 1 when $SOI_{i,t-1}$ and $OOI_{i,t-1}$ are positive, respectively, and -1 otherwise. Results are reported for main variables only but the regressions include all control variables. Standard errors are calculated with Newey-West adjustment to three lags. T-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)
$SOI+$	0.286***	
	(4.24)	
$SOI-$	-0.238***	
	(-3.63)	
$OOI+$	0.121	
	(0.71)	
$OOI-$	0.802***	
	(4.63)	
SOI		0.122**
		(2.53)
OOI		1.111***
		(5.75)
$S_1 \cdot SOI^2$		-0.001
		(-1.39)
$S_2 \cdot OOI^2$		-0.020***
		(-3.50)

I estimate the following equation:

$$Ret_{i,t} = \alpha + \beta_1 SOI_{MA_{i,t-1}} + \beta_2 SOI_{shock_{i,t-1}} + \beta_3 OOI_{MA_{i,t-1}} + \beta_4 OOI_{shock_{i,t-1}} + \theta X_{i,t-1} + \varepsilon_{i,t}, \tag{2.10}$$

Three MA periods of 3 days, 5 days, and 10 days are chosen, and the results are reported in table 2.10. Panel A shows that the time series averages of the daily cross-sectional correlations between the MAs and the shocks are large and negative for both SOI and OOI. For example, the 3-day correlations are -0.468 and -0.514 for SOI and OOI, respectively. As the length of the MA increases, the correlations become weaker.

Panel B reports the regression results of equation (2.10). The first column reports the result using a 3-day MA decomposition. Although the MA's and the shocks are negatively correlated, both of them are positively correlated with future stock returns. Neither the MA or the shock of SOI significantly predicts stock returns. For OOI, both the MA and the shock have significant coefficient estimates. But the shock component has a larger t-statistic (3.76) than the MA (1.79). As the MA horizon increases, the predictive ability of the OOI MA becomes slightly weaker, but the predictive ability of the OOI shock component stays unchanged.

6.4. Earnings announcement

This subsection investigates the relationship between order imbalances and earnings announcements. Several studies such as Skinner (1990) and Ho (1993) find that stocks with options exhibit smaller price reactions to earnings surprise, suggesting that option trading disseminates new information regarding earnings. Amin and Lee (1997) show that an increase in option open interest before an announcement is related to the direction of the earnings surprise. More recently, Roll, Schwartz, and Subrahmanyam (2010) and Johnson and So (2012) find that the option-to-stock volume ratio predicts the earnings surprise and the cumulative abnormal returns (CAR) around the announcement. If option trading contains private information, the return predictability from OOI should also be prominent around earnings announcements.

Table 2.10

Moving averages and shocks in order imbalances and return predictability

Panel A presents the time series averages of the cross-sectional correlations between the 3-day, 5-day, and 10-day moving averages and the shock components for stock order imbalance (SOI) and option order imbalance (OOI). Panel B reports the Fama-MacBeth regression results of the equation:

$$Ret_{i,t} = \alpha + \beta_1 SOI_{MA_{i,t-1}} + \beta_2 SOI_{shock_{i,t-1}} + \beta_3 OOI_{MA_{i,t-1}} + \beta_4 OOI_{shock_{i,t-1}} + \theta X_{i,t-1} + \varepsilon_{i,t},$$

where $Ret_{i,t}$ is stock i's return calculated using the midpoint of the last NBBO quote before market close on day t and $Xi,t-1$ includes the stock returns and equally-weighted option returns in the previous five days, the percentage bid-ask spread, the ratio of total stock volume to number of shares outstanding, log total stock and option volumes. Results are reported for main variables only but the regressions include all control variables. Standard errors are calculated with Newey-West adjustment to three lags. T-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Correlations			
	3day	5day	10day
SOI_{MA} and SOI_{shock}	-0.468	-0.381	-0.278
OOI_{MA} and OOI_{shock}	-0.514	-0.435	-0.343

Panel B: Regression Results			
	3day	5day	10day
SOI_{MA}	-0.033	-0.069	-0.142
	(-0.51)	(-1.00)	(-1.47)
SOI_{shock}	-0.014	-0.013	-0.015
	(-0.27)	(-0.26)	(-0.29)
OOI_{MA}	0.371*	0.460*	0.392
	(1.79)	(1.90)	(1.44)
OOI_{shock}	0.359***	0.360***	0.344***
	(3.76)	(3.79)	(3.58)

I collect earnings announcements data from I/B/E/S and merge that with the options sample. The final sample contains 21,103 observations on 2,647 firms. I follow Carhart (1997) and calculate risk-adjusted returns for each firm. Then I construct CAR around four alternative event windows: (0,2), (0,4), (-1,1), and (-2,2). For example, $CAR_{(0,2)}$ represents the three-day cumulative risk-adjusted returns from the announcement day t to day $t+2$. Mirroring previous tests, I estimate the following equation:

$$CAR_{(j,k)} = \alpha + \sum_{t=1}^{5+j} \beta_1^k SOI_{j-t} + \sum_{t=1}^{5+j} \beta_2^k OOI_{j-t} + \theta X_{j-1} + \varepsilon_{j,k}, \tag{2.11}$$

where the independent variables are the same as defined before. Table 2.11 presents cross-sectional regression results with standard errors clustered by firms. Across different event windows, OOI_{-5} always has significant coefficient estimates. For example, in a prediction of $CAR_{(0,2)}$, the OOI_{-5} coefficient is 4.289 (t-statistic = 2.64). The predictive ability of OOI_{-5} is largely the same in the other event windows. The SOI coefficients are mostly insignificant, except SOI_{-2} when predicting $CAR_{(-1,1)}$. The results indicate that the options market reveals information about scheduled earnings disclosure five days before the announcement.

Informed trading should be more active when potential profit is high. I use two proxies for potential profit: (i) earnings surprise, defined as the absolute value of the difference between actual earnings per share (EPS) and analyst consensus of EPS forecast immediately prior to the announcement, and (ii) analyst dispersion, defined as the standard deviation of analyst EPS forecast. Similarly to table 2.8, table 2.12 presents the coefficients and t-statistics of OOI_{-5} from subgroup regression results. The dependent variable in these regressions is $CAR_{(0,2)}$ because alternative event windows generate even stronger results. Panel A shows that the OOI_{-5} coefficient is statistically significant at the 1% level (8.214, t-statistic = 2.98) in the high surprise group, but not significant in the other two groups, suggesting that the predictive ability of OOI_{-5} comes mainly from those events with large earnings surprises. Panel B demonstrates that OOI_{-5} significantly predicts CAR for both positive and negative forecasting errors. However, the negative error group has a larger coefficient estimate and a larger t-statistic than the positive error group, possibly due to the short-

Table 2.11

Predicting earnings cumulative abnormal returns using order imbalances

This table presents the cross-sectional regression results of the following equation:

$$CAR_{(j,k)} = \alpha + \sum_{t=1}^{5+j} \beta_1^k SOI_{j-t} + \sum_{t=1}^{5+j} \beta_2^k OOI_{j-t} + \theta X_{j-1} + \varepsilon_{j,k}.$$

where $CAR_{j,k}$ is the cumulative abnormal return from event day j to event day k relative to the earnings announcement day and $Xi,t-1$ includes the stock returns and equally-weighted option returns in the previous five days, the percentage bid-ask spread, the ratio of total stock volume to number of shares outstanding, log total stock and option volumes. Firm subscription is omitted for all variables. Results are reported for main variables only but the regressions include all control variables. Standard errors are clustered by firms. T-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	$CAR_{(0,2)}$	$CAR_{(0,4)}$	$CAR_{(-1,1)}$	$CAR_{(-2,2)}$
OOI_{-1}	-1.653	-1.031		
	(-1.13)	(-0.66)		
OOI_{-2}	2.170	2.407	1.026	
	(1.43)	(1.45)	(0.79)	
OOI_{-3}	1.072	0.429	0.816	0.343
	(0.75)	(0.26)	(0.59)	(0.16)
OOI_{-4}	-0.650	-1.471	-0.389	-0.596
	(-0.53)	(-1.19)	(-0.31)	(-0.44)
OOI_{-5}	4.289***	4.860***	4.212***	4.31**
	(2.64)	(2.55)	(2.71)	(2.08)
SOI_{-1}	-0.101	0.224		
	(-0.19)	(0.43)		
SOI_{-2}	0.846*	0.701	1.265**	
	(1.76)	(1.25)	(2.54)	
SOI_{-3}	-0.527	-0.508	-0.429	-0.516
	(-1.23)	(-1.29)	(-1.12)	(-0.89)
SOI_{-4}	0.545	0.222	0.619	0.301
	(1.01)	(0.40)	(1.10)	(0.42)
SOI_{-5}	-0.590	-0.320	0.077	-0.533
	(-0.93)	(-0.49)	(0.15)	(-0.58)

sale constraints in the stock market. Panel C demonstrates that OOI_{-5} is not informative when analyst dispersion is low. When analysts exhibit divergent opinions, OOI_{-5} significantly predicts CAR. Panel D shows that the return predictability is stronger for firms with low analyst coverage, supporting the informational role of option trading. Using alternative measures of information asymmetry generates large the same results.

7. Conclusion

In this paper, I investigate whether option trading activity provides additional price information that is not in the stock market. By decomposing total stock order imbalance into an option-induced imbalance component and a residual component that is unrelated to option trading, I show that option-induced stock order imbalance positively and significantly predicts future stock returns and the predictive relation does not reverse at longer horizons. The remaining order imbalance generates only temporary price pressure on the stock price and exhibits a reversal effect later. The results hold at both daily and half-hour frequencies. The evidence suggests that a substantial amount of private information exists in options order flow. Specifically, permanent price information comes mainly from option contracts with relatively narrow bid-ask spreads (i.e., ATM and ITM options) but not contracts with higher leverage (i.e., OTM options).

I also investigate the link between information content in options order flow and the level of information asymmetry, short-sale costs, and market liquidity. Consistent with theoretical predictions, I find that return predictability from OOI is stronger for firms with high PIN, low analyst coverage, large bid-ask spread, large adverse-selection component of the spread, small market capitalization, and low institutional ownership. The concentration of informed traders and the short-sale constraints enhance the informational role of option trading in price discovery. The options market also becomes more informative when option trading is active, which is consistent with the prediction by Easley, O'Hara, and Srinivas (1998) that informed trading is more likely to occur in deep and liquid markets. In an event study, I also find that OOI significantly predicts CARs five

Table 2.12

Order flow information and potential profit from informed earnings trading

For each variable measuring potential profit from informed trading around earnings announcements, the full sample is divided into three groups: low ($<$ 30th percentile), medium (30th $\sim$ 70th percentile), and high ($>$ 70th percentile). Slope coefficients and t-statistics (in parentheses) are reported only for option order imbalance five days before the announcement (OOI_{-5}) whereas the regressions are based on the following equation:

$$CAR_{(0,2)} = \alpha + \sum_{t=1}^{5} \beta_1^k SOI_{-t} + \sum_{t=1}^{5} \beta_2^k OOI_{-t} + \theta X_{-1} + \varepsilon_{0,2}.$$

Surprise equals the difference between the actual quarterly earnings per share (EPS) minus the last analyst forecast consensus before the announcement. *Forecasting error* is the signed earnings surprise defined as actual EPS minus the analyst EPS forecast consensus. *Dispersion* is the standard deviation of analyst forecasts before the announcement. *Analyst coverage* is the number of analysts following. Standard errors are clustered by firm. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Surprise				*Panel B: Forecast error*			
	low	*medium*	*high*		*negative*	*neutral*	*positive*
OOI_{-5}	0.657	2.983	8.214***	OOI_{-5}	6.658***	1.321	4.221**
	(0.80)	(1.05)	(2.98)		(2.86)	(0.36)	(1.99)
Panel C: Dispersion				*Panel D: Analyst coverage*			
	low	*medium*	*high*		*low*	*medium*	*high*
OOI_{-5}	0.875	8.829***	3.907*	OOI_{-5}	5.403**	4.873*	1.379
	(0.34)	(3.23)	(1.95)		(2.25)	(1.92)	(0.48)

days before earnings announcements.

My findings indicate that the options market plays a key role in embedding private information into stock prices. However, this study does not examine investors' learning problem further. It is important to understand whether investors analyze the options order flow directly to extract information or the information is revealed through the hedging transactions transmitted from OOI. The results suggest that there can exist market frictions that prevent investors from immediately learning the stock price information in the options order flow. For example, the cost of monitoring and collecting order flow information can be too high for most investors. It is also possible that the options order flow is too complex for retail investors to understand.[8] As a result, the spot price does not fully adjust to the new information immediately, and the return predictability is preserved. For future research, it would be interesting to explore the learning behavior of stock market investors and investigate how such information in the options market can be learnt more efficiently.

The proposed OOI focuses on delta risk and has shown significant predictability regarding underlying stock price. Option contracts have other risk exposures that can be analyzed in similar ways. It would be interesting to investigate in a multidimensional framework the information links between the derivative markets and the underlying market because some risk can be traded only in the options market. Understanding that issue is important to both academicians and practitioners.

[8] In a related study, Henderson and Pearson (2011) show that retail investors persistently overprice a popular structured equity product.

Chapter 3

Option Listing and the Probability of Informed Trading in the Stock Market

Summary

This article empirically investigates listing effects of exchange-traded options on incentives of informed and uninformed traders and the level of information risk in the stock market. I find that investor order imbalance and institutional ownership are important determinants of option listing decisions in addition to firm size, stock trading volume, return volatility, and bid-ask spread. I then match the listing stocks with eligible but unselected stocks based on the predicted probability and examine the treatment effect on the stock market behaviors. I find that option listing increases the informed order arrival rate by 16.8% and the uninformed order arrival rate by 27.8%. Due to the disproportional changes in informed and uninformed trading, the overall probability of informed trading (PIN) of the listing firms decreases 11.5% more than that of the control firms. I also find significant option listing effects on reducing the bid-ask spread, investor order imbalance, realized volatility and volatility of volatility. The results indicate that option introduction leads to improvement in the stock market quality.

1. Introduction

In this paper, I empirically examine the impact of option introduction on incentives of heterogenously informed investors and the information environment in the stock market. This question has been theoretically analyzed by many researchers, e.g., Back (1993), Biais and Hillion (1994), and Cao (1999). Assuming less informed traders do not trade strategically, these authors show that option introduction increases the trading incentive of uninformed traders. The intuition is that uninformed traders now have better risk sharing opportunities as options help complete the market. The impact on the informed traders is more complicated. Informed traders can achieve greater leverage using options. Therefore, the level of informed trading may increase after option listing. However, strategic informed traders also worry about the information revealed by their orders and are likely to mimic uninformed traders. Biais and Hillion (1994) show that informed trading can decrease after option listing if uninformed traders now structure less extreme trades and hence make the uninformed order flow less attractive to informed traders. The option listing impact on informed trading and overall stock market information environment can therefore be either direction.

Empirical tests on this question face the challenge that neither informed nor uninformed trading is directly observable in the stock market. To address this issue, I infer the amount of informed and uninformed trading by estimating the sequence trade model of Easley and O'Hara (1992) (EO hereafter). The order arrival rates from the informed and uninformed traders are the two major parameters in the model. The other two parameters are the probability of information event and the probability of an information event being good news. Easley, Kiefer, and O'Hara (1996) show how these parameters can be estimated from the daily number of buy and sell trades. The EO model also elucidates how to calculate the unconditional probability of a trade being information based, which is a useful measure of information asymmetry as shown by Easley, Kiefer, and O'Hara (1996), Easley, Kiefer, and O'Hara (1997), and many others.

To gauge the option listing impact, I first estimate a logit model of option listing decision following Mayhew and Mihov (2004). Mayhew and Mihov find that options exchanges select stocks for listing based on firm size, trading volume, return standard deviation, and industry category. Danielsen, Ness, and Warr (2007) find that percentage bid-ask spread is also an important determinant. Hypothesizing that option exchanges may prefer to list stocks that are hard to sell short in the stock market, I include institutional ownership and investor order imbalance in the logit model as proxies for short sale constraints. Consistent with the hypothesis, the results show that both investor order imbalance and institutional ownership are significantly correlated with the listing decisions. After obtaining the propensity score of option listing, I match each listing firm to an eligible but non-listing firm at the same time to construct a control sample. I then investigate the difference of the EO model parameters between the listing and control firms as the treatment effect of option listing.

I find that the average daily informed orders increase by 40.7% for the listing firms in the first year after option listing. An increase in informed trading is also found in the control group, but the magnitude is much smaller at 23.9%. The treatment effect of option listing on the level of informed trading averages at 16.8%, and is significant at 1% level. Uninformed trading also becomes more active for both the listing and control firms. The average daily uninformed orders increase by 70.8% for the listing firms, and 42.9% for the control firms in the first year after listing. The listing effect on uninformed trading averages at 27.8%, also significant at 1% level. However, neither the listing group nor the control group exhibits significant difference in the probability of information event after option listing. Due to the disproportional increase in informed and uninformed trading, the probability of informed trading (PIN) becomes lower for both groups. The average PIN drops by 24.1% for the listing firms, and 12.6% for the control firms in the first year after option listing. The difference between the matched pairs averages at 11.5% with a t-statistic of 10.45.

The differences of PIN and informed and uninformed trading may come from the difference in firm characteristics as shown by Aslan, Easley, Hvidkjaer, and OHara (2008). For robustness check, I regress the treatment effects on the differences of firm size, sales growth rate, analyst coverage, stock turnover ratio, insider holding, institutional holding, return on assets, stock return standard deviation, and Tobin's Q between the two groups. The results show that the listing effects cannot be explained by these firm characteristics.

A recent study by Duarte and Young (2009) extends the original EO model to incorporate symmetric liquidity shocks in the uninformed order flow. Those authors show that PIN can be decomposed into an asymmetry component and a symmetric shock component. For robustness check, I also estimate the Duarte and Young (2009) model and find that the listing effect on PIN is indeed driven by the treatment effect on the asymmetry component rather than that on the liquidity shock component.

Concerning that the structural breaks may not happen right at the listing date, I investigate the dynamics of the EO model parameters and PIN estimated in event months in stead of event years. I find that the most significant treatment effects on both informed and uninformed trading occur in the first month after option listing. The listing effects drift into three months after listing, and become insignificant thereafter. The listing effect on PIN is only statistically significant in the first month (*t*-statistic=-3.61) and the third month (*t*-statistic=-2.16) after option listing. Hence, the results do not identify any other alternative event date for structural breaks.

Within the same sample of matched firms, I find that option listing also has significant impact on the stock bid-ask spread, order imbalance, and volatility of volatility. All these stock market metrics experience significant improvement in the listing group over the control group around the listing date. On the other hand, the listing effect on the stock trading volume is insignificant. Although the trading volume increases more for the listing firms than the control firms after listing, the volume difference between the two groups is mainly driven by a momentum created in the pre-listing period. The listing effect on stock return volatility is unclear. Long term volatility does not exhibit significant difference between the listing and control firms. However, I find that the average monthly volatility of the listing firms is significantly lower than the control firms seven months after listing.

This paper contributes to the finance literature in several ways. First, I provide evidence that both informed and uninformed traders benefit from option introduction. The trading volume from both types of traders increase after option listing. Second, I show that the information risk in the stock market is reduced by options. The listing effect on PIN is significantly negative, indicating that after option listing, an uninformed trader is less likely to bear loss from trading against an informed trader. Third, I find that the institutional ownership and investor order imbalance are both important determinants of the option listing decision. Finally, the simple estimation method adopted in the paper provides a solution to the numerical difficulty in estimating the EO model.

The remainder of the paper is organized as follows. In section 2, I review the literature on option introduction and informed trading. In section 3, I document the sample selection and matching procedure. Section 4 reports the main results of PIN analysis. Section 5 provides additional analysis on the listing effects on other stock market behaviors and section 6 concludes.

2. Related literature

If derivative securities are redundant assets, their introduction to the market should have no effect on the underlying asset prices or the trading volume. The informational role of options is first noted by Black (1975) as he argues that informed traders may use options to achieve higher leverage given institutional difference in the underlying and derivative markets. Grossman (1988) points out that even if options can be synthesized by a dynamic trading strategy, its real trading transmits volatility information that will not be revealed by the dynamic trading strategy. Back (1993) shows how information asymmetry and option introduction can cause stochastic volatility in the underlying market. Biais and Hillion (1994) formally investigate how options can change the incentive of informed and uninformed traders. In their model, liquidity and informed trades are interdependent and the informed traders mimic the liquidity traders. Information asymmetry imposes additional costs on the liquidity traders, which can lead to a no trade problem as in Milgrom and Stokey (1982). In an incomplete market with a single tradable security of stock, the market may break down even without presence of information asymmetry. With options to complete the market, the liquidity traders have better risk sharing opportunities and become more willing to trade. If the benefit of increased hedging opportunity exceeds the potential loss of trading against the informed traders, introducing options can resolve the no trade problem by inducing hedging-motivated liquidity trades. The informed traders could also trade more, and more private information is revealed to the market. However, if there is no breakdown in the incomplete market, option introduction may reduce the incentive and the profit of the informed traders because the liquidity traders can now structure less extreme trades to hedge their endowment shocks, making the liquidity orders less attractive to the informed traders. The impact of option introduction on the overall information risk and market efficiency is thus ambiguous in this setup.[1]

Although not explicitly stated, many other studies imply that the informational risk should increase after options are listed. For example, Figlewski and Webb (1993) argue that options relax the short sale constraint on the underlying stock. They show that the short interest increases after option listing. Even if investors only establish short positions on the options market, the option market makers will transfer the short positions to the underlying market through delta hedging. Since short selling is more likely to be information based (Diamond and Verrecchia (1987)), informed trading and the information risk will increase after option listing. In a partially rational expectation framework, Cao (1999) considers the cost of information collection. Options increase the incentive of informed traders for acquiring more precise signals and make the market price more efficient. Consequently, there will be more informed trading, and the information risk rises for uninformed traders in trading rounds. [2]

Easley and O'Hara (1992) explicitly model the learning problem of market makers, who re-

[1]Huang and Wang (1997) reach similar conclusions on information efficiency by noting that option introduction increases the informational trades but also changes the information content of the existing allocational trades.

[2]Introduction of options can add another type of information risk to the market, i.e. asymmetric information about stock return volatility. Since volatility is a key input of option pricing models, private information about volatility is also profitable in the options market, but not so in the stock market. In this paper, I do not regard volatility information risk as a main source of information risk in the stock market for three reasons. First, informed trading on volatility in the options market has neutral impact on the stock prices and order imbalance if volatility traders only trade delta-neutral products such as straddles, or if both volatility traders and option market makers dynamically hedge the delta exposure of their option positions. The hedging transactions on the stock market may increase the total trading volume but will not make the stock order flow imbalanced. Therefore, these hedging trades can be viewed as liquidity trades. Second, it is difficult to study the option listing effect on volatility trading because no observation is available before options are publicly listed and traded, even though these options may be actively traded in the over-the-counter market for a long time. Third, current options microstructure theories lack of a method that can empirically separate informed trading on stock return and volatility information. The identification problem cannot be solved without knowing the hedging strategies of both parties, which are hard to observe.

ceive orders from both informed and uninformed traders. The detailed setup can be found in appendix 1. One can empirically estimate the trading intensity of informed and uninformed traders from intra-day transaction data. Another distinguish feature of the model is that the probability of informed trading (PIN) can be calculated from the model parameters. PIN has been widely used as a measure for information asymmetry (e.g., Easley, Kiefer, and O'Hara (1996), Easley, Kiefer, and O'Hara (1997), Brown, Hillegeistb, and Lo (2004)). Moreover, Easley, Hvidkjaer, and O'Hara (2002) and Easley and O'Hara (2004) show that PIN also has a significant pricing effect in the cross section. [3] Unlike price-based measures of information asymmetry such as the bid-ask spread, PIN focuses on the information in the order flow and is a quantity-based measure. Compared to the spread-based measures of information asymmetry, PIN has two advantages. First, PIN is a direct measure of asymmetry while spread-based measures are indirect proxies and can cause many econometric and interpretation problems. Second, more importantly, the structural model of EO is able to shed light on the sources of information asymmetry. The risk of trading against informed traders can be higher because of (1) greater probability of information events, (2) more informed orders on information days, and (3) fewer uninformed orders. These three sources are captured by three different parameters of the model. However, PIN also has two disadvantages compared to the spread-based measures. The first disadvantage is that PIN cannot be updated continuously because it is not directly observable in the market. Practically, one needs to keep the model parameters constant over a period of time in order to estimate the parameters. Although pre-specified parameter dynamics can be incorporated into the model (Easley, Engle, O'Hara, and Wu (2008)), PIN is not as flexible as the spread-based information asymmetry measures. In this event study, I assume that the model parameters are constant within each examined event window. The other disadvantage is heavy computation requirement. To estimate the structural model, maximum likelihood estimation (MLE) method is usually applied to data series of daily number of buyer-initiated and seller-initiated trades. One needs to classify each transaction as buyer- or seller-initiated according to certain algorithm such as Lee and Ready (1991), and then estimate four model parameters from daily aggregate number of transactions. Accompanying the decimalization and the rise of

[3] In a recent extension of the model, however, Duarte and Young (2009) show that the pricing effect of PIN is closely related to illiquidity in the stock market.

electronic trading in the new century, the trading volume and number of transactions in the US stock market have both experienced tremendous growth. This change in the market microstructure not only magnifies effort needed to classify trade directions but also creates a serious numerical problem in estimating the parameters. With large numbers in the exponential function, computing software may terminate the MLE procedure due to data overflow problem. Factoring out common components may lead to an underflow problem. To address the issue, I provide a new numerical method of estimating the EO model, which guarantees convergency in estimation. The estimation strategy is discussed in details in appendix 1.

This paper is also related to a large literature that investigates the option introduction effects on the stock market behaviors. Majority of the empirical studies in this area focus on the option listing impact on the underlying price and volatility.[4] In a rare analysis on the market microstructure, Kumar, Sarin, and Shastri (1998) find that the stock market bid-ask spread narrows and the quote depth increases, suggesting that the market quality is improved after option listing. Danielsen, Ness, and Warr (2007) note that the structural break may not happen right on the listing date. Using a flexible method for detecting structural breaks, these authors find that the decrease of bid-ask spread is a major determinant of option listing rather than a resulting benefit. However, both of these two studies are subject to a criticism raised by Mayhew and Mihov (2004), that the option listing decisions are endogenous, and the selection bias must be carefully addressed when analyzing the listing effects. Following Mayhew and Mihov, I use the propensity score matching method to tackle the selection bias in this event study focusing on the option listing effect on the stock market microstructure.

Finally, this work is also related to a strand of empirical research that investigates the stock price response to earnings announcement after option listing. If option trading reveals private information, the stock price will become more efficient after option listing and the forecasting error will be smaller in equilibrium. As a result, price reaction to earnings surprise should also become less sig-

[4]Studies on the option listing impact on the underlying price include Branch and Finnerty (1981), Conrad (1989), Detemple and Jorion (1990), Sorescu (2000), and Danielsen and Sorescu (2001). Studies on the option listing impact on the underlying volatility include Conrad (1989), Skinner (1989), Fedenia and Grammatikos (1992), Kumar, Sarin, and Shastri (1998), Bollen (1998), and Mayhew and Mihov (2004).

nificant. Early empirical studies support this hypothesis.[5] However, Mendenhall and Fehrs (1999) provide evidence that after 1986, the stocks with options trading exhibit slightly larger immediate price reaction to earnings surprise than the stocks without options. The inconsistency of these empirical results reflects a difficulty of testing joint hypotheses. If the stock price always adjusts to its new fundamental value right after earnings announcement, the price reaction is equivalent to the forecasting error. However, if the price discovery takes time because of heterogeneous belief (Miller (1977)) or incompetency of learning as argued by Mendenhall and Fehrs (1999), the stock price will not necessarily adjust to its fundamental value immediately, and a post-announcement drift can exist. With options, the price reaction to earnings surprise may be larger because investors now have additional trading tools, and the speed of adjustment increases. Consequently, the post-announcement drift may become less significant after option listing. Empirical evidence from Jennings and Starks (1986) and Botosan and Skinner (1993) supports this theory. Therefore, testing price reaction to earnings surprise conditioning on the option availability is a joint test on information efficiency gained from options and immediate price discovery. Also focusing on the gain of information efficiency from options, my work is different from these studies because I take a different angel and directly examine the trading intensity of informed and uninformed traders.

3. Data and sample selection

3.1. Option listing stocks

I obtain data on option listing events between February 11, 2001 and February 28, 2010 from the Options Clearing Corporation's (OCC) website. I include only new listing stocks that have no options traded at any other option exchanges at the time of listing. If a security has more than one new listing events during the sample period, I use only the first record. I exclude options on index and ETF and focus on single name equity options. In order to construct explanatory variables for the listing selection, I require the firm to have valid price information in the CRSP database and

[5] See Skinner (1990) and Ho (1993).

Table 3.1
Number of option listings over time
The table reports the time series distribution of number of firms selected for option listing between February 11, 2001 and February 28, 2010 as well as for the pool of firm-month observations that are qualified for option listing but not selected. The selection criteria include: (1) the underlying must be listed on NYSE, AMEX, NASDAQ, or any other national stock exchange; (2) the stock price is not below $3.00; (3) there must be at least seven million publicly-held shares; (4) there must be at least 252 trading days prior to this date.

Year	Listing	Eligible Non-listing
2001	116	18117
2002	98	16660
2003	164	16516
2004	237	17409
2005	143	17372
2006	246	16525
2007	192	15127
2008	204	11931
2009	161	9792
2010	12	828
Total	1573	140277

valid transaction data in the NYSE TAQ database for at least 252 trading days prior to the option listing date. The final listing sample includes 1573 observations. Table 3.1 lists the number of new listings every year. As seen, the number gradually increases in the first part of the sample period and peaks at 246 in 2006. It decreases slightly afterward and is down to 161 in 2009. The sample covers only the first two months in 2010 and there are 12 new listings in those two months.

3.2. Eligible non-listing stocks

I follow Mayhew and Mihov (2004) and construct a control group of firms that are eligible but not selected for option listing. First, the eligible firms need to be identified for option listing. Mayhew and Mihov (2004) document detailed regulation changes on the option listing standards imposed by the SEC before 1997. Comparing the current requirements with the requirements at the end of their sample period, several differences are noted: (i) the requirement of minimum trading volume has been removed; (ii) a security can be listed on option exchanges five days after its initial public offering (IPO) now while previously it must be traded for at least twelve months after IPO; (iii) the minimum security price is reduced from $7.50 to $3.00. The other three requirements remain unchanged: (1) the security must be listed on a national exchange; (2) the security must have at least seven million publicly held shares; (3) there must be at least two thousand shareholders. Without detailed information about changes of the listing standard after 1997, I define eligible firms for option listing in the next month as those meeting requirements (1) and (2) at the end of each month with the price above $3.00, and having at least 252 trading days in the CRSP database. Mayhew and Mihov (2004) argue that it is practically impossible to filter on requirement (3) because many shareholders hold shares in street names, and this omission is unlikely to misclassify qualified firms. Therefore, I also do not rule out firms based on this criterion. I adopt the more strict criterion of the trading history because I need information from this period to construct the explanatory variables. The main difference between my definition of eligible firms and Mayhew and Mihov's is the minimum stock price. There is evidence that this requirement was relaxed at the beginning of my sample period. For example, Hecla Mining Co. (ticker: HL) was traded at $4.27 when its options were listed on January 22, 2003, and Golden Star Resources, Ltd (ticker: GSS) was traded at $3.77 when its options were listed on October 6, 2003. Therefore, I use the current minimum price of $3.00 as the listing criterion. I obtain data on insider holdings from Thomson Reuter's insider database and after-market data from the CRSP. The filtering generates 140,277 eligible firm month observations in all during the sample period. I report the annual breakdown in table 3.1 as well. As equity options become more and more popular, it can be seen that the pool of eligible firms shrinks over time with 18,117 observations in 2001 (1,647 firms per month on average) and only 9,792 observations in 2009 (816 firms per month on average).

3.3. Determinants of the option listing decision

Determinants of the option listing decision include firm size, trading volume, return standard deviation, industry (Mayhew and Mihov (2004)) and percentage bid-ask spread (Danielsen, Ness, and Warr (2007)). Other than these well studied variables, I suspect that three other firm characteristics may also be related to the listing decisions:

1. Volume order imbalance (VOI). Stock order imbalance may reflect the difficulty of short sale constraint. Pessimistic investors are less likely to participate in trading if short selling is not possible. As a result, stock order imbalance may become positively biased. Introducing options on such stocks can generate high trading volumes in the options market, and these stocks should be preferred by the option exchanges given everything else the same. Therefore, a positive relation is expected between the order imbalance and the likelihood of option listing. I measure the order imbalance in trading volume as $\frac{B-S}{B+S}$, where B and S are daily buyer-initiated and seller-initiated trading volume, respectively.

2. Absolute volume order imbalance (AVOI). The absolute value of order imbalance is approximately equal to PIN. Market making is easier for stocks and options with smaller probability of informed trading. Therefore, options exchanges may choose to list stocks with small absolute volume order imbalance to increase the options market liquidity. Estimating PIN for all stocks can be overwhelming, but absolute order imbalance is easy to compute.

3. Institutional ownership. Institutional ownership has contradicting effects on the likelihood of option listing. On one hand, institutional ownership is negatively related to the difficulty of short selling because institutional investors provide the most supply in the security lending market. Options exchanges may prefer to list stocks with low institutional ownership because short selling will be easier after listing and the options volume will be high. On the other hand, institutional investors are more likely to trade options for hedging purposes if they hold large positions of the underlying securities. Therefore, options exchanges may prefer stocks with high institutional ownership. The overall effect is thus unclear.

Previous studies, namely Mayhew and Mihov (2004) and Danielsen, Ness, and Warr (2007), use the long term average and a short term abnormal variable of each independent variable in the regression. For a better matching, I consider both level and trend effects. For each of the explanatory variables except return standard deviation and institutional ownership, I measure its average daily value in one year before the matching date (t-1, t-12), one month before the matching date (t-1), and twelve months before the matching date (t-12). Return standard deviation is calculated from daily realized returns for the three measurement windows too. Institutional ownership data is on a quarterly basis. Constructing three variables from scarce observations can lead to a serious multi-collinearity problem. Therefore, I only use the most recent value (t-1) and the one-year lag value (t-12) of institutional ownership. Using both current and lag variables is able to catch the dynamic effect in addition to the level effect captured by the long-run average. To investigate determinants of option listing, I estimate the following logit model in the pooled sample:

$$\begin{aligned} Listing = & \beta_0 + \beta_1 Size + \beta_2 Volume_{t-1,t-12} + \beta_3 Volume_{t-1} + \beta_4 Volume_{t-12} + \beta_5 STD_{t-1,t-12} \\ & + \beta_6 STD_{t-1} + \beta_7 STD_{t-12} + \beta_8 Spread_{t-1,t-12} + \beta_9 Spread_{t-1} + \beta_{10} Spread_{t-12} \\ & + \beta_{11} VOI_{t-1,t-12} + \beta_{12} VOI_{t-1} + \beta_{13} VOI_{t-12} + \beta_{14} AVOI_{t-1,t-12} + \beta_{15} AVOI_{t-1} \\ & + \beta_{16} AVOI_{t-12} + \beta_{17} Institution_{t-1} + \beta_{18} Institution_{t-12} + \omega INUDSTRY + \varepsilon, \end{aligned} \quad (3.1)$$

where the dependent variable *Listing* is a dummy variable equal to 1 for listing stocks and 0 otherwise, *Size* is the natural logarithm of the market capitalization, all volume and volatility variables are in natural logarithm for standardization, daily percentage spread is calculated as $\frac{2(ask-bid)}{ask+bid}$ at the market close, and *INDUSTRY* is a vector of 71 industry dummy variables based on two-digit SIC codes.

The estimation results are reported in table 3.2. The first model includes only independent variables that have been used in the previous studies. All variables have significant predictive ability. Specifically, large stocks with high volatility and low bid-ask spread are more likely to be listed. However, the impact of trading volume is not clear. One-year average trading volume has a negative coefficient but one-month average volume has a positive coefficient, both significant at

1% level. The second model includes lagged variables of trading volume, volatility, and spread at month $t-12$. All lag variables have significant coefficients and the long-term volatility becomes insignificant. The results reveal important information about the dynamics of trading volume, volatility, and spread of the selected listing stocks. At one year before the listing dates, these stocks have lower trading volume, higher volatility, and higher percentage spread than the eligible but non-listing stocks. During the time to the listing date, however, these stocks experience increase in trading volume and decrease in percentage spread while volatility remains high. The last column in table 3.2 tests the full specification of the model. With additional control variables, firm size becomes an insignificant variable in explaining the option listing decisions. Current trading volume still has a significantly positive coefficient but the lag volume becomes insignificant. The volatility and spread impact also reduces. More interestingly, it shows that lag volatility and long-term volatility are negatively related to the listing decisions while current volatility still has a positive coefficient. It seems the listing stocks also experience an increase in volatility prior to the listing date. Volume order imbalance, absolute volume order imbalance, and institutional ownership are important determinants of option listing because all these variables are statistically significant in the regression. Specifically, the listing stocks have less positive volume order imbalance, larger absolute volume order imbalance, and lower institutional ownership at one year prior to the listing date. However, at one month prior to the listing date, these stocks have more positive volume order imbalance, smaller absolute volume order imbalance, and higher institutional ownership than the non-listing stocks. This model generates the highest pseudo-R square and has the largest percentage of concordant predictions. Therefore, I will use the full model prediction in matching stocks later.

In summary, I find that when selecting stocks for option listing, exchanges consider both the level and trend of several variables. On average, the listing stocks are associated with recent increases in trading volume, volatility, volume order imbalance, and institutional ownership, and decreases in spread and absolute order imbalance.

Table 3.2

Determinants of option listing: logistic regression results

This table presents the regression results from a logit model of the option listing decision. The independent variables are log market capitalization (*Size*), log average trading volume in the past 12 months ($Volume_{t-1,t-12}$), log average volume in the last month ($Volume_{t-1}$), log average volume in month $t-12$ ($Volume_{t-12}$), log standard deviation of daily returns in the past year ($STD_{t-1,t-12}$), log standard deviation of daily returns in the last month (STD_{t-1}), log standard deviation of daily returns in month $t-12$ (STD_{t-12}), average percentage bid-ask spread in the past year ($Spread_{t-1,t-12}$), average spread in the last month ($Spread_{t-1}$), average spread in month $t-12$ ($Spread_{t-12}$), average volume order imbalance in the past year ($VOI_{t-1,t-12}$), average volume order imbalance in the last month (VOI_{t-1}), average volume order imbalance in month $t-12$ (VOI_{t-12}), average absolute volume order imbalance in the past year ($AVOI_{t-1,t-12}$), average absolute volume order imbalance in the last month ($AVOI_{t-1}$), average absolute volume order imbalance in month $t-12$ ($AVOI_{t-12}$), current institutional ownership ($Institution_{t-1}$), institutional ownership one year ago ($Institution_{t-12}$), and 71 two-digit SIC industry dummies. Standard errors are in parentheses. *,**,*** denotes statistical significance at 10%, 5%, and 1% level, respectively.

Variables	(1)	(2)	(3)
Intercept	-19.56	-19.769	-11.143
	(215.6)	(212.3)	(155.3)
Size	0.115***	0.139***	0.027
	(0.033)	(0.033)	(0.035)
$Volume_{t-1,t-12}$	-1.137***	-0.769***	-1.217***
	(0.061)	(0.081)	(0.093)
$Volume_{t-1}$	1.389***	1.297***	1.494***
	(0.056)	(0.058)	(0.064)
$STD_{t-1,t-12}$	0.676***	0.416***	-0.287***
	(0.092)	(0.116)	(0.050)
STD_{t-1}	0.240***	0.287***	0.335
	(0.064)	(0.067)	(0.124)
$Spread_{t-1,t-12}$	-0.132**	-0.294***	0.003
	(0.059)	(0.067)	(0.073)
$Spread_{t-1}$	-1.261***	-1.113***	0.289***
	(0.110)	(0.113)	(0.072)
$Volume_{t-12}$		-0.294***	0.077***
		(0.043)	(0.063)
STD_{t-12}		0.242***	-0.423***
		(0.070)	(0.100)
$Spread_{t-12}$		0.031**	0.037**
		(0.013)	(0.017)

Table 3.2 (continued)

Variables	(1)	(2)	(3)
$VOI_{t-1,t-12}$			2.366***
			(0.646)
VOI_{t-1}			2.256***
			(0.391)
VOI_{t-12}			-0.271
			(0.282)
$AVOI_{t-1,t-12}$			-7.264***
			(0.932)
$AVOI_{t-1}$			-5.873***
			(0.647)
$AVOI_{t-12}$			1.365***
			(0.458)
$Institution_{t-1}$			1.556***
			(0.200)
$Institution_{t-12}$			-0.697***
			(0.205)
INDUSTRY	yes	yes	yes
$Pseudo-R^2$	2.05%	2.09%	2.46%
Percent concordant	84.2	84.5	86.3
Percent discordant	13	12.9	11.5
Percent tied	2.7	2.7	2.2

3.4. Matching listing to eligible non-listing stocks

After obtaining the propensity score of option listing, I then match each listing stock to the eligible non-listing stock with the closest propensity score in the same calendar month. The matching is done without replacement and once a non-listing stock is matched, it drops out of the pool of non-listing stocks for one year before it becomes available again to avoid overlapping observation. I report the matching results in table 3.3. The average log size of the listing stocks reaches 20.23 (610 million). The matched non-listing stocks have similar average size but the standard deviation is larger than the listing stocks. The pattern is generally true for the rest of the variables as well, i.e., the means of the two groups are well matched but the standard deviation of the control group is larger than the standard deviation of the listing group. T-tests on the mean confirm that there does not exist systematic difference between the two groups based on the listing criteria. Finally, the last row in table 3.3 presents the mean and standard deviation of the predicted probability of option listing for the two groups. The average predicted probability is 6% for the listing group and 5.3% for the control group.

4. Option listing effect on the probability of informed trading

4.1. Basic test

Note that the options must be available to all market participants when its impact on the investors' trading incentive becomes effective. Therefore, the event day is the day on which actual option trading begins. Practically, not all options become available immediately on the listing days if the appointed market maker is not ready for providing quotes. There may be several days delay before the options are quoted and traded. Therefore, the true event day may be unknown. There may be another concern regarding the event window that the option market makers usually need to establish a reasonable inventory of the underlying stocks before the option listing days. Their purchase of the underlying stocks may lead to uninformative order imbalance and potentially

Table 3.3

Matching option listing stocks to eligible but non-listing stocks

This table reports the propensity score matching results. For each option listing firm, the eligible but non-listing firm with the closest propensity score in the same month is chosen without replacement. Statistics are reported for log market capitalization (*Size*), log average volume in the past 12 months ($Volume_{t-1,t-12}$), log average volume in the last month ($Volume_{t-1}$), log average volume in month $t-12$ ($Volume_{t-12}$), log standard deviation of daily returns in the past year ($STD_{t-1,t-12}$), standard deviation of daily returns in the last month log (STD_{t-1}), log standard deviation of daily returns in month $t-12$ (STD_{t-12}), average percentage bid-ask spread in the past year ($Spread_{t-1,t-12}$), average spread in the last month ($Spread_{t-1}$), average spread in month $t-12$ ($Spread_{t-12}$), average volume order imbalance in the past year ($VOI_{t-1,t-12}$), average volume order imbalance in the last month (VOI_{t-1}), average volume order imbalance in month $t-12$ (VOI_{t-12}), average absolute volume order imbalance in the past year ($AVOI_{t-1,t-12}$), average absolute volume order imbalance in the last month ($AVOI_{t-1}$), average absolute volume order imbalance in month $t-12$ ($AVOI_{t-12}$), current institutional ownership ($Institution_{t-1}$), institutional ownership one year ago ($Institution_{t-12}$), and fitted probability of listing ($Prob_{fit}$).

	Listing		Matched	
	mean	std	mean	std
Size	20.227	1.038	20.108	1.453
$Volume_{t-1,t-12}$	12.259	0.858	12.062	1.477
$Volume_{t-1}$	12.655	0.944	12.526	1.448
$Volume_{t-12}$	11.761	1.106	11.580	1.708
$STD_{t-1,t-12}$	-0.718	0.503	-0.738	0.546
STD_{t-1}	-0.738	0.591	-0.731	0.667
STD_{t-12}	-0.874	0.609	-0.884	0.626
$Spread_{t-1,t-12}$	0.626	0.779	0.730	1.019
$Spread_{t-1}$	0.385	0.450	0.431	0.421
$Spread_{t-12}$	0.897	1.587	1.043	1.910
$VOI_{t-1,t-12}$	0.022	0.064	0.022	0.062
VOI_{t-1}	0.027	0.085	0.031	0.082
VOI_{t-12}	0.012	0.116	0.012	0.125
$AVOI_{t-1,t-12}$	0.192	0.067	0.206	0.088
$AVOI_{t-1}$	0.157	0.060	0.162	0.067
$AVOI_{t-12}$	0.224	0.104	0.243	0.133
$Institution_{t-1}$	0.355	0.225	0.329	0.251
$Institution_{t-9}$	0.301	0.231	0.277	0.243
$Prob_{fit}$	0.060	0.073	0.053	0.057

contaminate the PIN measure. To address these two issues, I leave one-month window around the listing date recorded by the OCC, and define the event year before listing as trading day [-262, -11] and the event year after listing as trading day [11, 262]. I then estimate the PIN model in these two event windows for all the listing and control stocks.

The results are reported in table 3.4. Panel A presents the cross sectional average of the estimated PIN as well as the model parameters for the two groups. Average PIN of the listing stocks is 18.2% before listing and 13.9% after listing. Average PIN of the control stocks is 17.7% before listing and 15.2% after listing. The probability of information event, α, does not change a lot for either group after the listing date. The order arrival rates are reported in natural logarithm. In both groups, both informed and uninformed order arrival rates increase after listing. Panel B reports the t-test results on percentage change in PIN and model parameters as well as the treatment effects. On average, the listing stocks experience a 24.1% drop in PIN, no significant change in α, a 40.7% increase in μ, and a 70.8% increase in ε. The control stocks experience a 12.6% drop in PIN, no significant change in α, a 23.9% increase in μ, and a 42.9% increase in ε. The treatment effect is significant for PIN, μ, and ε but not for α. It appears that option listing significantly reduces the probability of informed trading in the stock market. Moreover, this change is not due to change in the probability of information event. As shown, both uninformed and informed trading become more active after option listing but uninformed trading increases more than informed trading. As a result, the relative likelihood of informed trading is reduced.

4.2. Robustness check controlling for determinants of PIN

Aslan, Easley, Hvidkjaer, and OHara (2008) investigate the determinants of PIN and find that several firm characteristics can explain the cross sectional variation in PIN. Specifically, they estimate the following model:

Table 3.4
Probability of informed trading before and after option listing

This table examines the option listing effect on the probability of informed trading. For each listing firm, a control firm is chosen based on the predicted probability of option listing from previous logistic regression. PIN is the estimated probability of informed trading based on Easley and O'Hara (1992) model. α is the probability of information event. μ is the daily order arrival rate of informed trades. ε is the daily order arrival rate of uninformed trades. For each variable, percentage change is calculated as the natural logarithm of the post-listing value minus the natural logarithm of the pre-listing value. The cross sectional mean and standard error (in parentheses) are then reported for the listing stocks, the control stocks, as well as the paired difference between these two groups. *,**,*** denotes statistical significance at 10%, 5%, and 1% level, respectively.

Panel A: Averages of PIN and model parameters:

	Listing stocks		*Control stocks*	
	Before	After	Before	After
PIN	0.182	0.139	0.177	0.152
α	0.253	0.255	0.250	0.254
μ	6.049	6.456	5.830	6.069
ε	5.419	6.126	5.222	5.652

Panel B: Option listing effect:

	Listing stocks	Control stocks	difference
PIN	-0.241***	-0.126***	-0.115***
	(0.009)	(0.009)	(0.011)
α	0.005	0.033*	-0.028
	(0.017)	(0.018)	(0.025)
μ	0.407***	0.239***	0.168***
	(0.021)	(0.022)	(0.029)
ε	0.708***	0.429***	0.278***
	(0.016)	(0.017)	(0.022)

$$PIN = b_0 + b_1 Size + b_2 Growth + b_3 Age + b_4 Analyst + b_5 Turnover + b_6 Insider$$
$$+ b_7 Institution + b_8 Accrual + b_9 ROA + b_{10} STD + b_{11} TobinQ + \Omega Industry + \eta, \quad (3.2)$$

where *Growth* is the annual growth rate in sales, *Age* is the history in CRSP, *Analyst* is the number of analysts following the company, *Turnover* is the annual stock market trading volume scaled by shares outstanding, *Insider* is the percentage ownership of company insiders, *Accrual* is the estimate of the discretionary component of total accruals, *ROA* is return on asset calculated calculated as net income after depreciation over total asset, *TobinQ* is the market value of equity plus book value of debt over total asset, and the other variables are the same as my definition. These authors find that all variables are significantly related to firm PIN except *Accrual*. It is possible that changes in the firm characteristics are driving the change in PIN after option listing, and the listing effect may be spurious. For robustness check, I test the treatment effect of option listing on PIN controlling for changes in firm characteristics and estimate the first order difference version of model 4.2:

$$dPIN = b_0 Listing + b_1 dSize + b_2 dGrowth + b_3 dAnalyst + b_4 dTurnover + b_5 dInsider$$
$$+ b_6 dInstitution + b_7 dROA + b_8 dSTD + b_9 dTobinQ + \eta, \quad (3.3)$$

where *Listing* is a dummy variable equal to 1 for listing stocks and 0 for control stocks, and all the other variables are annual changes after option listing. Since *Accrual* does not explain PIN, I drop it from the model. I also drop *Age* and *Industry* because the change in firm age is universal for all firms and industry classification is not likely to change.

The sample size decreases slightly after I merge the PIN estimates with the firm characteristics in COMPUSTAT. Analyst coverage is extracted from I/B/E/S and is assigned zero if no record is found for the firm. Insider holding data are from Thomson Reuters insider fillings. There are 1,256 listing stocks and 1,283 control stocks with valid firm characteristics after data merging. The OLS regression result is reported in table 3.5. The coefficient of the listing dummy is estimated to be -0.037 with a t-statistic of -16.44. Given the average PIN of all firms around 0.186, the listing treatment reduces PIN by nearly 20% controlling for the other PIN determinants. Changes in stock market turnover and institutional ownership also significantly explain the change in PIN. The negative correlations are consistent with the results in Aslan, Easley, Hvidkjaer, and OHara (2008). The other variables are not able to explain the change in PIN.

In the next three columns of table 3.5, I regress changes in the three main parameters on the same set of explanatory variables. The regression on probability of information event, α does not fit well as the adjusted R-square is only 0.23%. The listing dummy is only 0.007 and significant at 10% level. The change in α is negatively related to institutional ownership and return on asset, and positively related to stock turnover. But economic significance of explanatory power is negligible. Significant option listing effect is found on the arrival rate of informed orders μ as the listing dummy has a coefficient of 0.182 with a t-statistic of 5.23 in the third column. Change in μ is also negatively related to change in turnover and positively related to change in return on asset. The last column shows that option listing has a significant impact on the arrival rate of uninformed orders ε too as the coefficient of the listing dummy reaches 0.444 with a t-statistic of 20.94. Firm size, turnover, and institutional ownership also have significant explanatory power. The results in this table show that option listing has robust impact on the probability of informed trading in the stock market. The decrease is not due to lower probability of information event but disproportional increase in informed and uninformed trading.

Table 3.5

Option listing effects on PIN controlling for other determinants

This table reports the regression results of equation 4.2. All variables are the annual changes after option listing dates except *Listing*, which is a dummy variable equal to 1 for the listing stocks and 0 otherwise. *Size* is the natural logarithm of the market capitalization. *Growth* is the annual growth rate in sales. *Analyst* is the number of analysts following the company. *Turnover* is the annual stock market trading volume scaled by shares outstanding. *Insider* is the percentage ownership of company insiders. *Institutional* is the percentage of shares held by institutional investors. *ROA* is the return on asset calculated calculated as net income after depreciation over total asset. *STD* is the standard deviation of the daily returns. *TobinQ* is the market value of equity plus book value of debt over total asset. Pooled regression results are reported with t-statistics in parentheses. *,**,*** denotes statistical significance at 10%, 5%, and 1% level, respectively.

	$dPIN$	$d\alpha$	$dlog(\mu)$	$dlog(\varepsilon)$
Listing	-0.037***	0.007*	0.182***	0.444***
	(-16.44)	(1.84)	(5.23)	(20.94)
dSize	-0.004	-0.004	0.068	0.092**
	(-0.93)	(-0.54)	(1.04)	(2.32)
dGrowth	-0.000	-0.000	-0.001	-0.001
	(-1.27)	(-0.27)	(-0.79)	(-1.16)
dAnalyst	-0.001	-0.004	0.044	0.031
	(-0.31)	(-0.68)	(0.83)	(0.96)
dTurnover	-0.000***	0.000***	-0.005***	-0.001***
	(-2.39)	(2.40)	(-5.84)	(-2.59)
dInsider	-0.000	-0.000	0.003	0.002
	(-0.47)	(-0.53)	(0.71)	(0.81)
dInstitution	-0.058***	-0.040*	0.337	0.278**
	(-4.07)	(-1.73)	(1.53)	(2.07)
dROA	-0.009	-0.019*	0.205**	0.092
	(-1.34)	(-1.86)	(2.04)	(1.50)
dSTD	0.000	-0.002	-0.009	-0.004
	(0.11)	(-0.30)	(-0.12)	(-0.09)
dTobinQ	0.001	0.001	-0.016	-0.008
	(0.99)	(0.82)	(-1.28)	(-1.04)
Adj R-sq	0.1643	0.0023	0.0361	0.2263

4.3. APIN-PSOS model

One limit of the EO model is that the uninformed order arrival rate is a constant over different trading periods. Therefore, only one order series, either buy or sell, can increase on information events. Practically, it is observed that both buy and sell can increase on certain days in the data. To incorporate the positive correlation of buy and sell orders, Duarte and Young (2009) (DY) extend the EO model by allowing uninformed orders of buy and sell to increase symmetrically on certain days, defined as liquidity shocks. Those authors show that PIN can be then decomposed into an asymmetry component (APIN) and a liquidity shock component (PSOS). Only APIN measures the information asymmetry. The model and estimation strategy is described in details in appendix 2. For robustness check, I also estimate the DY model and report results in table 3.6. Panel A shows that APIN consists a smaller proportion of PIN than PSOS on average. For both listing and control groups, APIN and PSOS become smaller after the listing date.

Panel B investigates the listing effects in details. The drop of APIN is statistically significant at the 1% level in both groups. The treatment effect arrives at 7.4% (t-statistic = -6.73). However, although the drop of PSOS is significant in the listing group, the treatment effect is only marginally significant (t-statistic = -1.87). Consistent with previous results, panel B shows that both arrival rates of informed and uninformed orders become significantly larger after option listing for both listing and control firms. The treatment effects are also significantly positive with the increment of uninformed trading surpasses that of informed trading. However, the treatment effect on systematic liquidity shock, δ, is only marginally significant (t-statistic = 1.8). The results demonstrate that the drop of PIN after option listing is driven by declining APIN rather than PSOS and the probability of informed trading is reduced by the listing event.

4.4. Dynamic analysis

Comparing estimation results from one year event window before and after listing may not well capture the exact timing of structural breaks. To explore the dynamic impact of option listing

Table 3.6

Option listing effects on adjusted PIN and liquidity shocks

This table examines the option listing effect on the probability of informed trading of Duarte and Young (2009) model. For each listing firm, a control firm is chosen based on the predicted probability of option listing from previous logistic regression. PIN is the estimated probability of informed trading based on Easley and O'Hara (1992) model. α is the probability of information event. μ is the daily order arrival rate of informed trades. ε is the daily order arrival rate of uninformed trades. For each variable, percentage change is calculated as the natural logarithm of the post-listing value minus the natural logarithm of the pre-listing value. The cross sectional mean and standard error (in parentheses) are then reported for the listing stocks, the control stocks, as well as the paired difference between these two groups. *,**,*** denotes statistical significance at 10%, 5%, and 1% level, respectively.

Panel A: Averages of PIN and model parameters:

	Listing stocks		*Control stocks*	
	Before	After	Before	After
APIN	0.128	0.111	0.131	0.122
PSOS	0.312	0.236	0.290	0.250
α	0.359	0.382	0.369	0.385
θ	0.252	0.243	0.239	0.238
μ	5.406	5.826	5.177	5.415
δ	5.950	6.238	5.622	6.027
ε	5.002	5.844	4.831	5.333

Panel B: Option listing effect:

	Listing stocks	Control stocks	difference
APIN	-0.144***	-0.071***	-0.074***
	(0.008)	(0.008)	(0.011)
PSOS	-0.451**	-0.003	-0.447*
	(0.196)	(0.136)	(0.239)
α	0.089***	0.090***	-0.001
	(0.022)	(0.021)	(0.030)
θ	-0.084***	-0.010	-0.074*
	(0.029)	(0.031)	(0.041)
μ	0.422***	0.239***	0.183***
	(0.023)	(0.022)	(0.032)
δ	0.288	0.406***	-0.117
	(0.195)	(0.133)	(0.237)
ε	0.843***	0.502***	0.341***
	(0.019)	(0.021)	(0.025)

on PIN, I estimate the PIN model for six months prior to and after the option listing date. I define event month -1 as trading days [-31, -11], event month 1 as trading days [11, 31], and the rest in a similar way so that each event month contains 21 trading days.

Table 3.7 presents the results. The first column shows that along event months, significant monthly treatment effect on PIN occurs in month 1 (5.2%, t-statistic = -3.61) and month 3 (2.9%, t-statistic = -2.16). Option listing has no significant treatment effect on PIN in the other event months. Column 2 shows a similar pattern of the treatment effect on the probability of information event, α. However, the magnitude of the effect is much smaller in both statistical and economic sense. Column 3 presents the monthly treatment effect on informed trading. The treatment effect is significant in three event months, -4, 1, and 3. However, the most significant change is in month 1, when μ of listing firms increases by 13.5% over control firms (t-statistic = 3.9). The last column shows that the arrival rate of uninformed trading takes off half a year before the listing date, consistent with the fact that trading volume is an important determinant of the listing decision. However, the most significant structural break occurs in month 1 again, when ε of listing firms increases by 14.5% over control firms (t-statistic = 8.42). For all four parameters, the listing effect diminishes after three months.

To summarize the results in this section, I find that option listing increases both informed and uninformed trading. But it is neutral to the probability of information event. The most significant structural break occurs in the first month following listing. The treatment effect drifts into three months after listing and becomes insignificant thereafter. The increase in uninformed trading outweighs the increase in informed trading. As a result, PIN is reduced by option listing.

The results suggest that the informed traders trade more when options are available. This increase of information risk in the stock market, however, is offset and reversed by the increase of uninformed trades. Uninformed traders also trade more because they perceive the asset allocating benefit of options to be greater than the potential loss of trading against the informed traders. Both types of traders can benefit from introduction of exchange traded options, and the overall level of information risk decreases in the stock market.

Table 3.7

Monthly option listing effects on PIN

This table reports the probability of informed trading (PIN) in Easley and O'Hara (1992) around the option listing dates. For each listing firm, a control firm is chosen based on the predicted probability of option listing from previous logistic regression. Panel A reports the cross sectional averages of the listing stocks, the control stocks, the difference between the two groups, and t-statistics of paired t-tests in each event month. Panel B investigates the listing effect on the month-to-month change of PIN. For each firm, the percentage change is first calculated as the first order difference of the natural logarithm of PIN. The treatment effect is then calculated as the difference of the percentage change in PIN between the listing and the control firms.

month	PIN	α	μ	ε
-5	-0.021	0.007	-0.008	0.024*
	(-1.56)	(0.26)	(-0.26)	(1.75)
-4	0.018	-0.009	0.061**	0.031**
	(1.28)	(-0.31)	(1.99)	(2.17)
-3	-0.015	0.002	0.045	0.065***
	(-1.09)	(0.09)	(1.42)	(4.25)
-2	-0.021	-0.003	-0.042	-0.022
	(-1.55)	(-0.11)	(-1.28)	(-1.39)
-1	0.018	-0.002	-0.027	-0.051***
	(1.26)	(-0.08)	(-0.74)	(-2.89)
1	-0.052***	-0.05*	0.135***	0.145***
	(-3.61)	(-1.65)	(3.90)	(8.42)
2	0.022	0.039	0.007	0.021*
	(1.63)	(1.39)	(0.23)	(1.71)
3	-0.029**	-0.056*	0.067**	0.046***
	(-2.16)	(-1.94)	(2.14)	(3.66)
4	0.005	0.048	-0.030	0.013
	(0.33)	(1.61)	(-0.94)	(1.05)
5	-0.004	-0.004	0.010	0.011
	(-0.32)	(-0.14)	(0.32)	(0.88)
6	-0.008	0.002	0.016	0.026**
	(-0.60)	(0.08)	(0.52)	(2.11)

5. Other listing effects

It is interesting to examine the listing effects on other market quality metrics in this recent sample. For this purpose, I consider five stock market metrics. For each of these metrics, I calculate the value for one year before and after the option listing date. The statistics are reported in table 3.8:

(1) Spread

The average bid-ask spread decreases by 45.3% in the listing group and 28.1% in the control group after option listing. The treatment effect of option listing is 17.2% with a standard error of 1.7%. It is tempting to conclude that option listing reduces the bid-ask spread in the stock market. However, as pointed out by Danielsen, Ness, and Warr (2007), the structural break may not happen at the listing event, and the crude test on annual change does not have information on spread dynamics. For a more flexible test, I examine the time series of the daily average of percentage spread in the listing and control groups. To reduce noise, I run a local linear regression of each data series with a smoothing parameter of 0.25. The nonparametric regression takes into consideration that there can be more than one structural break in the event window examined. Both original and fitted values are plotted in figure 3.1. Panel A shows that for the listing stocks, the bid-ask spread significantly narrows prior to listing but stays at the same level after listing. Based on similar observation, Danielsen, Ness, and Warr (2007) conclude that the decrease in spread is a selection criterion but there is no listing effect on the spread. However, it is not clear whether there is any market wide effect that drives the change of the spread. Panel B shows that the control stocks also experience a decrease in the spread before the listing date, but the spread widens afterward. Panel C plots the mean differences between the listing and control groups, and it shows that compared to the control stocks, the listing stocks have much lower percentage spread after option listing. The plot of daily t-statistics from paired t-tests on the mean also shows the same result. It is clear that the bid-ask spread is an important listing criterion for option listing but the listing event also has a feedback effect as it nails the spread at a lower level than the control stocks after listing.

(2) Absolute volume order imbalance

Table 3.8
Stock market quality before and after option listing
Option listing effects on stock market quality are examined in this table. For each listing firm, a control firm is chosen based on the predicted probability of option listing from previous logistic regression. For each measure of market quality, X, the percentage change is calculated as the natural logarithm of the post-listing value minus the natural logarithm of the pre-listing value: $log(X_{after}) - log(X_{before})$. The cross sectional mean and the standard error (in parentheses) are then reported for the listing stocks, the control stocks, as well as the paired differences between these two groups. Spread is the one year average of daily percentage bid-ask spread. Volume is the one year average of stock trading volume. RVol is the one year average of daily realized volatility. AVOI is the one year average of absolute volume order imbalance. VVol is the one year standard deviation of daily realized volatility. *,**,*** denotes statistical significance at 10%, 5%, and 1% level, respectively.

	listing	control	difference
Spread	-0.453***	-0.281***	-0.172***
	(0.017)	(0.017)	(0.017)
AVOI	-0.234***	-0.133***	-0.101***
	(0.007)	(0.007)	(0.009)
RVol	-0.240***	-0.047	-0.193***
	(0.038)	(0.038)	(0.048)
VVol	-0.352***	-0.122**	-0.230***
	(0.059)	(0.059)	(0.077)
STD	-0.024**	-0.019	-0.005
	(0.010)	(0.012)	(0.012)

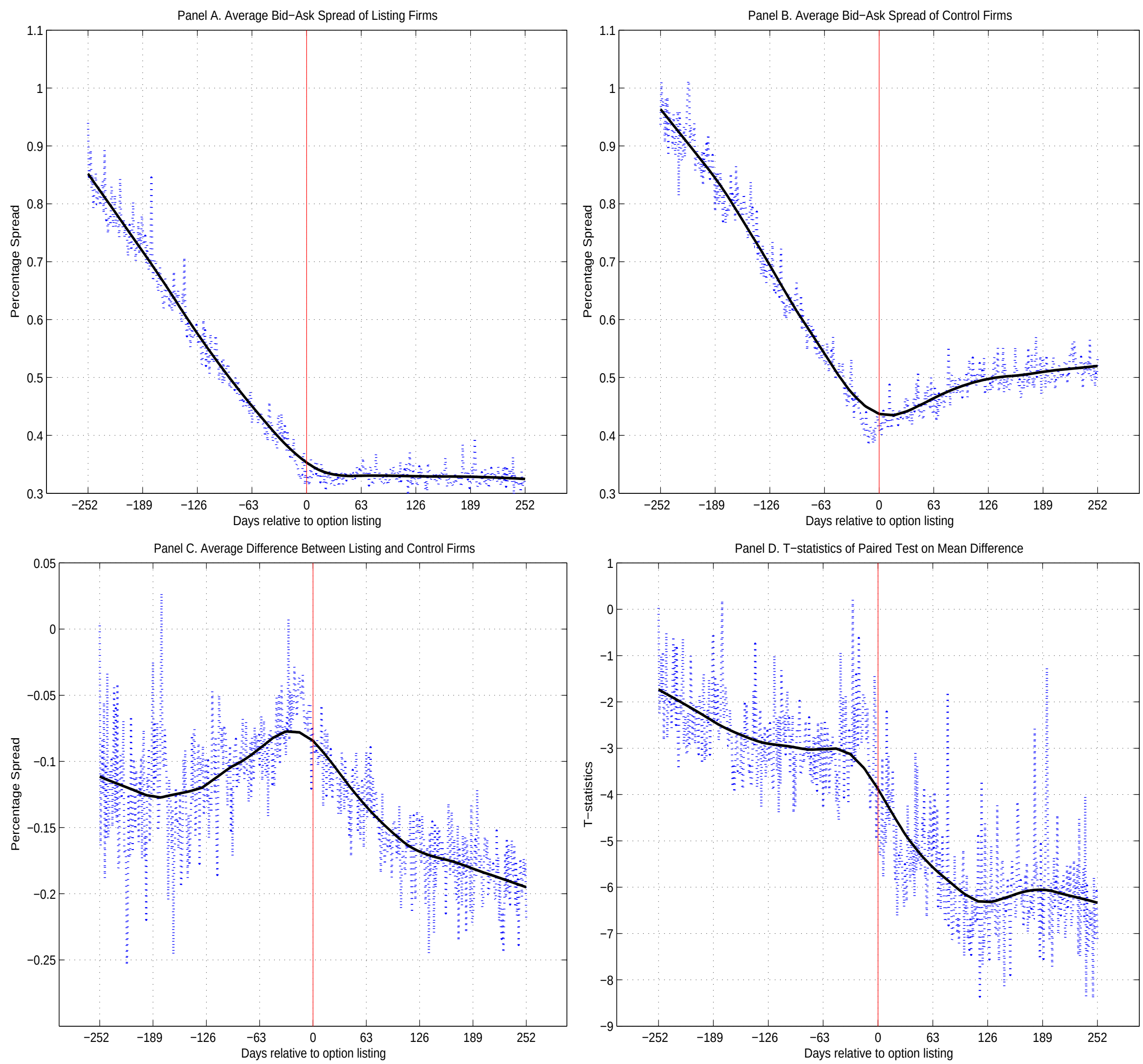


Figure 3.1

Option listing effect on spread

This figure plots the daily average percentage bid-ask spread in the stock market around the two-year event window of the option listing dates for the listing and the control stocks in panel A and B, as well as the mean differences between the two groups in panel C and t-statistics of paired t-tests in panel D. The blue dashed line plots the actual value of each statistics. The solid black line plots the smoothed value predicted from a local linear regression with smoothing parameter of 0.25.

The absolute volume order imbalance is approximately equal to the expected value of PIN. Table 3.8 shows that trading becomes more balanced for both listing and control stocks after the listing date. The average annual imbalance decreases by 23.4% for the listing stocks and 13.3% for the control stocks. The difference between the two groups averages at -10.1% and is significant at 1% level. Order imbalance dynamics are plotted in figure 3.2. For comparison with the PIN results, I calculate average order imbalance in event months. Panel A and B show that similar to spread, order imbalance also decreases sharply prior to the listing date for both listing and control stocks. After listing, order imbalance remains at the low level for listing stocks while it gradually increases for control stocks. Panel C and D show that the treatment effect on order imbalance is significant and negative after the listing date.

It is interesting to compare the dynamics of PIN and absolute order imbalance because the order imbalance is also supposed to reflect the probability of informed trading. Actually, PIN and order imbalance exhibit very different dynamics around the listing date. For the listing stocks, order imbalance mainly reduces before listing and stabilizes afterward. On the other hand, the most significant change in PIN occurs right after option listing although it starts decreasing half a year before that. Although absolute order imbalance is the expected value of PIN, the realized values can have quite different dynamics and potentially measure different things.

(3) Realized volatility

One important impact of option listing is that now investors can trade volatility in the options market. It is interesting to examine the impact of option listing on stock return volatility. I calculate daily realized volatility as the square root of the sum of five-minute return squares. Table 3.8 shows that the one-year average realized volatility reduces 35.2% for the listing firms (t-statistic = 5.97). However, there is no significant change in realized volatility of the control firms. As a result, the treatment effect is -19.3%, statistically significant at the 1% level (t-statistic = 4.02). Figure 3.3 clearly shows a declining trend of realized volatility in the listing group and a persistent pattern in the control group.

(4) Volatility of volatility

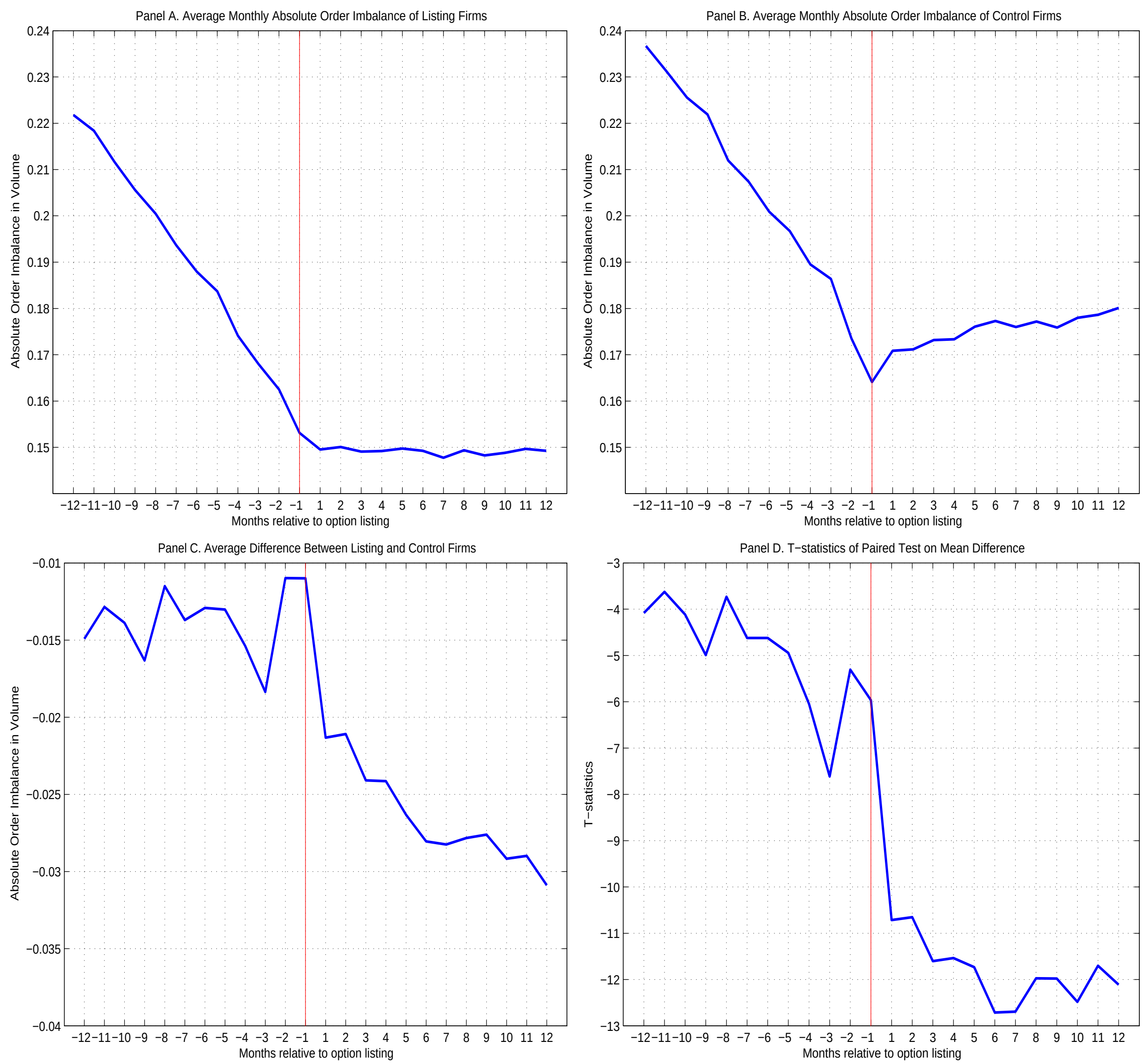


Figure 3.2

Option listing effect on monthly absolute volume order imbalance

This figure plots monthly average absolute volume order imbalance in the stock market around two-year event window of option listing dates for listing and control stocks in panel A and B, as well as the mean difference between the two groups in panel C and t-statistics of paired t-tests in panel D.

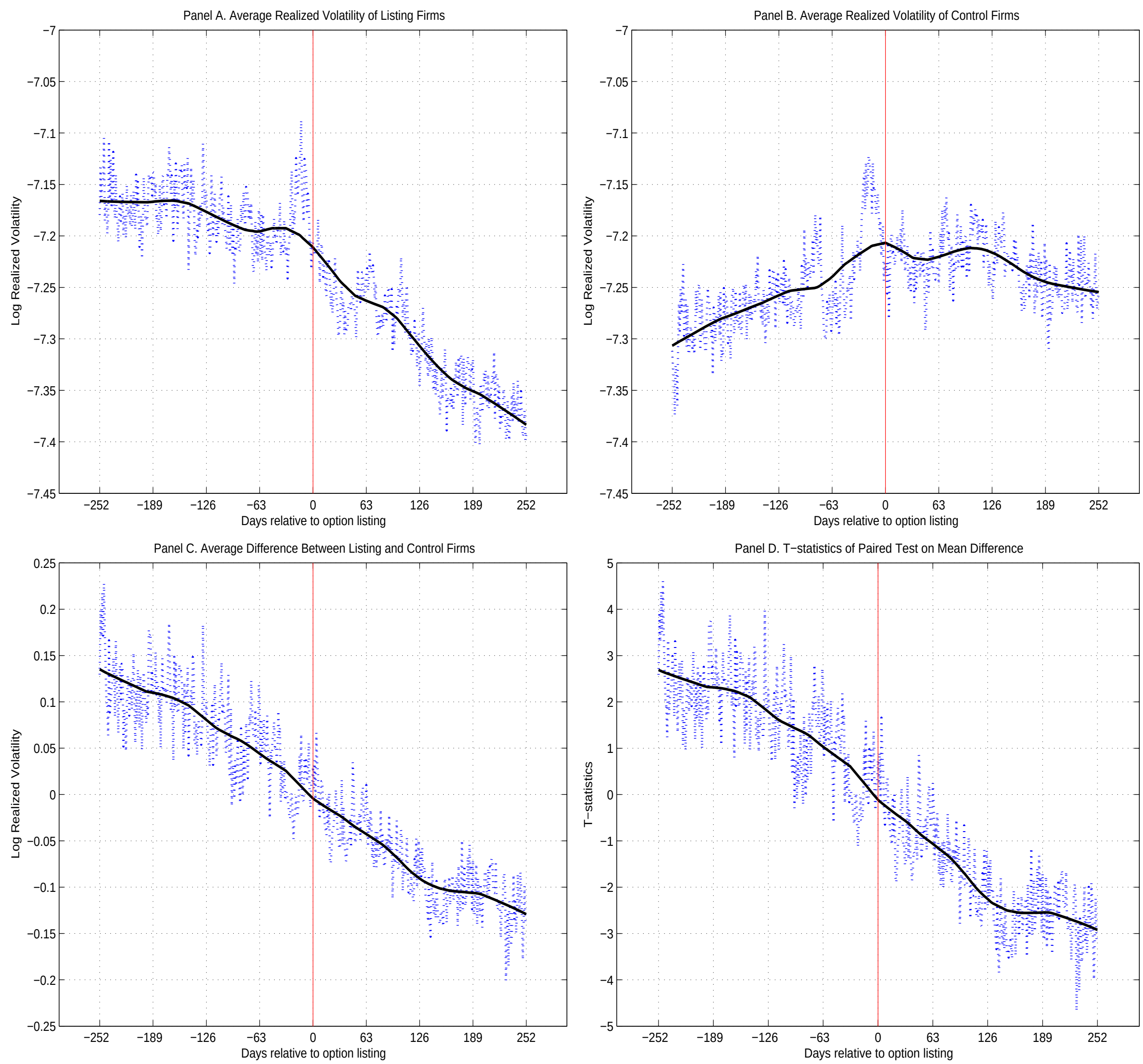


Figure 3.3

Option listing effect on realized volatility

This figure plots daily average realized volatility on the stock market around two-year event window of option listing dates for listing and control stocks in panel A and B, as well as the mean difference between the two groups in panel C and t-statistics of a paired t-test in panel D. The blue dashed line plots the actual value of each statistics. The solid black line plots the smoothed value predicted from a local linear regression with smoothing parameter of 0.25.

Given substantial change in monthly volatility, it would be interesting to investigate the listing effect on the volatility of volatility. To compute the volatility of volatility, I first compute the daily realized volatility as the square root of the sum of five-minute return squares. The volatility of volatility is then calculated as the standard deviation of the realized volatility one year before and after the listing date. Table 3.8 shows that volatility of volatility decreases by 35.2% for the listing stocks and 12.2% for the control stocks. The treatment effect, -23%, is significant at 1% level. I plot the monthly volatility of volatility in figure 3.4 to investigate the dynamic impact. For cross sectional standardization, I take natural logarithm. Panel A shows volatility of volatility significantly decreases for the listing stocks after listing. Similar effect is not observed in panel B for the control stocks. Panel C and D confirm that the treatment effect is significantly negative. The result indicates option listing reduces not only the level of return volatility but also the variation in volatility.

(5) Return standard deviation

Compared to large changes in spread and trading volume, table 3.8 shows that the change in return standard deviation is small after option listing. The listing stocks have a decrease of 2.4% on average, and the control stocks have a decrease of 1.9%. The mean difference is only 0.5% and is not significantly different from zero. The result is consistent with previous analysis by Mayhew and Mihov (2004). Figure 3.5 plots the dynamics of return standard deviation for the listing and control stocks in event months. Panel A shows that the return volatility spikes in the month right before option listing and drops to the normal level immediately after listing for the listing stocks. Volatility continues to decrease six months after listing. Panel B shows that although a similar spike and quick reversal pattern is also observed, volatility is more persistent for the control stocks after listing. Plots of the mean differences and t-statistics in panel C and D indicate that significant treatment effect on volatility exists seven months after option listing.

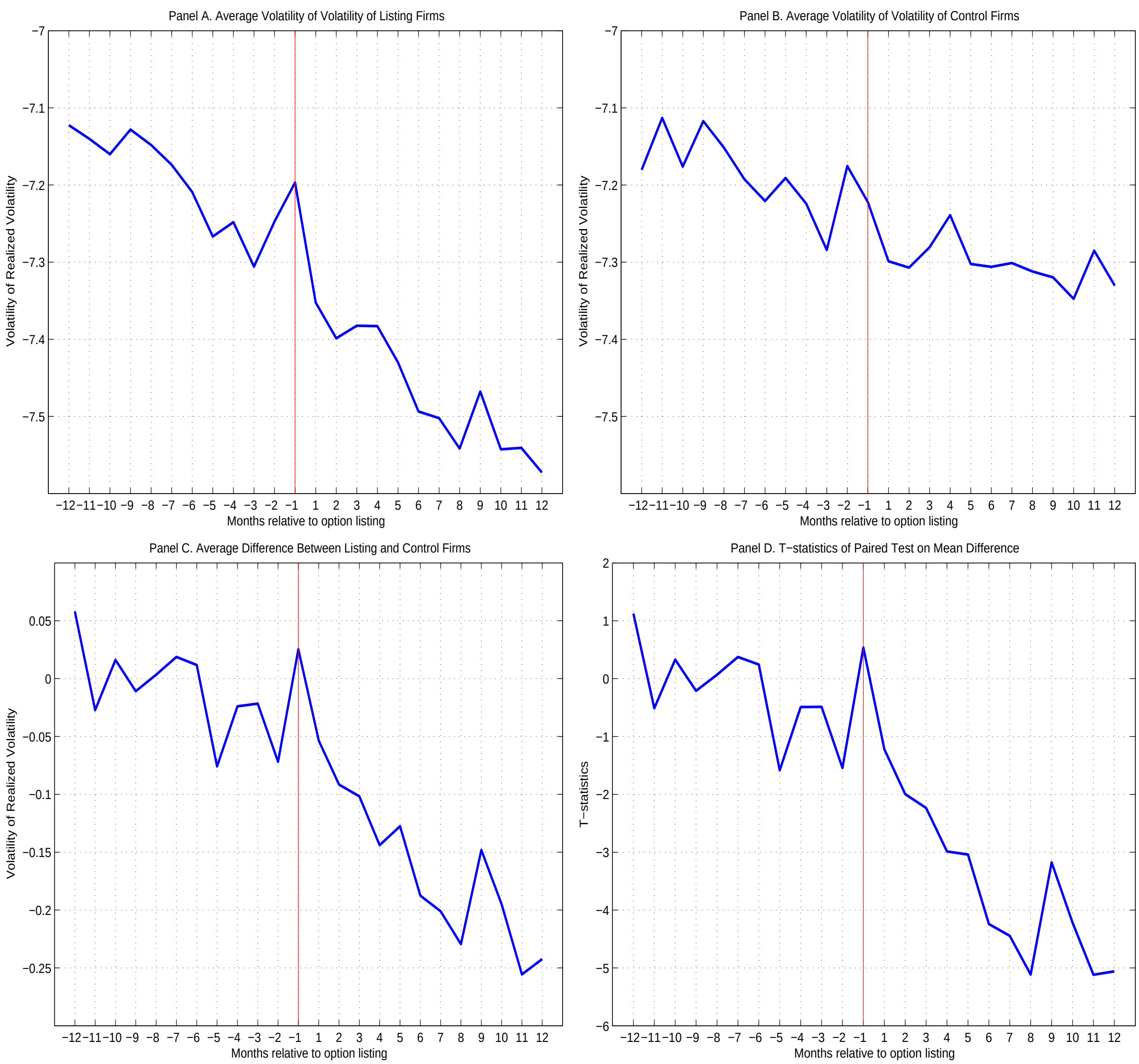


Figure 3.4

Option listing effect on volatility of volatility

This figure plots the monthly average standard deviation of the daily realized volatility in the stock market around the two-year event window of the option listing dates for the listing and the control stocks in panel A and B, as well as the mean differences between the two groups in panel C and t-statistics of paired t-tests in panel D.

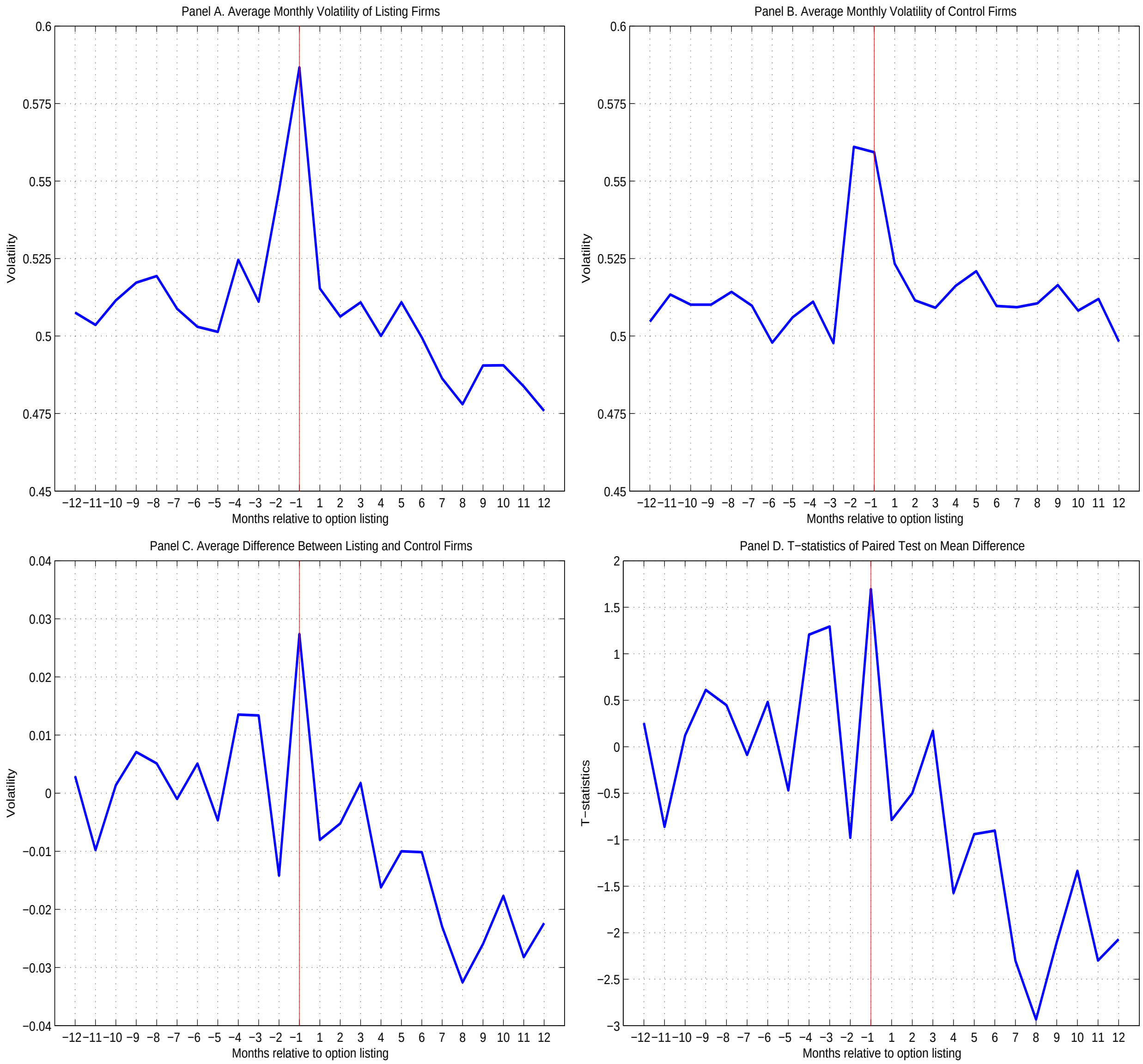


Figure 3.5
Option listing effect on monthly return standard deviation
This figure plots the monthly average standard deviation of daily stock returns around the two-year event window of the option listing dates for the listing and the control stocks in panel A and B, as well as the mean differences between the two groups in panel C and t-statistics of paired t-tests in panel D.

6. Conclusion

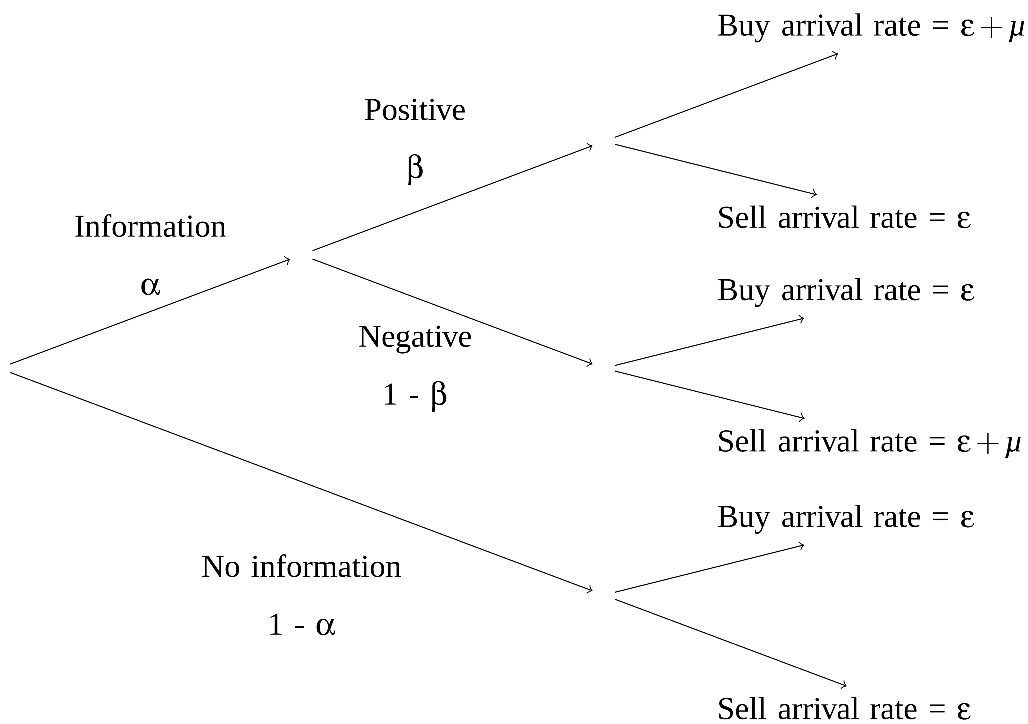
This paper investigates the impact of option introduction on the incentive of informed and uninformed traders, and the probability of informed trading in the stock market. The probability of informed trading is important because it is directly associated with firms' cost of capital as shown by Easley, Hvidkjaer, and O'Hara (2002) and Easley and O'Hara (2004). I match each option listing stock to a non-listing but eligible stock, and study the treatment effect on PIN and three main parameters of the sequence trade model proposed by Easley and O'Hara (1992). The results show that option listing significantly reduces PIN by 10% in the year after listing. Both informed and uninformed trading increase after option listing but the probability of information event does not change relative to the control group. The decline in PIN is due to disproportional increase in informed and uninformed trading because option listing increases informed trading by 12% and uninformed trading by 24%. The results are robust to controls of firm characteristics known as determinants of PIN. Dynamic analysis shows that the most significant structural break of the difference between the listing and control stocks occurs in the first month following option listing. The empirical results indicate that both informed and uninformed traders can benefit from option introduction. Biais and Hillion (1994) predict that uninformed traders become more willing to trade because options provide better risk sharing opportunities, and informed traders' profit can increase if option introduction mitigates the market breakdown problem. My findings are consistent with their model prediction. Additionally, I find significant option listing effects on the bid-ask spread, order imbalance, and volatility of volatility in the stock market. The listing stocks have narrower spread, more balanced order flows, and lower volatility of volatility than the control stocks after listing. There is no significant treatment effect on trading volume. Although listing stocks have higher post-listing volume, the pattern is more likely to be caused by a pre-listing trend. The listing effect on return volatility is ambiguous. The impact on long-term volatility is not significantly different from zero. However, monthly volatility dynamic shows that the listing stocks do experience significant decrease in the return standard deviation six months after listing compared to the control stocks. Collectively, the results suggest that the introduction of exchange traded options significantly improves the underlying stock market quality.

My results should also be interpreted with caution. Although more private information is revealed when options are available, the spot price is not necessarily more efficient. Option trading has an information externality as it blurs the information content in existing trades as recognized by Stein (1987). Biais and Hillion (1994) and Huang and Wang (1997) separately show that options may reduce information efficiency. Hu (2011) argues that the learning cost of uninformed traders can be so high that information in the options market does not immediately transmit to the underlying market. Future empirical research can better tackle this issue if robust dynamic measures of information asymmetry and price efficiency are uncovered.

Appendices

1. The PIN model

This appendix describes the sequence trade model of Easley and O'Hara (1992) and the estimation strategy. There are three types of participants in the model: informed traders who know exactly the future value of the tradable asset, uninformed traders who submit orders for liquidity reasons, and a risk neutral market maker. On any given day, nature determines whether an information event occurs. With probability α, an information event would occur and some traders acquire private information. With probability $1 - \alpha$, there is no information event. An information event is associated with good news with probability β and bad news with probability $1 - \beta$. The informed traders will only submit buy orders on the good news days and sell orders on the bad news days. The uninformed traders are modeled as exogenous liquidity traders, and they will submit both buy and sell orders regardless of the realization of the information event. Without loss of generality, the daily arrival rates of the uninformed traders' buy and sell orders are modeled to follow the same Poisson distribution with a mean of ε. On the information event days, the order arrival rate of the informed traders is modeled to follow another Poisson distribution with a mean of μ. The whole structure is illustrated in the following trading tree.



Given the model setup, the unconditional probability of a trade being information based can be calculated as

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon}. \tag{1.1}$$

Let B and S denote the daily number of buy and sell trades, respectively. Easley, Engle, O'Hara, and Wu (2008) show that the denominator in equation 1 is just the expected value of all orders $E[B+S]$, and the numerator is approximately equal to $|B-S|$. Therefore, PIN is approximately equal to the expectation of the order imbalance $E(\frac{|B-S|}{B+S})$. To classify trade directions, I use Lee and Ready (1991) algorithm with one-second quote lag $(t-1)$ instead of the original five-second quote lag $(t-5)$ because my empirical tests show that one second quote lag makes the largest proportion of transactions fall exactly on the bid or ask, and the smallest proportion of trades outside the quote bounds. When estimating the model, I make one deviation from the original model, and use buy and sell volume instead of number of trades. As argued by Easley, Prado, and O'Hara (2011), volume can be a better proxy for information content brought to the market than number of trades, especially in a market dominated by high frequency trading.

To calculate the probability of informed trading, one needs to estimate the parameters of the structural model first. There is no close form solution for the model, and maximum likelihood estimation (MLE) is used. Under the model specifications, the likelihood function of a particular day with B and S trades is

$$L(B,S|\theta) = \alpha\delta e^{-(\mu+\varepsilon)}\frac{(\mu+\varepsilon)^B}{B!}e^{-\varepsilon}\frac{\varepsilon^S}{S!} + \alpha(1-\delta)e^{-\varepsilon}\frac{\varepsilon^B}{B!}e^{-(\mu+\varepsilon)}\frac{(\mu+\varepsilon)^S}{S!} + (1-\alpha)e^{-\varepsilon}\frac{\varepsilon^B}{B!}e^{-\varepsilon}\frac{\varepsilon^S}{S!}. \tag{1.2}$$

The log likelihood function is

$$\begin{aligned} l(B,S|\theta) = &-2\varepsilon - log(B!S!) + log\{\alpha\delta exp[B\cdot log(\mu+\varepsilon) + S\cdot log(\varepsilon) - \mu] \\ &+ \alpha(1-\delta)exp[S\cdot log(\mu+\varepsilon) + B\cdot log(\varepsilon) - \mu] + (1-\alpha)exp[(B+S)\cdot log(\varepsilon)]\}. \end{aligned} \tag{1.3}$$

I drop $-log(B!S!)$ because this constant term does not affect the MLE outcome. The estimation suffers from a data overflow problem when daily number of buy or sell trades becomes too large. A popular treatment is to factor out a common component in the exponential terms. However, if the common factor is too large, it can lead to an underflow problem when all exponential terms shrink to zero. It will not solve the overflow problem if the common factor is too small. Since there is substantial variation in daily number of buy and sell trades, it is difficult to find a constant common factor to fully resolve the numerical issue. The daily log likelihood function is likely to bounce between the overflow and underflow problems. To circumvent the numerical difficulty, I factor out the largest exponential term in the curly brackets in equation (3) instead of a pre-determined constant term. Denote

$$a_1 = exp[B\cdot log(\mu+\varepsilon) + S\cdot log(\varepsilon) - \mu], \tag{1.4}$$

$$a_2 = exp[S\cdot log(\mu+\varepsilon) + B\cdot log(\varepsilon) - \mu], \tag{1.5}$$

$$a_3 = exp[(B+S)\cdot log(\varepsilon)], \tag{1.6}$$

$$max = max\{a_1, a_2, a_3\}. \tag{1.7}$$

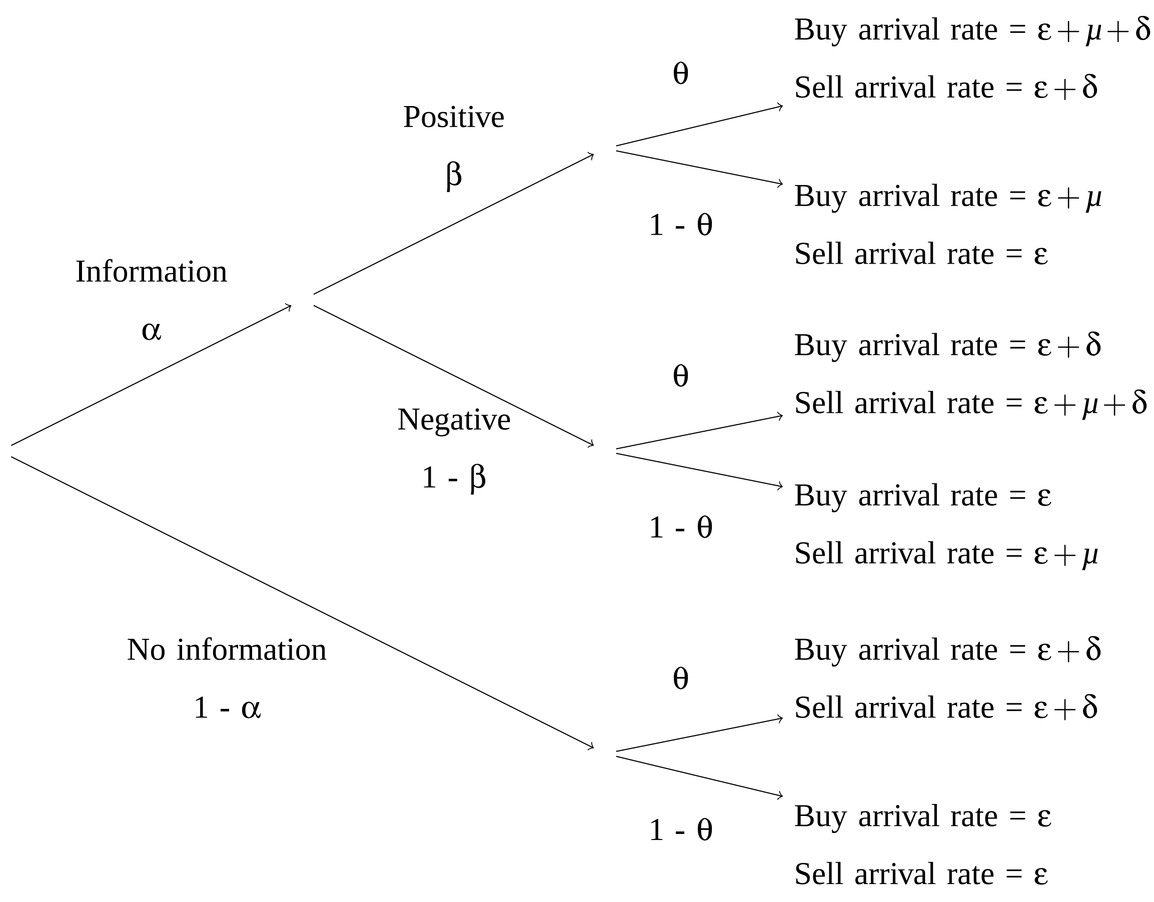
It can be shown that $max = a_1$ when $B \geq S$ and $B \geq \mu k$, where $k = log(\frac{\varepsilon}{\mu+\varepsilon})$; $max = a_2$ when $S \geq B$ and $S \geq \mu k$; and $max = a_3$ when $\mu k \geq B$ and $\mu k \geq S$. The estimation is then based on the following log likelihood function:

$$l(B,S|\theta) = -2\varepsilon + log(max) + log\{\alpha\delta\frac{a_1}{max} + \alpha(1-\delta)\frac{a_2}{max} + (1-\alpha)\frac{a_3}{max}\}. \tag{1.8}$$

Using this dynamic factor guarantees the log likelihood function can be calculated as long as the parameters α and δ do not touch the bounds. To ensure the estimation is the global maximum rather than a local maximum, I estimate the model using initial parameters from a $5 \times 5 \times 5 \times 5$ matrix, with $\alpha, \delta \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, and $\mu, \varepsilon \in \{0.1m, 0.3m, 0.5m, 0.7m, 0.9m\}$, where m is $max_{0,T}\{max\{B_t\}, max\{S_t\}\}$ within estimation period $[0,T]$. α and δ are bounded between 0 and 1, and μ and ε are positive. I use logit transformation of α and δ and natural logarithm of μ and δ to make sure the estimates are within the boundaries. The estimation is done using the *fminsearch* function in Matlab. Out of the 625 estimations from different initial parameter sets, I first drop all corner solutions and then choose the parameter set that yields the greatest log likelihood.

2. The APIN-PSOS model

This appendix describes the APIN-PSOS model of Duarte and Young (2009). The information structure and trading rules are the same as in the EO model. The difference is that the arrival rate of uninformed traders can symmetrically increase by δ on days of liquidity shock with probability of θ, unconditional on the information type of the day. The trading game is illustrated in the following figure.



The probability of informed trading in this adjusted model is now:

$$APIN = \frac{\alpha\mu}{\alpha\mu + 2\theta\delta + 2\varepsilon}. \tag{2.1}$$

The liquidity shock is of particular interest of this model. The unconditional probability of a trading being the result of a liquidity shock can be calculated as

$$APIN = \frac{2\theta\delta}{\alpha\mu + 2\theta\delta + 2\varepsilon}. \tag{2.2}$$

The likelihood function of the model is

$$\begin{aligned} L(B,S|\theta) =& \alpha\beta(1-\theta)e^{-(\mu+\varepsilon)}\frac{(\mu+\varepsilon)^B}{B!}e^{-\varepsilon}\frac{\varepsilon^S}{S!} + \alpha\beta\theta e^{-(\mu+\delta+\varepsilon)}\frac{(\mu+\delta+\varepsilon)^B}{B!}e^{-(\delta+\varepsilon)}\frac{(\delta+\varepsilon)^S}{S!} \\ &+ \alpha(1-\beta)(1-\theta)e^{-\varepsilon}\frac{\varepsilon^B}{B!}e^{-(\mu+\varepsilon)}\frac{(\mu+\varepsilon)^S}{S!} + \alpha(1-\beta)\theta e^{-(\delta+\varepsilon)}\frac{(\delta+\varepsilon)^B}{B!}e^{-(\mu+\delta+\varepsilon)}\frac{(\mu+\delta+\varepsilon)^S}{S!} \\ &+ (1-\alpha)(1-\theta)e^{-\varepsilon}\frac{\varepsilon^B}{B!}e^{-\varepsilon}\frac{\varepsilon^S}{S!} + (1-\alpha)\theta e^{-(\delta+\varepsilon)}\frac{(\delta+\varepsilon)^B}{B!}e^{-(\delta+\varepsilon)}\frac{(\delta+\varepsilon)^S}{S!}. \end{aligned} \tag{2.3}$$

The model can be estimated using the same strategy in the previous appendix.

Bibliography

Amin, K. I., and C. M. C. Lee, 1997, "Option Trading, Price Discovery, and Earnings News Dissemination," *Contemporary Accounting Research*, 14, 153–192.

Ané, T., and H. Geman, 2000, "Order Flow, Transaction Clock and Normality of Asset Returns," *Journal of Finance*, 55(5), 2259–2284.

Anthony, J. H., 1988, "The Interrelation of Stock and Options Market Trading-Volume Data," *Journal of Finance*, 43, 949–964.

Aslan, H., D. Easley, S. Hvidkjaer, and M. OHara, 2008, "Firm Characteristics and Informed Trading: Implications for Asset Pricing," *Journal of Empirical Finance*.

Asquith, P., P. A. Pathak, and J. R. Ritter, 2005, "Short interest, institutional ownership, and stock returns," *Journal of Financial Economics*, 78, 243–276.

Back, K., 1993, "Asymmetric Information and Options," *Review of Financial Studies*, 6, 435–472.

Battalio, R., B. Hatch, and R. Jennings, 2004, "Toward a National Market System for U.S. Exchange-Listed Equity Options," *Journal of Finance*, 59(2), 933–962.

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

Biais, B., and P. Hillion, 1994, "Insider and Liquidity Trading in Stock and Options Markets," *Review of Financial Studies*, 7, 743–780.

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., 1998, "Valuing Options in Regime-Switch Models," *Journal of Derivatives*, 6(1), 38–49.

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.

Botosan, C., and D. Skinner, 1993, "The Relation Between Post-Earnings-Announcement Drift and Options Listing," working paper, University of Michigan.

Branch, B., and J. E. Finnerty, 1981, "The Impact of Option Listing on the Price and Volume of the Underlying Stock," *Financial Review*, 16, 1–15.

Brown, S., S. A. Hillegeistb, and K. Lo, 2004, "Conference calls and information asymmetry," *Journal of Accounting and Economics*, 37, 343–366.

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

Cao, H. H., 1999, "The Effect of Derivative Assets on Endogenous Information Acquisition and Price Behavior in a Rational Expectations Equilibrium," *Review of Financial Studies*, 12, 131–163.

Carhart, M. M., 1997, "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52(1), 57–82.

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

Carr, P., and L. Wu, 2008, "Leverage Effect, Volatility Feedback, and Self-Exciting Market Disruptions: Disentangling the Multi-dimensional Variations in S&P 500 Index Options," working paper, Bloomberg and Baruch College.

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

Carr, P., and L. Wu, 2010, "Stock Options and Credit Default Swaps: A Joint Framework for Valuation and Estimation," *Journal of Financial Econometrics*, forthcoming.

Carr, P., and L. Wu, 2011, "A Simple Robust Link Between American Puts and Credit Protection," *Review of Financial Studies*, 24, 473–505.

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.

Chan, K., Y. P. Chung, and H. Johnson, 1993, "Why Option Prices Lag Stock Prices: A Trading-Based Explanation," *Journal of Finance*, 48(5), 1957–1967.

Chordia, T., R. Roll, and A. Subrahmanyam, 2002, "Order Imbalance, Liquidity and Market Returns," *Journal of Financial Economics*, 65, 111–130.

Chordia, T., and A. Subrahmanyam, 2004, "Order imbalance and individual stock returns: Theory and evidence," *Journal of Financial Economics*, 72, 485–518.

Conrad, J., 1989, "The Price Effect of Option Introduction," *Journal of Finance*, 44, 487–498.

Danielsen, B. R., B. F. V. Ness, and R. S. Warr, 2007, "Reassessing the Impact of Option Introductions on Market Quality: A Less Restrictive Test for Event-Date Effects," *Journal of Financial and Quantitative Analysis*, 42, 1041–1062.

Danielsen, B. R., and S. M. Sorescu, 2001, "Why Do Option Introductions Depress Stock Prices? A Study of Diminishing Short Sale Constraints," *Journal of Financial and Quantitative Analysis*, 36, 451–484.

Detemple, J., and P. Jorion, 1990, "Option listing and stock returns : An empirical analysis," *Journal of Banking and Finance*, 14, 781–801.

Diamond, D. W., and R. E. Verrecchia, 1987, "Constraints on Short-Selling and Asset Price Adjustment to Private Information," *Journal of Financial Economics*, 18(2), 277–312.

Dong, W., and N. Sinha, 2011, "Where do informed traders trade: Trading around news on Dow 30 options," working paper, University of Illinois at Chicago.

Duarte, J., and L. Young, 2009, "Why is PIN priced?" *Journal of Financial Economics*, 91, 119–138.

Easley, D., R. Engle, M. O'Hara, and L. Wu, 2008, "Time Varying Arrival Rates of Informed and Uninformed Trades," *Journal of Financial Econometrics*, 6, 171–208.

Easley, D., S. Hvidkjaer, and M. O'Hara, 2002, "Is Information Risk a Determinant of Asset Returns?" *Journal of Finance*, 57(5), 2185–2221.

Easley, D., N. M. Kiefer, and M. O'Hara, 1996, "Cream-Skimming or Profit-Sharing? The Curious Role of Purchased Order Flow," *Journal of Finance*, 51(3), 811–833.

Easley, D., N. M. Kiefer, and M. O'Hara, 1997, "One Day in the Life of a Very Common Stock," *Review of Financial Studies*, 10(3), 803–835.

Easley, D., and M. O'Hara, 1987, "Price, Trade Size, and Information in Securities Markets," *Journal of Financial Economics*, 19(1), 69–90.

Easley, D., and M. O'Hara, 1992, "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47(2), 577–605.

Easley, D., and M. O'Hara, 2004, "Information and the Cost of Capital," *Journal of Finance*, 59(4), 1553–1583.

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.

Easley, D., L. D. Prado, and M. O'Hara, 2011, "Flow Toxicity and Volatility in a High Frequency World," working paper, Cornell University.

Egloff, D., M. Leippold, and L. Wu, 2010, "The Term Structure of Variance Swap Rates and Optimal Variance Swap Investments," *Journal of Financial and Quantitative Analysis*, 45, 1279–1310.

Engle, R., 2004, "Risk and Volatility: Econometric Models and Financial Practice," *American Economic Review*, 94(3), 405–420.

Fama, E. F., and J. D. MacBeth, 1973, "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81(3), 607–636.

Fedenia, M., and T. Grammatikos, 1992, "Options Trading and the Bid-Ask Spread of the Underlying Stocks," *Journal of Business*, 65, 335–351.

Figlewski, S., and G. Webb, 1993, "Options, Short sales, and Market Completeness," *Journal of Finance*, 48(2), 761–778.

Finucane, T. J., 1991, "Put-Call Parity and Expected Returns," *Journal of Financial and Quantitative Analysis*, 26, 445–457.

Glosten, L., and L. Harris, 1988, "Estimating the Components of the Bid-Ask Spread," *Journal of Financial Economics*, 21, 123–142.

Glosten, L. R., and P. R. Milgrom, 1985, "Bid, Ask, and Transaction Prices in a Specialist Market with Heterogenously Informed Traders," *Journal of Financial Economics*, 14(1), 71–100.

Grossman, S. J., 1988, "An Analysis of the Implications for Stock and Futures Price Volatility of Program Trading and Dynamic Hedging Strategies," *Journal of Business*, 61, 275–298.

Henderson, B., and N. Pearson, 2010, "The Price Impact of Large Hedging Trades," working paper, University of Illinois at Urbana-Champaign.

Henderson, B., and N. Pearson, 2011, "The darkside of financial innovation: Acasestudy of the pricing of aretailfinancialproduct," *Journal of Financial Economics*, 100, 227–247.

Ho, L.-C. J., 1993, "Option Trading and the Relation between Price and Earnings: A Cross-Sectional Analysis," *Accounting Review*, 68, 368–384.

Ho, T., and H. Stoll, 1981, "The Dynamics of Dealer Markets under Competition," *Journal of Finance*, 38, 1053–1074.

Holowczak, R., J. Hu, and L. Wu, 2011, "Consolidating information in option transactions," working paper, Baruch College.

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.

Hu, J., 2011, “Does Option Trading Convey Stock Price Information?” working paper, Baruch College and National University of Singapore.

Huang, J., and J. Wang, 1997, “Market Structure, Security Prices and Informational Efficiency,” *Macroeconomic Dynamics*, 1, 169–205.

Izzeldin, M., 2007, “Trading volume and the number of trades: a comparative study using high frequency data,” working paper, Lancaster University Management School.

Jarnecic, E., 1999, “Trading Volume Lead/Lag Relations Between the ASX and ASX Option Market: Implications of Market Microstructure,” *Australian Journal of Management*, 24, 77–94.

Jennings, R., and L. T. Starks, 1986, “Earnings Announcements, Stock Price Adjustments and the Existence of Exchange-Listed Options,” *Journal of Finance*, 41, 107–125.

Johnson, T. L., and E. C. So, 2012, “The option to stock volume ratio and future returns,” *Journal of Financial Economics*, 106, 262–286.

Jones, C. M., G. Kaul, and M. L. Lipson, 1994, “Information, Trading, and Volatility,” *Review of Financial Studies*, 36(1), 127–154.

Kumar, R., A. Sarin, and K. Shastri, 1998, “The Impact of Options Trading on the Market Quality of the Underlying Security: An Empirical Analysis,” *Journal of Finance*, 53, 717–732.

Kyle, A. S., 1985, “Continuous Auctions and Insider Trading,” *Econometrica*, 53(6), 1315–1336.

Lakonishok, J., I. Lee, N. D. Pearson, and A. M. Poteshman, 2007, “Option Market Activity,” *Review of Financial Studies*, 20, 813–857.

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

Lin, J.-C., G. C. Sanger, and G. G. Booth, 1995, “Trade size and components of the bid-ask spread,” *Review of Financial Studies*, 8, 1153–1183.

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

Mayhew, S., and V. Mihov, 2004, “How Do Exchanges Select Stocks for Option Listing?” *Journal of Finance*, 59(1), 447–471.

McInish, T. H., and J. Upson, 2012, “Strategic Liquidity Supply in a Market with Fast and Slow Traders,” working paper, University of Memphis.

Mendenhall, R. R., and D. H. Fehrs, 1999, “Option listing and the stock-price response to earnings announcements,” *Journal of Accounting and Economics*, 27, 57–87.

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

Milgrom, P., and N. Stokey, 1982, “Information, Trade and Common Knowledge,” *Journal of Economic Theory*, 26, 17–27.

Miller, E. M., 1977, “Risk, Uncertainty, and Divergence of Opinion,” *Journal of Finance*, 32, 1151–1168.

Muravyev, D., N. D. Pearson, and J. P. Broussard, 2013, “Is There Price Discovery in Equity Options?” *Journal of Financial Economics*, 107(2), 259–283.

Newey, W. K., and K. D. West, 1987, “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55(3), 703–708.

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

Ni, S. X., N. D. Pearson, and A. M. Poteshman, 2005, “Stock Price Clustering on Option Expiration Dates,” *Journal of Financial Economics*, 78(1), 49–87.

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

Poteshman, A. M., 2006, “Unusual Option Market Activity and the Terrorist Attacks of September 11, 2001,” *Journal of Business*, 79(4), 1703–1726.

Roll, R., E. Schwartz, and A. Subrahmanyam, 2010, “O/S: The relative trading activity in options and stock,” *Journal of Financial Economics*, 96, 1–17.

Simaan, Y., and L. Wu, 2007, “Price Discovery in the U.S. Stock Options Market,” *Journal of Derivatives*, 15(2), 20–38.

Skinner, D., 1990, “Options Markets and the Information Content of Accounting Earnings Releases,” *Journal of Accounting and Economics*, 13, 191–211.

Skinner, D. J., 1989, “Options Markets and Stock Return Volatility,” *Journal of Financial Economics*, 23, 61–76.

Sorescu, S. M., 2000, “The Effect of Options on Stock Prices: 1973 to 1995,” *Journal of Finance*, 55, 487–514.

Stein, J. C., 1987, “Informational Externalities and Welfare-Reducing Speculation,” *Journal of Political Economy*, 95, 1123–1145.

Stephan, J. A., and R. E. Whaley, 1990, “Intraday Price Change and Trading Volume Relations in the Stock and Stock Option Markets,” *Journal of Finance*, 45, 191–220.