Essays on the microstructure of US equity options

Mengyu Zhang

A thesis submitted for the degree of Doctor of Philosophy

School of Business

University of Essex

January 2020

Acknowledgement

Undertaking this Ph.D was literally a long journey for me. From Bath, Newcastle, to Essex, I would not have been able to finish my Ph.D journey without the support and guidance that I received from many people.

First and foremost, I would like to express my special appreciation and thanks to my supervisor Dr. Thanos Verousis for all the support, guidance, and encourage he gave me. I am indeed fortunate to have the supervisor who is patient and knowledgeable. Without him guidance and constant feedback, I would not achieve this Ph.D.

Besides my supervisor, I greatly appreciate the collaborative work undertaken with Professor Alejandro Bernales, Professor Iordanis Angelos Kalaitzoglou, and Dr. Nikolaos Voukelatos. I have further learned how to become a professional researcher and work with others.

I would also like to thank all my friends who are always helpful in many ways – thank you to Dingshuang and He,

A very special thanks to my mother for always believing, supporting, and encouraging me throughout my Ph.D. This thesis is dedicated to her.

Abstract

This thesis is structured as three essays on market microstructure of US equity options. The appearance of high-frequency trading changed the process of trading and the structure of financial markets. Additionally, literature on individual equity options is relatively low because of the problems with data availability. Hence, the three essays investigate the market microstructure of US equity options from three aspects through a high-frequency dataset including all options contracts written the 30 components of the Dow Jones Industrial Average from January 2012 to June 2014. The first essay fulfils research gaps by investigating the intraday commonality in options liquidity in a quote-driven market. It finds that the commonality in option liquidity is driven by the inventory risks and represents a higher level at the beginning of the trading day.

The second essay investigates the informational content of options order flow through a new proxy of predictive information. It finds that the options trades contain predictive information about the future price movements of underlying stocks. This predictive power is varied across information contents of calls and puts. In particular, the informational contents of calls have the longer predictive horizon in stock volatility and those of puts have greater predictive power on stock returns.

The third essay focuses on price clustering and size clustering which has been limited to investigations into options markets. It provides supportive evidence that prices and sizes are clustered in the options market. The relationships between price and size clustering have different between calls and puts. For calls, price clustering has a positive impact on size clustering. For puts, size clustering is negatively related to price clustering. Moreover, the moneyness and maturity can also influence price and

size clustering because this essay finds out-of-the money contracts and near-to-maturity contracts show less price and size clustering.

Contents

List of Tables

Lists of Figures

Chapter 1: Introduction

1.1 Research background and motivations

Market microstructure was first introduced by Garman (1976) when he developed an article about market making and inventory costs. After years of development, market microstructure became a research field that focuses on examining how economic forces influence trades, quotes, and prices (Biais *et al.*, 2005). Madhavan (2000) indicated the development of market microstructure is primarily driven by changes in the structure, regulation, and technology in the securities industry. For example, O'Hara (2015) confirmed that the appearance of high-frequency trading changed the way of trading and the structure of financial markets. Although the central idea of this thesis is about market microstructure, it takes wide interests in the implications for market liquidity, trading strategy design, and asset pricing. In order to provide a better understanding of market microstructure, this thesis presents three particular empirical essays to investigate the following important questions:

1. What are the determinants of liquidity commonality in a quote-driven options market?
2. Do options order flows contain predictive information on future underlying stock price movement?
3. What is the relationship between price clustering and size clustering in the options market?

To address these research questions, this thesis uses a high-frequency dataset including all equity contracts written on the 30 components of the Dow Jones Industrial Average (DJIA) from January 2012 to June 2014. These equity contracts are listed on the Chicago Board Options Exchange (CBOE) which is one of the world's largest exchanges. Studies on individual equity options are relatively fewer than those on index options because of the problems with data availability and relatively low trading activities observed in equity options markets. However, the trading activity on equity options has increased in the recent years. For example, the annual total option contract volume in CBOE increased from around 1.11 billion in 2012 to over 1.27 billion in 2014. Particularly, the average daily volume of options increased by 14% from 4.44 million to 5.06 million during the period. The equity options were considered as the largest source of the trading volume that accounted for around 36% of the total volume. Hence, this thesis focuses on CBOE to examine the above research questions.

The first research question of this thesis focuses on the co-movement between liquidity in individual assets. Chordia *et al.* (2000) provided the first empirical study that shows this liquidity commonality in US stock markets. Huberman and Halka (2001), Galariotis and Giouvris (2007), and Moshirian *et al.* (2017) further provided evidence of liquidity commonality in stock markets around the world. Cao and Wei (2010) and Verousis *et al.* (2016a) extended the work of commonality in liquidity to options markets. However, few studies have explored the intraday pattern of commonality in liquidity is in the quote-driven markets. Since the quote-drive markets have market makers who have the obligation to maintain liquidity, the trading regimes may influence the existence and pattern of liquidity commonality (Brockman and Chung, 2002). Another motivation of this research is unclear fundamental sources of commonality in option liquidity. Moshirian *et al.* (2017) showed asymmetric information risks as the determinants of commonality in liquidity, while Huberman and Halka (2001) found that neither asymmetric information risks nor inventory risks could explain commonality in liquidity. Hence, we explore the determinants of liquidity commonality in the equity options market.

The second research explores whether options order flows contain predictive information about further underlying stock price movement. Previous literature (e.g. Schlag and Stoll, 2005; Pan and Poteshman,2006; Hsieh and He, 2014) investigated the information on options trades through options volumes and trade directions. However, Dufour and Engle (2000) and Engle (2000) found that trade duration also contained information about future financial returns. Manganelli (2005) and Wong *et al.* (2009) confirmed this finding by using the autoregressive conditional duration (ACD) model to capture information involved trade duration. Moreover, Kalaitzoglou and Ibrahim (2013) combined volume and duration as one variable which was used in the Smooth Transition Autoregressive Conditional Weighted Duration (STM-ACWD) model. Although these studies confirmed the information involved in duration, they did not estimate the predictability of captured information in the multiple-market setting. In this context, we conjecture that the underlying stock price movement can be predicted by the information extracted from trade duration.

The third research of this thesis investigates price and size clustering in the options market and the determinants of them. There is a large amount of literature that examines price clustering in different financial markets, such as equity markets (Lien *et al.*, 2019), futures markets (ap Gwilym and Alibo, 2003), and options markets (Capelle-Blancard and Chaudhury, 2007). Size clustering has been investigated in equity markets and futures market (ap Gwilym and Meng, 2010; Verousis and ap Gwilym, 2013b) but not in options markets so far. Hence, the third research focuses on the price and size clustering in the options market. Further, we investigate the relationship between price and size clustering and how the characteristics of options contracts influence the clustering. However, previous literature provided mixed results for these effects. For example, ap Gwilym and Meng (2010) and Meng *et al.* (2013)

showed a negative relationship between price and size clustering, while Verousis and ap Gwilym (2013b) found a positive relationship. Besides, Capelle-Blancard and Chaudhury (2007) and ap Gwilym and Verousis (2013) also found different effects of moneyness and maturity on price clustering. Therefore, we attempt to develop more understanding about the reasons of price and size clustering.

1.2 Primary findings and contributions

This thesis investigates the market microstructure of US equity options from three aspects: commonality in options liquidity, predictability of information conveyed in options trades, and options price and size clustering. Empirical findings of Chapter 4 (1st empirical chapter) show that the intraday commonality in options liquidity has the highest level at the beginning of the trading day. The level of liquidity commonality begins to reduce after the first and half hours of the opening of the market. These results can be potentially interpreted as the effects of information asymmetry and inventory risks. However, our empirical results show inventory risks as the main determinants of liquidity commonality because we find that market volatility and down market can lead to higher commonality in liquidity. Since market-wide volatility is related to multiple options, market makers would like to adjust the spreads and depth of these options simultaneously that contribute to commonality in liquidity. Besides, during market declines, the higher order imbalance occurs because of correlated trading that improves the inventory risk and thereby improve commonality in liquidity. With these findings, this essay provides a further understanding of the systematic liquidity risk. Since this risk is connected with asset pricing, the further understanding may contribute to the development of asset pricing models. The variation of systematic liquidity movement can have impacts on the paradigm of asset pricing models and support investors to develop intraday trading strategies. Additionally, the exploration on determinants of liquidity commonality contributes to understanding liquidity shocks in options markets.

Chapter 5 (2nd empirical chapter) focuses on the information contained in options trades. Our empirical results confirm that the informational content of option order flow can be captured by the volume-weighted duration and the STM-ACWD model. Although the volume-weighted duration can be used as the proxy of predictive information, it cannot directly reflect the level of the information. Hence, we create three sets of information measures based on the estimation results of STM-ACWD model. Our results confirm the predictive power of these information measures. In particular, we find that these information measures can significantly influence contemporaneous and future stock volatility and return. The information contained in calls and puts also show different levels of predictive power. Informational contents of call options transactions have the greater predictive ability in stock volatility and those of put options transactions show greater ability in predicting stock return. This essay makes important contributions to the literature by confirming the predictive power of options trade intensity on the underlying stock price movements. Additionally, these results suggest that informed traders are also active in options markets. They incorporate information into the options market through their trading activities.

Chapter 6 (3rd empirical chapter) investigates the price and size clustering in CBOE and provides supportive evidence that trades and quotes are clustered in the options market. To give more details, quoted prices are more clustered than traded prices, while quoted size is less clustered than traded sizes. We subsequently investigate the determinants of price and size clustering. The empirical results show one-way relationships between price and size clustering. Additionally, these relationships are reversed between calls and puts. For call options, there are positive effects of price clustering on size clustering. For put options, we find the negative effects of size clustering on price clustering. Our results also demonstrate that price and size are less clustered when trading out-of-the money contracts and near-to-maturity contracts with other factors controlled. This essay does not only document the relationship between price and size clustering but also highlights how the characteristics of options contracts impact the price and size clustering. Due to the trade-offs between price, quantity, and execution speed, the empirical findings have important implications for developing trading strategies

1.3 Structure of the thesis

This thesis consists of seven chapters and is organised as follows. Chapter 1 provides the background, motivation, primary findings, and contributions of this thesis. Chapter 2 offers further information about market microstructure which is the theoretical framework of this thesis. Chapter 3 presents a systematic literature review of equity options. Chapter 4 provides the data cleaning method. Chapter 5 to 7 are three empirical chapters that investigate the research questions proposed in Chapter 1. In particular, Chapter 5 (1^{st} empirical chapter) investigates the determinants of liquidity commonality in CBOE. Chapter 6 (2^{nd} empirical chapter) examines the predictability of information contained in options trades. Chapter 7 (3^{rd} empirical chapter) investigates the determinants of price and size clustering. Chapter 8 summarises this thesis.

Chapter 2: Theoretical framework: market microstructure theory

2.1 Introduction

Literature on market microstructure investigates the process and outcomes of trading financial assets under explicit trading regulations. In particular, the microstructure literature usually focuses on a select set of issues, such as price formation, market design, account liquidity (Madhavan, 2000). This chapter shows select parts of the theory of market microstructure that are related to the empirical chapters of this thesis. The next section shows that the issues of market design, followed by market liquidity.

2.2 Market design

To investigate price formation or market liquidity, the initial focus of the literature on market microstructure is logical centring the position of market design in the trading process. Previous studies show that market structure and design can influence the speed of price discovering, liquidity, and the cost of trading (Brockman and Chung, 2002; Madhavan, 2002). In this section, the discussions of market structure and design focus on the quote- and order-driven market systems.

It is useful to understand the taxonomy of market structure that can lead the subsequent discussion. Market structure, or market architecture, is constructed by a set of rules governing the trading process. Madhavan (2000) indicates that the decision of market architecture is related to a variety of attributes: market type, price discovery, order forms, protocols, and transparency.

- Market type includes three elements. The first element is related to the degree of continuity that leads to periodic and continuous systems. Trades in the former system can only occur at specific points in a time, while those in the latter system can occur at any point in trading hours. The second element is the dealer presence. Trades can occur between public investors without dealer intermediation in auction (or order-driven) markets, while a market maker is in the opposite side of each transaction in a dealer (or quote-driven) market. The last element is the degree of automation (floor or screen-based electronic systems) that reflects the role of technology in order submission.

- Price discovery shows whether the market can offer independent price discovery because some markets may use prices determined in other markets as the basis for transactions.
- Order forms permitted contain market order, limit order, hidden order, etc.
- Transparency reflects the quantity and quality of information which is received by market participants during the trading process. In transparent markets, the participants can receive relevant information before and after trades. Transparency is also reflected in the extent and speed of dissemination, degree of anonymity, and in whether after hours trading is permitted.

Considerable research is inspired by the diversity of systems. One group of these studies show the distinct values of the dealer market that cannot be provided by the auction markets (Madhavan, 2000; Frino et al., 2008; Anand et al., 2016). In a quote-driven market, investors trade through financial intermediaries (dealers, specialists, or market makers). Hence, asset prices are determined by bid and ask quotes made by these intermediaries (Harris, 2003). The major benefit of a quote-driven market is the flexible response of dealers to liquidity needs. Since market makers have the obligation to provide liquidity via their own inventory, they are able to respond quickly to changing market condition (Stoll, 2003).

The designed market makers are not existed in order-driven markets. Investors can submit the price and quantitates of an asset that they desire to buy or sell. All these buy and sell orders are shown in a limit order book and ranked by the trading system. The system will match the highest-ranking orders at the minimum order amount. If there is a remaining size in a buy (sell) order, the system will further match it with the next order in the rankings (Harris, 2003). Although all market participants can provide liquidity in an auction market, Stoll (2003) indicates that pure auction markets cannot offer sufficient liquidity, especially in less active assets.

2.3 Market Liquidity

Huberman and Halka (2001) indicate that it is difficult to define liquidity due to its abstract properties. Since trading is a bilateral search process in which buyers and sellers search for each other, the definition of liquidity is related to both sides of trading (Lehalle and Laruelle, 2013). During this search process, traders characterise search problems as cost, size, and time. These problems reflect three dimensions of liquidity which are price, quantity, and execution speed (Moulton, 2005; Hodrick and Moulton, 2009). In particular, the price dimension is represented by the bid-ask spread; the quantity dimension is represented by depth; the execution speed dimension is represented by trade duration. Regarding these dimensions, liquidity could be generally defined as the ability of markets to absorb large trading volume without significant impact on prices (Massimb and Phelps, 1994). In a perfect liquid market, traders can trade their desired quantity at any time without influencing the market price. However, the real financial markets are not perfect liquid. Traders should sacrifice one or more dimensions of liquidity in order to execute their transactions quickly. This phenomenon shows trade-offs between the three dimensions of liquidity (Moulton, 2005). In this context, a group of literature on market microstructure focus on price and size clustering. Fundamentally, this group of literature investigates the different dimensions that selected by traders to sacrifice.

Under the dimensions of liquidity, furthermore, microstructure research on market liquidity study the approach to explain the liquidity changes and, in turn, influence price formation (Madhavan, 2000). The early literature is concerned with the market maker who provides liquidity to the market and thereby support continuous trading by fulfilling gaps in the arrival of investor orders (Stoll, 2003). Hence, the literature proposes an inventory explanation that focuses on the inventory position of market makers (Stoll, 1978; Amihud and Mendelson, 1982). However, more recently, the information-based explanation has become more popular that integrate asymmetric information into the liquidity changes (Kyle, 1985; Glosten, 1994). The detailed information of these two explanations is provided in the following sections.

2.3.1 Inventory-based explanation

The inventory explanation is developed from the importance of market makers' inventory. Since market makers usually do not know the fundamental values of financial instruments, large inventory positions may expose them to losses when the market moves against them. Thus, market makers have the desired inventory level and use bid and ask prices to optimise their inventory levels. Particularly, these prices could influence the stochastic arrival rates of sellers and buyers (Huberman and Halka, 2001). In this context, market makers are not only active providers of immediacy, but also passively manage quoted bid and ask prices based on their inventory positions (Smidt, 1971). The bid-ask spread is a source of profit to compensate market makers for exposure to risks and order-processing costs (Stoll, 1978). Since market makers seek to adjust their inventory by changing the general level of the quoted bid and ask prices, this, in turn, increases the liquidity in the market.

2.3.2 Information-based explanation

Other studies, however, suggest that inventory explanation does not sufficiently recognise the importance of information in the market (Madhavan and Smidt, 1991; Huang and Stoll, 1997). De Jong and Rindi (2009) argue that the major limitation of inventory explanation is its boundedness. Frino and Fabre (2004) suggest that inventory holding costs are less relevant to order-driven markets due to the absence of market makers. More specifically, there are no intermediaries in such markets who the institutional obligation to offer liquidity continuously. In this context, researchers (e.g. Kyle, 1985) have developed an asymmetric-information-based explanation of liquidity.

In this information-based explanation, there are three types of traders, namely informed traders, noise traders, and market makers. Informed traders have superior information and expect to profit from their information in trades with uninformed traders. Noise traders have no such informational advantages, though they are liquidity-motivated (Madhavan, 2000). Market makers also possess no specific information. Since they should provide liquidity to the market continually, they need to obtain a fair return on their capital. By using their vantage point, market makers could adjust bid and ask prices to ensure that they purchase assets at lower prices than those for which they are sold (Huberman and Halka, 2001). Chordia *et al.* (2001) recognised that profitable spreads are created by this process. Moreover, market makers should keep the spread wide enough to obtain profit from trading with uninformed traders, and thereby offset the losses from trading with informed traders. They, in turn, provide a continuous market. With adjusting bid-ask spread, the information is conveyed by the trading process (Pan and Poteshman, 2006; Ryu and Yang, 2018).

2.4 Conclusion

Studies of market microstructure investigate the interrelation between institutional structure, investor behaviour, and prices. This chapter shows the theoretical framework applied in the empirical chapters of this thesis. In particular, this chapter introduces attributes (market type, price discovery, order forms, protocols, and transparency) related to market design and thereby brief describe quote- and order-driven markets. The different market design may influence the level of liquidity in each market. Hence, this chapter further discusses the dimensions of liquidity: price, quantity, and execution speed. In a well-functioning market, traders could pursue all these dimensions. However, in the real financial market, traders may need to sacrifice one or two dimensions during the trading process. Moreover, this chapter also provides inventory- and information-based explanations to support the understanding of liquidity.

Chapter 3: What do we know about individual equity options?

3.1 Introduction[1]

What is the current knowledge regarding individual equity options? In spite of the rapid growth of equity option markets since the first day of trading on the Chicago Board Option Exchange (CBOE) on April 26, 1973, there has been no effort in the financial economics literature to consolidate, in a literature review, the current knowledge and understanding of individual equity options. The number of empirical papers examining individual equity options has increased at a slower rate than the number examining index options. This is mainly due to problems with data availability and relatively low trading activity observed in individual equity option markets, compared with index options that have historically been highly traded. However, data availability and trading activity on individual equity options have increased in the last decades, such that the volume of empirical papers on individual equity options has now reached a level that merits a survey of this literature.

The objective of this study is to offer a systematic review of the empirical literature on individual equity options, by discussing questions examined, data sets used and main findings, and providing some avenues for future research. Our survey of the equity options literature shows several research areas that have emerged, ranging from topics of relative consensus and solid understanding, to areas where the evidence is rather mixed and more research is required.

Although a chronological literature review could highlight the historical changes in the research field, studies on equity options span several research topics, which could make a chronological review needlessly complicated. Thus, studies in this survey are primarily ordered thematically, to provide a big picture of the knowledge on

[1] This chapter is forthcoming at Journal of Futures Markets

individual equity options. We mainly focus our attention on empirical studies where individual equity option data are used. Thus, in this literature review, we do not consider theoretical studies, which can be applied to options with other underlying types, such as indexes, bonds or exchange rates. However, in some parts of this empirical literature review, we will briefly mention theoretical advances that apply to options, as a means to better understand the results obtained in the empirical studies on individual equity options.

Before starting our literature review on individual equity options, it is important to answer the following question: What can we learn from empirical studies on *individual equity* options that we cannot learn from empirical studies on *index* options? This is an important question, since its answer makes the current survey valuable and provides a motivation for its development. There are many reasons why it is useful to analyse individual equity options rather than index options. Firstly, there are analyses that can be performed more cleanly with individual equity options than with index options. For example, the analysis of the factors that affect the introduction and success (in terms of trading activity) of new options listed for the very first time is difficult to perform with index options. This is because there are not many listings of index options in option market history, while there are plenty of listings of individual equity options (e.g., Mayhew and Mihov , 2004; Danielsen *et al*., 2007; and Bernales ,2017). For instance, on the first day of option trading on the CBOE in 1973, individual equity options were traded on 16 stocks, and no option contracts were traded on indexes (the first index option was introduced only 10 years later in 1983). The large number of listings of individual equity options in the following years, compared with listings of index options, can be observed in the current option market status. For example, in 2018, individual equity options were traded on 4,337 stocks, while index options were traded on only 34 indexes in the United States.[2] Moreover, the large number of equity options is not only useful for studies on option listings, but also allows researchers to perform robust *cross-sectional* analyses of the impact of option listings on the underlying assets, by controlling for specific features of the option contracts and stocks (e.g. trading activity, market volatility, firm industry, amongst others).

Secondly, in addition to topics that are difficult to examine using index options due to the very low frequency at which certain events are observed (such as option listings as discussed above), there are other research questions where equity options can provide a more fertile ground for analysis, due to the specific type of information that is relevant to their trading. This is the case with some studies that examine potential information flows between the option market and the underlying asset market (e.g. Stephan and Whaley, 1990; Chan *et al*., 2002; Muravyev *et al*., 2013). For instance, we will describe studies that show that levels of *informed* trading in the underlying stock market are reduced after the introduction of equity options, which can improve the price discovery process (i.e. the process by which information is progressively incorporated into prices). In fact, we can expect informed agents to use their *private* information for trades in stocks in which they have informational advantages, which is captured in the trading activity of stocks by market microstructure models (e.g. Easley *et al*., 1996, 1997, 1998a; Odders-White and Ready, 2008; and Duarte and Young, 2009). The private information that agents may have on indexes is not the same in *nature* as the private information on a particular stock, which makes the analysis of information flows between the option market and the underlying asset market different. In particular, private information on a particular *stock* is mainly related to undisclosed

[2] Information obtained from the Option Clearing Corporation web page, www.optionsclearing.com.

news or events regarding the firm that issued the stock, while private information on *indexes* mainly reflects some anticipated *global* economic view of the market.

Thirdly, individual equity options and index options are dissimilar in the sense that they attract different types of investors, and thus demand for them reacts to different factors. For instance, Lemmon and Ni (2014) show that equity options (index options) are actively traded by individual investors (sophisticated institutional investors). Lemmon and Ni (2014) also present evidence that trading activity in equity options is related to individual investors' sentiment and past market returns, while trades of index options are motivated by a hedging demand. Along similar lines, Johnson *et al.* (2016) show that index options are mainly used for hedging purposes in relation to crash risks for the whole market. As a result, findings from the literature on index options in terms of agents' behaviours cannot necessarily be extended to the case of individual equity options.

For all the reasons described above, a systematic review of the current state of the literature on individual equity options (which is independent of the index option literature) is both timely and particularly important. We start this literature review in Section 2 by discussing the relationship between the equity option market and the underlying stock market. Under the Black and Scholes (1973) assumptions, options written on individual stocks represent redundant securities. For instance, the Black and Scholes (1973) option pricing framework is based on the property that the payoff of an option contract can be replicated by a portfolio consisting of the underlying stock and a risk-free bond. Thus, we should expect that the listing of equity options will not affect the underlying stock market. However, in Section 2, we discuss a number of studies which provide evidence that the introduction of individual equity options (i.e. when they are listed for the very first time on the option exchange) does have an impact on the underlying equity securities.

The reported effect of equity option listing on the underlying stock market is a natural starting point for this literature review, since it also suggests that the efficiency of the equity option market may be rejected, which is the second topic of our survey. Thus, we continue the literature review in Section 3 with studies concerned with analysing the efficiency of the equity option market. Here, we discuss papers providing evidence that the market efficiency hypothesis is rejected when equity option data are used. For instance, there is evidence of 'permanent' option mispricing and abnormal returns on individual equity options.

Nevertheless, as argued by Malkiel and Fama (1970), any test of market *efficiency* is based on a model that specifies the *nature* of the market, in which prices should fully reflect all available information. Thus, any test of market efficiency is a *joint* test of efficiency and a particular pricing model. This means that some tests may reject the market efficiency hypothesis because the model behind the test is not well specified. One way for asset pricing models not to be well specified is if they do not incorporate market frictions coming from trading mechanisms in the option markets and/or investors' behavioural biases. This is particularly important because frictions from trading mechanisms and behavioural biases violate the fundamental assumptions of the efficient market hypothesis, in relation to there being no market frictions and investors being rational. Thus, we firstly discuss in Section 4 papers related to the effect of trading mechanisms on equity options. In particular, in this section, we examine studies related to the impact of market microstructure on the equity option market, including liquidity determinants, the market-making process, and changes in tick size, amongst other things. Afterwards, at the end of Section 4, we discuss studies that analyse trading behaviours of investors who depart from the rational-investor paradigm.

An alternative argument for why option markets are not efficient, other than the models potentially not being well specified, is that option markets may not be efficient under a strong form, but may be so under a semi-strong form. A market is semi-strong efficient when the current price only reflects information contained in past prices and all *public* knowledge (e.g. financial statements and news reports). Thus, a semi-strong form of efficiency recognizes that there are agents with *private* information that is not yet reflected in prices. Therefore, in Section 5, we discuss papers that analyse where informed investors trade (i.e. in the equity option market and/or in the underlying stock market), and the information flows between equity options and their underlying stock, which can help the price discovery process.

Afterwards, in Section 6, we discuss the type of *private* information revealed by informed investors through equity option prices (i.e. option-implied information). In this section, we describe studies that suggest some option-implied features of the equity option market that forecast underlying stock prices and returns. We also present papers about the option-implied information contained in option prices, regarding the future underlying stock volatility, and discuss studies suggesting that option-implied information can be extracted from equity options, in relation to the credit quality of the companies issuing the underlying stock.

Therefore, our overall objective is to provide a useful framework for understanding the current, wide scope of the empirical literature on equity options. Finally, Section 7 concludes and discusses potential directions for future research.

3.2 The impact of option listing on the underlying stock market

This section discusses the empirical literature on the impact option listing has on the underlying stock market, with Table 2-1 describing the related studies. We split the literature into three groups. In Section 2.1 (see Table 2-1 Panel A), we discuss empirical studies on the determinants of the introduction of new equity options into the market, and the ex-post success of such equity option listings. In Section 2.2 (see Table 2-1 Panel B) and Section 2.3 (see Table 2-1 Panel C), we analyse empirical studies that report short-term and long-term effects, respectively, of equity option listing, on the underlying stock market.

---------------Insert Table 2-1----------------

3.2.1 The option-listing decision

We begin the literature review by discussing papers related to the listing process for equity options (Table 2-1 Panel A). The listing process for equity options is very different to that of the Initial Public Offerings (IPOs) that take place in the underlying stock market, with the decision to conduct IPOs being taken by the company that issues the stocks, while the option-listing decision is taken directly by the option exchange. The option-listing process is examined by Mayhew and Mihov (2004) and Danielsen *et al.* (2007). In particular, they report that stock volatility and stock liquidity are the most important *ex-ante* (before the listing date) selection factors used by option exchanges to choose a stock to be used as the underlying asset for an option listing. This is expected, since equity options are more attractive when the volatility increases, given that investors may use options either to reduce their volatility exposure or to exploit changes in the levels of volatility by using option portfolios such as straddles.

In addition, stocks with high liquidity are more likely to have the attention of more market participants, who will also be willing to trade options with such stocks as the underlying.

Additionally, for each new option contract there is no initial *established number* of contracts that have to be traded. This again differs from IPOs, where the number of assets is exogenously determined by the issuer. Conversely, the number of option contracts is established through an endogenous process, based on the willingness of investors to participate in and trade the newly listed securities. Thus, a new call option contract is created (with a given moneyness and time-to-maturity, and with stock *S* as the underlying) when an investor wants to sell this contract and another investor is simultaneously willing to buy the same contract. In this sense, we can judge the success or failure of a new option's introduction based on the number of option contracts. For instance, Bernales (2017) examines factors that predict the *ex-post* success of stock option introductions, and shows that measures of information asymmetries predict option adoption rates. Informed traders will want stocks about which they have access to superior information to be optioned. This is because options offer cheap ways to effectively turn private information into profits, due to the leverage inherent in option contracts. Thus, the trading activity of informed investors, especially when an equity option has only recently been listed, may trigger the whole *ex-post* demand for the new option.

3.2.2 Short-term impact of option introductions on the equity market

After new equity options have been listed, it is interesting to analyse their impact on the underlying stocks (Table 2-1 Panel B). The empirical evidence on the impact of option listing on the underlying stock market is mixed. In one of the earliest studies on this topic, Detemple and Jorion (1990) examine individual equity options in the US market during 1973-1986, and document significantly positive stock returns during a two-week window surrounding option introduction. They also show that option introduction resulted in a temporarily lower stock volatility over this period. However, the impact of individual stock option listings on the underlying stock market seems to have disappeared in the later part of their sample period. Gjerde and Saettem (1995) and Watt *et al.* (1992) argue that this positive impact might be driven by liquidity suppliers charging a premium for their services, and by the hedging-related demands of dealers. Bollen (1998), Hamill *et al.* (2002), Gjerde and Saettem (1995), and Watt *et al.* (1992) further confirm the positive impact of option listing on the underlying stock prices, an effect that nevertheless gradually disappeared during the post-1980 period.

Freund *et al.* (1994) examine a similar sample and confirm the results of Detemple and Jorion (1990) in terms of lower stock volatility following option listing during the earlier part of the sample period. However, Freund *et al.* (1994) find that the introduction of individual stock *put* options has a negative effect on the underlying stock prices, a finding consistent with the hypothesis that put options allow investors to trade on negative information more efficiently than when they can only trade in stocks. Another potential explanation for the negative effect of option introduction on the underlying stock is related to the short-sale constraints that some financial institutions face. When there is no option market for a stock, traders with additional, negative information about the stock cannot take *bearish* positions if the costs of short-selling are too high. However, after the introduction of options on such a stock, the negative views of traders can be exploited, since short positions can be generated by buying puts or writing calls. Danielsen and Sorescu (2001) present evidence that the negative impact of the introduction of equity options on the underlying stock prices is consistent with the mitigation of short-sale constraints.

3.2.3 Long-term impact of option introduction on the equity market

Panel C of Table 2-1 shows the main empirical findings on the longer-term impacts of equity option introduction. We can observe from this table that there is a variety of long-term potential effects of new equity option listing on the underlying stock market, and a lack of consensus about the nature of such relationships.

Conrad (1989) is the first study to show that introducing option contracts causes a permanent price increase in the underlying stock, as evidenced by significantly higher stock prices in the long run. This empirical finding is supported by Detemple and Selden (1991), who develop a theoretical framework to link the incomplete primary market to the derivatives market, in which the prices of the underlying stocks are expected to change in response to the listing of new derivative contracts. However, Mayhew and Mihov (2000) challenge whether this effect is consistently positive. After correcting for the endogeneity of the decision to list options, Mayhew and Mihov (2000) show that the impact of option listing on stock prices was positive pre-1981 but turned negative post-1981. Furthermore, Ni *et al.* (2005) show evidence of a negative impact of equity option introductions on the underlying stock market, through a significant tendency for the prices of stocks on which options are written to cluster around strike prices on option expiration dates.

In terms of the long-term effects of option introduction on stock volatility, Skinner (1989) and Damodaran and Lim (1991) find that the volatility of the underlying stock market decreases significantly after the listing of options. Jennings and Starks (1986) argue that this effect is due to option contracts allowing stock prices to absorb new information more efficiently. Conversely, Faff and Hillier (2005) and Liu (2010), focusing on the UK and Japan, respectively, find that stocks on which options are written tend to have higher levels of return volatility than non-optioned control groups. Thus, the empirical findings of Faff and Hillier (2005) and Liu (2010) cast some doubt on the effect option listing has on volatility. Moreover, Mayhew and Mihov (2004) find no evidence that volatility changes with option introduction, using a control-sample methodology designed to correct for the endogeneity of option listing. In addition, Mazouz (2004) suggests that there are changes in market-wide volatility that should be considered in tests of the impact of equity options on the underlying stock market. Using a conditional volatility model, Mazouz (2004) shows that option listing has no significant effect on stock volatility after accounting for the time variation in stock variances.

Option listing has also been found to have a long-term effect on the market performance of the underlying stock market. For instance, Agyei-Ampomah and Mazouz (2011) show that stocks with options written on them co-move more strongly after the decision to list has been made, which induces a reduction in diversification benefits for the underlying optioned stocks. Fedenia and Grammatikos (1992) find that stocks with options written on them have significantly higher liquidity. Sahlström (2001) examines the Helsinki Stock Exchange and finds that option listing results in tighter bid-ask spreads for the underlying stocks. Kumar *et al.* (1998) confirm the positive effect of option listing on stock liquidity, while also reporting lower information asymmetry and improved price efficiency for the optioned stocks. Nevertheless, Danielsen *et al.* (2007) show evidence that equity options do not systematically improve the market liquidity of the underlying security; rather, the market liquidity of the underlying security improves before the listing decision is made.

Furthermore, Bernales (2017) confirms the results of Kumar *et al.* (1998) in terms of a reduction in asymmetric information after option listing. He shows that successful listings end up improving market quality in terms of a reduction in the levels of asymmetric information observed in the underlying stock market.

In summary, whilst the decision to introduce options on the underlying stocks is a function of stock volatility and stock liquidity, the success of equity option introductions is also related to information asymmetries. Moreover, a consensus has emerged on the positive effect of option introductions on stock returns and in terms of reductions in asymmetric information; however, the literature is clearly not conclusive on the impact of these introductions on stock volatility and stock liquidity.

In relation to similar studies on index options, as with the effect of equity option introductions on underlying stocks, the literature on index options also shows mixed results. For instance, Rahman (2001) shows that the introduction of derivatives trading on the Dow Jones Industrial Average (DJIA) index futures and futures option contracts is not associated with any changes in the volatility of the underlying stock components of the DJIA index. Harris (1989) shows that, since the start of trading in index options, stock components of the S&P 500 index have been relatively more volatile, with a difference that is statistically (but not economically) significant. Furthermore, Kumar *et al.* (1995) show that trading volume, volatility, and bid-ask spreads decline for the stocks contained in the Nikkei 225 index after the listing of the index options.

3.3 Market efficiency

Is the individual equity option market efficient? In this section, we discuss the papers that attempt to answer that question. As we explained in the introduction, under the Black and Scholes (1973) assumptions, individual equity options are redundant securities (i.e. they can be replicated by a portfolio consisting of the underlying stock and a risk-free bond). However, as we highlighted in the previous section, there are studies showing some evidence that the introduction of equity options has an impact on the underlying stock market, which suggests that the efficiency of the equity option market may be rejected. Hence, in this section, we present the studies that explicitly test the hypothesis of market efficiency among individual equity options. The studies discussed in this section are presented in Table 2-2.

---------------Insert Table 2-2----------------

We start, in Section 3.1 (see Table 2-2 Panel A), by discussing option mispricing in the individual equity option market, reporting on studies providing evidence that 'permanent' arbitrage opportunities exist. If the option market were efficient, investors should detect and trade options that were mispriced, which should move option prices to their 'correct' values and potential option mispricing should disappear. However, in this section, we discuss studies reporting evidence of 'permanent' option mispricing, which suggests that the equity option market may not be efficient. Furthermore, in Section 3.2 (see Table 2-2 Panel B), we examine studies reporting anomalies in the returns of equity options.

3.3.1 Option mispricing

To commence our survey on option mispricing in individual equity options, in an early study, Galai (1978) shows that, in contrast to theoretical predictions, closing prices of stocks and options do not satisfy no-arbitrage conditions. Galai (1978) then develops a trading strategy that exploits mispricing opportunities, resulting in profits that are on average positive, albeit relatively small when compared to their variability. In a similar vein, Castagna and Matolcsy (1982) develop a two-stage approach for testing market efficiency, based on the returns offered by strategies that exploit potential option mispricing. Using Black and Scholes (1973) implied volatilities to detect potential mispricing, Castagna and Matolcsy (1982) find that a portfolio that is long in underpriced options and short in overpriced ones offers abnormal positive profits. However, these profits are eliminated when transaction costs are taken into account. In addition, Nordén (2001) shows that equity option prices do not move as expected after stock price changes. For instance, the prices of calls and puts may move in the wrong direction compared to what the movement of the underlying stock would suggest, or the price changes in different option contracts written on the same stock may be uncorrelated with one another.

Additionally, Battalio and Schultz (2006, 2011) test for option mispricing during periods of short-sale constraints, since traders can generate synthetic short-sale positions by buying puts or writing calls. Using a sample of options written on US stocks during the peak of the internet bubble in 2000, when short-selling restrictions had been put in place, Battalio and Schultz (2006) find no evidence of tradeable arbitrage opportunities in the option market. However, in a later study, Battalio and Schultz (2011) show that the 2008 short-sale restrictions resulted in a significant dislocation between actual and synthetic prices of banned stocks. More specifically, Battalio and Schultz (2011) find that option-based synthetic prices of banned stocks were significantly lower than the actual prices, suggesting that market makers had increased the ask prices for puts, and lowered the bid prices for calls, potentially because the short-sale ban had substantially reduced their ability to hedge their inventory. As a result, trading in the option market became much more costly for investors, with Battalio and Schultz (2011) suggesting there was a $500 million extra liquidity cost during that period.

3.3.2 Abnormal returns on individual equity options

This section surveys the literature on the abnormal returns observed on individual equity options. In an early study, Sheikh and Ronn (1994) examine the intraday patterns of individual stock option returns in the CBOE and their relationship to trading patterns in the underlying stocks. For instance, they find that option returns are, on average, negative between 9:00 and 10:00, while on Tuesdays and Thursdays option returns are significantly positive. Moreover, there are some differences between the patterns in the returns on call versus put options (e.g. put option returns are positive on Fridays and negative over the weekend, but this is not observed in call option returns). This suggests that informed and discretionary liquidity traders do exhibit strategic trading behaviour in the option market.

Years later, although Coval and Shumway (2001) find that theoretical properties of option returns are confirmed in the historical returns of S&P 500 index options under mild assumptions (i.e. expected call returns exceed those of the underlying security and increase with the strike price), Ni (2008) shows that this is not the case for individual equity options. Examining the returns of options written on the constituent stocks of the S&P 500 from 1996 to 2005, Ni (2008) finds that the returns of out-of-the-money calls are on average negative. In addition, Ni (2008) finds that call options with high strike prices tend to offer lower returns, on average, than call options with low strike prices.

There is also evidence of a relationship between individual equity option returns and the volatility risk premium (VRP), i.e. the difference between implied volatility and realized volatility. For instance, Goyal and Saretto (2009) focus on the VRP as an element that affects the cross-section of individual equity option returns. Treating large values of the VRP as indicative of option mispricing, they show that going long in options with the highest values of VRP and short in options with the lowest values of VRP generates statistically and economically significant returns. Importantly, the profitability of this trading strategy cannot be explained by a set of common risk factors or idiosyncratic characteristics, and it remains significant even after accounting for transaction costs and margin requirements. However, recently, Bernales (2017) have offered a theoretical explanation for the relationship between the VRP and option returns. They use learning to explain both why implied volatility deviates from historical volatility and how this deviation generates predictive dynamics in the returns of option portfolios due to the recursive process induced by learning mechanisms.

Cao and Han (2013) find that dealers charge a higher premium for options written on stocks with higher idiosyncratic volatility, due to higher arbitrage costs. As a result, delta-hedged option returns are shown to be abnormal and negatively related to the idiosyncratic volatility of the underlying stocks. Vasquez (2017) shows that option returns are positively related to the slope of the implied volatility term structure, and abnormal option returns are observed from strategies based on the implied volatility term structure.

Boyer and Vorkink (2014) focus on the third moment of the returns distribution and report a negative relationship between individual equity option returns and ex-ante skewness. This negative cross-sectional relationship is not subsumed by the option's moneyness, and it results in abnormal returns for option portfolios with high ex-ante skewness. Additionally, Driessen *et al.* (2009) investigate the effects of market-wide correlation shocks on expected option returns. Increases in correlation at the aggregate market level are expected to have a negative impact on investor welfare through a reduction in diversification benefits and an increase in market volatility. Using data on options written on the S&P 100 constituent stocks, Driessen *et al.* (2009) find evidence of a significant correlation risk premium in the cross-section of individual option returns, with trading strategies exploiting priced correlation risk generating abnormal returns.

Overall, the empirical studies above show that there are some market inefficiencies in individual equity options. In particular, market inefficiencies have been documented with regard to 'permanent' option mispricing and abnormal returns on these options. Moreover, we show that our understanding is very limited regarding the determinants of the abnormal returns observed in individual equity options.

In relation to the index option market, and in line with some results for the individual equity option market, Evnine and Rudd (1985), Sim *et al.* (2016), and Ackert and Tian (2001) show that index options often violate the no-arbitrage conditions. Regarding index option returns, Coval and Shumway (2001) show that, while returns on index options follow theoretical properties under mild assumptions, returns on index option portfolios do not respect theoretical features under strong assumptions (i.e., under strong assumptions, expected option returns should vary linearly with option betas). For instance, Coval and Shumway (2001) document strong negative average returns in zero-beta at-the-money straddles using index options, which should not be observed. Bondarenko (2014) looks at simple trading strategies involving naked puts, and shows that they yield large profits for the options' sellers. In a multi-factor analysis, Jones (2006) finds high abnormal negative returns to be associated with short-term out-of-the-money puts. Constantinides *et al.* (2013) show that a single-factor model fails to explain the cross-section of index option returns. They also report a decreasing trend in the magnitude of leverage-adjusted put option returns as the moneyness and time-to-maturity rise, for the case of index options.

However, some potential explanations for the abnormal behaviour of index options have appeared in the last several years. For instance, Broadie *et al.* (2009) show that index option returns can be explained by models that can generate jump risk premiums or an estimation risk (i.e., when agents cannot estimate parameters and state variables obtained from short samples). Broadie *et al.* (2009), though, do not provide an economic justification for the *existence* of such models. However, recently, Bernales *et al.* (2018b) have shown that one potential economic explanation for the abnormal returns observed in index options may be the learning process followed by investors.

3.4 The market microstructure and behavioural biases in individual equity options

In this section, we focus on the market microstructure properties and behavioural biases of investors observed in the case of individual equity options, which may explain the results seemingly showing that the equity option market is not efficient (see Section 2). Importantly, the fundamental assumptions of the efficient market hypothesis, in relation to no market frictions and investor rationality, do not hold when considering market microstructure properties and behavioural biases. Therefore, given that tests of market efficiency depend on the option pricing model (and its assumptions) used to describe the market, then some option pricing models might not be well specified, in the sense of not considering market frictions related to trading mechanisms, and/or investors' irrational behaviours. With this in mind, in this section we survey the literature on the effect of market microstructure on equity options, and then trading behaviours of investors that depart from the rational-investor paradigm. Table 2-3 presents the empirical studies that have examined these topics.

--------------Insert Table 2-3---------------

As a first step, in Section 4.1, we focus on the empirical studies that investigate the liquidity of equity option contracts (see Table 2-3 Panel A). This section starts with a review of the empirical studies that investigate option liquidity patterns. Afterwards, we discuss the studies that investigate the determinants of option liquidity. In Section 4.2, we discuss studies that investigate the impact of market microstructure changes on option liquidity (see Table 2-3 Panel B). Finally, in Section 4.3 (see Table 2-3 Panel C), we discuss some studies that analyse potentially irrational behaviours of agents, which may also explain why the equity option market is inefficient.

3.4.1 Option liquidity patterns and determinants

In Table 2-3 Panel A, we present the empirical studies that discuss intraday patterns in individual equity option liquidity. In the earlier literature, Aggarwal and Gruca (1993) and Chan *et al.* (1995) find evidence that the bid-ask spreads of equity options follow an L-shaped pattern during the trading day (i.e. a pattern in which bid-ask spreads decline sharply after trading opens, and then level off), which is different from the U-shaped pattern observed in the underlying stock market. Aggarwal and Gruca (1993) and Chan *et al.* (1995) suggest that potential explanations for this L-shaped pattern are related to the competition in market making, and the informed trading observed in equity option markets, which cause an increase in activity at the beginning of the day. Segara and Sagara (2007) further confirm this finding for the Australian Options Exchange.

A number of studies have also explored the determinants of liquidity across different equity option markets. A large part of this literature focuses on the effect of market microstructure on option liquidity. For instance, in one of the earliest studies in this research area, Vijh (1990) compares the liquidity of options listed on the CBOE to the liquidity of the underlying stocks that are traded on the New York Stock Exchange

(NYSE). Vijh (1990) shows that options and stocks have comparable bid-ask spreads. However, due to having multiple dealers per contract, options exhibit a significantly greater market depth than the underlying stocks, suggesting they are more efficient at absorbing larger trades.

Berkman (1993) also analyses the role of market microstructure characteristics in option liquidity by examining the hybrid market of the Options Exchange in Amsterdam, which is characterized by the existence of market makers and a limit order book. Berkman (1993) highlights the importance of limit orders for option liquidity, in the sense that limit orders supply liquidity more cheaply than market makers. Maberly *et al.* (2010) and ap Gwilym and Verousis (2013) analyse the levels of liquidity for different contracts (i.e. with different degrees of moneyness and times-to-maturity) written on the same underlying asset. They find clustering for particular strike prices and times-to-maturity.

A related part of the literature attempts to explain the bid-ask spreads observed in options through the prism of hedging costs and asymmetric information, which should increase the spreads. For instance, Lakonishok *et al.* (2007) and Flint *et al.* (2014) find that the bid-ask spreads of individual equity options are affected by the cost of hedging the option contracts. Goyenko *et al.* (2015) examine intraday data on options written on the S&P 500 constituents, and they also document a significant impact of market makers' hedging costs, with the future rebalancing cost dominating the initial hedging cost. Cao and Wei (2010) show evidence that asymmetric information is one of the main components of the bid-ask spread. Verousis *et al.* (2016a) further show that volume and volatility are positively related to the bid-ask spreads of individual equity options, consistent with information asymmetry and hedging-cost effects on option liquidity. Christoffersen *et al.* (2017) confirm the previous literature by showing that proxies for asymmetric information and hedging costs (and also stock illiquidity, inventory risk, and option order imbalances) are significant drivers of effective spreads.

Wei and Zheng (2010) find evidence that the bid-ask spread is affected by maturity-substitution and moneyness-substitution in option trading, with these effects driven by expiration cycles and stock return volatility. Examining a large sample of options written on US stocks, Wei and Zheng (2010) show that, due to the structure of the expiration cycles in the option market, demand shifts predictably from medium-term to short-term options when the third expiration month is too far away. Moreover, higher levels of stock return volatility seem to shift demand for options from in-the-money to out-of-the-money contracts. These substitution effects have significant impacts on the bid-ask spreads of different option contracts.

Furthermore, Mayhew *et al.* (1999) show that the liquidity of individual equity options is significantly related to the underlying stock market. More specifically, Mayhew *et al.* (1999) find that equity option liquidity is positively related to the price volatility, trading volume, and firm size of the underlying stocks. In addition, this relationship seems to be bi-directional, since stocks with liquid options tend to exhibit higher liquidity themselves. This significant relationship between the liquidity of options and that of the underlying stocks is further supported by the empirical findings of Verousis and ap Gwilym (2013a).

3.4.2 Impact of market microstructure changes on option liquidity

Panel B of Table 2-3 presents the studies that investigate the impact of changes in market microstructure on the liquidity of individual equity options. Neal (1987) examines option trading in the US in the late 1980s and finds that the listing of options on multiple exchanges significantly lowers transaction costs. Mayhew (2002) confirms this finding in an extended sample period (1986-1997). However, Battalio *et al.* (2001) find that option transactions executed in multiple exchanges are usually executed at economically inefficient prices. More specifically, Battalio *et al.* (2001) show that the introduction of a national market system for equity options in the US has led to an improved quality of execution and lower option spreads.

Pinder (2003) reports that options traded in an order-driven system are characterized by a lower bid-ask spread. Anand and Weaver (2006) investigate option trading on the CBOE and find that the introduction of a designated primary market maker has led to a reduction in the quoted and effective spreads. Anand *et al.* (2016) further show that the make-take structure could reduce execution costs and, as a result, increase the quote competition among liquidity suppliers. In particular, the introduction of this structure could encourage liquidity suppliers to be more competitive when providing liquidity, and to post better prices that benefit liquidity-demanding traders.

Focusing on the short-selling restrictions that were imposed on a number of US stocks after the financial crisis, Battalio and Schultz (2011) find that the spreads increased significantly for such stocks. This result is further confirmed by Verousis and ap Gwilym (2013a) for option trading in Europe. More specifically, Verousis and ap Gwilym (2013a) show that equity option liquidity dropped substantially after the short-sale ban of 2008, consistent with the hypothesis that, when market makers cannot hedge their inventories easily, trading costs in the option market increase significantly. Moreover, Verousis *et al.* (2016b) explore the effect of a change in tick size on the liquidity of individual equity option trading, in the NYSE LIFFE. Although introducing a smaller tick size is found to have had a positive effect on option liquidity, as evidenced by smaller spreads, the lower depths observed after the tick size reduction are indicative of a deterioration in the market's ability to absorb larger trades.

3.4.3 Investor behaviours departing from rationality

The analysis and tests of option market efficiency assume rationality, and thus some results may be affected by non-rational trading behaviours of investors. Therefore, Panel C of Table 2-3 presents a number of studies that focus on the behavioural properties of trading in individual equity options. These studies do not explicitly test for the efficiency of option markets, but they examine certain types of behaviour that could potentially deviate from the rational-investor paradigm.

In this context, Poteshman and Serbin (2003) and Hao *et al.* (2010) investigate the early-exercise decision in exchange-traded options on individual stocks. They identify a large number of early exercises as irrational, even without using a particular option pricing model. These irrational early exercises of individual equity options appear to be triggered by the underlying stocks reaching their peak level from the previous year and/or by having high stock returns. Poteshman and Serbin (2003) further show that this irrational behaviour is not uniform across all investor types, being exhibited mostly by customers of discount brokers and those of full-service brokers, rather than by traders from large investment institutions.

Lemmon and Ni (2014) find that demand for individual equity options that increase exposure to the underlying is positively related to individual investor sentiment and past market returns. This finding is not observed in index options, which suggests that there are behavioural biases among individual 'unsophisticated' investors. This is because individual equity options (index options) are actively traded by individual investors (sophisticated institutional investors). Moreover, Lemmon and Ni (2014) find that individual equity options in which a higher proportion of trading is carried out by less sophisticated investors have prices that are more sensitive to the individual sentiment, which is consistent with their behavioural arguments.

In a more recent study, Bernales *et al.* (2016) find evidence of herding behaviour in the US equity option market, with investors suppressing their own beliefs in favour of the market consensus, during periods of market stress. This herding behaviour is reflected in investors being heavily influenced by the contemporaneous returns of index options when they are pricing individual stock options, resulting in a price clustering that reduces the ability to hedge positions efficiently.

In summary, there is a relative consensus about the significant impact of the option market microstructure on equity option markets. There is evidence of an L-shaped pattern of option bid-ask spreads, and empirical studies show that competition in exchanges and liquidity supply improves market efficiency. There is also evidence that option liquidity is related to the price volatility, trading volume, firm size and short-sales constraints of the underlying stock. In addition, option liquidity is negatively affected by hedging costs, inventory costs and adverse selection costs. Furthermore, in terms of behavioural biases, there is some evidence of irrational early exercise in equity options, a relationship between individual investor sentiment and equity option demand, and herding behaviours being exhibited by investors in periods of stress. In fact, understanding investors' behavioural biases in relation to individual equity options could constitute an important field for future research, since very little is known about potentially irrational behaviours of agents in the equity option market.

Regarding index options, ap Gwilym *et al.* (1997) show that the bid-ask spreads of index options also follow an L-shaped pattern during the trading day, which is consistent with the evidence reported for individual equity options. Moreover, Cho and Engle (1999) and Wu *et al.* (2014) show that the hedging activities of market makers are the most important determinant of option bid-ask spreads for the index option market. Finally, in relation to potential behavioural biases of agents in the option market, there is little evidence of investor behaviours departing from rationality in the case of index options. This is probably because index options are mainly traded by sophisticated institutional investors, rather than individual investors as is the case with equity options. Probably the paper that comes closest to looking at behavioural biases in index options is that of Diz and Finucane (1993). They present evidence of inefficient early exercise of S&P 100 option contracts, a result that is in line with the findings for equity option contracts.

3.5 Price discovery

As described in Section 3, the evidence does not seem to support the market efficiency hypothesis in equity option markets. In Section 4, we discussed that a possible explanation may be related to option pricing models not being well specified, since they do not consider frictions from the market microstructure and/or behavioural biases, which also violate the fundamental assumptions of the efficient market hypothesis. An alternative explanation for potential market inefficiencies is that equity option markets are efficient only in a semi-strong form, in the sense that there are agents with *private* information that is not yet reflected in option prices. Informed investors with private information might prefer to trade in option markets given the leverage inherent in options, which means they need less capital to exploit their private information than they do in the underlying stock market. Therefore, in this section, we review and discuss the papers that examine the price discovery process (i.e. the process whereby information is gradually incorporated into prices), with Table 2-4 reporting the related studies.

---------------Insert Table 2-4----------------

In Section 5.1 (Table 2-4 Panel A), we firstly discuss 'agent-driven' studies, meaning studies examining the impacts of different *types of participant* (i.e. informed and uninformed agents) on the price discovery process in the equity option market. Afterwards, in Section 5.2 (Table 2-4 Panel B), we discuss 'event-driven' studies, being studies that analyse the option price discovery process in the light of *corporate events or announcements*.

3.5.1 Price discovery: Agent-driven studies

Table 2-4 Panel A presents the set of studies that investigate the role of agents in the price discovery process. Chakravarty *et al.* (2004) find that the information share of options in price discovery varies across different underlying stocks, and they suggest that informed investors trade in both the option market and the stock market. Kaul *et al.* (2004) show evidence that informed investors trade strategically in the equity option market, taking into account the leverage and transaction costs of different option contracts. In addition, Anand and Chakravarty (2007) present evidence of stealth trading in option markets, while Bernales *et al.* (2018a) report liquidity-searching behaviour exhibited by informed investors in option markets as a means to hide their informed-trading strategies.

Conversely, a number of other studies have challenged the hypothesis that informed investors prefer to trade in the option market. Stephan and Whaley (1990) is one of the earliest studies to have argued that the equity market in fact leads the option market in price discovery. Analysing intraday data on firms whose options are traded on the CBOE, Stephan and Whaley (1990) report that the equity market leads the option market by fifteen minutes when the lead/lag relationship is estimated using price changes, with the equity market's lead being even longer when trading volumes are used. Furthermore, Chan *et al.* (2002) find that net stock trading volume has predictive ability for both stock and option quote revisions, but net option trading volume has no incremental predictive ability. Based on this finding, Chan *et al.* (2002) argue that informed traders are actually more likely to initiate trades in the stock market than in the option market.

Holowczak *et al.* (2006) and Muravyev *et al.* (2013) also support the hypothesis that price discovery is led by the underlying stock market rather than by the option market. They argue that this is due to higher transaction costs in the option market, and the increasing use of automated quoting algorithms by option market makers. O'Connor (1999) provide further evidence of the stock market leading the option market.

3.5.2 Price discovery: Event-driven studies

Table 2-4 Panel B presents the studies that investigate the relative contribution of equity options to the price discovery process around corporate events. In this literature, several studies have focused on a particularly important type of corporate news, namely earnings announcements, and examined how the option market incorporates this information into prices. For instance, Patell and Wolfson (1979) provide evidence that option prices reflect the anticipation of a temporary increase in the volatility of the underlying stock due to earnings announcements. This empirical finding is further supported by Levy and Yoder (1993) and Donders *et al.* (2000).

In addition, Ajinkya and Gift (1985) show that option prices reflect contemporaneous information about the dispersion of analysts' earnings forecasts that is incremental to the information already incorporated in the underlying stock prices. Jennings and Starks (1986) find that the prices of stocks that have options written on them can adjust to earnings announcements more efficiently than the prices of non-optioned stocks. Furthermore, Amin and Lee (1997) show that option traders participate in price discovery around earnings announcements, with individual equity options containing incremental information on top of that contemporaneously available in the underlying equity market. Other empirical studies that report evidence of incremental price discovery in the option market around earnings announcements include Roll *et al.* (2010), Billings and Jennings (2011), and Atilgan *et al.* (2015).

Hayunga and Lung (2014) examine the relative contributions of the option and underlying equity markets in terms of price discovery around financial analysts' consensus revisions. Examining individual equity options trading in the US market during 2000-2009, Hayunga and Lung (2014) show that the option market leads the stock market in price discovery when analysts revise their recommendations, and option investors trade in the direction consistent with the upcoming revision approximately three days prior to the announcement. This empirical finding is further confirmed by Lung and Xu (2014), who also argue that informed trading in the option market could be driven by information leakage rather than superior stock-picking skills.

Dong and Sinha (2011) examine a broader set of firm-specific news items associated with underlying stocks and find evidence of the option market leading the stock market in price discovery. More specifically, they show that the information share increases much more substantially in the option market than in the equity market around corporate news events, with this difference being even more pronounced after the imposition of short-sale restrictions that followed the 2008 crisis. Moreover, a number of studies document significant changes in the option trading volume around corporate announcement dates. Anthony (1988) and Arnold *et al.* (2006) find evidence of abnormal trading volumes observed sooner in the option market than in the equity market after corporate announcements are released. In the same vein, Easley *et al.*

(1998a) show that the option trading volumes around announcement dates lead stock price changes over the next few days.

Overall, the debate about whether the option market leads the stock market in price discovery, or vice versa, is far from settled. As described in Section 5.1, the evidence shows that, in *normal* times, the underlying stock market in general leads the option market in the price discovery process. Nevertheless, when there are corporate announcements, there is evidence that individual equity options are used by traders who are informed about such events, which also contributes to the price discovery process.

In relation to index options, it is important to note that informed investors' private information about *indexes* is mainly related to an anticipated *global* economic view of the market, while informed agents' *private* information about a particular stock is related to undisclosed corporate news. Thus, there are some studies that investigate whether investors who are 'informed' on the index option market can anticipate global market changes, although with mixed results. For instance, Kang and Park (2008) and Hsieh and He (2014) present evidence about information revealed by index options regarding index changes. However, Chen and Gau (2009), Chiang and Fong (2001), Schlag and Stoll (2005), and Ryu (2015) present opposing evidence, by showing that index options do not provide substantially more information about the movements of indexes.

3.6 Option-implied information in individual equity options

In this section, we discuss the type of private information revealed by informed investors in individual equity options. Thus, as a first step, and following on from our discussion in the previous section of studies describing the price discovery contribution made by individual equity options, where new private information is incorporated (and revealed) through equity option prices, we examine the *type* of option-implied information revealed in the equity option market. We divide this section into three types of option-implied information that can be captured from individual equity options: firstly (Section 6.1, Table 2-5), information about stock prices and returns; secondly (Section 6.2, Table 2-6), information about stock return volatility; and thirdly (Section 6.3, Table 2-7), information about probability of default.

--------------Insert Table 2-5---------------

--------------Insert Table 2-6---------------

--------------Insert Table 2-7---------------

3.6.1 Option-implied information about stock prices and returns

As described above, a number of studies view option prices as measures related to investors' expectations (based on public and 'private' information) about the future prices and returns of the underlying stocks (Table 2-5). In this research area, Manaster and Rendleman (1982) are among the earliest researchers to have directly compared the option-implied stock price to the actual price of the stock observed in the underlying equity market. They show that option prices contain *additional* fundamental information not contemporaneously reflected in the stock market. This information is reflected in the stock market on average 24 hours later, suggesting that option prices have significant ability to predict future stock prices. Diltz and Kim (1996) confirm the empirical findings of Manaster and Rendleman (1982) regarding the predictive ability of option prices, suggesting that stock prices tend to adjust to the level of option-implied prices over the course of two trading days. Conversely, Bhattacharya (1987) suggests that option prices' ability to predict stock prices is economically insignificant.

Although option-implied prices are indeed found to contain information not contemporaneously available in stock prices, Bhattacharya (1987) shows that exploiting this information is not possible when trading costs and other market frictions are considered.

Later studies show that individual equity options' forecasting regarding stock prices and returns can also be derived from higher moments of the risk-neutral distribution. For instance, Govindaraj *et al.* (2014) and Lin and Lu (2015) find that the volatility of the risk-neutral distribution has significant forecasting power for future stock returns, especially during important firm-specific events. However, Bali and Hovakimian (2009) show that sorting stocks into portfolios based on the volatility of their risk-neutral distribution results in statistically insignificant stock returns; they suggest that it is the call-put risk-neutral volatility spread that is actually predicting future stock returns.

Conrad *et al.* (2013) use the framework developed by Bakshi and Madan (2000) and Bakshi *et al.* (2003) to extract the volatility, skewness, and kurtosis of the underlying stock's risk-neutral distribution, and they show that these higher moments can forecast future stock returns. After accounting for risk factors being priced in the cross-section, Conrad *et al.* (2013) find that the risk-neutral skewness obtained from option prices remains significantly negatively related to future stock returns. Using a different approach to extract risk-neutral skewness from option prices, Rehman and Vilkov (2012) confirm the significant relationship between the skewness of the risk-neutral distribution and future stock returns, but find that this relationship is in fact positive. In a similar spirit, Van Buskirk (2011) finds that the skewness of the risk-neutral distribution has significant ability to predict future stock returns, but only in relatively short windows around earnings announcements. This ability of the skewness of the risk-neutral distribution to predict future stock returns is further confirmed by Xing *et al.* (2010), Jin *et al.* (2012), Liu *et al.* (2014), and Fu *et al.* (2016). In a more recent study, Fan *et al.* (2018) extract forecasts for the return distribution of individual stocks using option prices and high-frequency stock returns. After looking at several combinations, Fan *et al.* (2018) find that the most accurate forecast of the future return distribution of the underlying stock is obtained by transforming a simple Black and Scholes (1973) risk-neutral density into a real-world density. Importantly, Fan *et al.* (2018) provide further support for the hypothesis that option-implied information is superior in forecasting future stock returns to the information contained in historical returns.

Another stream of the related literature explores the predictive ability of other implied measures that can be extracted from option prices to predict stock prices and returns. For instance, Cremers and Weinbaum (2010), Liu *et al.* (2014), and Fu *et al.* (2016) focus on deviations from put-call parity. They show that such deviations are significantly related to future stock returns, with stocks with relatively expensive calls outperforming those with relatively expensive puts. Furthermore, Jin *et al.* (2012) show that the forecasting power of deviations from put-call parity is particularly high during important firm-specific information events. Borochin and Yang (2017) argue that the predictive ability of the skewness of the risk-neutral distribution and deviations from put-call parity stems from the fact that they reflect anticipated future net leverage changes which, in turn, impact future stock returns.

Han and Zhou (2012) investigate the difference between the risk-neutral implied variance and the realized variance, typically referred to as the volatility risk premium (VRP), as a potential predictor of future stock returns. Using a sample of 500 stocks, they find evidence of the VRP being significantly and positively related to future stock returns. Fu *et al.* (2016) further show that the ability of the VRP to predict stock returns persists before and after the 2008 crisis. In addition, Bernales and Valenzuela (2016) use the market-aggregate implied correlation to predict stock returns. They show that the implied correlation obtained from options written on the constituent stocks of the S&P 100 index is an indicator of market-wide risk and contains information on future market returns. This predictive ability of implied correlation is particularly strong over quarterly and semi-annual forecasting horizons.

Another part of the related literature examines whether trading volumes in the option market also contain information about the future returns of the underlying stocks. In an early study, Easley *et al.* (1998b) find evidence against the hypothesis that option trading volumes have unconditional predictive ability over stock returns. However, they also show that the volumes of specific types of option trades, which could be classified as informed trades, are significantly related to future stock returns. Similarly, Cao *et al.* (2005) also reject the hypothesis of the unconditional predictive ability of option volume, but find that trading-volume imbalances in the option market can forecast stock returns around takeover announcements.

Pan and Poteshman (2006) is the first study to have provided strong evidence on the information contained in option trading volumes about future stock prices. Focusing on new positions opened by investors in the option market, they find that a stock's put-to-call ratio is significantly negatively related to that stock's returns over the next week. Moreover, they suggest that this forecasting power of the put-to-call trading volume ratio stems from informed investors trading on non-public information. Blau and Wade (2013) confirm the significant ability of put-to-call ratios to predict the future returns of individual stocks, but they find that the ratio of short-sales to the total trading volume in the equity market partly subsumes the informational content of the put-to-call ratio. Goyenko *et al.* (2015) find that option-induced order flows can predict the future returns of the underlying stocks. However, this forecasting power of option trading activity is significant only during periods of decreased option liquidity, when abnormal order flows are more likely to be driven by trading on private information than by liquidity trading.

Roll *et al.* (2010) introduce the option-to-stock trading volume ratio (O/S) and find that it can be used to forecast future stock returns around earnings announcements. More specifically, they find that stocks with higher O/S levels tend to offer higher returns in the few days after earnings announcements, supporting the hypothesis that a large part of the pre-announcement trading in options can be classified as informed. Johnson and So (2012) develop an asymmetric information model to show that, theoretically, the O/S ratio and future stock returns are related. They argue that the above relationship is driven by equity short-sale costs, and present a set of empirical results that confirm this theoretical prediction.

3.6.2 Option-implied information about the stock return volatility

As mentioned at the beginning of this section, a growing part of the literature has focused on the forward-looking nature of option contracts regarding the future realized volatility and/or the future option-implied volatility (Table 2-6). In one of the earliest studies, Latane and Rendleman (1976) use the Black and Scholes (1973) model to extract stock return volatilities implied by option prices. Using a weighted average of implied standard deviations, they find that implied volatility outperforms historical volatility measures in forecasting future realized volatility. Lamoureux and Lastrapes (1993) also show that implied volatility helps to predict future volatility. Mayhew and Stivers (2003) find that the relative predictive ability of implied volatility depends on the option trading volume. More specifically, implied volatility outperforms historically based volatility estimates for stocks with the most actively traded options, but for stocks with lower option trading volumes the information content of implied volatility is subsumed by information contained in the time-series of past returns. The ability of Black and Scholes (1973) implied volatilities to predict individual stocks is further confirmed by Dennis *et al.* (2006) and Cao *et al.* (2006).

Taylor *et al.* (2010) examine whether the model-free approach of Britten-Jones and Neuberger (2000) can produce more accurate volatility forecasts than standard Black and Scholes (1973) implied volatility. When considering short-term forecasting horizons, historical models are generally found to produce more efficient forecasts of future volatility than option-implied estimates. However, Taylor *et al.* (2010) show that implied volatility measures extracted from individual equity options outperform historically based estimates for longer forecasting horizons, with simple at-the-money Black and Scholes (1973) estimates being more informative than model-free implied volatilities. Furthermore, Bernales and Guidolin (2014) focus on forecasting features of the implied volatility surface of equity options. In contrast to Black and Scholes' (1973) assumptions, the volatilities implicit in option contracts written on one underlying asset differ across strike prices and times-to-maturity (which was observed for the very first time by Rubinstein , 1985). This phenomenon is known as the implied-volatility surface (henceforth IVS). Bernales and Guidolin (2014) provide evidence that the IVS for individual equity options can be forecasted using vector autoregressive models, while Bernales and Guidolin (2015) suggest that a potential explanation for the forecasting property derives from the recursive learning process followed by option investors.

3.6.3 Option-implied information and the probability of default

A number of studies have also explored the extent to which option-implied information extracted from individual equity options is associated with the likelihood of a firm's default (Table 2-7). In this context, Cao *et al.* (2006) find that the volatility implied by option prices is a significant determinant of credit default swap (CDS) spreads. Analysing more than 1,000 US firms, Cao *et al.* (2006) show that the informational content of the simple Black and Scholes (1973) at-the-money implied volatility, regarding CDS spreads, is particularly important for firms with lower credit ratings, higher option volumes, and higher option open interest. Benkert (2004) and Da Fonseca and Gottschalk (2014) confirm this strong relationship between option-implied volatility and credit spreads using international data. Cremers *et al.* (2008) use options written on individual stocks to extract volatility and jump measures, and they find that both measures are significantly related to a firm's credit spread, which is further confirmed by Kita (2012). In a similar spirit, although from a theoretical perspective, Chen and Kou (2009) develop a model of credit risk with two-sided jumps, and show that the resulting implied volatility and credit spreads would be expected to move in the same direction. In addition, Wang *et al.* (2013) find that the difference between implied and realized volatilities, i.e. the VRP described earlier, has significant explanatory power for credit spreads, especially when implied volatility is measured as the Britten-Jones and Neuberger (2000) model-free expectation.

Another strand of this literature uses the prices of options written on a firm's stock to explicitly estimate the risk-neutral probability of default. For instance, Capuano (2008) develops a methodology for extracting the risk-neutral probability of default from individual equity options using the principle of minimum cross-entropy, without making any assumptions about the underlying stock's distribution or the recovery rate. Furthermore, Vilsmeier (2016) proposes some technical modifications to the original Capuano (2008) methodology to address issues of accuracy and numerical stability. As an illustrative example, Vilsmeier (2016) uses data on options written on the Bank of America to show that this methodology would have produced implied default probabilities that could have served as an early-warning signal before the bank's downgrading by Moody's in 2011.

Following a different approach, Câmara *et al.* (2012) use a simple lognormal distribution augmented with a probability of default to model stock returns, and they show that the resulting implied probability of default tends to outperform a set of standard credit risk measures. Taylor *et al.* (2014) propose modelling a stock's risk-neutral distribution as a mixture of two lognormal densities with a default probability. Based on empirical evidence of a closer fit to realized stock return distributions, they suggest that this model allows for a more accurate estimation of the risk-neutral probability of default using prices of individual stock options.

Carr and Wu (2011) develop a theoretical framework that uses the prices of out-of-the-money American put options to compute the value of a synthetic credit insurance contract on the firm's stock. They show that the implied probabilities of default extracted from out-of-the-money puts closely match those embedded in CDS spreads. Chang and Orosi (2017) extend their modelling assumption by incorporating a positive expected equity recovery into the framework. They show that this adjustment results in a more accurate estimation of the implied probability of default using options on individual stocks. Conrad *et al.* (2017) argue that the Carr and Wu (2011) approach requires data on deep out-of-the-money put options, which are not always available for individual stocks. In order to address this limitation, Conrad *et al.* (2017) propose an alternative framework that uses all available options to infer the implied probability of default, and they find that these option-implied default probabilities are very close to the ones provided by CDS spreads.

In summary, there is consensus on the forecasting features of individual equity options for the prediction of future stock returns, volatility and probability of default. This is due to the forward-looking nature of option-implied information, since options should reflect agents' expectations about future market conditions (i.e. at the time when the option contracts will be exercised). Nevertheless, regarding studies of index options, in contrast to the individual equity option literature that is mostly concerned with firm-specific information contained in equity option contracts, the index option literature focuses on assessing whether option-implied information can be captured about aggregate market conditions. For instance, Faccini *et al.* (2018) show that option-implied information from index options can be used to predict US real economic activity. Christensen and Prabhala (1998) present evidence that the implied volatility of index options can forecast future aggregate market volatility. Finally, Goncalves and Guidolin (2006) and Bernales and Guidolin (2015) offer evidence that the implied volatility surface from index options, regarding the aggregate market volatility, can be predicted by vector autoregressive models.

3.7 Conclusion and future research

Over the last few decades, the literature on individual equity options has been growing consistently, in tandem with the increasing trading activity in these derivative contracts in global financial markets. This paper provides a comprehensive review of this literature, highlighting the main empirical findings regarding equity option markets. Our review of the equity option literature identifies several themes that have emerged, ranging from areas of relative consensus and solid understanding, to areas where the evidence is rather mixed and more research is required.

Across the numerous empirical studies on individual equity options, we observe that there is some consensus on the rejection of the classical view of equity options as redundant securities. On this issue, the empirical evidence suggests that introducing options on individual stocks generally has a significant short-term and long-term impact on the underlying equity market, although the precise nature of this impact seems to vary. In addition, empirical studies show that there are market inefficiencies in the equity option market, which are reflected in 'permanent' option mispricing and abnormal option returns. In fact, we seem to understand very little about the determinants of the returns on equity options themselves. Although some idiosyncratic characteristics have been found to be informative in this respect, the literature has yet to develop a credible model for equity option returns.

Furthermore, liquidity in the equity option market seems to depend on market microstructure issues, while equity options are consistently found to contribute substantially to the price discovery process. Nevertheless, the debate about whether the equity option market leads the stock market in price discovery, or vice versa, is far from settled.

Another area of consensus is the forecasting power of option-implied information regarding the future state of the underlying stock market. Given that equity options are forward-looking by design, it is not particularly surprising that a substantial body of empirical studies shows that information extracted from equity options has significant ability to predict future stock returns, volatility, and the probability of default.

In terms of potential future research topics, the area of expected equity option returns could constitute an important field for future research. Compared to the vast body of literature on the cross-section of stock returns, our limited understanding of the cross-section of equity option returns seems somewhat surprising. In addition to examining the role of idiosyncratic characteristics, future research could potentially examine the impact of market-wide factors, such as liquidity, short-sale constraints, and market microstructure, on the dynamics of the returns observed on individual equity options.

Besides the lack of clear evidence as to whether the equity option market leads the stock market (or vice versa) in terms of information flows, more research is also needed on the topic of price discovery. A focus on high-frequency data, in particular, could potentially help tackle the question of which market leads the other in this process. Finally, additional research needs to be developed regarding the irrational behaviour of investors regarding equity options, while the area of algorithmic trading in equity options also remains underexplored, and both may provide interesting research topics for future empirical studies.

Table 3-1 Impact of option listing on the equity market

Panel A: Introduction of new equity options into the market			
	Market	Period	Main findings
Mayhew and Mihov (2004)	US	1973 – 1996	Exchanges tend to list options on stocks with high trading volume, volatility, and market capitalization.
Danielsen *et al.* (2007)	US	1993 – 2002	The size of a stock's bid-ask spread is the single most important option-listing determinant.
Bernales (2017)	US	1996 – 2009	A high level of asymmetric information predicts option adoption rates.
Panel B: Short-term impact of option listing on the equity market			
	Market	Period	Main findings
Detemple and Jorion (1990)	US	1973 – 1986	Positive impact on individual stock returns and volatility. The positive impact decreases after index options are introduced.
Watt *et al.* (1992)	UK	1978 – 1989	Positive impact on stock returns.
Freund *et al.* (1994)	US	1973 – 1990	Positive impact on volatility. Negative impact of put listing on stock returns.
Gjerde and Saettem (1995)	Norway	1990 – 1994	Positive impact on stock returns.
Bollen (1998)	US	1987 – 1992	Positive impact on stock returns.
Danielsen and Sorescu (2001)	US	1973 – 1995	The negative impact of the introduction of equity options on the underlying stock prices is consistent with the mitigation of short-sale constraints.

Panel C: Long-term impact of option listing on the equity market

	Market	Period	Main findings
Jennings and Starks (1986)	US	1981 – 1982	Volatility decreases after option listing.
Conrad (1989)	US	1974 – 1980	Positive impact on stock prices.
Skinner (1989)	US	1973 – 1985	Volatility decreases after option listing.
Damodaran and Lim (1991)	US	1973 – 1983	Volatility decreases after option listing.
Fedenia and Grammatikos (1992)	US	1970 – 1988	Optioned stocks have higher liquidity.
Kumar *et al.* (1998)	US	1983 – 1989	Higher liquidity, lower information asymmetry, and improved price efficiency for optioned stocks.
Mayhew and Mihov (2000)	US	1973 – 1996	Impact on stock prices was positive pre-1981 and turned negative post-1981.
Sahlstrom (2001)	Finland	1992 – 1995	Narrower bid-ask spreads for option stocks.
Mayhew and Mihov (2004)	US	1973 – 1966	No evidence that volatility declines with option introduction, using control-sample methodology designed to correct for the endogeneity of option listing.
Mazouz (2004)	US	1973 – 2001	No impact on stock volatility after accounting for changes in market-wide volatility.
Faff and Hillier (2005)	UK	1973 – 1995	Stocks with options tend to exhibit higher volatility.
Ni *et al.* (2005)	US	1996 – 2002	Negative impact, with stock prices clustering around options' strike prices on expiration dates.
Danielsen *et al.* (2007)	US	1993 – 2002	Options do not systematically improve the market liquidity of the underlying security; rather, the market liquidity of the underlying security improves before the decision to list is made.
Liu (2010)	Japan	1997 – 2007	Stocks with options tend to exhibit higher volatility.
Agyei-Ampomah and Mazouz (2011)	UK	1986 – 2007	Optioned stocks co-move more, leading to reduced diversification benefits.
Bernales (2017)	US	1996 – 2009	Levels of asymmetric information are reduced after equity option introduction.

Table 3-2 Market efficiency issues

	Market	Period	Main findings
Panel A: Option mispricing			
	Market	Period	Main findings
Galai (1978)	US	1973	Simultaneous prices of stocks and options are not fully synchronized.
Castagna and Matolcsy (1982)	Australia	1976 – 1977	No opportunity for arbitrate profits after accounting for transaction costs.
Norden (2001)	Sweden	1995 – 1996	Equity option prices do not move as expected after stock price changes.
Battalio and Schultz (2006)	US	2000	No evidence of arbitrage opportunities during the 2000 short-sale ban.
Battalio and Schultz (2011)	US	2008	Significant arbitrage opportunities during the 2008 short-sale ban.
Panel B: Abnormal returns on individual equity options			
	Market	Period	Main findings
Sheikh and Ronn (1994)	US	1986 – 1987	Intraday option returns have patterns that show evidence of informed trading.
Ni (2008)	US	1996 – 2005	Option returns deviate significantly from theoretical predictions. Option traders are seeking idiosyncratic skewness.
Driessen *et al.* (2009)	US	1996 – 2003	Trading strategy exploiting priced correlation risk generating abnormal returns.
Goyal and Saretto (2009)	US	1996 – 2006	The volatility gap can explain the cross-section of option returns. Significant mispricing detected.
Boyer and Vorkink (2014)	US	1996 – 2009	Option returns are negatively related to ex-ante skewness. Abnormal returns of option portfolios with high ex-ante skewness.
Cao and Han (2013)	US	1996 – 2009	Delta-hedged option returns behave abnormally since they are shown to be negatively related to the idiosyncratic volatility of the underlying stocks.
Vasquez (2017)	US	1996 – 2012	Abnormal option returns from strategies based on the implied volatility term structure.

Table 3-3 Option market microstructure and investor behaviour departing from rationality

Panel A: Liquidity patterns			
	Market	Period	Main findings
Vijh (1990)	US	1988	Options have greater market depth than stocks, due to having multiple dealers per contract.
Berkman (1993)	Europe	1989	Competition in the limit order book improves liquidity.
Aggarwal and Gruca (1993)	US	1986	L-shaped pattern of option bid-ask spread.
Chan *et al.* (1995)	US	1986	L-shaped pattern of option bid-ask spread.
Mayhew *et al.* (1999)	US	1993	Option liquidity is related to the price volatility, trading volume, and firm size of the underlying stock.
Segara and Sagara (2007)	Australia	2000	L-shaped pattern of option bid-ask spread.
Lakonishok *et al.* (2007)	US	1990 – 2001	Liquidity is driven by market makers' hedging costs, but not by volatility trading.
Cao and Wei (2010)	US	1996 – 2004	Information asymmetry drives liquidity.
Wei and Zheng (2010)	US	1996 – 2007	Evidence of maturity substitution and moneyness substitution among different options.
Maberly *et al.* (2010)	US	1973 – 2008	Market microstructure issues (e.g. price thresholds) have a significant impact on liquidity.
ap Gwilym and Verousis (2013)	Europe	2005	The market-maker scheme drives price clustering.
Verousis and ap Gwilym (2013)	Europe	2008 – 2010	Option liquidity is negatively related to stock volatility.
Flint *et al.* (2014)	Australia	2007	Bid-ask spreads of equity options are affected by the cost of hedging.
Christoffersen *et al.* (2017)	US	2004 – 2012	Option liquidity is driven by asymmetric information, hedging and inventory costs, stock illiquidity, and option order imbalances.
Goyenko *et al.* (2015)	US	2004 – 2013	Liquidity is driven by market makers' hedging costs.
Verousis *et al.* (2016)	Europe	2008 – 2010	Information asymmetry and hedging costs drive liquidity, with volume and volatility positively related to the bid-ask spread.

Panel B: Impact of market microstructure changes

	Market	Period	Main findings
Neal (1987)	US	1985 – 1986	Listing on multiple option exchanges lowers transaction costs.
Battalio *et al.* (2001)	US	2000	Trading in a national market system leads to improved quality of execution and lower spreads.
Mayhew (2002)	US	1986 – 1997	Listing on multiple option exchanges lowers transaction costs.
Pinder (2003)	Australia	1995 – 1999	Order-driven system results in lower bid-ask spreads.
Anand and Weaver (2006)	US	1999	Designating primary market makers leads to lower quoted and effective spreads.
Battalio and Schultz (2011)	US	2008	Spreads increased for stocks that were the object of the short-sale ban.
Verousis and ap Gwilym (2013)	Europe	2008 – 2010	Option liquidity dropped after the short-sale ban.
Verousis *et al.* (2015)	Europe	2009 – 2010	Reducing the tick size resulted in smaller spreads but lower depths.
Anand *et al.* (2016)	US	2007 – 2013	A make-take structure increases quote competition among market makers, reducing execution costs.

Panel C: Investor behaviour departing from rationality

	Market	Period	Main findings
Poteshman and Serbin (2003)	US	1996 – 1999	Evidence of irrational early exercise of American-style options.
Hao *et al.* (2010)	US	2003	Option investors regularly fail to exercise options rationally before ex-dividend dates.
Lemmon and Ni (2014)	US	1990 – 2010	The demand for individual equity options that increases exposure to the underlying is positively related to the individual investor sentiment and past market returns.
Bernales *et al.* (2016)	US	1996 – 2012	Option investors herd around the consensus during periods of market stress.

Table 3-4 Price discovery

Panel A: Agent-driven studies			
	Market	Period	Main findings
Kaul *et al*. (2004)	US	1995	Informed investors trade strategically in the equity option market.
Anand and Chakravarty (2007)	US	1999	Evidence of stealth trading in equity option markets.
Bernales *et al*. (2018)	US	1996 – 2009	The option bid-ask spread may still be a good proxy for informed trading, despite the liquidity-searching behaviour of informed agents.
Stephan and Whaley (1990)	US	1986	The equity market leads options in price discovery.
Chan *et al*. (2002)	US	1995	The stock trading volume can predict option quote revisions, but the option trading volume has no predictive ability.
Holowczak *et al*. (2006)	US	1990 – 2001	The equity market leads the option market in price discovery.
Muravyev *et al.* (2013)	US	2003 – 2006	The equity market leads the option market in price discovery.
O'Connor (1999)	US	1990	The equity market leads the option market in price discovery.
Chakravarty *et al.* (2004)	US	1988 – 1992	Informed traders trade in both markets.

Panel B: Event-driven studies

	Market	Period	Main findings
Patell and Wolfson (1979)	US	1974 – 1978	Options reflect anticipated stock volatility increases prior to earnings announcements.
Levy and Yoder (1993)	US	1982 – 1985	Options reflect anticipated stock volatility increases prior to earnings announcements.
Donders *et al.* (2000)	US	1991 – 1993	Options reflect anticipated stock volatility increases prior to earnings announcements.
Atilgan *et al.* (2015)	US	1996 – 2008	Options contain incremental information around earnings announcements.
Ajinkya and Gift (1985)	US	1977 – 1978	Options contain incremental information about the dispersion of analysts' forecasts of earnings per share.
Jennings and Starks (1986)	US	1981 – 1982	Optioned stocks adjust to earnings announcements more efficiently than non-optioned stocks.
Amin and Lee (1997)	US	1988 – 1989	Options contain incremental information around earnings announcements.
Roll *et al.* (2010)	US	1996 – 2007	Options contain incremental information around earnings announcements.
Billings and Jennings (2011)	US	1996 – 2006	Options contain incremental information around earnings announcements.
Hayunga and Lung (2014)	US	2000 – 2009	Options lead the price discovery process during analysts' revisions.
Lung and Xu (2014)	US	2009 – 2011	Options lead the price discovery process during analysts' revisions, driven by information leakage.
Dong and Sinha (2011)	US	2003 – 2009	Options lead the price discovery process around corporate news.
Anthony (1988)	US	1982 – 1983	Abnormal trading volume in options around corporate announcements.
Arnold *et al.* (2006)	US	1994 – 2000	Abnormal trading volume in options around corporate announcements.
Easley *et al.* (1998a)	US	1990	Abnormal trading volume in options around corporate announcements, leading stock price changes.

Table 3-5 Option-implied information about stock prices and returns

	Market	Period	Main findings
Manaster and Rendleman (1982)	US	1973 – 1976	Option prices contain information about future movements of stock prices.
Bhattacharya (1987)	US	1977 – 1978	Informational content of option prices regarding future stock prices is economically insignificant.
Diltz and Kim (1996)	US	1988	Stock prices adjust to the level implied by option prices within two trading days.
Easley *et al.* (1998b)	US	1990	Informed option trading volume can predict stock returns, but general trading volume cannot.
Cao *et al.* (2005)	US	1986 – 1994	Option trading volume can predict stock returns around takeover announcements.
Pan and Poteshman (2006)	US	1990 – 2001	The put-to-call trading volume ratio is negatively related to future stock returns.
Bali and Hovakimian (2009)	US	1996 – 2004	Deviations from put-call parity can predict future stock returns.
Cremers and Weinbaum (2010)	US	1996 – 2005	Deviations from put-call parity can predict future stock returns.
Xing *et al.* (2010)	US	1996 – 2005	The skewness of the risk-neutral distribution can predict future stock returns.
Roll *et al.* (2010)	US	1996 – 2007	The option-to-stock trading volume ratio is positively related to future stock returns.
Van Buskirk (2011)	US	1996 – 2009	The skewness of the risk-neutral distribution can predict stock returns around earnings announcements.
Han and Zhou (2012)	US	1996 – 2009	The variance risk premium can predict future stock returns.
Jin *et al.* (2012)	US	1996 – 2010	The skewness of the risk-neutral distribution and deviations from put-call parity can predict stock returns.

	Market	Period	Main findings
Rehman and Vilkov (2012)	US	1996 – 2011	Risk-neutral skewness is positively related to future stock returns.
Johnson and So (2012)	US	1996 – 2008	The option-to-stock trading volume ratio is negatively related to future stock returns.
Blau and Wade (2013)	US	Not specified	The put-to-call trading volume ratio predicts stock returns, but the short-sales to total stock trading volume ratio subsumes that information.
Conrad *et al.* (2013)	US	1996 – 2005	Risk-neutral skewness can predict future stock returns.
Govindaraj *et al.* (2014)	US	1996 – 2011	The volatility of the risk-neutral distribution can predict future stock prices, especially during firm-specific events.
Liu *et al.* (2014)	US	1996 – 2011	The skewness of the risk-neutral distribution and deviations from put-call parity can predict stock returns.
Lin and Lu (2015)	US	1996 – 2010	The volatility of the risk-neutral distribution can predict future stock prices, especially during firm-specific events.
Goyenko *et al.* (2015)	US	2004 – 2013	Option-induced order flows can predict stock returns during periods of option illiquidity.
Fu *et al.* (2016)	US	1996 – 2014	The skewness of the risk-neutral distribution, deviations from put-call parity, and the variance risk premium can predict future stock returns.
Bernales and Valenzuela (2016)	US	1996 – 2010	The option-implied correlation obtained from 100 stock options (where the underlying stocks are part of the S&P 100 index) is an indicator of market-wide risk and contains information on future market returns.
Borochin and Yang (2017)	US	1996 – 2012	The skewness of the risk-neutral distribution and deviations from put-call parity can predict future stock returns, due to expected leverage changes.
Fan *et al.* (2017)	US	2003 – 2012	The risk-neutral distribution can predict future realized distributions.

Table 3-6 Option-implied information about the stock return volatility

	Market	Period	Main findings
Latane and Rendleman (1976)	US	1973 – 1974	At-the-money (ATM) implied volatility forecasts future volatility better than historical measures.
Lamoureux and Lastrapes (1993)	US	1982 – 1984	Implied volatility helps to predict future volatility.
Mayhew and Stivers (2003)	US	1988 – 1995	ATM implied volatility forecasts future volatility efficiently only for stocks with liquid options.
Cao *et al.* (2006)	US	1996 – 2004	ATM implied volatility forecasts future volatility better than historical measures.
Dennis *et al.* (2006)	US	1988 – 1995	ATM implied volatility forecasts future volatility better than historical measures.
Taylor *et al.* (2010)	US	1996 – 1999	ATM implied volatility outperforms model-free implied volatility and historical measures in forecasting future volatility for longer horizons.
Bernales and Guidolin (2014)	US	1996 – 2006	Evidence of strongly predictive features of the implied volatility surface in the cross-section of equity options.
Bernales and Guidolin (2015)	US	1996 – 2007	Learning is a potential reason for the predictive features of the implied volatility surface, which is suggested by a learning model that generates the forecasting characteristics of the implied volatility surface observed empirically from equity options.

Table 3-7 Option-implied information about the probability of default

	Market	Period	Main findings
Benkert (2003)	International	1999 – 2002	ATM implied volatility is a significant determinant of CDS spreads.
Cao *et al.* (2006)	US	1996 – 2004	ATM implied volatility is a significant determinant of CDS spreads, especially for firms with lower credit ratings, higher option trading volumes, and higher option open interest.
Capuano (2008)	US	2008	The option-implied probability of default increases before credit events.
Cremers *et al.* (2008)	US	1996 – 2002	Implied volatility and implied skewness are related to credit spreads.
Camara *et al.* (2012)	US	1996 – 2008	The option-implied probability of default outperforms standard credit risk measures.
Carr and Wu (2011)	US	2005 – 2007	Option-implied probabilities of default closely match those extracted from CDS contracts.
Kita (2012)	US	2010 – 2011	Implied volatility and implied skewness are related to credit spreads.
Wang *et al.* (2013)	US	2001 – 2008	The volatility risk premium has explanatory power regarding credit spreads.
Da Fonseca and Gottschalk (2014)	Europe	2007 – 2012	ATM implied volatility is a significant determinant of CDS spreads.
Taylor *et al.* (2014)	US	2007 – 2009	Assuming stock returns follow a mixed lognormal distribution results in more accurate implied probabilities of default.
Chang and Orosi (2016)	US	2008 – 2009	Adjusting for positive expected recovery improves the accuracy of implied default probabilities.
Vilsmeier (2016)	US	2011	The option-implied probability of default increases before credit events.
Conrad *et al.* (2017)	US	2001 – 2012	Option-implied probabilities of default closely match those extracted from CDS contracts.

Chapter 4: Data cleaning

4.1 Introduction

With the development of the financial market in recent years, high-frequency data (HFD) was applied to study a wide range of issues in financial markets, such as commonality (Verousis *et al.*, 2015), market efficiency (Conrad *et al.*, 2015), and price clustering (Davis *et al.*, 2014). This trend may be driven by two reasons. First, the low-frequency data, such as daily closing price data, cannot efficiently investigate high-frequency trading (HFT) which is responsible for over 70% of trading volume in the US in 2009 (Fabozzi *et al.*, 2011; Hendershott *et al.*, 2011). Second, the high-frequency data include more information that can further explore what happens to the intraday financial market and thereby help shed light on a better understanding of market microstructure (Liu, 2009).

Although HFD is beneficial to analysing financial markets, there are two potential challenges in using HFD (Fabozzi *et al.*, 2011). The first challenge comes from the characteristics of HFD which is an irregularly spaced time series. In particular, trading and quotes occur at random times that should be considered when developing research models (Aït-Sahalia and Jacod, 2014). The second challenge is driven by the vast amount of data offered by HFD. Florescu (2016) indicated the importance of identifying and removing the tick records which are not providing useful information or reflecting the trading process. Regarding the first challenge, we will adopt an appropriate research design in the following empirical chapters. This chapter will focus on the second challenge and thereby develop a data filtering procedure to clean and manage our high-frequency data. In general, we will exclude the error data, missing values, outliers, and expiration effect. Besides, we also manage dataset based on moneyness of options contracts.

The remainder of this chapter is organised as follows. Section 2 discusses the characteristics of HFD and highlights the importance of data cleaning in HFD. Section 3 outlines the market structure of Chicago Board Options Exchange (CBOE). Section 4 shows the data cleaning procedure and the way of managing dataset. Section 5 provides a brief statistical description of the dataset.

4.2 Literature review

In this section, we review literature on HFD and indicate the three characteristics of HFD: irregular spaces, diurnal patterns, and price discreteness. Besides, we also discuss why HFD should be cleaned before data analysis.

4.2.1 Characteristics of high-frequency data

4.2.1.1 Irregular spaced data

Engle (2000, p.1) defined HFD as "*a full record of transactions and their associated characteristics*". Hence, HFD is relative to each transaction which occurs at random times. This implies that the first characteristic of HFD is irregular spaced. Most statistical models are developed for homogeneous time series[3]. This creates a complication for data analysis (Gençay *et al.*, 2001; Russell and Engle, 2009). It also should consider whether the sampling method could capture the dynamic nature of the market because the stochastic process is not stable all the time. Hence, researchers should make a choice regarding the time intervals over which to analyse the data.

Previous literature (e.g. Hasbrouck and Seppi, 2001; Russell and Engle, 2009) showed that the fixed interval and the stochastic interval could be used to analyse HFD. If fixed intervals were selected, Russell and Engle (2009) suggested that the analysis

[3] Homogeneous time series is considered as the regularly spaced time series, while inhomogeneous is irregularly spaced series.

should follow some interpolation rules to regulate the interval. For example, Hasbrouck and Seppi (2001) allocated transactions of each stock into 15-minute intervals and estimated price changes by midpoints of the quotes at the beginning and end of each interval. Alternatively, if stochastic intervals are used, the spacing of the data should be taken into account. Trades occur at varying time intervals, which may lead to intraday ‘seasonals’ in the trading volume, volatility, and bid-ask spreads (Goodhart and O'Hara, 1997). These issues of time intervals are important when examining the micro-behaviour of markets. Market microstructure studies directly depend on access to high frequency data that reflect the second-by-second movement of the market operational details (Fabozzi *et al.*, 2011).

4.2.1.2 Diurnal patterns

High frequency data typically present strong diurnal or periodic patterns. This could be viewed as a key characteristic that differ high-frequency data from lower-frequency data (Goodhart and O'Hara, 1997). This characteristic usually presents a time-of-day phenomenon. For most stock and option markets, the intraday seasonality in high-frequency data arises from market opening and closing. Researchers discover that the volatility, volume, and spreads generally exhibit a U-shaped or a reversed U-shaped pattern during the trading day (Stephan and Whaley, 1990; Segara and Sagara, 2007; Russell and Engle, 2009). For example, Lee *et al.* (1993) and Chan *et al.* (1995) suggested that bid-ask spreads in NYSE reached the highest point at the opening, rapidly dropped to lower levels around mid-day, and then increased again at the closing.

4.2.1.3 Price discreteness

Since the financial data is discrete, the variance of this discrete process is generally greater than the magnitude of the minimum movement when investigating long time horizons (Russell and Engle, 2009). However, this phenomenon is not shown in HFD which is transaction-by-transaction data. The changes in each transaction price may only take a handful of values called a tick. More specifically, financial markets usually require prices to fall on a pre-specified set of values which is called the tick size. Hence, price movements should fall on multiples of the tick size. Russell and Engle (2009) indicated that the price movements usually fell on a very small number of possible outcomes. They used a sample of transaction prices for 10 months and discovered that over half of the transaction prices did not change as compared with previous prices. In this context, HFD tends to show a high level of kurtosis (Engle, 2000; Russell and Engle, 2009).

4.2.2 Significance of data cleaning

HFD provides all trading and quotes occurred in the market and thereby have a significant challenge in data handling (Brownlees and Gallo, 2006). Fabozzi *et al.* (2011) indicated that the increase in the speed of transactions directly improves the probability of error records in reporting trading information. A record is defined as an error record when it cannot represent a correctly reported transaction price or a possible price at the reported time, such as decimal errors and errors raised by test ticks (Gençay *et al.*, 2001). Verousis and ap Gwilym (2010) directly defined an error record as an observation that could bot reflect the trading process. In this context, these records are unable to connect market participants with recorded observation. The error record cannot be eliminated by adopting automatic trading systems because Falkenberry (2002) found errors in an automatic and partly automatic trading system. Verousis and ap Gwilym (2010) confirmed that almost every high-frequency dataset included some error records. Brownlees and Gallo (2006) showed that the sequence of quotes usually had some incorrect records. For example, the quotes are not time-ordered or represent some anomalous behaviour driven by certain market conditions (e.g. opening, closing, etc.). If these records are used, the empirical results are unreliable and unusable. Thus, data cleaning is necessary for the research with high-frequency data.

Data cleaning methodologies can be used to eliminate error records. Dasu and Johnson (2003) defined data cleaning as a process of detecting and removing errors and inconsistencies from data. With data cleaning, researchers could improve the quality of data and thereby avoid incorrect results. Gençay et al. (2001) divided errors into two types: human errors which are directly caused by human data contributors, and systems errors which are caused by computer system interactions and failures. Although it is impossible to discover the exact reason for data errors, Gençay et al. (2001) still guessed the cause of errors through the particular behaviour of the bad ticks. A possible explanation is that they believed the knowledge of the error mechanism that could contribute to filtering and correcting bad ticks. Hence, five types of error records should be highlighted by researchers:

(1) Decimal errors are the errors that fail to change a decimal digit of the quote. For example, a bid price of 19.0579 is followed by a real quote 19.0592, but the bid quote is 19.0492 when publishing. This error type could result in a long series of error quotes (Breymann *et al.*, 2003).

(2) Test ticks can also cause errors. More specifically, data contributors sometimes deliver test ticks to the systems when the market is not liquid. For example, a contributor may deliver a tick at the beginning of the trading day to test whether he/she is connected to the market. If the market is inactive overnight, traders would be not concerned about this test tick.

(3) The third error type is repeated ticks that are driven by repeating the last tick at regular time intervals. The repeated ticks are harmful when repeating with high frequency because it can obstruct the validation of the few good ticks.

(4) The fourth error is tick copying that is caused by copying and resending the ticks among contributors. This type is damaging when some contributors produce slightly modified ticks through adding small random corrections (Gençay *et al.*, 2001). To give more details, these slightly varied copy ticks could obstruct other contributors who intend to identify fake monotonic or repeated series.

(5) The last error type is the scaling problem that quoting regulation are varying defined across markets. For example, some contributors may quote the value of 1 unit, while others quote 100 units.

Based on these explanations, error records could influence the statistical analysis and thereby impact on the reliability of empirical results. Therefore, a data cleaning procedure should be included in empirical studies to deal with error records.

4.3 Sample of data

In this section, we introduce the market structure of CBOE and then describe the dataset. The fundamental information of firms in the dataset is also shown in this section that further provides a background of our dataset.

4.3.1 Institutional background

The modern era of exchange-based option trading began in April 1973 with the foundation of Chicago Board Options Exchange (CBOE) which is an extension of the Chicago Board of Trade and owned by CBOE Holdings, Inc. (Mayhew, 2002). CBOE provides a single system for trading options. This system integrates electronic trading and traditional open outcry trading on CBOE's trading floor in Chicago that is known as Hybrid Trading System (CBOE, 2017b). As the US's largest options exchange, its annual trading volume was around 1.13 billion contracts at the end of 2016 (CBOE, 2017a). CBOE operates one trading session each day from 9:30 a.m. to 16:00 p.m.[4] Since CBOE is a quote-driven market, traders cannot directly place orders with the CBOE. All orders are placed by the exchange's Hybrid system that allows traders to trade electronically or through open outcry (CBOE, 2017b).

CBOE has two types of liquidity providers, and they are market-makers and Designated Primary Market Makers (DPMs). Market-makers are considered to be the backbone of the CBOE trading system, providing continuous liquidity in the marketplace. More specifically, they are risking their own capital by making bids and offers on their own accounts when lacking public buy or sell orders (CBOE, 2017b). DPMs are a market-maker obligated to provide continuous quotes in all option series in their appointed option classes. Unlike traditional market-makers, DPMs could act as both a broker and a dealer on the same day. Furthermore, each option class is assigned to one DPM who needs to maintain the limit order book. Before assigning the option to a specialist, an exchange employee maintains and views the limit order book and trade for their own account. After the change in the system, specialists are able to view the limit order book and base their trading decisions on the book's contents (Vijh, 1990; Mayhew, 2002; Anand and Weaver, 2006). With the support of these liquidity providers, CBOE can resolve uncertainty about investor preference and can sufficiently improve market liquidity, as confirmed by the empirical results of Anand and Weaver (2006).

4.3.2 Structure of dataset

--------------Insert Table 3-1---------------

[4] The trading hours is from 8:30 a.m. to 15:00 p.m. Central Time (Chicago time). After adjusting to Eastern Time, this paper uses the trading hours as 9:30 a.m. to 16:00 p.m.

This thesis use Thomson Reuters Tick History dataset created by SIRCA. The dataset comprises real-time tick-by-tick data for all options written on 30 individual equities that were traded on the CBOD from January 2012 to June 2014. The 30 individual equities are components of the Dow Jones Industrial Average (DJIA) during the sample period. Table 3-1 represents the number of options contracts written on each ticker. All tickers have similar numbers of call and put contracts except INTC which has 3,664 call contracts and 3,666 put contracts[5]. Although these are options contracts written on the same underlying stock, they have different characteristics in terms of the strike price, expiration, and option type (call or put).

---------------Insert Table 3-2----------------

Furthermore, the dataset includes Reuters instrument code (RIC)[6], bid/ask price, bid/ask volume, and trading price and volume. The total number of tick records is over 4 billion as shown in Table 3-2. More specifically, the numbers of call and put observations among JNJ and JPM are over 100 million, while the numbers among TRV are around 17 million.

4.3.3 Fundamental information

---------------Insert Table 3-3----------------

Table 3-3 represents the fundamental information of the collected tickers. During the sample period, JPM had the highest earnings before interest and taxes (EBIT) that over $50,000 million. A possible explanation is that this high earning is driven by the size of the company. Particularly, JPM also had the highest total assets,

[5] Since option writers do not have to create same number of call and put option contract, the number of call and put options is inconsistent in INTC.
[6] In the dataset, each ticker consists of several sub-tickers. These sub-tickers are option contracts that are written on the same underlying stock but have different characteristics. RIC can identify these characteristics that includes strike price, time-to-maturity, and contract type (call or put).

liabilities, and equities. Regarding the book value per share, however, GS (around $150) was three times higher than JPM (around $50). This may imply that the stock of GS is undervalued as compared with JPM.

4.4 Data filtering procedure and data management

This section will explain in detail the procedure of data cleaning and the approach of data management. The summarised procedures are shown as follows.

(1) We excluded errors in tick records which included incorrect trading time and irrelevant tick information.

(2) The problem of missing values was solved in the second step.

(3) Outliers contained in the dataset were removed.

(4) We dealt with the problem of expiration effect as suggested by Cao and Wei (2010) and Verousis *et al.* (2015a).

(5) The options contracts were categorised into five types in order to avoid the potential problems raised by moneyless.

4.4.1 Error in tick records

---------------Insert Table 3-4----------------

Since this research investigates market microstructure which is related to bid and ask information and trading information, we only kept quote and trading records and remove other records, such as correction messages[7] and option interest messages[8]. Besides, we also found that not all records were located in the training time of CBOE equity options markets (from 09:30:00 to 16:00:00) (CBOE, 2016). Hence, we further excluded such out-of-hour data. Table 3-4 shows the percentage of records marked as

[7] Correction messages provide a corrected or inserted price and volume at the time of the message.
[8] Option interest messages provide the volume of outstanding contracts at the time of the message.

error records. On average, 0.75% and 0.77% of call and put options were identified as error tickers. Besides, Table 3-4 also shows that AXP has the highest percentage of error records (1.38% for calls and 1.49% for puts) and the percentage of error records in other tickers are generally lower than 1%.

4.4.2 Missing value

Missing value is widely represented in the financial dataset and influences the quality of the dataset. The missing value in this study is defined as both ask and bid information lacked in a quote record and both trading price and volume lacked in a trade record. If only ask or bid information is missing, we use the previous ask or bid information to replace the missing value.

Unlike error records, a missing value could be converted to a new value through statistical methods such as listwise deletion and mean imputation method. However, these methods can lead to serious inference problems (Little and Rubin, 2014). With this concern, we applied a simplistic way to deal with missing values that the observations with the missing values were deleted as suggested by Farhangfar *et al.* (2008). Although this method directly reduces the sample size, this reduction is not a potential issue for this study considering the large sample size of our high-frequency dataset. As shown in Table 3-5, around 2% of records were evaluated as the missing value for calls and puts, respectively. In particular, MMM reports the highest percentages of missing value that are 4.94% for calls and 5.15% for puts.

---------------Insert Table 3-5----------------

4.4.3 Outliers

Han *et al.* (2011) defined outliers as noise that results in estimation biases. Particularly, the existence of an outlier can create a disproportionate influence on statistical analyses. To eliminate outliers, firstly, extreme and negative bid-ask spread values[9] were removed because the extreme value would significantly impact on the distribution of data and the negative spread is impossible in option markets. Based on previous studies (e.g. Chung and Van Ness, 2001; Verousis et al., 2015), the limit of the percentage the bid-ask spread is 150%. Therefore, we deleted any quotes that the bid-ask spread is negative, zero, or exceeding 150%. Besides, all messages with negative and zero price and volume were deleted. Table 3-6 shows that 9.69% and 10.09% of quote records in call and put options were defined as outliers based on the range of bid-ask spreads. Besides, 9.27% and 9.49% of trade records in calls and puts were marked as outliers because the traded prices and sizes were lower than zero. Besides, V, VZ, WMT, and XOM show higher percentages of outliers than other tickers.

---------------Insert Table 3-6----------------

4.4.4 Expiration effect

Previous studies (e.g. Corredor et al., 2001) found evidence of abnormal behaviour in the underlying assets close to the expiration date. This effect has been recognised as the expiration effect that is related to improved volatility, volume, and price reversals in the underlying markets (Chiang, 2014; Xu, 2014). In order to avoid expiration effects, Cao and Wei (2010) suggested that it should avoid options with too long or too short maturity cycles. Hence, options with a maturity less than 7 days and over 365 days were deleted. As shown in Table 3-7, 22.06% and 20.9% of records in calls and puts were marked because of the expiration problem.

[9] Bid-ask spread is calculated as the difference between the contemporaneous bid and ask prices divided by the mean of the bid and ask prices.

---------------Insert Table 3-7----------------

4.4.5 Moneyness

We further managed our dataset by using moneyness to classify options contracts. Moneyness is defined as the difference between the stock price and the strike price. It is positive for in-the-money[10] (ITM) options, zero for at-the-money[11] (ATM) options and negative for out-the-money[12] (OTM) options (Alexander, 2008). Based on a common formulation for moneyness, S/K[13], we classified options into five categories: DOTM, OTM, ATM, ITM, and DITM based on 0.9, 0.95, 1.05 and 1.1. To give more details, DOTM call contracts have moneyness smaller than 0.9, OTM call contracts have moneyness between 0.9 and 0.95, ATM call contracts have moneyness between 0.95 and 1.05, ITM call contracts have moneyness between 1.05 and 1.1, and DITM call contracts have moneyness over 1.1. Put contracts are based on the opposite classification.

The reason of managing data in this way is derived from the effects of moneyness. For example, Chan *et al.* (2009) discovered that informed traders were more likely to use OTM options to capitalise on their information because OTM options could provide the highest possible leverage. Chang (2011) confirmed that OTM options offered the highest leverage for investors and thereby strike prices of OTM options had

[10] In-the-money indicates that the strike price of a call option is lower than the market price of the underlying assets, or that the strike price of a put option is higher than the market price of the underlying assets.

[11] At-the-money refers to the strike price of an option is identical to the market price of the underlying security.

[12] Out-of-the-money refers to a call option' strike price that is higher than the market price of the underlying asset, or the strike price of a put option that is below than the market price of the underlying assets.

[13] S refers to the contemporaneous mid-quote price of the underlying assets but be not adjusted for dividend payments, and K refers to the option strike price (Verousis et al., 2015).

a greater probability to be exercise. Hence, we classified options contracts based on moneyness and thereby support the following empirical chapters.

---------------Insert Figure 3-1----------------

---------------Insert Figure 3-2----------------

Figure 3-1 and 3-2 show the distributions of each moneyness types for call and put options contracts, respectively. It could clearly identify that call options generally have a higher level of moneyness compared with put options. Besides, the percentage of ATM contracts is higher than other moneyness types in most tickers.

4.5 Statistical description

---------------Insert Table 3-8----------------

Table 3-8 shows the percentage of remaining observations after data cleaning. It reports that over half of the observations were removed from our dataset. This result may confirm the view that the probability of error records are increased with the reporting frequency (Fabozzi *et al.*, 2011). On average, 42.82% and 42.53% of observations remained after data cleaning. In particularly, most tickers remained over 40% of the observations, while less than 40% of observations in TRV and V were kept. The following empirical chapters may not use all these remaining observations because they have their own research aims. Hence, the empirical chapters will have specific sections to indicate their sample selections.

Table 4-1 Numbers of option contracts

Name	RIC	Call	Put
AMER EXPRESS CO	AXP	3,424	3,424
BOEING CO	BA	3,892	3,892
CATERPILLAR INC	CAT	3,487	3,487
CISCO SYSTEMS	CSCO.O	3,523	3,523
CHEVRON	CVX	3,692	3,692
DU PONT CO	DD	2,426	2,426
WALT DISNEY CO	DIS	2,847	2,847
GENERAL ELEC CO	GE	3,594	3,594
GOLDM SACHS GRP	GS	3,678	3,678
HOME DEPOT INC	HD	3,483	3,483
INTL BUS MACHINE	IBM	4,287	4,287
INTEL CORP	INTC.O	3,664	3,666
JOHNSON JOHNSON	JNJ	3,033	3,033
JPMORGAN CHASE	JPM	4,798	4,798
COCA-COLA CO	KO	3,872	3,872
MCDONALD'S CORP	MCD	3,348	3,348
3M COMPANY	MMM	3,562	3,562
MERCK & CO	MRK	3,288	3,288
MICROSOFT CP	MSFT.O	4,208	4,208
NIKE INC CL B	NKE	3,596	3,596
PFIZER INC	PFE	3,859	3,859
PROCTER & GAMBLE	PG	3,089	3,089
AT&T	T	4,214	4,214
THE TRAVELERS CO	TRV	519	519
UNITEDHEALTH GP	UNH	3,211	3,211
UNITED TECH CP	UTX	3,520	3,520
VISA INC	V	3,861	3,861
VERIZON COMMS	VZ	4,621	4,621
WAL-MART STORES	WMT	3,193	3,193
EXXON MOBIL	XOM	3,306	3,306

Note: This table shows the numbers of contracts in call and put options

Table 4-2 Numbers of observations

	Call		Put		
RIC	Quote	Trade	Quote	Trade	Total
AXP	98.41%	0.78%	98.45%	0.68%	109,002,721
BA	98.70%	1.30%	98.97%	1.03%	175,896,673
CAT	98.14%	1.86%	98.38%	1.62%	173,179,289
CSCO.O	97.59%	2.41%	98.43%	1.57%	114,101,804
CVX	98.83%	1.17%	99.09%	0.91%	127,563,593
DD	99.22%	0.78%	99.43%	0.57%	113,097,562
DIS	98.96%	1.04%	99.28%	0.72%	126,561,869
GE	97.61%	2.39%	98.60%	1.40%	112,399,872
GS	97.44%	2.56%	98.24%	1.76%	147,377,844
HD	98.86%	1.14%	98.99%	1.01%	172,320,582
IBM	97.93%	2.07%	98.25%	1.75%	172,980,120
INTC.O	97.49%	2.51%	98.31%	1.69%	134,108,666
JNJ	98.79%	1.21%	99.10%	0.90%	217,800,000
JPM	98.08%	1.92%	98.54%	1.46%	236,000,000
KO	98.84%	1.16%	99.17%	0.83%	109,823,013
MCD	98.62%	1.38%	98.96%	1.04%	121,088,558
MMM	99.16%	0.84%	99.27%	0.73%	115,515,287
MRK	98.94%	1.06%	99.35%	0.65%	119,577,851
MSFT.O	97.66%	2.34%	98.43%	1.57%	192,951,146
NKE	99.19%	0.81%	99.30%	0.70%	127,746,438
PFE	98.78%	1.22%	99.21%	0.79%	125,910,817
PG	98.73%	1.27%	98.86%	1.14%	113,740,716
T	98.49%	1.51%	98.99%	1.01%	138,689,875
TRV	99.37%	0.63%	99.49%	0.51%	35,322,214
UNH	99.35%	0.65%	99.48%	0.52%	108,698,955
UTX	99.43%	0.57%	99.56%	0.44%	125,275,614
V	98.72%	1.28%	99.04%	0.96%	152,479,381
VZ	98.87%	1.13%	99.26%	0.74%	171,517,961
WMT	98.96%	1.04%	99.07%	0.93%	121,210,949
XOM	97.91%	2.09%	98.33%	1.67%	136,018,028
Total					4,147,957,398

Note: This table shows the percentage of quotes and trades in calls and puts, respectively. It also shows the total number of observations in each ticker. AXP contains quote records, trade records, and other records. Hence, the total number of quote and trade in AXP is lower than 100%.

Table 4-3 Fundamental information

RIC	Earnings Before Interest and Taxes	Total Assets	Total Liabilities	Total Equity	Book Value Per Share	Total Dividends	Dividends per Share
AXP	8,739	155,206	135,521	19,685	18.00	977	0.87
BA	6,845	93,586	83,668	9,802	13.33	1,737	2.21
CAT	8,452	86,311	67,882	18,363	29.33	1,491	2.27
CSCO	10,708	99,361	43,665	55,687	10.67	2,856	0.54
CVX	27,317	250,920	102,770	146,888	76.67	7,415	3.87
DD	4,491	50,370	37,089	12,975	14.00	1,652	1.77
DIS	9,951	80,108	34,013	43,382	24.33	1,320	0.74
GE	24,292	663,412	529,384	127,250	12.67	8,127	0.77
GS	17,645	902,101	822,695	71,460	153.33	1,281	2.02
HD	9,177	40,516	27,309	13,207	9.33	2,172	1.53
IBM	19,634	120,989	103,014	17,840	17.00	4,032	3.75
INTC	14,270	89,555	34,143	55,108	11.33	4,413	0.89
JNJ	19,152	128,383	58,839	69,544	24.67	7,223	2.58
JPM	53,735	2,449,319	2,233,548	202,344	53.67	6,325	1.36
KO	10,970	89,417	57,028	32,094	7.33	4,971	1.12

RIC	Earnings Before Interest and Taxes	Total Assets	Total Liabilities	Total Equity	Book Value Per Share	Total Dividends	Dividends per Share
MCD	8,282	35,431	20,713	14,719	14.67	3,076	3.09
MMM	6,761	32,898	16,522	16,062	24.33	2,076	2.77
MRK	9,869	103,371	51,177	50,477	17.33	5,154	1.72
MSFT	27,652	145,362	66,998	78,364	9.33	7,895	0.91
NKE	3,703	19,259	7,697	11,562	13.00	826	0.91
PFE	16,699	175,724	99,084	76,256	11.33	6,580	0.96
PG	16,080	138,591	71,018	67,108	24.67	6,523	2.29
T	19,281	280,977	190,610	89,907	17.00	9,805	1.80
TRV	4,732	103,943	78,930	25,012	71.33	723	1.97
UNH	9,717	83,050	49,561	31,927	32.67	1,079	1.09
UTX	8,394	90,431	59,317	29,664	32.33	1,903	2.20
V	7,210	38,179	10,875	27,304	42.33	822	1.27
VZ	21,902	244,009	179,134	28,097	9.67	6,910	2.08
WMT	26,027	203,105	120,848	76,343	23.00	5,361	1.59
XOM	49,881	333,795	162,135	165,863	37.00	10,092	2.18

Note: This table shows the fundamental information of the firms in our sample. The results are in million US dollar except for book value per share and dividends per share.

Table 4-4 Percentage of error records

RIC	Call	Put
AXP	1.38%	1.49%
BA	0.51%	0.54%
CAT	0.59%	0.58%
CSCO.O	0.86%	0.90%
CVX	0.82%	0.76%
DD	0.74%	0.76%
DIS	0.70%	0.77%
GE	0.71%	0.73%
GS	0.49%	0.48%
HD	0.57%	0.64%
IBM	0.79%	0.76%
INTC.O	0.65%	0.67%
JNJ	0.69%	0.71%
JPM	0.53%	0.55%
KO	0.77%	0.78%
MCD	0.75%	0.74%
MMM	1.01%	0.99%
MRK	0.79%	0.77%
MSFT.O	0.64%	0.66%
NKE	0.70%	0.73%
PFE	0.83%	0.84%
PG	0.73%	0.72%
T	0.87%	0.82%
TRV	0.91%	1.02%
UNH	0.73%	0.76%
UTX	0.81%	0.77%
V	0.79%	0.85%
VZ	0.74%	0.71%
WMT	0.80%	0.81%
XOM	0.66%	0.64%
Average	0.75%	0.77%

Note: This table shows the percentages of error records in total observations of call and put options, respectively. The error records include the irrelevant records and out-of-hour records.

Table 4-5 Percentage of missing values

RIC	Call	Put
AXP	2.10%	2.33%
BA	1.02%	1.12%
CAT	1.37%	1.39%
CSCO.O	1.98%	2.07%
CVX	1.65%	1.61%
DD	1.90%	1.98%
DIS	1.67%	1.88%
GE	1.15%	1.22%
GS	0.99%	0.99%
HD	1.12%	1.30%
IBM	1.03%	1.00%
INTC.O	1.10%	1.13%
JNJ	0.73%	0.78%
JPM	0.87%	0.92%
KO	0.93%	0.97%
MCD	2.01%	2.06%
MMM	4.94%	5.15%
MRK	2.01%	2.11%
MSFT.O	1.69%	1.78%
NKE	1.79%	1.97%
PFE	2.43%	2.52%
PG	1.79%	1.88%
T	2.27%	2.21%
TRV	3.18%	3.49%
UNH	1.72%	1.86%
UTX	1.84%	1.88%
V	3.26%	3.50%
VZ	3.31%	3.37%
WMT	2.96%	3.17%
XOM	2.78%	2.76%
Average	1.92%	2.01%

Note: This table shows the percentages of missing value in total observations of call and put options, respectively. The missing value is defined as both ask and bid information missed in a quote record and both trading price and volume missed in a trade record.

Table 4-6 Percentage of outliers

	Call options		Put options	
RIC	Quote	Trade	Quote	Trade
AXP	3.76%	3.46%	4.88%	4.25%
BA	3.00%	2.63%	3.23%	2.90%
CAT	3.72%	3.40%	3.23%	2.81%
CSCO.O	6.45%	6.36%	6.87%	6.77%
CVX	5.58%	4.51%	2.86%	2.33%
DD	2.93%	2.28%	3.83%	2.83%
DIS	2.23%	2.04%	3.67%	3.27%
GE	7.27%	7.16%	7.99%	7.79%
GS	4.97%	4.42%	3.47%	2.94%
HD	2.55%	2.40%	2.72%	2.55%
IBM	7.64%	6.76%	4.66%	4.04%
INTC.O	6.17%	6.07%	5.58%	1.94%
JNJ	3.67%	3.26%	4.73%	4.20%
JPM	4.03%	3.86%	5.81%	5.55%
KO	5.27%	5.19%	4.47%	4.39%
MCD	5.06%	4.71%	3.95%	3.64%
MMM	4.18%	3.57%	3.46%	2.97%
MRK	4.06%	3.57%	5.65%	4.90%
MSFT.O	5.91%	5.79%	6.45%	6.29%
NKE	2.96%	2.62%	3.56%	3.09%
PFE	7.50%	7.19%	7.29%	6.74%
PG	4.24%	4.11%	3.61%	3.46%
T	7.23%	7.11%	5.56%	5.46%
TRV	4.10%	4.05%	5.11%	5.02%
UNH	4.16%	3.65%	3.96%	3.20%
UTX	4.34%	3.84%	3.05%	2.65%
V	20.08%	19.51%	21.64%	20.94%
VZ	32.13%	32.00%	34.93%	34.74%
WMT	30.91%	30.77%	34.90%	34.72%
XOM	29.51%	29.24%	30.66%	30.45%
Average	7.85%	7.52%	8.06%	7.56%

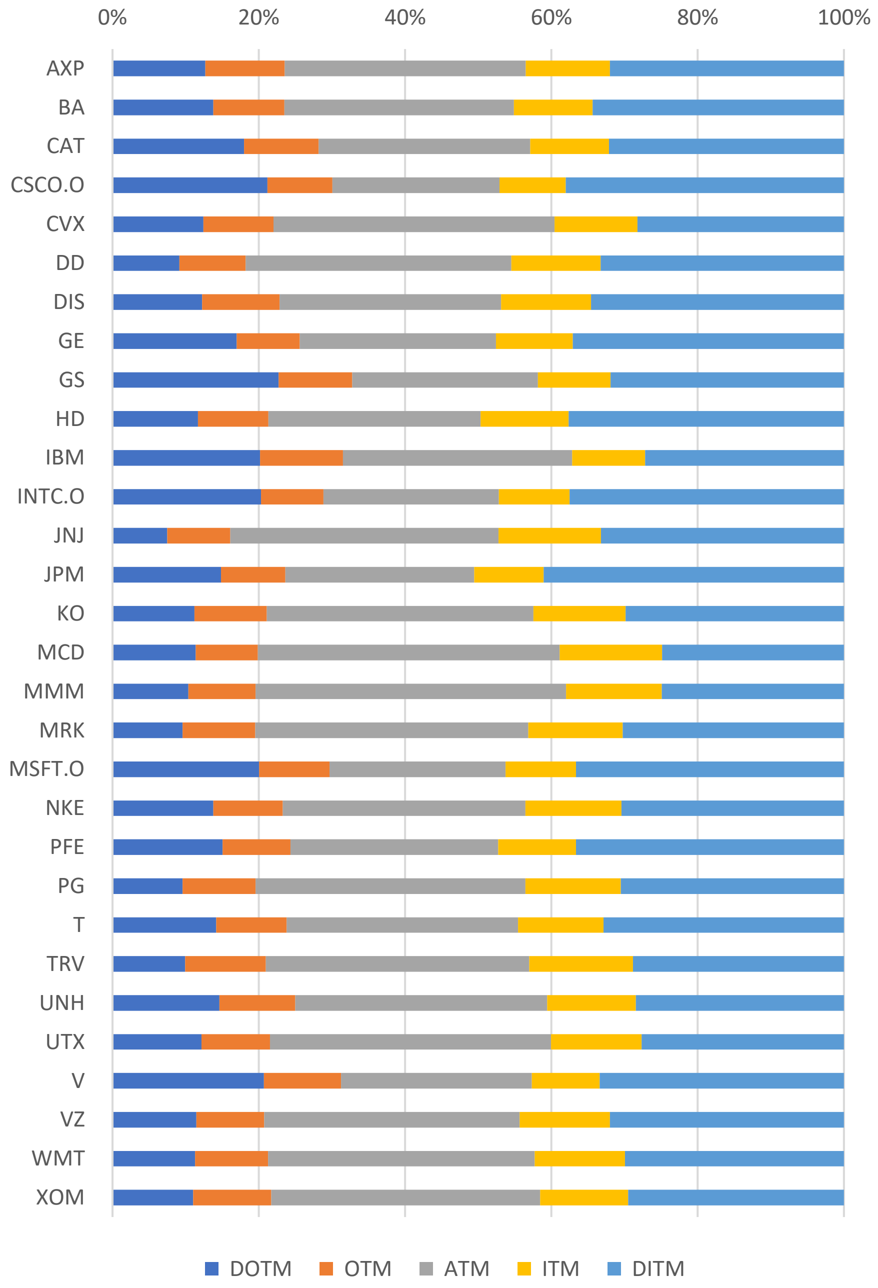
Note: This table shows the percentages of outlier in total observations of call and put options, respectively. For quotes, a record is defined as an outlier when the bid-spread of the record is higher than 150% or lower than 0%. For trades, a record is defined as outliers when the price or volume of the records is lower than zero.

Table 4-7 Percentage of records with expiration problem

RIC	Call	Put
AXP	25.99%	24.62%
BA	23.62%	23.17%
CAT	23.88%	22.55%
CSCO.O	18.66%	18.32%
CVX	23.27%	22.37%
DD	18.98%	17.36%
DIS	21.41%	19.27%
GE	18.06%	16.96%
GS	23.74%	23.17%
HD	23.70%	21.16%
IBM	22.17%	21.42%
INTC.O	18.39%	17.86%
JNJ	25.68%	24.83%
JPM	19.81%	19.23%
KO	21.05%	20.02%
MCD	23.40%	22.36%
MMM	24.87%	23.40%
MRK	20.29%	18.44%
MSFT.O	18.74%	18.35%
NKE	24.88%	23.10%
PFE	17.16%	16.11%
PG	24.93%	23.51%
T	19.42%	17.80%
TRV	18.91%	17.83%
UNH	22.47%	21.25%
UTX	24.28%	22.93%
V	22.36%	21.71%
VZ	19.04%	17.43%
WMT	25.22%	23.82%
XOM	27.48%	26.60%
Average	22.06%	20.90%

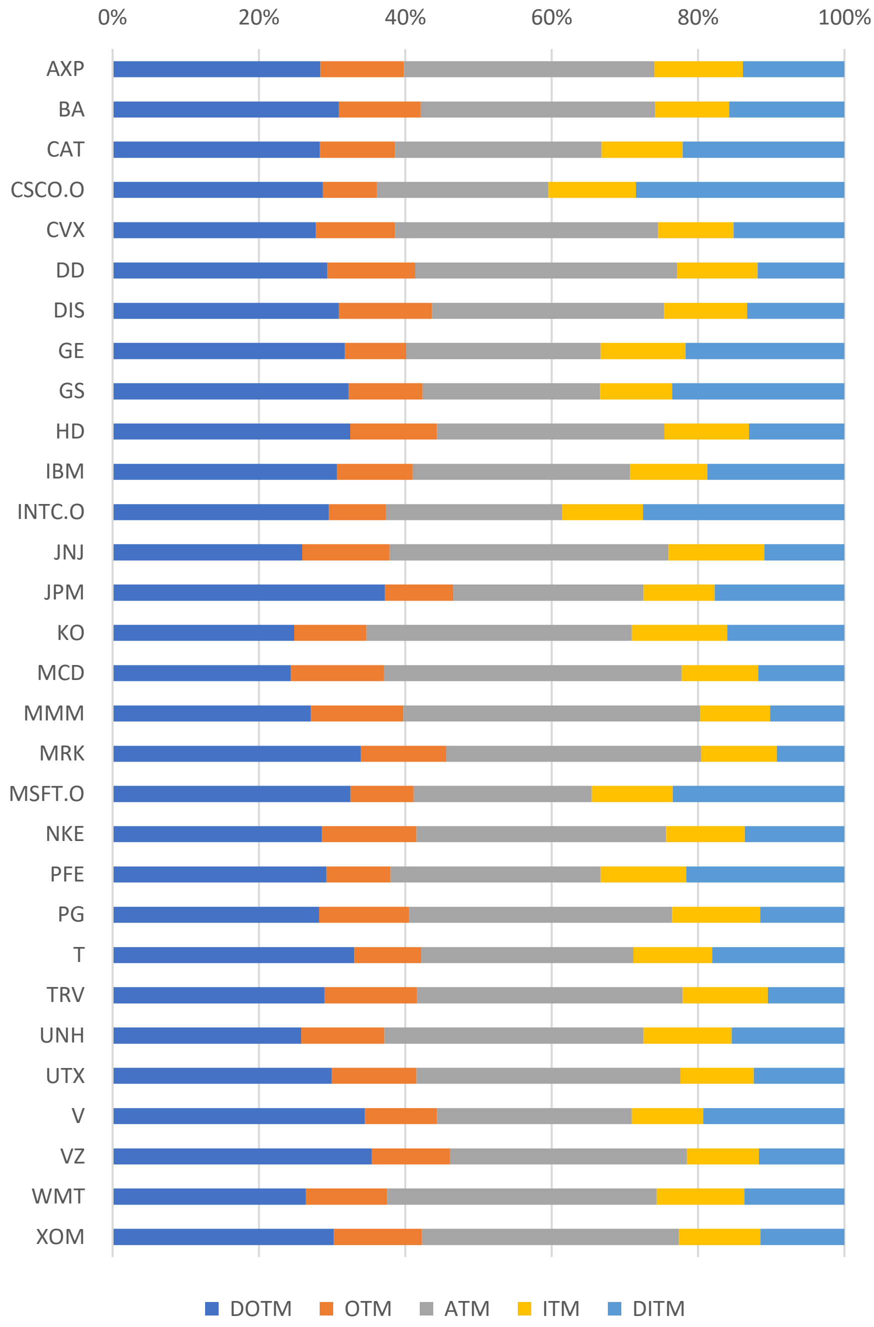
Note: This table shows the percentages of records with expiration problem in total observations of call and put options, respectively. The expiration problem is defined as options contracts that will be expired within 7 days or after over 365 days.

Figure 4-1 Moneyness distribution of call options contracts



Note: This plot shows the distribution of call options contracts based on their moneyness.

Figure 4-2 Moneyness distribution of put options contracts



Note: This plot shows the distribution of call options contracts based on their moneyness.

Table 4-8 Percentage of remaining observations

RIC	Call	Put
AXP	42.20%	41.79%
BA	45.45%	45.35%
CAT	41.00%	42.28%
CSCO.O	41.63%	40.89%
CVX	44.07%	44.87%
DD	49.51%	47.86%
DIS	48.17%	46.66%
GE	42.46%	40.26%
GS	40.92%	42.06%
HD	41.69%	41.46%
IBM	39.59%	41.78%
INTC.O	41.55%	41.21%
JNJ	44.46%	43.69%
JPM	42.28%	40.34%
KO	44.50%	44.42%
MCD	43.26%	42.87%
MMM	44.93%	45.58%
MRK	48.37%	46.02%
MSFT.O	42.68%	41.00%
NKE	43.90%	44.86%
PFE	40.84%	40.45%
PG	46.68%	47.52%
T	41.11%	40.69%
TRV	33.55%	33.07%
UNH	43.09%	43.94%
UTX	41.26%	41.77%
V	36.08%	36.13%
VZ	46.08%	43.30%
WMT	41.52%	42.30%
XOM	41.89%	41.45%
Average	42.82%	42.53%

Note: This table shows the percentages of remaining observations in total observations of call and put options, respectively

Chapter 5: What are the determinants of liquidity commonality in a quote-driven options market?

5.1 Introduction

It is accepted that liquidity plays a crucial role in all financial markets. A financial market with high liquidity allows traders to apply their trading strategies cheaply and reduce volatility (Harris, 2003). Hence, the variation of liquidity is a major concern to any market participants. There is an extensive body of literature that investigates the liquidity of individual assets and its determinants. Since the work of Chordia *et al.* (2000), research on liquidity began to focus on a market-wide phenomenon. The patterns and sources of liquidity commonality have important implications for the market participant as a priced source of risk (Coughenour and Saad, 2004; Rösch and Kaserer, 2014). Although the empirical evidence confirms the existence of co-movement between liquidity in individual assets, the intraday pattern and underlying fundamental source of liquidity commonality are yet to be determined.

Using high-frequency data from the Chicago Board Options Exchange (CBOE), we provide evidence of time-series liquidity co-variation in the options market and explore the common factors that drive the liquidity commonality. In particular, we consider two aspects of option liquidity: the behaviours and determinants of equity liquidity commonality. Specifically, the determinants are developed from two alternative strands of market microstructure theory: information- and inventory-based models.

Our study contributes to the literature in two ways. First, we study the intraday patterns of equity option liquidity commonality in CBOE which is a quote-driven market[14]. Cao and Wei (2010) first confirmed the existence of commonality in liquidity

[14] A quote-driven market is an electronic financial market system that allows market makers, dealers, or specialists to determine price through their bid and ask quotations. Additionally, dealers fill orders from their own inventory or through matching them with other orders (Harris, 2003).

for US options markets at the daily level, and Verousis *et al.* (2015b) employed high-frequency dataset to further show the liquidity co-movement pattern for order-driven options markets[15] at the intraday level. Brockman and Chung (2002) and Frino *et al.* (2008) argued that the behaviour of liquidity could be changed by the market structure. Bai and Qin (2015) suggested that the different roles of liquidity providers in each market structure have various impacts on liquidity. Since previous studies have failed to explore the intraday co-movement pattern of equity options liquidity in quote-driven markets, we employ principal component analysis to show the distribution of common liquidity factors during a trading day of CBOE.

Second, we extend the work of Verousis *et al.* (2015b) to the underlying fundamental source of commonality in options liquidity. Verousis *et al.* (2015b) displayed the intraday pattern of liquidity commonality but did not explore the determinants of liquidity commonality. They demonstrated evidence which confirmed asymmetric information and inventory risk as important components of liquidity. Huberman and Halka (2001) argued that commonality in stock liquidity cannot be explained by these two components due to the presence and effects of noise traders. However, Moshirian *et al.* (2017) found information asymmetry as a market-level determinant of liquidity commonality in stock markets. Hameed *et al.* (2010) confirmed these adverse effects but did not support the effect of inventory risk on stock liquidity commonality.

The empirical findings of our study can be summarised as follows. First, principal component analysis (PCA) provides evidence of liquidity commonality on the quote-driven market. The first three principal component factors can explain 54.8%

[15] In an order-driven market, all buyers and sellers display the prices at which they wish to buy or sell a particular security. In contrary to a quote-driven market, order-driven market also displays the amounts of the security desired to be bought or sold (Harris, 2003).

(55.1%) and 25% (25.2%) of the total variation in calls' (puts') bid-ask spread and depth. Second, we further use PCA to explore the variation of liquidity commonality during the trading day. The results show that commonality in bid-ask spread and depth is highest at the beginning of the trading day and levels off during the trading day. At the end of the trading day, commonality in depth has a clear increasing trend. Third, the time-series regression results show that commonality in liquidity is influenced by inventory risk but not by information asymmetry. The effects of inventory risk proxy show that commonality in liquidity is greater when the market is volatile or in a downward trend. Such market conditions could lead to order imbalance and increase the inventory risks of market makers who may adjust bid-ask spreads and depth across many options. This can contribute to liquidity commonality. The results of information asymmetry proxy are inconsistent with previous studies which shows a connection between information asymmetry and liquidity commonality.

With these empirical results, we not only offer academic contributions but also provide implications for market practitioners. First, we use the high-frequency dataset to study the intraday behaviour of equity option liquidity commonality. As a priced source of risk, intraday commonality in liquidity can contribute to the understanding of intraday price formation and thereby support investors to develop an intraday trading strategy. In particular, commonality in liquidity is a component of liquidity risks that can contribute to the asset pricing model. Additionally, commonality in liquidity shows an undiversifiable source of price risks. Therefore, it adds an additional dimension of risk for the investor when investing portfolios. Second, we examine the determinants of intraday liquidity commonality in a quote-driven options market so that can offer more understanding about how market liquidity shocks arise. To give more details, the empirical results of this chapter support the inventory risk explanation of liquidity commonality and thereby have an implication for crisis management. Since the inventory risks could lead to liquidity shock during the market downtrend, the regulator could develop macroeconomic policies to remit the inventory risks of market makers during the crisis.

The rest of the paper is organised as follows: Section 2 reviews previous studies on liquidity commonality and its determinants. Section 3 describes the institutional background, the data, and measurements. Section 4 shows the employed empirical methodologies. Section 5 represents the empirical results, followed by a conclusion in Section 6.

5.2 Literature review

5.2.1 Commonality in liquidity

The commonality in liquidity refers to the proportion that market-wide liquidity factors can partly influence the liquidity of individual firm (Brockman and Chung, 2002). Such commonality was firstly demonstrated by Chordia *et al.* (2000) on the US stock market and then widely investigated in the context of different market structures (Brockman and Chung, 2002), different countries (Galariotis and Giouvris, 2007), and different assets (Cao and Wei, 2010). Hence, in this section, we discuss the empirical literature on liquidity commonality within these different contexts.

The concept of commonality in liquidity was first introduced by Chordia *et al.* (2000) through investigating NYSE stocks during the course of 1992. By employing a market model time series regression, their results showed that liquidity is not an asset-specific phenomenon because the liquidity co-movement across assets can be found after controlling for individual sources of liquidity such as volatility. Huberman and Halka (2001) examined a similar sample and also found evidence confirming the existence of the systematic, time-varying component of liquidity. However, they argued that the systematic component of liquidity might not be driven by inventory and asymmetric information risks which are against Chordia *et al.* (2000). Hasbrouck and Seppi (2001) provided evidence of liquidity commonality with a sample of 30 stocks from the Dow Jones Industrial Average. Unlike previous studies, they computed liquidity into 15-minute intervals and adopted principal component analysis and canonical correlation analysis to test commonality in liquidity. As to the results, they found that the common factors in liquidity varies across liquidity measures. Quote based proxies of liquidity represent significant common factors, while price impact measures of liquidity show less significant factors.

To further explore commonality, Brockman and Chung (2002) changed the focus from the quote-driven markets to the order-driven markets. They used the trading data of all firms listed on the Hong Kong Stock Exchange to show that commonality has less effect on liquidity than commonality reported in the quote-driven market. A possible explanation is the market structure, namely that no market makers have the obligation to maintain liquidity-provision services in order-driven systems during market-wide liquidity shocks. Hence, such systems are more susceptible to commonality. Coughenour and Saad (2004) supported the effects of market structure arguing that, in quoted-driven markets, the specialists within a firm share capital and information and thereby provide correlated liquidity. Based on these arguments, Galariotis and Giouvris (2007) investigated the effect of trading regimes by providing first-time evidence on the existence of liquidity commonality on the London Stock Exchange. Particularly, their data spans the period between 1996 and 2001 which covers a quote-driven and an order-driven trading regime. Unlike the above-mentioned studies, their findings indicated that there is no significant difference in commonality between different trading regimes. Furthermore, Brockman *et al.* (2009), Karolyi *et al.* (2012), and Moshirian *et al.* (2017) show commonality in liquidity in most of the world's stock markets.

From the above literature, the evidence on the existence of commonality in liquidity is based on the stock markets. To fulfil the research gap, Cao and Wei (2010) found evidence from the US equity options market. Using daily data from 1996 to 2004, Cao and Wei (2010) showed that daily liquidity measures of the individual option (bid-ask spreads, volume, and price impact of an individual option) are susceptible to market-wide option liquidity and underlying stock liquidity. These results are robust for call and put options individually. Based on the hypothesis that commonality is influenced by market structure, Verousis *et al.* (2015b) extended the search for commonality to an order-driven market for a high-frequency dataset of equity options. Using intraday data from 2008 to 2010, they employed principal component analysis to identify the intraday pattern of systematic liquidity. The results suggested that systematic movements are linked to asymmetric information factors. This is supported by the pattern that the peak in systematic liquidity at the opening of the markets corresponds to overnight information arrival and the opening of the US equity market. By using the same approach, Verousis *et al.* (2016a) further investigated the time series of systematic liquidity for European options markets[16] at the high-frequency level. As to the results, they discovered a higher commonality in liquidity during the period of liquidity deterioration that implies a connection between systematic liquidity and financial crisis.

[16] Their dataset includes all equity options traded on the ICE exchanges in Amsterdam, London, and Paris from 2008 to 2010.

In summary, the aforementioned publications reveal that the intraday pattern of liquidity commonality has been limited to investigations into the US quote-driven options markets. Since the market structure may influence commonality in liquidity, it is important to explore commonality in liquidity for quote-driven options markets. In addition, the intraday pattern can efficiently contribute to the development of market microstructure theories. Particularly, the price formation can be improved by the in-depth understanding of inventory and asymmetric information risks in liquidity commonality which is a priced source of risk.

5.2.2 Determinants of commonality in liquidity

Although extensive research has confirmed commonality in liquidity among stocks or options, we know relatively little about the underlying fundamental source of liquidity commonality. Previous studies have proposed a mixture of explanations around commonality in liquidity that will be reviewed in this section.

Demand- and supply-side explanations are proposed by Hameed *et al.* (2010), Karolyi *et al.* (2012), and Rösch and Kaserer (2013). From the demand-side explanation, commonality in liquidity is related to correlated trading behaviours caused by institutional investors and investor sentiment. Kamara *et al.* (2008) showed that institutional investing can explain the increase in liquidity commonality among large-cap US equities from 1985 to 2005. To give more detail, institutional investing uses a security basket as a possible means of trading and concentrates on large-cap stocks in order to satisfy the "prudent man" rule. Karolyi *et al.* (2012) used a global sample to confirm these arguments by finding a positive relationship between liquidity commonality and the presence of institutional investors. They further provided evidence that the investor sentiment can contribute to co-movement among individual securities, consistent with the conjecture of Huberman and Halka (2001) that commonality in liquidity is driven by the presence of noise traders.

From the supply-side explanation, commonality in liquidity is driven by funding constraints which is explained by the previous literature in different ways. Hameed *et al.* (2010) found greater liquidity commonality on the NYSE during market decline because the inventory constraints prevent market-makers from absorbing temporary liquidity shocks. This explanation is supported by Coughenour and Saad (2004) who argue that the co-movement of NYSE stocks is driven by the same specialist firm. Karolyi *et al.* (2012) and Rösch and Kaserer (2013) explained funding liquidity through the view of financial intermediaries. In particular, the intermediaries may reduce the provision of liquidity because they bear losses in collateral values during a market decline. However, these researchers found mixed results regarding the relationship between commonality and the behaviour of financial intermediaries.

Moshirian *et al.* (2017) investigated the sources of commonality through another strategy that is based on market- and firm-level factors. They argued funding supply and investors' demand proposed by Karolyi *et al.* (2012) as the partial components of market-level determinants. Their results for market-level factors further showed the effects of economic and financial environments on the commonality which supports the findings of Brockman *et al.* (2009). Particularly, with a sample of 47 stock exchanges, Brockman *et al.* (2009) found that liquidity commonality is positively influenced by the macroeconomic announcement. In addition, Moshirian *et al.* (2017) showed the lack of publicly available information and the cultural background of investors as other market-level factors that can lead to commonality in liquidity. In the firm-level factors, liquidity commonality is not only driven by correlated trading but caused by firm-level transparency and stock performance (Moshirian *et al.*, 2017). In particular, greater firm-level transparency can contribute to stock price informativeness and meanwhile reduce the level of information asymmetry. Lang and Maffett (2011) confirmed that liquidity commonality can be reduced by a lesser degree of information asymmetry during financial crises.

Our contribution to commonality in liquidity is to understand the fundamental source of commonality in a quote-driven options market through two alternative strands of market microstructure theory. Chordia et al. (2000) suggested inventory costs and information asymmetry as possible sources of commonality in liquidity, while empirical studies find mixed results around the effects of inventory risks and asymmetric information on liquidity commonality. For example, Huberman and Halka (2001) did not find evidence that inventory risk and information asymmetry can explain systematic liquidity. Conversely, the results of Coughenour and Saad (2004) and Moshirian *et al.* (2017) support these two sources, whilst not directly examining the proxy of inventory risk and information asymmetry.

5.3 Data and variable construction

5.3.1 Sample selection

This chapter uses the high-frequency dataset that contains all options contracts written on the 30 components of the Dow Jones Industrial Average Index (DJI) from 03 Jan 2012 to 30 Jun 2014. In order to collect a more accurate estimate of liquidity and determinant variables, a set of filters is applied to form the final sample, following on from Cao and Wei (2010) and Verousis *et al.* (2016a). First, we removed options contracts with a maturity of more than 90 days. Second, we excluded observations if their moneyness was out of the range between 0.95 and 1.05[17]. Finally, we exclude observations between 09:30 and 09:35 to avoid overnight effects. After applying these filters, our final sample can focus on short-term at-the-money (ATM) contracts and consist of 348 million and 345million quote records for call and put options.

5.3.2 Variable construction

5.3.2.1 Liquidity measures

As with Verousis *et al.* (2016a), we construct intraday time-series for each ticker by averaging the liquidity measure across all options written on a stock at millisecond *t*. For example, we compute a liquidity measure of ticker *q* at millisecond *t* by averaging the liquidity measure across the respective sub-tickers at that millisecond. We create liquidity measures in two ways, following on from the researches of Chordia *et al.* (2000) and Cao and Wei (2010). The simplest proxies of liquidity are the quoted spread (hereafter spread) and quoted depth (hereafter depth) which can reflect the different aspects of liquidity (Chordia *et al.*, 2000). Quoted spread refers to the difference between the contemporaneous bid and ask prices, and quoted depth refers to the quantity dimension of liquidity. These liquidity measures for sub-ticker *i* at millisecond *t* can be computed as follows:

$$SPD_{i,t} = 100\% \times \frac{P_{i,t}^A - P_{i,t}^B}{Mid_{i,t}} \tag{1}$$

$$DPT_{i,t} = \frac{Q_{i,t}^A + Q_{i,t}^B}{2} \tag{2}$$

where $SPD_{i,t}$ refers to the proportional quoted spread of sub-ticker *i* at millisecond *t*; $P_{i,t}^A$ and $P_{i,t}^B$ refer to ask and bid price of sub-ticker *i* at millisecond *t*,

[17] Moneyness is the spot-to-strike ratio calculated by the underlying's average bid/ask price over the option's strike price.

respectively; $Mid_{i,t}$ indicates the midpoint of the quote spread that is the mean of the quoted ask and bid prices; $DPT_{i,t}$ refers to quoted depth of sub-ticker *i* at millisecond *t*; $Q_{i,t}^{A}$ and $Q_{i,t}^{B}$ refer to ask and bid volume of sub-ticker *i* at millisecond *t*, respectively.

--------------Insert Table 4-1---------------

We present descriptive statistics for liquidity measures in Table 4-1 that report bid-ask spread and depth for each call and put options series. The average and standard deviation are calculated on a per sub-tickers basis. Our descriptive statistics reveal considerable cross-sectional variation across the tickers. For calls (Panel A of Table 4-1), the range of SPD is from 4.55% (MSFT) to 12.63% (TRV) and that of DPT is from 133.57 (GS) to 1280.7 (INTC). For puts (Panel B of Table 4-1), the ranges of SPD and DPT are from 3.6% (JPM) to 10.75% (TRV) and from 141.35 (GS) to 1296.42 (INTC), respectively. The results suggest that puts are more liquid than calls during the sample period.

--------------Insert Figure 4-1---------------

Following Chan *et al.* (1995), we further standardise all liquidity measures through the daily mean and daily standard deviation of each sub-ticker. This could control cross-sectional differences across tickers. Additionally, the measures are also winsorised at 99% level for removing the extreme value. Figure 4-1 shows the distributions of the measures for calls and puts during the trading day that displays an L-shaped pattern. The distribution of spread is consistent with Chan *et al.* (1995) and ap Gwilym *et al.* (1998a) who conclude that spreads are widest after the open and steadily decline within one hour because of accumulated overnight information and informed traders. However, the distribution of quoted depth at the beginning does not support those who argue for a negative relationship between depths and information asymmetry. Furthermore, the changes in spreads at the end of the trading day is also inconsistent with Chan *et al.* (1995) who observe lower spread at the close on the CBOE. In particular, the inventory management theory suggests that market makers reduce spreads in order to achieve their desired level of inventory. A possible explanation is that within the multiple market-makers system it is easier for market-makers to avoid accumulating inventory risk during peak trading periods (Chan *et al.*, 1995). Moreover, previous studies also do not mention temporary changes in liquidity around 10:00 and 14:00. Particularly, this study observes that the bid-ask spread suddenly increases during these time intervals and reverts to its pre-level during the next time interval.

5.3.2.2 Determinants and control variables

Regarding the determinants of liquidity commonality, we focus on inventory risks and asymmetric information risks. Microstructure theory indicates these risks can determine the liquidity of assets. Hence, common components of liquidity may also connect with these risks. For example, Chordia *et al.* (2000) showed that the correlation in inventory holding costs has a greater impact on liquidity. This correlated inventory holding costs are driven by the market-wide swings in trading activity. However, previous studies not only conclude mixed results from these risks but also provide very little evidence on options markets (Chordia *et al.*, 2000; Huberman and Halka, 2001; Cao and Wei, 2010; Moshirian *et al.*, 2017). Using a group of proxies, we contribute to the literature by testing what risks have contributed to the variations in liquidity commonality.

The market trend and options volatility are used to proxy the inventory risk. When the market declines or has increased uncertainty, market-makers have a greater aversion to inventory risk that reduces liquidity across many assets significantly. As for proxies for information asymmetry, the trade volume, trade duration and number of trade transactions are adopted. The informational role of options volume is supported by Easley *et al.* (1998b) and Lee and Cheong (2001) who argue that informed traders prefer to trade in large volumes to maximise their trading profits. Additionally, informed traders may also trade in a rush in order to take advantage of their private information (Easley and O'hara, 1992). This will increase the number of trades and reduce trade duration. Hence, high volume, a high number of trades, and low trade duration may reflect high information asymmetric risks.

Similar to Cao and Wei (2010), we separate the market trend into up and down market movements. More specifically, R_k^+ (R_k^-) refers to the average return of 30 tickers at interval *k* when the average return is positive (negative). If the average return is negative (positive), R_k^+ (R_k^-) equals to zero. The return of each ticker during the interval *k* is estimated by the principle of daytime return, as follows:

$$R_{i,k} = \frac{Mid_{i,k}^{last}}{Mid_{i,k}^{first}} \tag{3}$$

where $R_{i,k}$ refers to the return of sub-ticker *i* during the interval *k*; $Mid_{i,k}^{last}$ and $Mid_{i,k}^{first}$ refers to the last and the first midpoint of the quote spread during the interval *k*.

The volatility for sub-ticker *i* during the interval *k* is computed by the absolute value of the return thus:

$$Volt_{i,k} = \left| R_{i,k} \right| \tag{4}$$

Regarding proxies of information asymmetry, we compute each proxy for a given ticker *q* during the interval *k* by averaging the proxy across its respective sub-tickers on that interval. The trade volume (Vlm) and the number of transactions (Tr) can be directly estimated from the final dataset. The trade duration (Dur) is defined as the volume-weighted time between two same-day transactions:

$$Dur_{i,t} = Vlm_{i,t} \times S_{i,t} \tag{3}$$

where $S_{i,t}$ refers to how many seconds have elapsed between this transaction of the sub-ticker *i* at millisecond *t* and the previous same-day transaction.

---------------Insert Table 4-2----------------

Table 4-2 represents the summary statistics of determinant variables. It is interesting to note that the average return for the sample is negative. This may explain why puts are more liquid than calls, as puts are generally bearish. Regarding volatility and volume, calls and puts have no big difference. In terms of trading frequency, calls represent a higher frequency than puts.

---------------Insert Table 4-3----------------

Since Moshirian *et al.* (2017) showed that the commonality in liquidity is also sensitive to the economic and financial environment, we employ the dummy variable of macroeconomic announcements to control for this effect. Table 4-3 represents the selected announcements with their category and announcement time. The selection of announcement is primarily based on Bloomberg's relevance index and is supported by Frijns et al. (2015). As to the results, we selected 17 events which were announced during the trading hours of CBOE.

5.4 Methodology

5.4.1 Evidence of commonality in liquidity

To extract the commonality in liquidity, the market regression model and principal component analysis (PCA) were widely used by previous studies. Chordia *et al.* (2000) and Cao and Wei (2010) adopted the former method to identify liquidity commonality in equity markets and derivative markets. However, Hasbrouck and Seppi (2001) adopted PCA at the daily level and argued that the explanatory power of the typical regression model is not impressive because large components of noise and other influences on changes among individual options liquidity constructs can influence the quality of results.

PCA method for examining common factors mainly depends on the covariance matrix of the standardised variables and assigns no special role to the market portfolio. Unlike factor analysis, PCA is a data reduction technique that allows a large number of variables to be combined into a few variables. This method is able to construct factors to maximise explanatory power within a set of related variables. Since PCA is able to extract the common information across liquidity measures, principal components can be considered as ‘liquidity factors’ for individual equity options. A small set of extracted variables can represent most of the variations in the liquidity measure. In this context, we adopted PCA to extract market-wide liquidity.

Before conducting PCA, as mentioned before, we separated call and put options and standardise all liquidity measures through the daily mean and daily standard deviation of each sub-ticker in order to address the problem of overweight. Korajczyk and Sadka (2008) suggest that the units across measures could vary by several orders of magnitude when conducting PCA to perform a factor decomposition across the selected liquidity measures. The scale of some measures could make them more variable than other measures and thereby lead to overweighting these measures. Furthermore, we balance the number of observations n and the number of tickers i as we attempt to apply PCA in the intraday level. In principle, n should higher than i as avoiding erroneous conclusions.

We obtain one time-series of liquidity per ticker through averaging the liquidity measure across its respective sub-tickers at that millisecond because each ticker contains several sub-tickers. Hence, the number of observations n should be higher than 30 which is the number of tickers i in our dataset. When deciding the number of observations n, the time interval k of applying PCA should be also considered. For example, if the time interval k is 10 minutes, we should estimate liquidity measures for every 20 seconds in order to provide 30 observations for applying PCA. However, not all tickers can estimate liquidity measures for every 20 seconds of each trading day. Since a large number of liquidity measures is missing, PCA cannot provide reliable results. To improve the quality of results, we conduct a set of pre-tests for balancing the time interval k, the number of observations n, and the number of tickers i.

As a result, we can obtain sufficient results when the time interval k is 30 minutes and the number of observation n is 60. In particular, the liquidity measures of each ticker are averaged across each 30-second intervals. Finally, we have 30 time-series of liquidity and each of them has 60 observations over a 30 minutes interval k. Hence, we can apply PCA separately for calls and puts in each 30-minute interval. Regarding the results, we extract the first three principal components for each liquidity measure on each 30-minute interval and adopt them as the proxy of the commonality in liquidity, $\mathrm{CiL}_{\mathrm{t}}$. Since the number of available tickers per day varies, we exclude the

PCA results if the factor of results is less than 30. Besides, we also estimate the liquidity commonality for the entire dataset that provides quantitatively similar results.

5.4.2 Determinants of liquidity commonality

--------------Insert Table 4-4----------------

After providing evidence of the systematic liquidity, we further investigate its determinants by examining inventory risk and asymmetric information risk. We examine these determinants by performing the analysis for the common factors created by each liquidity measure. The expected effects of these determinants are represented in Table 4-4. The following regression summarises the potential relationship between the common factors, their determinants, and control variables:

$$CiL_k = \beta_0 + \beta_1 \text{ Inventory}_k + \beta_2 \text{Information}_k + \beta_3 ANN_k + \varepsilon_t \quad (5)$$

where CiL_{k} refers to the proportion of variance explained by the common factor of SPD and DPT during the 30-minute interval k. Inventory $_k$ refers to the proxies of inventory risks that include volatility ($Volt_k$) and market trend ($R_k^+, R_k^-, PR_k^+,$ and PR_k^-). In particular, $Volt_k$ refers to quote volatility during the 30-minute interval k. R_k^+ (R_k^-) refers to the average return of 30 tickers at 30- minute interval k when the average return is positive (negative) and zero otherwise. PR_k^+ (PR_k^-) refers to positive (negative) market returns at the interval k-1. Information_k refers to the proxies of information asymmetry that includes trade volume (Vlm_k), volume-weighted duration (Dur_k), and the number of transactions Tr_k. In particular, Vlm_k refers to the logarithmic trading volume averaged across the 30-minute interval k. Dur_k refers to the volume-weighted time between two same-day transactions averaged across the 30-minute interval k. Tr_k refers to the total number of options trade transaction during the 30-minute interval k. $ANN_{i,t}$ refers to the dummy variable of macroeconomic announcements at the 30-minute interval k. It takes the value of zero if there are no macroeconomic announcements and one if there is an announcement. Table 4-5 shows the correlation results between variables.

---------------Insert Table 4-5----------------

5.5 Empirical results

5.5.1 Commonality in liquidity and variation during the trading day

Before discussing the determinants of liquidity commonality, we show the results of PCA for calls and puts based on the daily time-series of spread and depth in Table 4-6. It reports the first three eigenvalues for the call and put options, respectively, and the cumulative proportion of liquidity explained by the first three principal components. For all liquidity measures in Table 4-6, the results display strong evidence of liquidity commonality in the options market as suggested by Cao and Wei (2010) and Verousis *et al.* (2016a).

---------------Insert Table 4-6----------------

Corwin and Lipson (2011) indicated that each eigenvalue will equal one and the first three principal components can explain 3/N of the total variation if there are no common components in the original variables. For bid-ask spread (Panel A of Table 4-6), the first eigenvalues for the call and put options are 14.806 and 14.923, respectively. Since the sampled options are written on the 30 Dow Jones stocks, the results of eigenvalues imply that 14.806/30=49.4% and 14.923/30=49.7% of the total variation in call and put bid-ask spreads can be explained by the first common factor. Although the relatively low values of eigenvalues in second and third principal component imply that the additional factors are ineligible, the high proportion of liquidity is explained by the first principal component factor and could be viewed as strong evidence of commonality in liquidity. In total the first three principal components explain 54.8% and 55.1% of the total variation in the spreads of calls and puts. Furthermore, the PCA results are weaker in the case of depth. The first eigenvalues in depth (Panel B of Table 4-6) for calls and puts are 5.316 and 5.399, respectively. Although the second and third eigenvalues for them are much smaller than the first eigenvalues, they are still higher than one. In total, the first three principal components can explain 25% of the total variation in calls' depth and 25.2% of puts' depth liquidity. Hence, the results for depth also suggest liquidity commonality.

---------------Insert Figure 4-2----------------

PCA is replicated separately for each 30-minute interval on each trading day that offers the intraday behaviour of commonality in liquidity. The first three common factors for calls and puts can also explain the liquidity variance at a high level as shown in Figure 4-2. The figure shows the time-series of liquidity variance in bid-ask spreads and depth explained by the first three common factors[18]. For commonality in spreads, it represents an L-shaped pattern that the first three components explain the highest (65%) of liquidity variation at the beginning of the trading day. The explanatory power of spread commonality drops to around 38% after 10:30am and maintains this level until the end of the trading day. This pattern is consistent with Verousis *et al.* (2015b) who argued that the high commonality at the open may correspond to the accumulated overnight information. For commonality in depth, it represents a weak U-shaped pattern. The first three factors can explain nearly 50% of the variation of depth at the beginning of the trading day and the explanation power maintain around 36% during

[18] The vertical axis displays the cumulative proportion of liquidity explained by the first three principal components

the day. At the end of the trading day, there is a small increase in the depth commonality that reaches over 40%.

Both information asymmetry and inventory risk may explain the higher commonality during the beginning of the trading day. As aforementioned, the high spreads at the open may be driven by the information asymmetry (Chan *et al.*, 1995). In particular, the bid-ask spread is widened by market-makers who can obtain compensation for trading with informed traders. Since the market-wide information could influence multiple firms simultaneously (Chordia *et al.*, 2000), the spread of these firms will become wider because of higher asymmetry in market-wide information. This adverse selection risk across options can result in liquidity commonality. Since informed traders may use private information to take advantage at the same time and thereby trade in the same direction (Hirshleifer *et al.*, 1994; Chordia *et al.*, 2005), this implicates herding behaviours that not only directly contribute to commonality in spreads but also result in order imbalances across options. In this context, the inventory problem faced by market-makers is exacerbated in that market-makers are encouraged to adjust spreads and depth similarly across options (Chordia *et al.*, 2002; Li *et al.*, 2005). To give more detail, the order imbalances across options could be considered as market-wide swings in trading activity that will influence the inventory holding costs. As a result, the correlated inventory holding costs contribute to commonality in liquidity (Chordia *et al.*, 2000). To further explore commonality in liquidity, the later section will use regression analysis to investigate the determinants of liquidity commonality.

5.5.2 Determinants of liquidity commonality

The previous section confirms that the liquidity of individual options can be explained by systematic liquidity, whilst only implying that the liquidity commonality is driven by the inventory risk and the risks associated with information asymmetry. Hence, this section focuses on providing empirical evidence of the fundamental source of liquidity commonality.

--------------Insert Table 4-7---------------

Table 4-7 reports the relationship between commonality in liquidity and our proposed determinant variables. We find significant effects from all key determinants, while the signs are not fully consistent with our expectations as shown in Table 4-4. Volatility (Volt) and the market downturn (R^-) are the proxies of inventory risks. They have significant coefficients with the expected sign. With greater market volatility and down markets, the market is associated with greater liquidity commonality. This finding is consistent with Hameed *et al.* (2010) and Karolyi *et al.* (2012).

The asset's price uncertainty is considered as one of the most important factors that influence inventory risks. The asset's value has a high probability of decrease when its price is volatile. In this context, market-makers experience higher inventory risks of holding this asset and thereby adjust spreads and depth to manage their inventory risk. Since market-wide volatility is associated with multiple options, market-makers may adjust spreads and depth of these options simultaneously that lead to liquidity commonality. According to Chan *et al.* (1995), the volatility of CBOE is higher at the beginning of the trading day. This may explain why commonality in liquidity is higher at the beginning. During market declines, the greater order imbalance raised by the correlated trading increase the inventory risk that also lead to commonality in liquidity. In addition, market-makers may also face funding constraints during the period. Hence, they cannot continually provide liquidity that further contributes to liquidity commonality (Hameed *et al.*, 2010).

As for proxies for information asymmetry, the signs of trading volume (Vlm), value-weighted duration (Dur), and the number of transactions (Tr) have significant impact on liquidity commonality but completely differ from our expectations. Based on Chordia *et al.* (2000), Cao and Wei (2010) and Verousis *et al.* (2016a), information asymmetry can increase liquidity commonality. Regarding our proxies, high volume, short duration, and a greater number of transactions implicate greater information asymmetry that should in turn lead to high commonality in liquidity. However, our results in Table 4-7 represent the opposite result, namely that liquidity commonality is increased by low volume, long duration, and a low number of transactions. A possible explanation is that large volume no longer reflects the high information asymmetry. Informed traders prefer to trade in small sizes the better for hiding their information (Blau *et al.*, 2009). Hence, small volume implies high information asymmetry. This corresponds to the negative relationship between volume and liquidity commonality. However, this possible explanation conflicts with the negative relationship between the number of transactions and commonality in liquidity. When informed traders split their orders, the high number of transactions reflects adverse selection risks that should result in liquidity commonality.

In this context, a more reliable conclusion is that information asymmetry cannot explain the commonality in liquidity. This conclusion is consistent with Huberman and Halka (2001), namely that liquidity co-movement is not driven by information asymmetry. A possible explanation is that uninformed traders may create noise trading and do not engage in herd trading because they cannot detect informed traders who hide their trading activities by trading in small sizes.

5.6 Conclusion

Liquidity is an important feature of financial markets and has been widely researched by previous publications, while limited studies on liquidity have investigated the market-wide phenomenon of liquidity, especially in options markets. Using high-frequency CBOE data from 3 January 2012 to 30 June 2014, we contribute to the literature on the liquidity co-movement of individual options through investigating the intraday liquidity commonality of equity options and the determinants of the commonality.

Our main findings offer several new insights. First, we provide evidence of stronger liquidity commonality at the beginning of the trading day, indicating stronger co-movement between liquidity in individual equity options and the aggregate market-wide liquidity during the first and half hour of the trading day. Consistent with Verousis *et al.* (2015b), this liquidity commonality pattern supports the argument that the overnight information arrival may contribute to the commonality. Second, our empirical results confirm that inventory risks contribute to liquidity commonality. With high market volatility or market declines, market-makers experience greater inventory risks and thereby tend to adjust spreads and depth across options that lead to liquidity commonality. Besides, the funding constraints of market makers may also contribute to explain the positive relationship between the inventory risks and liquidity commonality. Regarding the proxies of information asymmetry, they are significantly related to commonality in liquidity but with unexpected signs. In line with Huberman and Halka (2001), we conclude that information asymmetry explanation with regards to liquidity cannot explain the commonality in liquidity. However, this study has only focused on market-level determinants of liquidity commonality and has not considered firm-level determinants. For example, firm-specific information can encourage more individual trading of an asset that decreases commonality in liquidity.

The results we document have important implications for understanding the systematic liquidity risk which is connected with asset pricing. We represent how commonality varies over the day which may have important implications for investors. For example, the time-series pattern of systematic liquidity movement could influence the paradigm of asset pricing models and change the trading behaviours of investors. In addition, we have uncovered the determinants of liquidity commonality that can contribute to understanding pervasive liquidity shocks. Policymakers and market regulators may be interested in developing a useful framework for reducing inventory risks of market makers during the period of market stress.

Table 5-1 Statistic results of liquidity measures

Panel A: Call options

		SPD		DPT	
RIC	N	Mean	Std	Mean	Std
AXP	10,149,907	7.45	10.99	245.04	249.72
BA	16,604,598	5.66	8.08	240.10	269.61
CAT	13,299,048	5.34	8.41	265.83	278.22
CSCO.O	6,579,449	6.22	10.06	1205.94	1206.01
CVX	14,227,248	9.06	15.20	352.25	380.62
DD	12,910,435	9.53	14.65	328.39	338.98
DIS	11,315,369	5.10	7.53	328.88	342.10
GE	7,806,625	8.09	11.90	1134.82	929.72
GS	9,815,618	6.03	8.33	133.57	127.85
HD	11,165,141	5.95	9.94	293.99	330.24
IBM	14,302,865	4.97	8.61	134.64	144.08
INTC.O	8,042,271	5.88	9.27	1280.70	1029.13
JNJ	11,285,699	9.36	15.72	426.75	417.86
JPM	15,688,134	4.58	8.10	316.59	381.85
KO	11,025,673	8.29	13.01	556.79	464.32
MCD	13,691,992	7.57	12.77	244.23	256.77
MMM	14,353,023	8.35	12.71	206.14	201.67
MRK	12,475,644	8.21	12.41	431.52	424.19
MSFT.O	11,838,807	4.55	7.12	822.51	833.68
NKE	12,575,783	6.96	8.58	273.50	270.63
PFE	9,258,877	8.20	12.79	717.77	688.47
PG	11,929,913	7.25	12.37	318.24	318.77
T	10,160,281	7.10	11.16	725.21	679.60
TRV	2,987,831	12.63	13.83	315.58	227.48
UNH	10,772,831	7.53	10.80	215.70	243.79
UTX	13,351,063	8.82	13.89	214.92	191.80
V	10,196,406	6.28	7.75	144.84	136.76
VZ	15,523,378	5.99	10.14	491.42	507.65
WMT	12,276,917	7.79	13.67	321.36	319.35
XOM	12,909,617	7.33	13.25	411.22	456.08

Panel B: Put options

RIC	N	SPD Mean	SPD Std	DPT Mean	DPT Std
AXP	9,884,950	6.62	8.44	251.41	251.15
BA	16,153,892	5.03	5.47	251.75	279.02
CAT	13,131,053	4.43	5.41	278.77	282.96
CSCO.O	6,541,873	5.20	7.27	1,225.46	1,248.05
CVX	14,085,238	6.91	8.89	347.57	368.37
DD	12,637,975	7.77	9.66	332.18	342.41
DIS	11,104,832	4.57	5.19	340.17	352.57
GE	7,724,535	6.65	8.08	1,184.48	968.19
GS	9,518,284	5.50	5.94	141.35	129.01
HD	11,097,252	4.93	6.23	296.40	324.46
IBM	14,016,615	4.02	4.75	142.75	147.36
INTC.O	7,953,936	4.83	6.31	1,296.42	1,065.35
JNJ	11,430,027	7.20	10.04	433.87	413.02
JPM	15,404,817	3.60	4.47	339.71	411.42
KO	11,217,141	6.45	8.93	549.00	452.71
MCD	13,754,879	6.02	8.30	239.32	249.56
MMM	14,271,378	6.68	7.55	200.64	197.73
MRK	12,345,782	6.75	9.07	443.03	433.14
MSFT.O	11,693,264	3.66	4.47	819.35	836.56
NKE	12,468,825	6.80	7.29	268.70	267.14
PFE	9,241,351	6.76	9.67	723.50	697.15
PG	11,994,510	5.23	6.60	328.38	320.47
T	10,264,903	5.55	7.58	728.94	692.95
TRV	2,996,344	10.75	11.01	306.39	226.32
UNH	10,636,275	6.91	8.55	219.01	247.45
UTX	13,241,326	6.82	8.11	203.60	183.32
V	9,857,067	5.96	6.63	154.49	140.85
VZ	15,537,730	4.64	6.74	508.78	533.18
WMT	12,542,368	6.72	10.28	316.14	311.69
XOM	12,680,327	5.81	8.46	409.22	449.72

Note: This table represents the statistic results (the mean, standard deviation, median) of calls and puts options' liquidity measures, which are separated in Panel A and B. The statistic results are based on the final dataset that contains 348,520,443 and 345,428,749 observations for call and put options, respectively. SPD refers to the proportional quote bid-ask spread. DPT refers to the quoted depth.

Table 5-2 Statistic results of determinant variables

	Call options				Put options			
Variables	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Return	-0.00071	0.03	-0.21	0.20	-0.0017	0.03	-0.27	0.21
Volt	0.04	0.02	0.000057	0.15	0.04	0.02	0.00	0.15
Vlm	13.99	7.98	4.88	243.19	13.20	6.44	4.18	157.40
Dur	824.76	473.50	68.46	2798.55	1141.54	677.31	66.09	4031.65
Tr	6.24	2.63	1.84	33.39	5.05	2.00	1.59	34.18

Note: This table represent the average value, standard deviation, minimum value, and maximum value of determinant variables. Return refers to return of quote during a 30-mintues interval. Volt refers to the option volatility calculated by the absolute value of quote return. Vlm refers to the trade volume. Dur refers to the trade duration that is estimated as how many seconds have elapsed between a transaction and its previous same-day transaction. Tr refers to the amount of trade transactions during a 30 minutes interval.

Table 5-3 Macroeconomic announcement

Category	Announcement name	Announcement time	Announcement frequency
Consumption	Consumer Credit	15:00	First week of each month
Investment	Wholesale Inventories MoM	10:00	Second week of each month
	Business Inventories	10:00	Third week of each month
	Construction Spending MoM	10:00	First week of each month
	Factory Orders	10:00	First week of each month
	Philadelphia Fed Business Outlook	10:00	Thursday at Third week of each month
Housing sector	New Home Sales	10:00	Fourth week of each month
	New Home Sales MoM	10:00	Fourth week of each month
	Existing Home Sales	10:00	Fourth week of each month
	Existing Home Sales MoM	10:00	Third or Fourth week of each month
	Pending Home Sales MoM	10:00	Fourth or Fifth week of each month
	Pending Home Sales NSA YoY	10:00	Fourth or Fifth week of each month
Forward looking	Conf. Board Consumer Confidence	10:00	Tuesday at the last week of each month
	Leading Index	10:00	Thursday at Third or Fourth week of each month
	Chicago Purchasing Manager	09:45	Last week of each month
	Minutes of FOMC Meeting	14:00	First week of January, April, October; Second week of February, May, July, November; and Third week of August
Government	Monthly Budget Statement	11:00	Second or Third week of each month

Note: all announcements are downloaded from Bloomberg and their times are Eastern Time.

Table 5-4 Expected effects of determinant variables

Variable	Description	Expected effects
Volt	$Volt_k$ refers to option quote volatility during the 30-minture interval *k*.	Positive
R_k^-	R_k^- refers to the average return of 30 tickers at 30-minture interval *k* when the average return is negative and zero otherwise	Positive
Vlm_k	Vlm_k refers to the logarithmic trading volume averaged across the 30-minture interval *k*.	Positive
Dur_k	Dur_k refers to the volume-weighted time between two same-day transactions averaged across the 30-minture interval *k*.	Negative
Tr_k	Tr_k refers to the total number of option trade transaction during the 30-minture interval *k*.	Positive

Table 5-5 Correlation between variables

Panel A: Call options

	CiL (spread)	CiL (depth)	Volt	R^{+}	R^{-}	PR^{+}	PR^{-}	Vlm	Dur	Tr
CiL (depth)	0.777									
Volt	0.343	0.159								
R^{+}	0.081	0.018	0.435							
R^{-}	-0.100	-0.057	-0.509	0.367						
PR^{+}	0.034	0.004	0.143	0.059	-0.031					
PR^{-}	-0.052	-0.010	-0.149	-0.068	0.046	0.371				
Vlm	-0.010	-0.043	0.048	-0.002	-0.011	0.034	-0.004			
Dur	-0.048	0.108	-0.235	-0.066	0.050	-0.063	0.079	0.587		
Tr	-0.094	-0.316	0.307	0.140	-0.086	0.093	-0.044	0.133	-0.362	
Ann	0.531	0.220	0.257	0.070	-0.066	0.031	-0.015	0.056	-0.118	0.174

Panel B: Put options

	CiL (spread)	CiL (depth)	Volt	R^{+}	R^{-}	PR^{+}	PR^{-}	Vlm	Dur	Tr
CiL (depth)	0.771									
Volt	0.314	0.153								
R^{+}	0.059	0.037	0.505							
R^{-}	-0.108	-0.033	-0.490	0.355						
PR^{+}	0.024	0.006	0.157	0.068	-0.066					
PR^{-}	-0.045	-0.003	-0.118	-0.005	0.062	0.357				
Vlm	-0.014	-0.096	0.094	0.015	-0.038	0.038	-0.005			
Dur	-0.098	0.056	-0.246	-0.086	0.055	-0.104	0.089	0.505		
Tr	-0.073	-0.295	0.316	0.203	-0.076	0.140	-0.033	0.190	-0.331	
Ann	0.553	0.229	0.241	0.066	-0.067	0.014	-0.024	0.061	-0.148	0.168

Note: This table represents the correlation results between variables of calls and puts options, which are separated in Panel A and B. CiL (spreads) and CiL (depth) refer to the commonality in quoted spreads and quoted depth. Volt refers to quote volatility calculated by the absolute value of option return during the interval *k*. R^{+} (R^{-}) refers to the market return when the return is positive (negative) and zero otherwise. PR^{+} (PR^{-}) refers to positive (negative) market returns at the interval *k*-1. Vlm refers to logarithmic trading volume. Dur refers to the volume-weighted time between two same-day transactions. Tr refers to the total number of option trade transactions during the interval *k*. ANN refers to the dummy variable of macroeconomic announcements at the interval *k*. It takes the value of zero if there are no macroeconomic announcements and one if there is an announcement.

Table 5-6 Principal component analysis for liquidity

Panel A

		Call			Put		
		Eigenvalue	Proportion	Cumulative proportion	Eigenvalue	Proportion	Cumulative proportion
Bid-ask spread	Factor 1	14.806	0.494	0.494	14.923	0.497	0.497
	Factor 2	0.842	0.028	0.522	0.837	0.028	0.525
	Factor 3	0.779	0.026	0.548	0.765	0.026	0.551

Panel B

		Call			Put		
		Eigenvalue	Proportion	Cumulative proportion	Eigenvalue	Proportion	Cumulative proportion
Depth	Factor 1	5.316	0.177	0.177	5.399	0.180	0.180
	Factor 2	1.155	0.038	0.216	1.146	0.038	0.218
	Factor 3	1.018	0.034	0.250	1.010	0.034	0.252

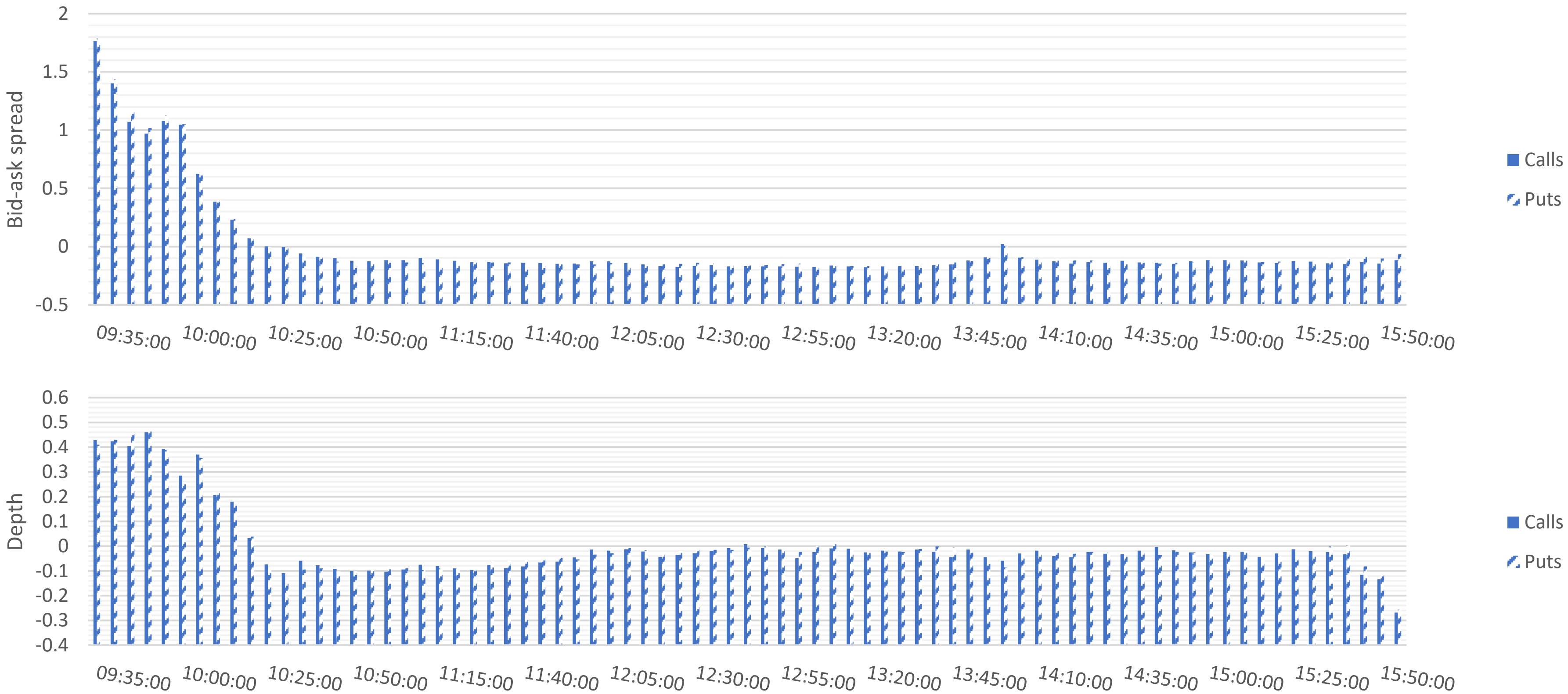
Note: Panel A and B represent the results of commonality in liquidity through applying Principal Component Analysis to the sample period. Bid-ask spread refers to quoted spread. Depth refers to the quantity of quoted depth. Both spread and depth per tickers are averaged across intra-daily 30-minutes intervals. All tickers are at-the money and have a maturity date between 7 to 90 days. Cumulative explained variance refers to cumulating the proportion of variance in liquidity by adding the proportion of each factor.

Table 5-7 Regression results

	CiL (SPD)		CiL (DPT)	
Variable	Call	Put	Call	Put
Volt (+)	4.89***	5.18***	3.06***	3.40***
R^{+}	-2.13***	-2.58***	-1.36***	-1.52***
R^{-} (+)	2.22***	2.40***	1.36***	1.65***
PR^{+}	-0.075	0.0014	-0.044	0.11*
PR^{-}	-0.0056	0.056	0.068	0.073
Vlm (+)	-0.022***	-0.022***	-0.033***	-0.046***
Dur (-)	0.00000070***	0.00000033**	0.0000015***	0.0000012***
Tr (+)	-.013***	-0.016***	-0.014***	-0.017***
Ann	0.28***	0.30***	0.098***	0.10***
Obs.	7308	7067	7320	7075
R-square	0.43	0.43	0.25	0.24

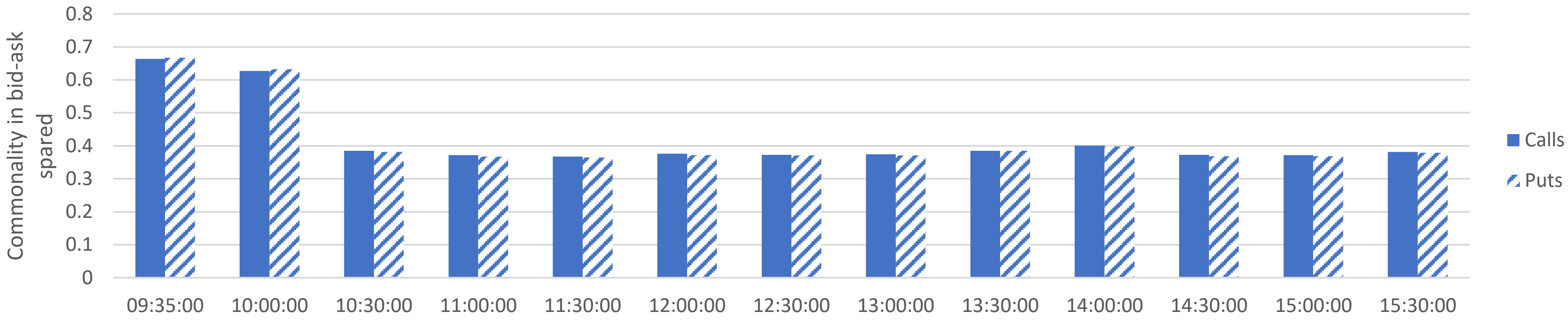
Note: ***, **, * refer to the significant levels at 1%, 5%, and 10%. This table represents the regression results of determinants on liquidity commonality, which are separated in call and put options. CiL (SPD) and CiL (DPT) refer to the commonality in quoted spread and in quoted depth, respectively. Volt refers to quote volatility calculated by the absolute value of option return during the interval *k*. R^{+} (R^{-}) refers to the market return when the return is positive (negative) and zero otherwise. PR^{+} (PR^{-}) refers to positive (negative) market returns at the interval *k*-1. Vlm refers to logarithmic trading volume. Dur refers to the volume-weighted time between two same-day transactions. Tr refers to the total number of option trade transaction during the interval *k*. ANN refers to the dummy variable of macroeconomic announcements at the interval *k*. It takes the value of zero if there are no macroeconomic announcements and one if there is an announcement.

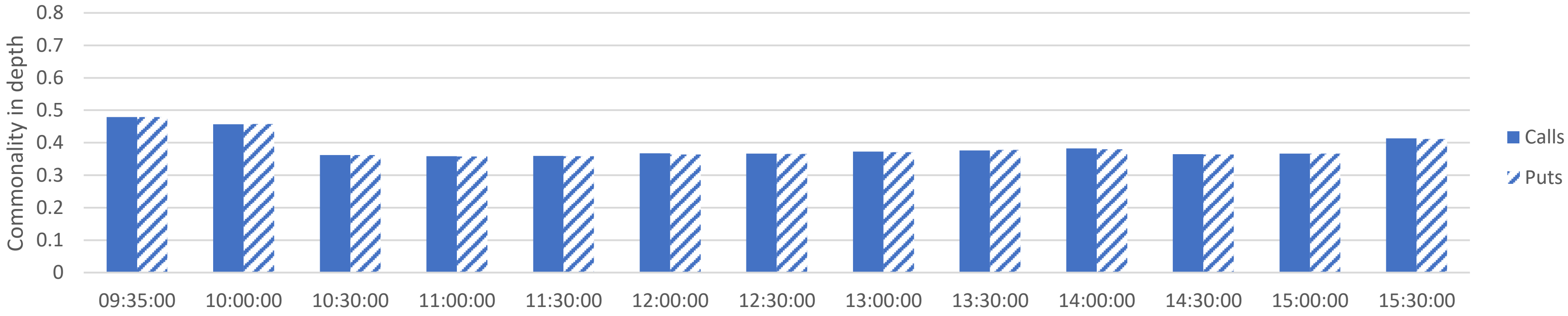
Figure 5-1 Intraday time series of liquidity



Note: This figure contains two graphs that represent the intraday time series distribution of liquidity measures for calls and puts. The first graph displays the average bid-ask spread for calls and puts per 5-min interval, and the second graph displays the average depth for calls and puts per 5-min interval. The quoted spread and quoted depth are standardised by subtracting the daily mean and dividing with the daily standard deviation. Since the trading hour of CBOE is from 09:30 to 16: 00, there are 78 successive 5-minute intervals at the trading day. In order to avoid the overnight effects, the time interval between 09:30 and 09:35 has been removed. The estimations are based on ATM options expired between 8 and 90 days.

Figure 5-2 Intraday time-series of commonality in liquidity





Notes: This figure contains two graphs that represent the intraday time series distribution of commonality in quoted spreads and quoted depth for calls and puts. The first graph displays the commonality in spreads for calls and puts per 30-min interval, and the second graph displays the commonality in depth for calls and puts per 30-min interval. Commonality in liquidity refers to the proportion of variance in liquidity explained by the first three component factors. Each component factor is extracted from applying Principal Component Analysis to spreads or depth on each intraday 30-minture interval, such as 09:35:00-10:00:00, 10:00:00-10:30:00, and so on.

Chapter 6: Predictive information of options order flows

6.1 Introduction

What is the information role of equity options in underlying stock markets? On a theoretical level, options would be redundant because options trading conveys no additional information to market participants. On a practical level, in contrast, options trading can predict the future price movement of underlying stocks because of market incompleteness. This predictive power may be driven by informed traders (Pan and Poteshman, 2006; Ryu and Yang, 2018). Informed traders are attracted by options markets rather than by the underlying stock market due to volatility trading, higher liquidity, greater implicit leverage, and lower transaction costs (Black, 1975; Back, 1993; Mayhew *et al.*, 1995; Blau *et al.*, 2014). While Easley *et al.* (1998b), Chakravarty *et al.* (2004), and Chen *et al.* (2005) found empirical evidence that informed traders are willing to trade in both stock and options markets. Thus, the stock and options markets are in the pooling equilibrium. In particular, Easley *et al.* (1998b) found that the participation of informed traders in the options market is influenced endogenously within an equilibrium framework. The researchers also propose a separating equilibrium where informed traders are active in the stock market. The findings of Chan *et al.* (2002) and Chang *et al.* (2009) are consistent with this separating equilibrium, and Vijh (1990) directly shows that options trades convey less information compared with stock trades.

Motivated by the literature, we directly test whether options transactions contain information contents and disseminate the information to stock markets. In particular, with the sample of the components of the Dow Jones Industrial Average Index (DJI) and their traded options on the Chicago Board Options Exchange (CBOE), we constructed three sets of information measures by a proxy of predictive information based on options volume and options trade duration. Using the information measures, we examined whether the information content of the options order flow can predict the contemporaneous and future underlying stock price movements.

Our study contributes to the literature in three ways: The first contribution of this paper refers to the employing of the Smooth Transition Autoregressive Conditional Weighted Duration (STM-ACWD) model and the volume-weighted duration to measure the information content in the options order flow. Kalaitzoglou and Ibrahim (2015) used this method to identify the informational role of trade duration and volume simultaneously in a futures market. Pan and Poteshman (2006) and Chakravarty *et al.* (2012) confirmed the informational role of transactions, but reached no consensus regarding how to detect the information contained in the trade flow. Easley *et al.* (1997a) supported the information contained in the trade size and argues that informed investors prefer to trade in large volumes in order to maximise trading profits, while Pan and Poteshman (2006) and Chakravarty *et al.* (2012) show that informed investors shift to small-size trades to better hide their information. Dufour and Engle (2000), Xu *et al.* (2006) and Furfine (2007) focus on the role of time between trades and take the view that short durations are associated with more information. To the best of our knowledge, existing studies observe a relationship between the information contained in the duration and volume (Manganelli, 2005; Wong *et al.*, 2009) or compare the information contained in them (Eom and Hahn, 2005), while no studies are extended to combine volume and duration as a proxy for capturing predictive information in options markets. Since both duration and volume can convey information, we expect that a proxy based on them should also contain predictability about future stock price movements.

The second contribution of this paper is to identify the percentage of options trades that contain information in the CBOE equity options market. As mentioned previously, there are mixed results about where informed traders trade. For example, Chang et al. (2009) report that informed traders are only active in stock markets, while Hu (2014) found informed trading in the options market that drives the return predictability from options order flow. With STM-ACWD model and volume-weighted duration, we classify trades in CBOE into trades with and without information. In particular, a set of information measures is created to directly show the percentage of options traded with information in each trading day. The presence of informed trading in the options market not only has important practical implications but is also a necessary condition for testing whether the options order flow can predict the underlying stock price.

The third contribution this paper makes is its examining the predictive ability of the options order flow via the new proxy of information. Although Pan and Poteshman (2006), Hu (2014), Ryu and Yang (2018) found the empirical evidence to prove the predictive power of the options order flow, Chan *et al.* (2002), Schlag and Stoll (2005) and Tsai *et al.* (2015) report different results. These mixed results may be driven by the proxies for information content in the order flow used by these studies. For example, Tsai *et al.* (2015) did not find the predictive information on options volume imbalance but in options quote changes. Considering the variables relating to the trading volume is the general approach employed for examining the predictive power of options trade flow (Pan and Poteshman, 2006; Johnson and So, 2012; Hu, 2014). For example, Pan and Poteshman (2006) show that the signed option volume can predict future stock prices. However, very little effort has been made in the previous studies to examine the price predictability stems from variables relating to options transaction duration. Hence, we employed the approach of Kalaitzoglou and Ibrahim (2015) to capture information in the options order flow and examined the predictability of captured information.

Empirical findings of our study are summarised as follows: First, with the STM-ACWD model, we provide supportive evidence that the volume-weighted duration can be used as the proxy to identify whether a trade contains information. According to the estimation results of the STM-ACWD model, we report that the trading activity would be accelerated when observing large and fast trading and the activity would be prolonged when observing small and slow trading. In particular, trades seem to be accelerated if they contain information, and prolonged if they do not contain information (Engle, 2000; Furfine, 2007; Kalaitzoglou and Ibrahim, 2015).

We use the volume-weighted duration, the estimation results of STM-ACWD model, and trade direction to create three sets of information measures. These information measures successfully capture information involved in the options order flow because we find that they can influence the contemporaneous and future stock price movements. In particular, the percentage of trades with information is linked to stock volatility and the directional information measures is associated with stock returns. In addition, we report that the levels of information contained in each information measures are different. Additionally, as compared with calls information measures, puts information measures contain more predictability about future stock return but less predictability about future stock volatility. Although the levels of information vary across measures and option types, our findings provide supportive evidence that informed traders are not only active in stock markets. Since they are trading in options markets, some option trades contain information.

With the empirical results, this chapter provides implications for market practitioners. Our findings may offer supports to investors, especially for uninformed traders. In general, investors could obtain trading profits in stock markets by using the predictability of options order flow. For uninformed traders, moreover, they can obtain some clues from option order flows. In particular, they may avoid trading options during high intensity period in order to minimise their losses to informed traders.

The rest of the empirical chapter is organised as follows: Section 2 reviews the empirical literature on information content of the options order flow. Section 3 describes the data and discusses the empirical methodology. Section 4 presents and discusses the empirical results, and Section 5 concludes the empirical chapter.

6.2 Literature review

In this section, we review empirical studies regarding how the options order flow provides information about the price movement of the underlying stocks. As we explained in the introduction, options markets are used by informed traders as an alternative trading venue. Hence, options transactions contain important information that can make a contribution towards predicting the price movements of underlying stocks. Since previous studies used different methods and proxies to capture the information, we start by discussing studies that employ variables based on the trade volume and then analyse those that capture the information by trade durations.

6.2.1 Information on trade volume

Focusing on the private information involved in options order flow, Easley *et al.* (1998b) developed the first research to investigate the predictive role of options volume in future stock price movements. As we mentioned, they propose two equilibrium frameworks in which informed traders tend to trade in the stock or options markets. There is a pooling equilibrium when informed traders are active in both markets due to the sufficient high leverage and liquidity of options. Additionally, options may be more attractive to informed traders because the existence of multiple option contracts leads to some issues around understanding to uninformed traders. In this context, the options order flow can include information about the fundamental values of the underlying stocks. However, Chakravarty *et al.* (2004) provide evidence in the opposite direction, finding that the information share of options in CBOE is on average, around 17%. Holowczak *et al.* (2006) confirm a decrease trend on the information share of options during their sample period.

In order to explore the information role of options order flow, Easley *et al.* (1998b) developed a multimarket sequential trade model to test a sample of 50 firms over 44 trading days in 1990. They confirmed the predictive power of option volumes which are indicative of information-based trading. However, this predictive power comes from information-based option volumes rather than from standard trading volumes. In particular, the options order flow can predict stock price changes when the volumes are sorted on the basis of good and bad information. Without directional information, standard option volumes do not lead to stock price movements. Hu (2014) proposed an option order imbalance based on risk exposure to the underlying stock price and found that it could predict the next-day stock returns. However, this predictive ability disappears on longer horizons. Chan *et al.* (2002) also did not find predictive power from signed option trade volume even if the proxy of information content is constructed by trading volume and directional information. In particular, these authors studied the interrelation of order flows and price movements for 60 stocks traded on the NYSE and their CBOE traded options. They found only the strong predictive power of stock net trade volume on option returns and concluded that stock trades contain more information than options trades. With a dataset from Eurex, Schlag and Stoll (2005) showed that signed options volume is significantly informative about the contemporaneous underlying asset price but becomes silent for future price changes. Tsai *et al.* (2015) also did not find the predictive information from volume imbalances in VIX options but confirmed the predictive ability of quote changes on the future price movement of VIX index.

To further explore the information in option trades, Chen *et al.* (2005) constructed trading value as a new measure of information content. Like previous studies, they confirm that the market expectations cannot be reflected by option volume alone. But this new measure uses options premiums to reflect good and bad news information. With all firms included in the S&P500 from 1995 to 2002, Chen *et al.* (2005) found that the call-put trading value ratios can predict stock returns. However, this predictive power is only found in OTM options. This may imply that informed traders prefer to trade OTM options because they have higher liquidity and lower premiums. Moreover, their findings are also consistent with the pooling equilibrium assumption as showing that stock returns can predict option trading value ratios. Pan and Poteshman (2006) used the volumes of put and call contracts to reflect the predictive information and thereby construct the put-call ratios that is widely used by following studies. Using all CBOE listed options between 1990 and 2001, they confirm the predict power of the put-call ratios on future stock prices. After deconstructing the option volume into public and non-public components, they found that the predictability is mainly driven by valuable non-public information. Moreover, the predictive power of this information measure varies across option moneyness as observing the greatest level of predictive power from deep out-of-the-money (OTM) options.

Chang *et al.* (2009) and Hsieh and He (2014) used the put-call ratio as an information variable to examine the predictive ability of options traded on the Taiwan Futures Exchange. The researchers not only observe the predictive power varying across options moneyness but also identify foreign investors as informed traders. In particular, OTM options traded by foreign investors have a significant predictive ability on the spot index changes. However, this information variable may not efficiently capture the influence of order sizes and the position changes of option traders (Ryu and Yang, 2018). In addition, Johnson and So (2012) argued that the ratio of call to put volume is not an indication of the sign of private information when there is no information about trade directions.

To address these concerns, Roll *et al.* (2010) developed a ratio of option trading volume to stock trading volume (O/S) to explore how investors use derivatives markets to execute informed trades. Using this ratio, they detected informed trading in the options market by finding that the variation in O/S ratio may be driven by informed trades. Johnson and So (2012) provided further evidence by examining the predictive ability of the O/S ratio. They found that the O/S ratio was a negative signal for the future returns of underlying stocks. And, in particular, that the O/S ratio can predict stock returns over a one-week horizon. This negative relationship may be driven by the short-sale costs in equity markets. Hence, options markets become an attractive venture for traders with negative news.

Blau *et al.* (2014), Ge *et al.* (2016), and Ryu and Yang (2018) further investigated the predictive power of O/S ratio. Blau *et al.* (2014) directly compared the level of information contained in the put-call ratio and the O/S ratio. They found that the put-call ratio shows greater predictive power than O/S ratio at the daily level. In particular, both signed and unsigned put-call ratios have predictive power. In terms of the weekly and monthly level, O/S ratios contain more predictability about future stock returns. Ge *et al.* (2016) only focus on O/S ratio and believe that the position changes of options investors many influence the types of information reflected in calls and puts. Hence, they disaggregated the information into eight signed measures by option positions and found that all these measures predicted weekly stock returns. Ryu and Yang (2018) pursued the investors' position and decomposed O/S into call and put O/S ratios. They showed that call O/S ratios have better predictive power than put O/S ratios. In particular, they found that the call O/S ratios from open-buy [19] trades can significantly predict the next-day underlying spot returns. Moreover, regarding O/S ratio, Bergsma *et al.* (2019) developed a composite option trading score that shows stock return predictability at the intraday level.

6.2.2 Information on trade duration

Apart from put-call ratio and O/S ratio, other researchers (e.g Hu, 2014; Tsai et al., 2015) also developed other information measures by which to examine the relationship between options trade and underlying stock returns. However, these measures are generally based on trade direction and trade volume. Dufour and Engle (2000) and Engle (2000) showed that the duration between consecutive trades contains information about future financial returns. Dufour and Engle (2000) modelled the trading intensity using the autoregressive conditional duration (ACD) model proposed by Engle and Russell (1998). With the bivariate vector autoregressive (VAR) model, they find that the trade duration can influence the price formation process. 144 stocks on the NYSE, showed negative impacts regarding the time duration on the price of trades and the speed with which prices adjust to trade-related information. Engle (2000) also used the ACD model to estimate the conditional duration but applied a GARCH model to study the effects of the duration between trades. This empirical study shows that long durations are negatively associated with stock returns and variances.

[19] Open-buy refers to options bought to establish new long positions.

Although above literature confirms that duration conveys information, there are arguments about the effects of duration on price formation. Similar to Dufour and Engle (2000), Spierdijk (2004), Xu *et al.* (2006) and Furfine (2007) find a negative relationship between the price impact of a trade and its duration by employing extended VAR models. Their results suggest that a greater risk of informed trading is associated with a shorter duration between trades. This conclusion corresponds with Easley and O'hara (1992) who indicated that informed traders trade in a hurry in order to obtain profits from their timely information. Manganelli (2005) used both ACD and GARCH models to provide further evidence that high trading intensity reflects more information. However, Hafner (2005), Wong *et al.* (2009) and Beltran-Lopez *et al.* (2012) reported an opposite result, namely that a long duration between trades leads to higher price impacts. In particular, both Hafner (2005) and Wong *et al.* (2009) used a log ACD model to estimate the conditional stock trade durations. Hafner (2005) found that volume keeps its return predictability for longer durations. Wong *et al.* (2009) show that slower arriving trades move prices more than faster arriving trades. They explained that price adjustment takes a longer time to incorporate information when there is significant new information arrival.

These mixed results may be a result of the limitations of the ACD model. Kalaitzoglou and Ibrahim (2013) showed that duration alone cannot efficiently describe the intensity of trading in all situations. Hence, they suggested that the transaction time should be modelled with other variables. Taking this into consideration, the ACWD model is used to model volumes and duration that are sufficient at capturing information involved in transactions. Kalaitzoglou and Ibrahim (2013) found that ACWD significantly outperforms the two existing models – the Autoregressive Conditional Multinomial (ACM)-ACD model and the Autoregressive Conditional

Multinomial Duration (ACMD) model. In addition, the states of volume should be pre-specified under the ACM-ACD and ACMD models. The possibility of incorrectly specifying the states could be a potential source of error, while the volume and duration are combined under one variable within the ACWD model that avoids this potential error. Additionally, ACWD can identify relative levels of trading intensity, while the identifying function cannot be achieved via the ACM-ACD and ACMD models. Hence, Kalaitzoglou and Ibrahim (2013) identified different regimes of trading intensity in the European carbon futures market. Each regime can reflect the level of information contained in each transaction.

Previous studies not only offer mixed results on the role of trade duration but also include limited investigation regarding the effect of trade duration in the multiple-market setting. In this context, Collver (2009) used a sample of 40 NYSE-listed firms to study the relationship between stock and options market activity. He showed that option trade durations influence spreads and depths in the stock market. Furthermore, Cartea and Meyer-Brandis (2010) investigated how the trade duration of underlying stocks influences option prices. Their empirical results showed that the trade duration can be used to support the calculation of option prices. These studies helped to develop the assumption that the trade duration contains price related information, because it enables us to create a relationship between option trade duration and underlying stock price dynamics.

6.3 Methodology

This section discusses the sample and the data clearing procedure. Furthermore, the ACWD model will be described. With this model, we use volume-weighted duration to capture the information contained in transactions and thereby examine its return predictability.

6.3.1 Sample

The sample for this chapter includes the high-frequency trading data of the components of the Dow Jones Industrial Average Index (DJI) from 03 Jan 2012 to 30 Jun 2014. Apart from the filters mentioned in the Data Cleaning, we further diurnally adjust durations to address intraday seasonality as suggested by Kalaitzoglou and Ibrahim (2015). In addition, we remove options contracts with expiration times of within 7 days or over 37 days in order to eliminate the effect of option expiration. Deep-in-the-money (DITM) and deep-out-of-the-money (DOTM) options are also excluded. Moneyness is calculated by S/K, where *K* refers to the strike price of the options contracts and *S* refers to the underlying concurrent mid-quote price. DITM options are defined as the observations with moneyness ≥ 1.1 and DOTM is defined as the observations with moneyness ≤ 0.9. Our final sample used in this chapter consists of 8 million and 6 million observations for call and put options, respectively.

6.3.2 Autoregressive Conditional Weighted Duration

Kalaitzoglou and Ibrahim (2015) used the STM-ACWD model to measure non-price trading information and recognise the informational content of each transaction. We follow Kalaitzoglou and Ibrahim (2015) approach to modelling the informational content of options order flow. As mentioned, the trade duration cannot efficiently reflect the trading intensity. Hence, they shift from event (transaction) time to events defined by a unit change in associated variables of interest. These variables of interest are also called as marks which could be trade price or trade volume. Following

Kalaitzoglou and Ibrahim (2015), we model transaction time with the option trade volume m_i.

---------------Insert Table 5-1----------------

Table 5-1 reports the average time duration between trades and the average trading volume per contract of 30 components of DJI. Call options have a shorter trade duration than the put options, while not all call options represent a high volume than the puts counterparty. Additionally, the trading duration and trading volume vary significantly across the option series.

Since we focus on a unit quantity of associated marks, the first step is time rescaling. Let $\{t_0,\ t_1, \dots ,\ t_n, \dots\}$ with $t_0 < t_1 < \cdots < t_n < \cdots, t \in T$, as the sequences of arrival times of a trade of an option contract A. *N*(*t*) is the counting function that counts the number of events occurred by time *t* and $m_i \sim (\overline{m},\ \sigma_m^2), m \in M$ denotes an associated mark. It further assumes that the realisation time and the mark *m* are conditional on past history $F_i = (\check{t}_i,\ \check{m}_j)$, where $\check{t}_i,\ \check{m}_j$ is the history of *t* and *m* up to event *i*, and they formulate a temporal market point process $\{(\check{t}_i,\ \check{m}_j)\}$ on $\{\Omega, F_i, P\}$ where $\mathrm{P}: F \to [0, 1]$ is a mapping of F on [0, 1]. The conditional joint density function of $\{(\check{t}_i,\ \check{m}_j)\}$ is $(t_i,\ m_i)|F_{i-1} \sim f(t_i,\ m_i|\check{t}_{i-1},\ \check{m}_{j-1}; c)$, where c is a vector of parameters. The marginal density functions could be used to derive the conditional expectations of arrival time and the associated market: $\mathrm{E}[t_i|F_{i-1}] = \int t \cdot f(t_i|F_{i-1})dt$ and $\mathrm{E}[m_i|F_{i-1}] = \int m \cdot f(m_i|F_{i-1})dm$.

Furthermore, let $d_i = t_i - t_{i-1}$ denote the (raw) duration of transaction *i*, measured by the time (*t*) elapsed since the preceding mark at $i-1$. Let x_i denote the diurnally adjusted duration. The normalised mark $\mathrm{g}(m_i) = K(u_i)$ is employed as a scaling factor for the duration realisation process to formulate a new weighted duration variable z_i, with $\theta_i = E(z_i|F_{i-1})$. In this context, the temporal marked point process $(x_i,\ m_i)$ is transformed into a temporal point process $(\mathrm{Y}\{x_i,\ m_i\})$, with density function $(\mathrm{Y}\{x_i,\ m_i\}|F_{i-1}) = (z_i|F_{i-1}) \sim f(z_i|\check{z}_{i-1}; \varphi)$, where $\check{z}_{i-1}$ is the history of z up to time $i-1$ and φ is a vector of parameter. To give more details, with the following kernel density, it could define z_i as:

$$z_i = x_i * K(u_i) \tag{1}$$

$$K(u_i) = exp(-u_i/2) \tag{2}$$

$$u_i = (m_i - \overline{m})/\sigma_m) \tag{3}$$

This transformed variable z_i is a measure of trading intensity which is a proxy for predictive information. It can measure the waiting time for a single contract to be traded. The autoregressive conditional weighted z_i could be modelled as:

$$z_i = \theta_i \varepsilon_i \tag{4}$$

$$\theta_i = \theta(z_{i-1}, \dots, z_1; \varphi_1) \tag{5}$$

$$\varepsilon_i|J_i \sim i.i.d., \text{with density } f(\varepsilon_i|J_i; \varphi_2) \text{ and } E(\varepsilon_i|J_i; \varphi_2) = \mathrm{E}(\varepsilon_i) = 1 \tag{6}$$

where z_i is the volume-weighted duration which is estimated in Eq. (1), θ_i is the expected value of trading intensity, ε_i is the error term, J_i is an economically relevant threshold variable, and φ is parameters. With this model, empirical specifications for the conditional mean and the distribution can efficiently reflect the stylised facts of the investigated market. According to Kalaitzoglou and Ibrahim (2013) and Kalaitzoglou and Ibrahim (2015), the conditional mean is specified as:

$$\theta_i = \omega + \sum_{j=1}^{n} \alpha_j z_{i-j} + \sum_{j=1}^{q} \beta_j \theta_{i-j} \tag{7}$$

Where θ_i is the expected value of z_i, ε_i is given by the ratio z_i/θ_i that is the so-called standardised trading intensity. The conditional density function is a smooth transition mixture of Weibull:

$$f(z_i|J_i;\tau) = (h(J_i\colon\tau)z_i)[z_i\Gamma(1+1/h(J_i\colon\tau))/\theta_i]^{h(J_i\colon\tau)}\exp\left(-[z_i\Gamma(1+1/h(J_i\colon\tau))/\theta_i]^{h(J_i\colon\tau)}\right) \quad (8)$$

where

$$h(J_i\colon\tau) = \gamma_1 * \left(1 - G_1(J_i\colon g_1, j_1)\right) + \gamma_2 * G_1(J_i\colon g_1, j_1) \quad (9)$$

$$G_1(z_i\colon g_1, j_1) = (1 + \exp\{-g_k * (J_i - j_1)\})^{-1} \quad (10)$$

The shape parameter, $h(J_i\colon\tau)$, of the mixture of Weibull distribution is a function, h, of the threshold variable, J_i, which we represent by $\log(\frac{1}{z_i})$, and a vector of parameter coefficients $\tau = (\gamma_1, \gamma_2, g_1, j_1)$, where γ_1 and γ_2 are the shape parameters of the Weibull distribution in two regimes of different trade intensity determined by the threshold value j_1 of the threshold variable J_i, and g_1 is the smoothness parameter between regimes. For each trade duration, the overall shape parameter of the Weibull distribution $h(J_i = z_{i-1}\colon\tau)$ is the weighted average of γ_1 and γ_2 that is c. The weights are determined by a smooth transition function, G_1. Hence, the conditional intensity (hazard function) of weighted durations is revised after each transition that provides a measure of information content in the options order flow. In particular, the values of shape parameters can directly explore the shape of the hazard rate and thereby classify trades into the trades with predictive information and those without predictive information.

6.3.3 Predicative power of options order flow

--------------Insert Table 5-2---------------

With the STM-ACWD model, we use volume-weighted duration as a proxy of predictive information. However, we do not directly use this proxy to measure the predictive information in the options order flow. According to Kalaitzoglou and Ibrahim (2015), the informational content of each options trade is identified by estimation results of the STM-ACWD model. Therefore, we develop a group of dummy variables to reflect the predictability of an options trade. These dummy variables are created by the estimation results for the STM-ACWD model as shown in Table 5-2. For each options series, this table shows the estimated model parameters (ω, α, and β) and the vector of shape parameter coefficients ($\gamma_1, \gamma_2, g_1, j_1$). All results are statistically significant at 1% level. In addition, the sum of α and β in all option series are close to one and that implies stationarity with high persistence, underlying autoregressive dynamics. In this context, intensity shocks have prolonged subsequent effects. This follows the view that a shock cannot be adjusted immediately but persists over time (Kalaitzoglou and Ibrahim, 2015).

The vector of shape parameter coefficients can be used to identify different levels of trading intensity. Particularly, the distribution of error can be explained by the threshold value j_1 as a mixture of two Weibull distributions which are defined as high and low levels of trade intensity. $J_i > j_1$ refers to high trade intensity levels in the form of low weighted duration per contract, and $J_i < j_1$ refers to low trade intensity in the form of high weighted duration per contract. We construct the information dummy variable based on these interpretations. Moreover, the two levels of trade intensity are connected by a different Weibull distribution with distinct shape parameter γ_1 and γ_2 which represents low and high intensity level. Estimates of γ_1 for low levels of trading intensity are consistently close to one (e.g. 1.033 and 1.001 for AXP call and put options series, respectively) that implies a flat hazard function for the arrival of subsequent option contracts. Under this regime, we can observe a slow and/or a small transaction. Hence, we can expect that the arrival rate of a single contract does not vary over time which is one characteristic of trade without information. Regarding γ_2, the estimates are always lower than one (e.g. 0.222 and 0.215 for AXP call and put options series, respectively) that implies a decreasing hazard function. In this regime, the arrival rate of contracts accelerates because a large and/or fast trade is observed in the market. This suggests that the probability of trading a single contract decreases over time.

With the estimation results, we develop a group of dummy variables to identify the predictability of an options trade. Since the information content can be reflected by comparing threshold value j_1 with the threshold variable J_i, we contracture the information dummy variable I^1 and I^2. In addition, γ_1 and γ_2 could reflect the levels of trading intensity that relate to the levels of information contained in the order flow. According to Engle (2000) and Furfine (2007), large and fast transactions contain more information. Hence, we further develop the information dummy variable I^3 and I^4.

1. $I^1_{k,i,t}$ equals one if $J_{k,i,t} > j_1$, otherwise zero;
2. $I^2_{k,i,t}$ equals one if $J_{k,i,t} > j_1 + 3 * SE_{j_1}$, otherwise zero;
3. $I^3_{k,i,t}$ equals one if $h_{k,i,t}$ is lower than 30th percentile of the shape parameter $h_{i,t}$, otherwise zero[20];
4. $I^4_{k,i,t}$ equals one if $h_{k,i,t} < \gamma_2 + 3 * SE_{\gamma_2}$, otherwise zero;

where $I_{i,t,k}$ refers to information dummy variables for *k*th trade of options written on stock *i* on date *t*. $J_{k,i,t}$ refers to the threshold value for *k*th trade of options

[20] As aforementioned, γ_2 is lower than 1 and represents high level of trading intensity. Since $h_{k,i,t}$ is the weighted average of γ_1 and γ_2, the low level of $h_{k,i,t}$ could represent high level of trading intensity. In this context, we define the low level of $h_{k,i,t}$ as the situation that the value is lower than its 30th percentile.

written on stock *i* on date *t*, $h_{k,i,t}$ refers to the overall shape parameter of the Weibull distribution is the weighted average of γ_1 and γ_2.

These information dummy variables can identify whether an options trade conveys predictive information on future stock price movement. Based on this, we create three sets of informational measures that will be used in empirical analyses. Since we only focus on predictability at the daily level, the first set of informational measures uses the number of trades per day to create the percentage of trades with predictive information in a trading day (PER).

$$PER_{i,t}^{m} = \frac{\sum_{k=1}^{n} I_{k,i,t}^{m}}{Num\_trading} \tag{11}$$

We further create a set of measures based on the directional information content of options order flow because previous studies (e.g. Easley et al., 1998; Johnson and So, 2012; Ryu and Yang, 2018) show the importance of options trade direction in playing an important role in predicting future stock returns. According to Easley *et al.* (1998b), the different types of option trade contains differential information about stock price movements. Since each option trade includes both a writer and a buyer, Blasco *et al.* (2010) suggested that the buyer and seller are essentially "averaged" when investigating overall volume. Hence, the overall volume cannot recognise the active side of the trade. As they failed to find the predictive power of overall option volume, Easley *et al.* (1998b) indicated the importance of recognising the directional value implications for different types of option trades. The value of directional information content was further confirmed by Pan and Poteshman (2006). Therefore, the second set of measures is the directional information content of options trading (X) that includes trade directions in order to capture more predictive information. This set of measures is defined as the sum of directional information dummy variables.

$$X_{i,t}^{m} = \sum_{k=1}^{n} Dir_{k,i,t} * I_{k,i,t}^{m} \tag{12}$$

where $Dir_{i,t,k}$ refers to the dummy variable of trade direction for *k*th trade of options written on stock *i* on date *t*. Similar to Hu (2014), this research identified trade direction by adopting the approach of Lee and Ready (1991) which contains two steps: The first step is to compare the trade price with the prevailing quote midpoint. The transaction will be assigned as buyer (seller) initiated if the trade price is higher (lower) than the mid-quote price. For those trades executed at the quote midpoint, the second step is to assign their trade directions based on the most recent trade. If the previous trade is defined as buyer-initiated trade, the current trade is also defined by buyer-initiated trade. Based on these steps, $Dir_{i,t,k}$ equals one (negative one) if the trade price is higher (lower) than the most recent mid-quote price that is classified as a buyer-initiated (seller-initiated) trade.

Since $h_{k,i,t}$ is the overall shape parameter of the Weibull distribution which is the weighted average of γ_1 and γ_2, we create the last information measure as the value of directional overall shape parameter, DSP.

$$DSP_{i,t} = \sum_{k=1}^{n} Dir_{k,i,t} * \frac{1}{h_{k,i,t}} \tag{13}$$

After creating the informational measures, our empirical test was designed to explore the predictive ability of these information measures on future stock price movements. To achieve this aim, we creative two variables, return (*R*) and volatility (*V*), to reflect the price changes and its magnitude, respectively. The daily return for stock *i* on date *t* is the is the logarithmic change in the successive daily closing prices

$$R_{i,t} = \ln\left(\frac{p_{i,t}^{close}}{p_{i,t-1}^{close}}\right) \tag{14}$$

The volatility for stock *i* on date *t* is calculated by the absolute value of the daily return.

$$V_{i,t} = \left|R_{i,t}\right| \tag{15}$$

Motivated to a large extent by the empirical model of Easley *et al.* (1998b), Pan and Poteshman (2006), and Hu (2014), we investigate the information of options order flow around price movement of underlying stocks through the following models. We separate calls and puts when estimating their predictability. Additionally, we also include the asset and time fixed effects which are not shown in the following models.

$$V_{i,t+\tau} = \alpha + \beta_1 PER_{i,t}^{m} + \varepsilon_{i,t+\tau}, \tau = 0 \ to \ 5 \tag{16}$$

$$R_{i,t+\tau} = \alpha + \beta_1 IM_{i,t} + \varepsilon_{i,t+\tau}, \tau = 0 \ to \ 5 \tag{17}$$

$$R_{i,t+\tau} = \alpha + \beta_1 CIM_{i,t} + \beta_2 PIM_{i,t} + \varepsilon_{i,t+\tau}, \tau = 0 \ to \ 5 \tag{18}$$

where $R_{i,t+\tau}$ refers to the stock return (close-to-close return and overnight return) for stock *i* on date *t*; $V_{i,t}$ refers to the volatility for stock *i* on date *t*; $PER_{i,t}^{m}$ refers to the percentage of trades with predictive information in a trading day; $IM_{i,t}$ refers to the information measures ($X_{i,t}^{m}$, and $DSP_{i,t}$) of options written on stock *i* on date *t*; $CIM_{i,t}$ and $PIM_{i,t}$ refer to the information measures of call and put options written on stock *i* on date *t*.

In this study, we create three sets of informational measures based on the estimation results of the STM-ACWD model. With these measures, we can examine the predictive ability of call and put options trades, respectively.

6.4 Empirical results

This section will demonstrate the empirical findings that examine the predictive ability of options order flow. Before investigating this ability, we provide the descriptive statistics of information measures. To estimate the predictive ability, this section will show the effects of information measures on underlying stock volatility and return, respectively.

6.4.1 Summary statistics

--------------Insert Table 5-3---------------

Table 5-3 shows the percentage of trades with predictive information identified by different information measures. We can find that the percentage vary significantly between measures. PER^1 and PER^2 show that over 50% of transactions may contain predictive information, while PER^3 and PER^4 show lower percentages. In particular, PER^1 and PER^2 for all calls and puts are around 56% and 57%, respectively. For PER^3 and PER^4, the proportion for all calls and puts are reduced to around 25% and 11%, respectively. These results are driven by the different criteria used to create information measures. PER^1 and PER^2 are based on the comparison between volume-weighted duration and the threshold value j_1, while PER^3 and PER^4 are created entirely by the estimated model parameters. From the theoretical concerns, PER^1 and PER^2 are ideal because they purely follow the principle of the STM-ACWD model. In this context, we still include PER^3 and PER^4 as robustness tests.

--------------Insert Table 5-4---------------

Table 5-4 shows the descriptive statistic of directional information measures. The mean value of all measures is relatively low (below 0.01) and this implies a balance between buyer- and seller-initiated trades during the sample period. Regarding the positive results for all options series, they are consistent with Vijh (1990) that option investors are more likely to be buyers than sellers. Moreover, these directional information measures for call options are higher than for put options which suggests more buyer-initiated trade for call options. Since I^1 and I^2 identify more trades as trading with predictive information than the other two dummy variables (I^3 and I^4), the mean value of the directional information measures X^1 and X^2 for calls are 0.0098 and 0.0099, respectively and are higher than the mean values of X^3 (0.0066) and X^4 (0.0045). The mean values of these directional information measures for puts are lower than the values for calls but represent a similar tendency. Regarding DSP, the mean value of calls is 0.0072 which is lower than the mean value of puts (0.0334).

---------------Insert Figure 5-1----------------

---------------Insert Figure 5-2----------------

---------------Insert Figure 5-3----------------

Figure 5-1 displays the time-series distribution of the percentage of trades with predictive information identified by different information measures. Across the sample period, PER^1, PER^2, and PER^3 are less volatile after the third quarter of 2013, and PER^4 is slightly increased after this time point. Furthermore, Figures 5-2 and 5-3 display the time-series distribution of directional information dummy variables and directional overall shape parameter, respectively. It could be that the variations are around zero. This further confirms the balance between buyer- and seller-initiated trades. In addition, the volatility of these variables has no significant change during the sample period.

6.4.2 Underlying asset volatility and options trade flow

--------------Insert Table 5-5---------------

Before discussing the predictive power of information measures, we investigate the effects of the proportion of trades with predictive information on the underlying stock volatility. Table 5-5 reports the effects of the proportion of calls and puts, respectively. As can be seen in this table, the results vary across the proportion estimated by different information dummy variables. PER^4 represents the strongest predictive ability as it can predict volatility five days ahead. For PER^1 , PER^2 , and PER^3 , their predictive horizons are between three and four days. Additionally, calls transactions show a stronger predictive power than puts in terms of PER^1 , PER^2 , and PER^3 which have longer predictive horizons. Moreover, these three proportion measures also report different results as compared with PER^4 . In particular, these three measures show that the volatility of underlying stock increases with the number of trades conveying information, while PER^4 reports an opposite effect – that the volatility of underlying stock prices increases with more options trades with predictive information.

The predictive ability of the options order flow confirms the connection between options and stock market and is consistent with the work of Sarwar (2005) and Ni *et al.* (2008). Namely, that options trading is informative regarding the future volatility of underlying stock. Moreover, these results show that the relationship between volatility and the percentage of trade with predictive information varies across information dummy variables. However, it is difficult to say which relationship is more precise because previous literature (e.g. French and Roll, 1986; De Long *et al.*, 1990) provided mixed results regarding the relationship between informed trading and volatility. For example, French and Roll (1986) suggested that the appearance of private information is associated with higher volatility, while Wang (1994) and Blasco *et al.* (2012) showed that uninformed trading creates volatility. In particular, options market traders are better informed than underlying stock traders and thereby tend to predict greater future volatility when they posit that the stock market fluctuates due to a significant influence of uninformed traders (Blasco *et al.*, 2012).

6.4.3 Underlying asset returns and options trade flow

In this section, we demonstrate and discuss the predictive ability of options order flow for stock returns. Unlike the previous section, this section uses directional information measures as a proxy of predictive information. Table 5-6 shows the regression results of stock returns on lagged call and put options proxies, respectively. In particular, this table allows us to differentiate between the coefficients estimated from each directional information measures. Each panel shows the ability of directional information measures to predict contemporaneous stock returns and extend the daily predictive horizon from date t+1 to date t+5.

--------------Insert Table 5-6---------------

It is clear from Table 5-6 that all directional information measures can influence the underlying stock returns. There is not only a strong contemporaneous effect, but also a significant predictive ability. Therefore, we could conclude that the options' volume-weighted duration contains predictive information on underlying stock price movement. With the STM-ACWD model, we can capture such information from the options order flow.

In Table 5-6, all calls (puts) directional information measures positively (negatively) significantly impact contemporaneous underlying stock returns The contemporaneous impact of information content in the option order flow is consistent with Chan *et al.* (2002) and Hu (2014). Chan *et al.* (2002) explains the reversed effects of calls and puts directional information measures by the underlying signals of call and put trades. The positive calls directional information measures signal positive information, while the positive puts directional information measures signals negative information. Hence, the contemporaneous stock returns are positively related to the call directional information measures, but negatively related to the puts directional information measures.

Regarding longer predictive horizons, both calls and puts directional information measures can predict the next-day stock returns. Moreover, the predictive power of puts measures does not lose significance over a longer period. In particular, X^1, X^2, X^3, and DSP can positively, significantly influence the stock returns on *t*+2 days. X^4 and DSP are positively, significantly associated with the stock returns on *t*+5 days. These significant effects are consistent with Pan and Poteshman (2006), Hu (2014), and Ryu and Yang (2018) who reject the separating equilibrium. Informed traders are active in the options market. They can submit market orders in the options market and execute orders immediately before valuable information spreads across the options and stock markets (Tsai *et al.*, 2015). The empirical results in Table 5-6 show that the valuable information revealed in option trades is captured by our directional information measures.

--------------Insert Table 5-7---------------

Table 5-7 presents regression results by including calls and puts directional information measures at the same time. These results could further explore the relationship between the information in the options order flow and underlying stock returns. By comparing the results in Table 5-6 and 5-7, we observe similar results that calls directional information measures have a significantly, positively impact on the stock returns on date *t* and date *t*+1 and put directional information measures have opposite effects over the same period. In addition, put directional information measures have significant positive impacts on the underlying stock returns over a longer predictive horizon. For example, in Table 5-7, Models 4 and 5 show that put directional information measures can predict the underlying stock returns on date *t*+5.

Essentially, our results suggest that the shorter duration between options trades and larger options trade size are informative regarding the future underlying stock price movement. This is inconsistent with Kyle (1985) who argued that no information is conveyed over the duration of the trade and Beltran-Lopez *et al.* (2012) who argued that information is only conveyed over a longer trade duration. The findings in our study support another strand of research, namely that high trade intensity and large trading volume in options markets contains information about future stock price movements (Easley *et al.*, 1997a; Xu *et al.*, 2006; Furfine, 2007). They argue that informed traders tend to trade in a hurry in order to take advantage of their timely information and maximise their profits. This argument corresponds to the previous explanation that informed traders are likely to submit market orders.

6.5 Conclusion

Easley *et al.* (1998b), Pan and Poteshman (2006), and Johnson and So (2012) capture information from the options order flow and show that the information content of option transactions is related to the underlying stock price movements. Blau *et al.* (2014) show that the level of information varies across each information measure. Previous literature (e.g. Dufour and Engle, 2000; Meyer-Brandis, 2010; Tsai et al.,

2015) used trade direction, trade volume and trade duration as information measures to capture information in options trades. Kalaitzoglou and Ibrahim (2013) demonstrated that the information can be efficiently captured by the volume-weighted duration which is modelled by the STM-ACWD model.

To examine the predictive ability of option trading flow, we employ the STM-ACWD model and construct volume-weighted duration as a new proxy of information content in option trades. Our study sheds some light on the predictive power of options trade intensity on the underlying stock price movements. The empirical results may contribute further study of the predictive ability of options order flow in multiple-market context.

Our results indicate that the information contained in the options order flow can be efficiently captured by volume-weighted duration with the STM-ACWD model. In particular, we created three sets of information measures based on volume-weighted duration, estimation results of the STM-ACWD model, and trade directions. These measures provide supportive evidence that some options trades contain information. Hence, we reject the separating equilibrium that informed traders are only active in stock markets.

We further studied the predictability of these options trades and identified that they can significantly influence contemporaneous and future stock price movement. In particular, we report that all information measures on date *t* are related to the stock volatility and return on date *t* and *t*+1. Our results also show that there are differences in the level of information contained in each information measure. Regarding the percentage measures, PER^4 contains more predictability about future stock volatility. Regarding the directional information measures, X^4 and DSP can predict stock returns over a longer predictive horizon. Besides, the level of information is also varying across options types. As compared with calls information measures, puts information measures show a stronger predictive ability in terms of return but a weaker ability in predicting volatility.

Our findings strongly confirm previous findings that volume-weighted duration can capture the information content of the options order flow (Kalaitzoglou and Ibrahim, 2015). Most importantly, our evidence suggests that the captured information has the ability to predict future underlying stock price movements.

Table 6-1 Summary statistics of the options market

Panel A: Call options

		Trade duration		Volume	
RIC	N	Mean	Std Dev	Mean	Std Dev
AXP	118894	1040.06	2872.03	9.59	26.30
BA	353005	693.19	2346.18	8.26	43.81
CAT	522326	369.13	1593.71	8.14	22.98
CSCO	362504	356.46	1452.96	30.52	147.71
CVX	205242	828.91	2633.57	10.60	44.37
DD	108012	1157.06	3152.20	15.71	114.59
DIS	183033	875.99	2580.46	11.59	36.65
GE	336118	432.79	1656.02	27.79	162.22
GS	650816	277.94	1286.41	6.47	16.60
HD	292328	492.15	1942.44	12.99	78.88
IBM	616612	404.49	1565.48	6.96	17.69
INTC	427500	348.54	1422.39	27.39	177.14
JNJ	326924	431.87	1927.75	14.73	91.21
JPM	665879	354.73	1570.01	16.01	74.95
KO	146544	987.10	2773.94	19.97	85.70
MCD	243623	689.77	2292.43	10.87	57.88
MMM	87147	1450.88	3570.06	8.83	19.05
MRK	150486	934.25	2763.45	21.23	134.02
MSFT	634897	331.82	1404.02	25.36	122.83
NKE	151714	979.11	2865.37	9.56	29.18
PFE	194522	706.21	2296.28	29.68	199.24
PG	177950	811.11	2572.71	13.96	59.06
T	247378	639.66	2045.29	28.49	348.15
TRV	18409	2048.96	4101.78	12.54	52.57
UNH	91623	1225.91	3276.79	13.26	67.78
UTX	78048	1511.48	3669.38	11.27	30.07
V	327158	539.45	1883.50	5.91	12.55
VZ	245356	870.97	2584.70	24.36	374.31
WMT	170097	815.65	2590.69	16.36	135.53
XOM	418507	412.49	1720.48	13.28	65.52

Panel B: Put options

		Trade duration		Volume	
RIC	N	Mean	Std Dev	Mean	Std Dev
AXP	84933	1262.77	3264.13	10.30	33.19
BA	242009	884.51	2752.61	7.32	15.03
CAT	441258	412.63	1717.09	8.34	23.68
CSCO	215319	557.92	1923.36	25.87	95.52
CVX	163470	918.46	2842.08	9.54	23.74
DD	73376	1516.30	3688.02	12.38	31.89
DIS	109974	1238.78	3208.83	13.28	34.44
GE	169288	757.59	2350.62	23.86	76.61
GS	394981	416.92	1682.74	6.04	14.19
HD	224951	578.40	2175.30	14.10	87.06
IBM	519020	463.61	1728.60	7.16	17.50
INTC	281714	483.51	1779.35	27.56	134.42
JNJ	241580	543.13	2232.08	13.40	60.78
JPM	455450	474.06	1902.92	14.89	53.94
KO	104551	1275.01	3231.82	17.41	54.79
MCD	183866	867.04	2610.87	9.35	20.92
MMM	75417	1570.17	3816.79	9.36	18.02
MRK	87007	1348.47	3438.15	16.93	51.00
MSFT	409270	475.35	1756.61	24.84	102.40
NKE	117671	1106.11	3129.55	10.32	32.87
PFE	107695	1080.12	2982.84	25.02	101.44
PG	159345	867.58	2753.30	14.64	53.36
T	172458	857.35	2521.83	22.68	101.32
TRV	11604	2666.39	4907.57	9.86	25.14
UNH	60244	1546.39	3767.66	13.66	63.01
UTX	57797	1817.37	4111.26	10.93	20.28
V	191138	796.55	2399.57	5.98	11.36
VZ	160766	1160.76	3053.85	17.99	61.79
WMT	150644	913.96	2817.49	13.51	44.81
XOM	358437	442.26	1822.27	12.83	43.65

Note: The table present the option market activity from 03 Jan 2012 to 30 Jun 2014. Panel A represents the statistic results for call options and Panel B represent the statistic results for put options. N refers to the number of observations. There are 8,552,652 and 6,025,233 observations for call and put options, respectively. Trade duration refers to the time duration between trades which is estimated in seconds. Volume refers to the average option volume per trade.

Table 6-2 Estimation results of STM-ACWD

Panel A: Call options

RIC	ω	α	β	γ_1	γ_2	g_1	j_1
AXP	0.082	0.094	0.910	1.033	0.222	1.328	0.516
BA	0.064	0.098	0.912	0.990	0.228	1.233	0.636
CAT	0.042	0.082	0.932	1.017	0.243	1.361	0.611
CSCO	0.034	0.077	0.944	0.994	0.231	1.231	0.637
CVX	0.071	0.077	0.926	1.037	0.226	1.279	0.503
DD	0.084	0.093	0.913	0.986	0.218	1.227	0.574
DIS	0.041	0.089	0.930	1.018	0.222	1.220	0.626
GE	0.036	0.073	0.946	1.015	0.229	1.227	0.583
GS	0.054	0.096	0.918	0.972	0.244	1.359	0.638
HD	0.032	0.075	0.951	0.904	0.214	1.088	0.731
IBM	0.049	0.092	0.922	1.014	0.239	1.328	0.615
INTC	0.036	0.067	0.948	1.045	0.240	1.365	0.516
JNJ	0.033	0.065	0.965	0.900	0.196	0.956	0.698
JPM	0.053	0.098	0.922	0.948	0.231	1.212	0.694
KO	0.069	0.077	0.928	1.059	0.223	1.379	0.459
MCD	0.076	0.094	0.911	1.018	0.229	1.271	0.557
MMM	0.092	0.074	0.919	1.018	0.215	1.297	0.503
MRK	0.102	0.097	0.907	1.002	0.218	1.229	0.522
MSFT	0.041	0.087	0.933	0.983	0.232	1.228	0.664
NKE	0.041	0.082	0.934	0.967	0.220	1.185	0.687
PFE	0.068	0.099	0.919	0.981	0.217	1.112	0.619
PG	0.070	0.097	0.918	0.971	0.218	1.139	0.624
T	0.047	0.072	0.941	1.046	0.228	1.352	0.492
TRV	0.133	0.118	0.878	1.043	0.206	1.244	0.513
UNH	0.088	0.084	0.918	0.960	0.213	1.200	0.581
UTX	0.106	0.077	0.912	0.994	0.211	1.248	0.544
V	0.051	0.096	0.916	1.015	0.237	1.324	0.629
VZ	0.061	0.086	0.925	1.026	0.221	1.257	0.546
WMT	0.063	0.083	0.927	1.023	0.223	1.226	0.547
XOM	0.060	0.091	0.924	0.995	0.230	1.204	0.596

Panel B: Put options

RIC	ω	α	β	γ_1	γ_2	g_1	j_1
AXP	0.102	0.097	0.903	1.001	0.215	1.253	0.533
BA	0.058	0.093	0.921	0.968	0.221	1.186	0.645
CAT	0.050	0.092	0.924	0.982	0.236	1.288	0.648
CSCO	0.039	0.083	0.940	0.976	0.220	1.145	0.651
CVX	0.051	0.079	0.934	1.008	0.220	1.225	0.577
DD	0.128	0.099	0.894	0.957	0.207	1.131	0.590
DIS	0.086	0.093	0.912	0.980	0.213	1.217	0.582
GE	0.075	0.072	0.934	1.054	0.221	1.304	0.409
GS	0.060	0.092	0.918	0.983	0.239	1.340	0.618
HD	0.046	0.088	0.941	0.841	0.206	1.020	0.813
IBM	0.064	0.104	0.908	0.978	0.234	1.264	0.654
INTC	0.052	0.074	0.939	1.036	0.232	1.312	0.496
JNJ	0.049	0.081	0.953	0.843	0.191	0.915	0.775
JPM	0.057	0.096	0.922	0.927	0.224	1.163	0.703
KO	0.091	0.083	0.918	1.057	0.215	1.317	0.436
MCD	0.060	0.078	0.929	0.996	0.223	1.286	0.566
MMM	0.087	0.081	0.924	0.980	0.208	1.159	0.541
MRK	0.158	0.094	0.890	0.972	0.207	1.177	0.514
MSFT	0.057	0.090	0.927	0.967	0.225	1.183	0.635
NKE	0.076	0.099	0.912	0.911	0.212	1.100	0.716
PFE	0.108	0.098	0.906	0.943	0.208	1.096	0.612
PG	0.089	0.108	0.906	0.874	0.208	1.032	0.757
T	0.071	0.078	0.932	1.015	0.218	1.257	0.477
TRV	0.219	0.096	0.858	1.058	0.200	1.182	0.443
UNH	0.186	0.083	0.885	0.983	0.205	1.154	0.485
UTX	0.151	0.076	0.898	0.981	0.205	1.191	0.512
V	0.070	0.093	0.913	1.020	0.227	1.289	0.565
VZ	0.100	0.092	0.910	1.021	0.213	1.201	0.496
WMT	0.058	0.087	0.929	0.971	0.216	1.157	0.618
XOM	0.077	0.123	0.897	0.902	0.222	1.128	0.759

Note: This table represent the estimation results of STM-ACWD. Panel A represents the results of call options, and Panel B represents the results of put options. All results are statistically significant at 1% level.

Table 6-3 Summary statistics of percentage of informed trades

Panel A: Call options

RIC	PER^1	PER^2	PER^3	PER^4
AXP	0.552	0.546	0.222	0.150
BA	0.541	0.538	0.248	0.090
CAT	0.557	0.555	0.261	0.085
CSCO	0.583	0.581	0.298	0.078
CVX	0.560	0.556	0.232	0.123
DD	0.538	0.531	0.211	0.132
DIS	0.525	0.522	0.224	0.102
GE	0.591	0.589	0.294	0.104
GS	0.573	0.572	0.272	0.067
HD	0.606	0.604	0.335	0.067
IBM	0.542	0.540	0.247	0.088
INTC	0.588	0.586	0.288	0.102
JNJ	0.697	0.696	0.449	0.048
JPM	0.578	0.577	0.293	0.062
KO	0.557	0.552	0.225	0.171
MCD	0.540	0.537	0.223	0.109
MMM	0.546	0.539	0.197	0.157
MRK	0.571	0.567	0.233	0.134
MSFT	0.574	0.573	0.291	0.053
NKE	0.509	0.504	0.211	0.091
PFE	0.566	0.563	0.266	0.088
PG	0.563	0.559	0.244	0.099
T	0.576	0.573	0.263	0.151
TRV	0.531	0.519	0.183	0.188
UNH	0.553	0.547	0.216	0.132
UTX	0.546	0.537	0.208	0.161
V	0.524	0.522	0.240	0.085
VZ	0.558	0.554	0.239	0.122
WMT	0.561	0.556	0.238	0.125
XOM	0.591	0.589	0.293	0.101
All	0.563	0.560	0.255	0.109

Panel B: Put options

RIC	PER^1	PER^2	PER^3	PER^4
AXP	0.556	0.550	0.219	0.157
BA	0.552	0.548	0.240	0.084
CAT	0.564	0.562	0.260	0.081
CSCO	0.581	0.578	0.284	0.081
CVX	0.556	0.553	0.239	0.114
DD	0.548	0.542	0.216	0.146
DIS	0.544	0.540	0.219	0.138
GE	0.616	0.611	0.276	0.168
GS	0.561	0.558	0.252	0.081
HD	0.619	0.618	0.325	0.051
IBM	0.542	0.540	0.240	0.078
INTC	0.595	0.592	0.279	0.111
JNJ	0.702	0.701	0.429	0.039
JPM	0.585	0.583	0.283	0.063
KO	0.576	0.570	0.225	0.185
MCD	0.546	0.542	0.219	0.119
MMM	0.565	0.557	0.210	0.146
MRK	0.589	0.583	0.228	0.170
MSFT	0.580	0.578	0.284	0.063
NKE	0.530	0.527	0.226	0.090
PFE	0.578	0.574	0.263	0.124
PG	0.573	0.570	0.249	0.059
T	0.595	0.592	0.267	0.163
TRV	0.564	0.546	0.179	0.187
UNH	0.583	0.575	0.211	0.165
UTX	0.567	0.559	0.212	0.183
V	0.540	0.537	0.236	0.106
VZ	0.574	0.569	0.249	0.143
WMT	0.566	0.562	0.239	0.108
XOM	0.590	0.588	0.287	0.061
All	0.575	0.570	0.252	0.115

Note: The table presents the summary statistic of the percentages of informed trades. Panel A represents the statistic results for call options and Panel B represent the statistic results for put options. PER^1 , PER^2 , PER^3 , and PER^4 refer to percentages of informed trades in the trading day and the information trades are evaluated by different criteria.

Table 6-4 Summary statistics of directional information variables

Panel A: Call options

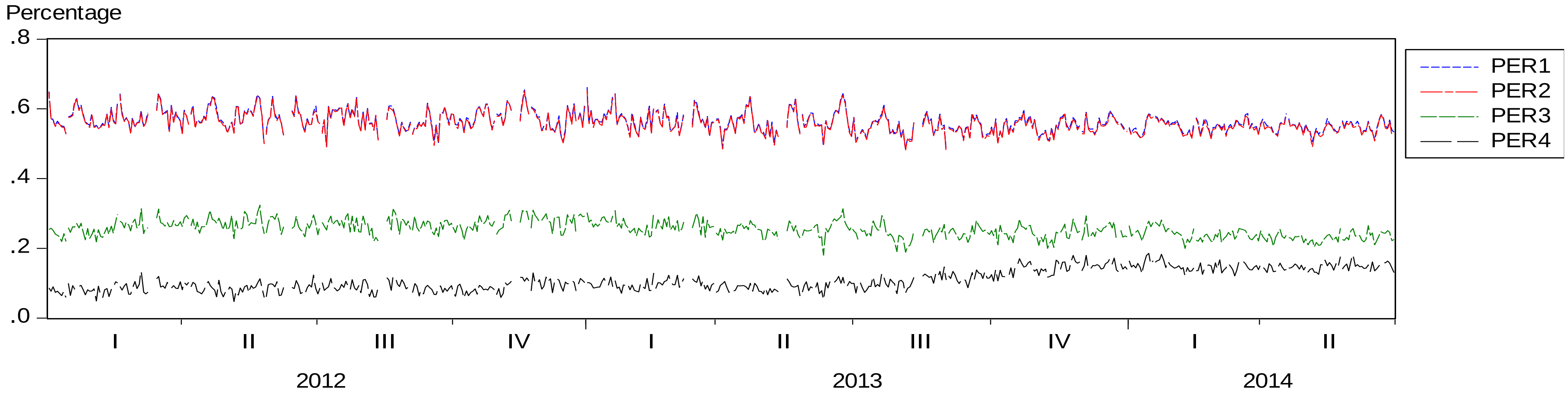
RIC	X^1	X^2	X^3	X^4	DSP
AXP	0.0148	0.0143	0.0086	0.0063	0.0059
BA	0.0101	0.0107	0.0049	0.0026	0.0006
CAT	0.0051	0.0055	0.0040	0.0026	0.0008
CSCO	0.0182	0.0180	0.0145	0.0107	0.0120
CVX	0.0118	0.0116	0.0063	0.0042	0.0097
DD	-0.0090	-0.0076	-0.0032	-0.00003	-0.0030
DIS	-0.0249	-0.0240	-0.0046	-0.0018	-0.0084
GE	0.0075	0.0079	0.0066	0.0048	0.0054
GS	0.0341	0.0341	0.0169	0.0025	0.0085
HD	0.0188	0.0196	0.0107	0.0068	0.0145
IBM	0.0044	0.0041	0.0030	0.0021	0.0071
INTC	0.0175	0.0176	0.0092	0.0040	0.0092
JNJ	0.0044	0.0045	0.0074	0.0018	0.0065
JPM	0.0052	0.0049	0.0056	0.0004	-0.0046
KO	0.0343	0.0331	0.0112	0.0097	0.0168
MCD	0.0140	0.0139	0.0079	0.0054	0.0005
MMM	0.0139	0.0155	0.0057	0.0047	0.0130
MRK	-0.0102	-0.0108	0.0011	0.0005	0.0088
MSFT	0.0132	0.0132	0.0103	0.0055	0.0079
NKE	-0.0366	-0.0367	-0.0028	-0.0009	-0.0220
PFE	-0.0093	-0.0086	-0.0009	-0.0003	-0.0051
PG	0.0298	0.0299	0.0153	0.0097	0.0215
T	0.0287	0.0290	0.0110	0.0132	0.0299
TRV	0.0152	0.0113	0.0047	0.0045	0.0272
UNH	-0.0008	-0.0003	-0.0004	0.0015	-0.0070
UTX	0.0082	0.0083	0.0065	0.0071	-0.0025
V	-0.0126	-0.0123	-0.0019	0.0013	-0.0117
VZ	0.0064	0.0076	0.0081	0.0078	0.0100
WMT	0.0341	0.0336	0.0112	0.0115	0.0250
XOM	0.0491	0.0491	0.0196	0.0079	0.0383
All	0.0098	0.0099	0.0066	0.0045	0.0072

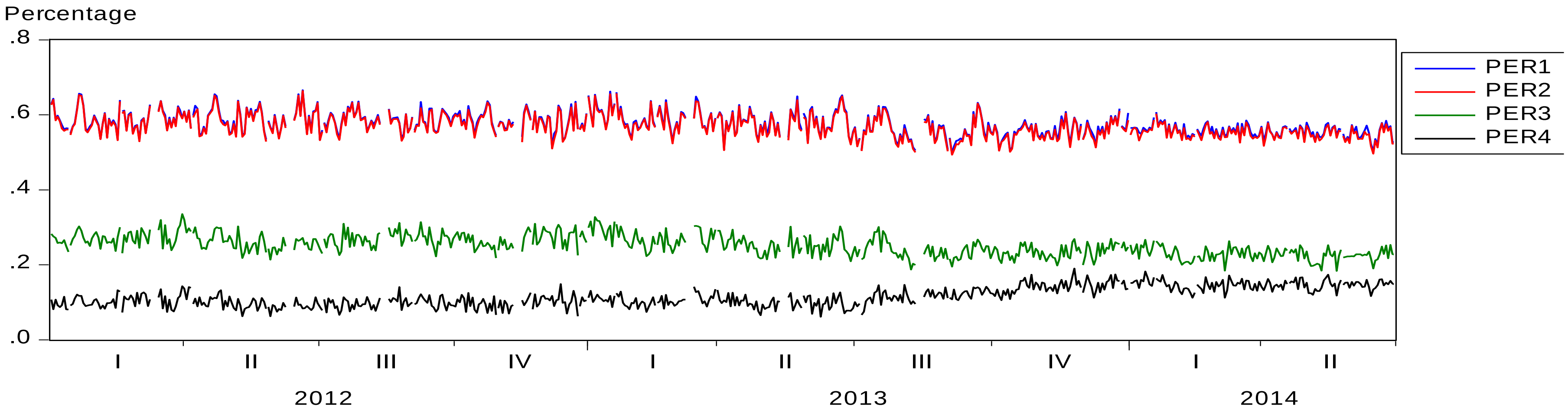
Panel B: Put options

RIC	X^1	X^2	X^3	X^4	DSP
AXP	0.0041	0.0024	0.0010	-0.0031	0.0040
BA	0.0065	0.0068	0.0027	0.0005	0.0342
CAT	0.0015	0.0020	0.0033	0.0010	0.0169
CSCO	0.0357	0.0362	0.0195	0.0070	0.1055
CVX	-0.0176	-0.0176	-0.0091	-0.0047	-0.0342
DD	0.0133	0.0126	0.0015	0.0023	0.0507
DIS	-0.0024	-0.0024	-0.0018	-0.0022	0.0079
GE	-0.0065	-0.0060	-0.0020	-0.0065	-0.0062
GS	0.0191	0.0190	0.0092	0.0033	0.0496
HD	0.00007	-0.00005	-0.0029	-0.0026	0.0009
IBM	0.0027	0.0024	0.0026	0.0021	0.0197
INTC	0.0277	0.0270	0.0079	0.0024	0.0656
JNJ	0.0172	0.0174	0.0121	-0.0003	0.0721
JPM	0.0085	0.0087	0.0039	0.0022	0.0201
KO	-0.0098	-0.0107	-0.0064	-0.0060	-0.0258
MCD	0.0427	0.0417	0.0126	0.0077	0.0989
MMM	0.0009	0.0008	0.0027	0.0015	0.0122
MRK	-0.0043	-0.0042	0.0008	0.0010	0.0078
MSFT	0.0294	0.0295	0.0157	0.00015	0.0893
NKE	0.0147	0.0148	0.0126	0.0057	0.0543
PFE	-0.0005	0.0008	0.0006	0.0014	0.0121
PG	0.0148	0.0151	0.0128	0.0044	0.0439
T	0.0156	0.0147	0.0075	0.0038	0.0582
TRV	0.0589	0.0532	0.0220	0.0223	0.1903
UNH	0.0217	0.0228	0.0078	0.0053	0.0463
UTX	-0.0037	-0.0050	-0.0020	-0.0040	-0.0005
V	-0.0219	-0.0220	-0.0073	-0.0043	-0.0499
VZ	-0.0168	-0.0166	-0.0082	-0.0063	-0.0205
WMT	0.0118	0.0114	0.0100	0.0070	0.0594
XOM	0.0078	0.0076	0.0045	0.0017	0.0266
All	0.0089	0.0086	0.0044	0.0014	0.0334

Note: The table presents the summary statistic of the directional information variables. Panel A represents the statistic results for call options and Panel B represent the statistic results for put options. X^1, X^2, X^3, and X^4 refer to directional information dummy variables and the information dummy variables are evaluated by different criteria. DSP refers to the directional overall shape parameter

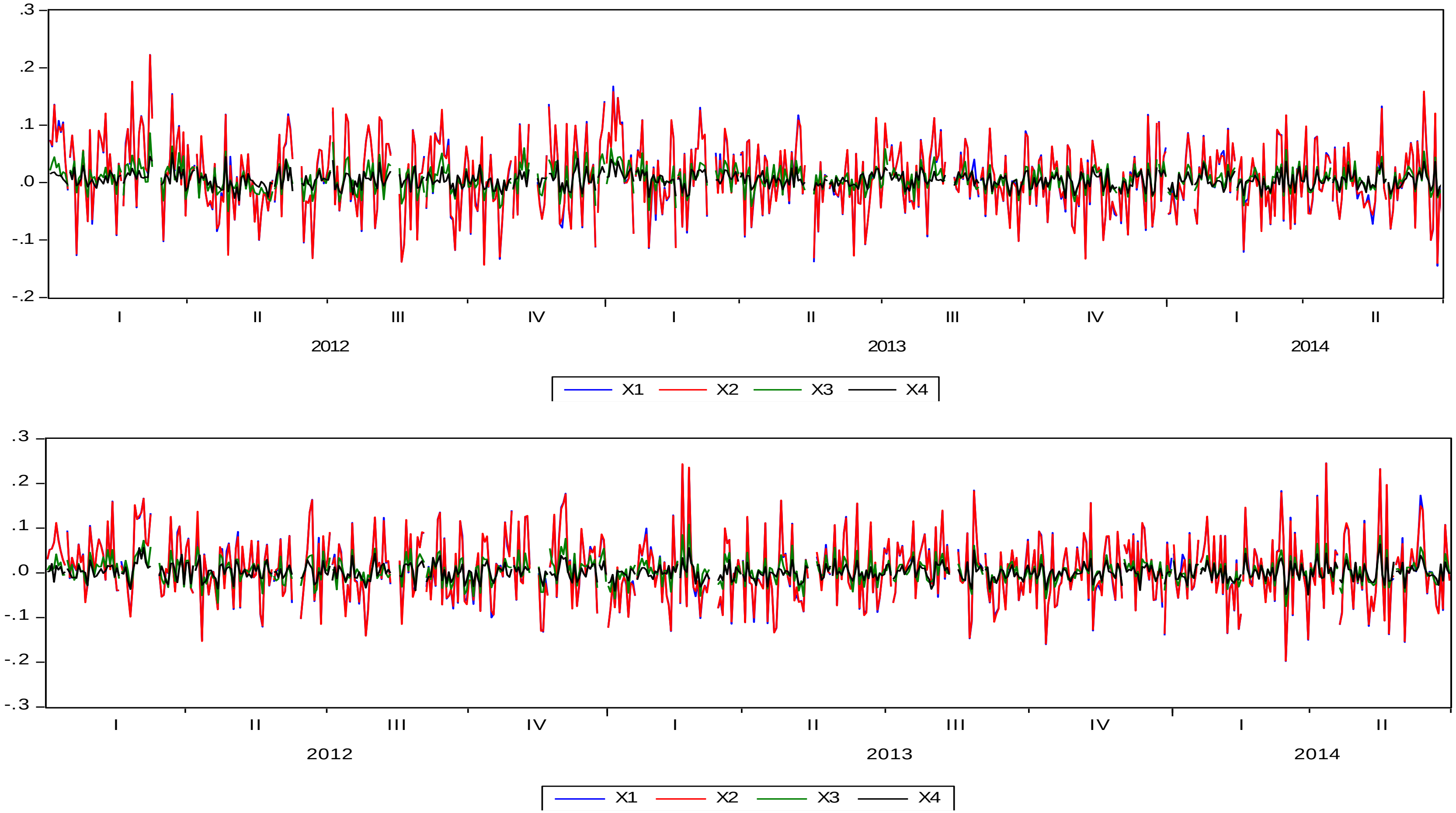
Figure 6-1 Percentage of informed trades





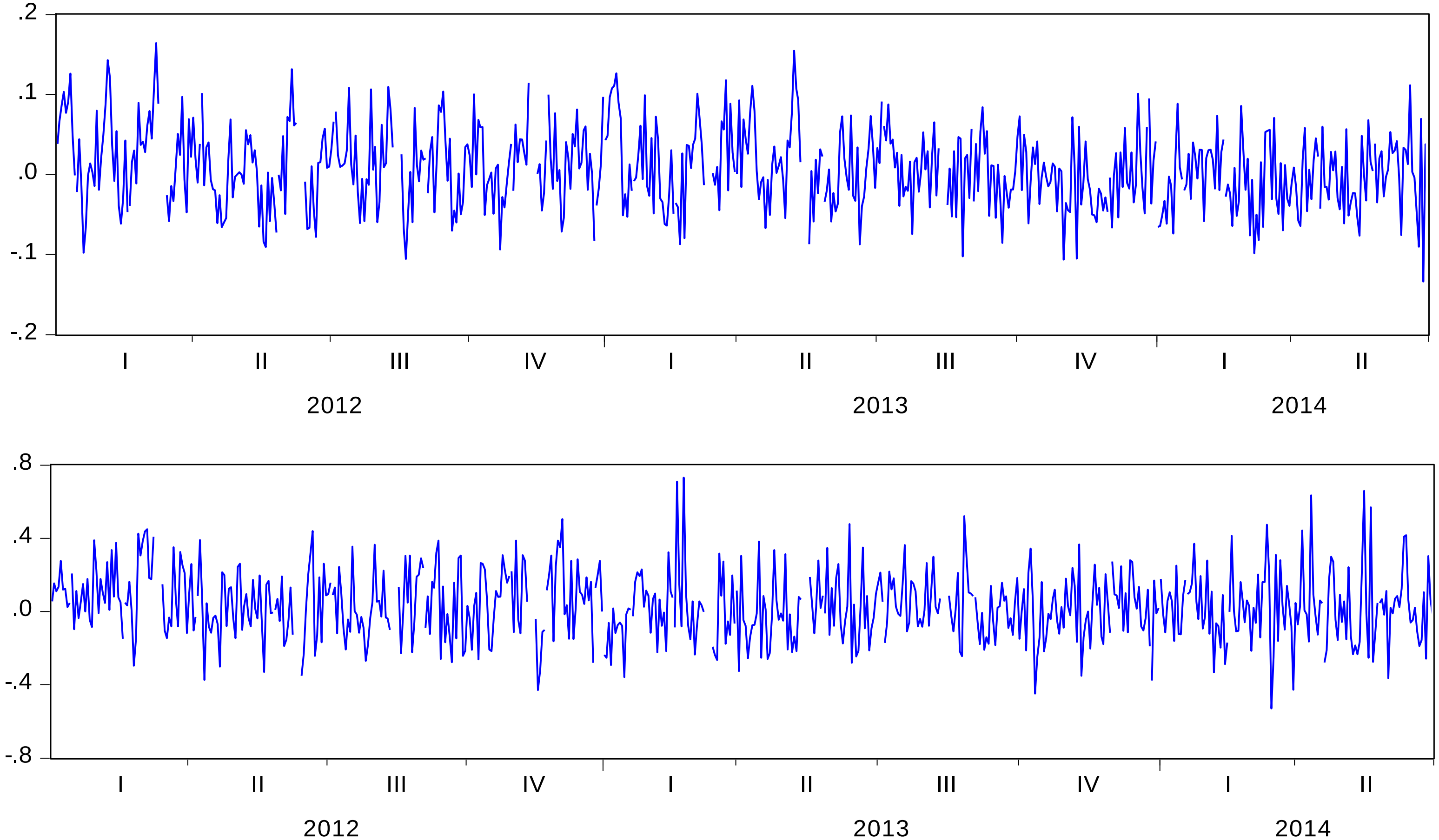
Note: the plots represent time-series percentages of informed trades based on different criteria of informed trades. The first plot represents the distribution of call options and the second plot represent the distribution of put options. The percentage of each informed trades is averaged across all option contracts.

Figure 6-2 Distribution of directional informed dummy variables



Note: the plots represent time-series of informed dummy variables based on different criteria of informed trades. The first plot represents the distribution of call options and the second plot represent the distribution of put options. The value of each directional informed dummy variable is averaged across all option contracts.

Figure 6-3 Distribution of directional overall shape parameter



Note: the plots represent time-series of directional overall shape parameter. The first plot represents the distribution of call options and the second plot represents the distribution of put options. The value of directional overall shape parameter is averaged across all option contracts.

Table 6-5 Daily regressions of stock volatility on lagged call and put options order information

Panel A: Information variable from call options order flow

τ -Days ahead	PER^1	PER^2	PER^3	PER^4
0	0.0179207***	0.0178092***	0.0066795***	-0.001877*
1	0.0072873***	0.0073087***	0.004635***	-0.0025877**
2	0.0017447**	0.0017809**	0.0017306*	-0.0056015***
3	0.0016961**	0.0018003**	-0.0000176	-0.0052517***
4	-0.0007329	-0.0008521	0.0003718	-0.0058686***
5	0.0002753	0.0002602	-0.000597	-0.0076657***

Panel B: Information variable from put options order flow

τ -Days ahead	PER^1	PER^2	PER^3	PER^4
0	0.0123461***	0.0122961***	0.0034671***	-0.0026019***
1	0.0051172***	0.0051217***	0.0016059*	-0.0025982**
2	0.0017484*	0.0017835*	-0.0004067	-0.0050164***
3	0.0006775	0.0006218	-0.0011792	-0.0058947***
4	0.000306	0.0005437	-0.0009175	-0.0060905***
5	0.0008258	0.0008441	0.0003603	-0.0042214***

Note: ***, **, * refer to the significant level at 1%, 5%, and 10%. The number of observations is 18,097. This table represents the estimation regression results of stock volatility on the proportion of informed trades. The time lag of volatility is from 1 day to 5 day. Panel A represents the results of call options and Panel B represents the results of put options. PER^1 , PER^2 , PER^3 , and PER^4 refer to percentage of informed trades in the trading day and the information trades are evaluated by different criteria. For regression models in Panel A and B, R-square is around 3%.

Table 6-6 Daily regressions of stock returns on lagged call and put options order information

Panel A: Information variable from call options order flow					
τ -Days ahead	X^1	X^2	X^3	X^4	DSP
0	0.00507***	0.00507***	0.00639***	0.00699***	0.00218***
1	0.00239***	0.00242***	0.00448***	0.00461***	0.00114***
2	-0.00011	-0.00009	-0.00012	-0.00064	0.00001
3	-0.00007	-0.00008	-0.00019	-0.00102	-0.00009
4	-0.00002	-0.00002	-0.00006	0.00133	-0.00004
5	0.00012	0.00014	0.00044	0.00159	0.00010
Panel B: Information variable from put options order flow					
τ -Days ahead	X^1	X^2	X^3	X^4	DSP
0	-0.00325***	-0.00326***	-0.00333***	-0.00472***	-0.00135***
1	-0.00243***	-0.00245***	-0.00533***	-0.00574***	-0.00111***
2	0.000699*	0.000718*	0.0017978*	0.00120	0.000257*
3	-0.00002	0.00001	0.00004	-0.00041	0.00008
4	0.00016	0.00013	-0.00036	-0.00014	-0.00004
5	0.00056	0.00057	0.00151	0.0024561*	0.0002678*

Note: ***, **, * refer to the significant level at 1%, 5%, and 10%. The number of observations is 18,097. This table represents the estimation regression results of stock returns on the directional information variables. The time lag of return is from 1 day to 5 day. Panel A represents the results of call options and Panel B represents the results of put options. X^1 , X^2 , X^3 , and X^4 refer to directional information dummy variables and the information dummy variables are evaluated by different criteria. DSP refers to the directional overall shape parameter. For regression models in Panel A and B, R-square is around 3%.

Table 6-7 Daily regressions of stock returns on lagged options order information

	(1)		(2)		(3)		(4)		(5)	
τ -Days ahead	$\mathrm{C}X^1$	PX^1	$\mathrm{C}X^2$	PX^2	$\mathrm{C}X^3$	PX^3	$\mathrm{C}X^4$	PX^4	$CDSP$	$PDSP$
0	0.005165***	-0.003306***	0.005173***	-0.003323***	0.006618***	-0.003607***	0.007259***	-0.004919***	0.002226***	-0.001386***
1	0.002519***	-0.002462***	0.002560***	-0.002486***	0.004818***	-0.005494***	0.004942***	-0.005806***	0.001180***	-0.001139***
2	-0.000094	0.0006802*	-0.000077	0.0007009*	-0.000149	0.001776*	-0.000571	0.001156	0.000016	0.0002521*
3	-0.000103	0.000004	-0.000098	0.000029	-0.000210	0.000078	-0.001069	-0.000331	-0.000097	0.000086
4	0.000040	0.000133	0.000027	0.000104	-0.000003	-0.000356	0.001424	-0.000159	-0.000024	-0.000044
5	0.000185	0.000556	0.000203	0.000574	0.000422	0.001506	0.001650	0.0024393*	0.000105	0.000265*

Note: ***, **, * refer to the significant level at 1%, 5%, and 10%. The number of observations is 18,097. This table represents the estimation regression results of stock returns on the directional information variables. (1), (2), (3), (4), and (5) refer to five regression models with different directional information measures. For each regression model, both call and put directional information measures are included. The time lag of return is from 1 day to 5 day. X^1, X^2, X^3, and X^4 refer to directional information dummy variables and the information dummy variables are evaluated by different criteria. DSP refers to the directional overall shape parameter. C and P before the information variable refer to the calls and puts information variables, respectively. R-square of regression models in this table is around 3%.

Chapter 7: What is the relationship between price clustering and size clustering in the options market?

7.1 Introduction

Investors usually perceive liquid markets as those that allow them to reduce trading costs. For example, negotiation costs can be minimised because investors can trade a desired quantity immediately without influencing the market price. In contrast, investors need to compromise on price, quantity, or execution speed when trading in an illiquid market (Meng *et al.*, 2013). Moulton (2005) and Hodrick and Moulton (2009) define price, quantity, and execution speed as the three dimensions of liquidity. By investigating these dimensions, they identify that investors not only have time-price trade-offs and also demonstrate the trade-offs between quantity and immediacy. Hence, investors may trade at frequently quoted prices or at certain quoted sizes in order to pursue the timing dimension of liquidity.

This chapter studies price clustering and size clustering in the options market using a sample of options traded on the Chicago Board Options Exchange (CBOE). This sample includes all options contracts written on the components of the Dow Jones Industrial Average Index (DJI) from January 2012 to June 2014. With trade and quote records of these options, we investigate presence of price and size clustering, and the determinants of clustering. Thus, we contribute to the literature in two ways.

To the best of our knowledge, this chapter is the first empirical study that investigates size clustering in the options market. Price clustering has been found in many different financial markets and explained by different hypotheses. For example, the price resolution hypothesis, the negotiation hypothesis, and the attraction hypothesis are used to explain price clustering in equity markets (Lien *et al.*, 2019), futures markets (ap Gwilym and Alibo, 2003), and options markets (Capelle-Blancard and Chaudhury, 2007). Regarding size clustering, this has been investigated in equity markets and futures markets (ap Gwilym and Meng, 2010; Verousis and ap Gwilym, 2013b) but not yet in options markets. The limited literature on size clustering may be a result of assumptions made by previous theoretical frameworks (e.g. Kyle, 1985; Easley and O'Hara, 1987), namely that traders always trade their desired quantities. So, quantity demand and quantity supply are always in equilibrium. In order to trade desired quantities, these traders should either sacrifice price to trade immediately or sacrifice time to trade at a good price (Black, 1971). In order to contribute to the literature, we investigate the presence of price and size clustering in an options market to further explore whether size clustering is the concentration of the quantity or whether it is to trade at specific amount.

As compared with price clustering, size clustering lacks well-developed theories to explain its causes. Researchers (for example, Alexander and Peterson, 2007) show that the explanations for price clustering are applicable to size clustering. Apart from these explanations, Moulton (2005) finds a quarter-end hypothesis to explain size clustering during the end of a quarter. Moreover, ap Gwilym and Meng (2010), Meng *et al.* (2013), and Verousis and ap Gwilym (2013b) provide evidence to support the argument that size clustering is driven by price clustering. These studies find that there is an interaction between price and size clustering, whilst showing opposite results on the relationship between price and size clustering. In particular, ap Gwilym and Meng (2010) and Meng *et al.* (2013) identify a negative relationship between price and size clustering. However, Verousis and ap Gwilym (2013b) show that price clustering and size clustering tend to occur simultaneously. Due to the mixed results, we employ Three Stage Least Squares Regression (3SLS) to explore the causes of size clustering. Apart from trading frequency, volatility, price level, and quarter-end factor, we also include moneyness and maturity as determinants of clustering. Capelle-Blancard and

Chaudhury (2007) and ap Gwilym and Verousis (2013) show the effects of moneyness and maturity on price clustering but demonstrate mixed results. Thus, we are interested in examining how the characteristics of options influence size clustering.

Our results are summarised as follows: We find supportive evidence of price and size clustering. In particular, the results demonstrate that the level of price and size clustering is similar between calls and puts. The extent of quote price clustering is higher than trade price clustering, while the size clustering shows no large difference between quote and trades.

The estimation results of the 3SLS model confirms the relationship between price and size clustering. However, unlike previous studies, we do not report the mutual and consistent relationship between them. In particular, the results of call options show positive effects of price clustering on size clustering and those of put options show negative effects of size clustering on price clustering. The results indicate that investors desire different dimensions of liquidity when trading calls and puts. In addition, the estimation results show that price and size are less clustered when trading out-of-the money contracts and near-to-maturity contracts after controlling other factors. The empirical findings have important implications for trading strategy design. In particular, investors should be aware of the sacrifice of prices and quantity when speeding up order execution. For example, informed traders should sacrifice desired trading price and quantity if they want to take advantage of their timely information.

The remainder of this chapter is organised as follows: Section 2 reviews the literature on price and size clustering. The clustering on options markets are also discussed in this section. Section 3 describes our sample and defines price and size clustering. Section 4 shows the methodology. Section 5 provides the results and discusses the findings, and Section 6 concludes.

7.2 Literature review

In financial markets, we expect trade prices and sizes to be uniformly distributed around the possible values. However, empirical studies (Gwilym et al., 1998; Aşçıoğlu et al., 2007) find that trade prices and sizes are usually rounded to certain amounts. This concentration of trade price or size at certain amounts is defined as price clustering and size clustering, respectively. For example, Meng *et al.* (2013) finds that traders are more likely to set the final digits of quoted and traded prices as “0” and “5”. Clustering at specific final digits implies that traders tend to take advantage of the greater degree of liquidity at these amounts. Hence, this section begins with a connection between liquidity and clustering. Since there is extensive literature around price clustering, this section further reviews four hypotheses to explain its causes. Finally, this section reviews the theories relating to size clustering.

7.2.1 Liquidity and clustering

It is difficult to give liquidity a clear definition because of its abstract properties (Huberman and Halka, 2001). The level of liquidity in an asset is related to both sides of trading. Buyers and sellers on each side are searching for the other and that creates a bilateral search process. If they spend more time solving search problems, they could arguably obtain better prices and desired quantity in trading (Lehalle and Laruelle, 2013). So, price, quantity, and time can be viewed as the three dimensions of liquidity. Based on these dimensions, in a perfectly liquid market, traders are able to trade their desired quantity at any time without changing the market price. In an illiquid market, however, traders need to sacrifice one or more dimensions of liquidity in order to execute a transaction. Hence, there are trade-offs between time, price, and quantity.

Clustering is considered as a process wherein traders select dimensions to sacrifice. For example, price clustering occurs when traders want to reduce execution costs in an illiquid market (Ball *et al.*, 1985; Harris, 1991; Aşçıoğlu *et al.*, 2007). Furthermore, a price-quantity trade-off relationship is confirmed by Moulton (2005) who shows that traders use rounded prices if they want to trade precise quantities. The detailed explanations of the trade-offs will be provided in the following sub-sections.

7.2.2 Literature on price clustering

The literature on price clustering begins with Osborne (1962) who provides the first empirical evidence by which to develop a framework for price clustering. Closing stock prices on the New York Stock Exchange (NYSE) are observed to cluster on certain values, such as whole numbers and quarters. Niederhoffer (1965) also confirms non-randomness in transaction prices. Niederhoffer shows that stock prices are near round numbers by investigating the books of specialists on the NYSE.

Ball *et al.* (1985) confirm that price clustering is pervasive by examining the London Gold Market between 1975 to 1981. In order to explain the reason behind price clustering, they develop the price resolution hypothesis which suggests a connection between clustering and uncertainty. Market participants tend to round their equilibrium price due to the existence of uncertainty. With higher uncertainty about prices, investors face higher market volatility and are therefore likely to set certain trade prices for facilitating immediacy. Hence, Ball *et al.* (1985) indicate the degree of price resolution as a function of the amount of information about the true price of the asset in the market. With this hypothesis, Grossman *et al.* (1997) suggests that the degree of price clustering is higher for market makers when faced with higher costs of market making. Their empirical results show that prices are less clustered in the liquid market which allows traders to easily ascertain the value of the asset and thereby complete transactions quickly. ap Gwilym *et al.* (1998b) and Lien *et al.* (2019) provide further supportive evidence for the price resolution hypothesis by identifying a positive relationship between volatility and price clustering on the London International Financial Futures and Options Exchange (LIFFE) and on the Taiwan Stock Exchange (TSE), respectively. Aşçıoğlu *et al.* (2007) investigate price clustering on the Tokyo Stock Exchange and show that high-risk stocks are more likely to exhibit price clustering than low-risk stocks. Unlike previous studies (ap Gwilym et al., 1998), their observed price clustering cannot be explained by the negotiation hypothesis because traders in TSE cannot directly negotiate with each other.

Harris (1991) proposes the negotiation hypothesis as another explanation of the causes of price clustering. A limited number of price points could reduce the number of bids and asks effectively lowering the cost of negotiation. For example, traders have different assessments of the value of assets if the assets are infrequently traded. This will increase the time taken to consummate a transaction and thereby lead to a trade-off between price and time. Thus, traders tend to use fewer price points in order to simplify the negotiation and quickly complete the transactions. In this context, the probability of a round-price trade rises. Harris' (1991) empirical results confirm that stock price clustering is positive related to volatility and price level but negatively related to transaction frequency. ap Gwilym and Alibo (2003) investigate the changes in price clustering when LIFFE changes from floor trading to an electronic trading system. Their result confirms a higher level of price clustering under floor trading than under an electronic trading platform. The negotiation hypothesis provides a possible explanation that traders are willing to use the full range of prices under the electronic system because of lower negotiation costs. This explanation indicates that price clustering could be lower for traders who use electronic trading system to trade. With 120 stocks from NASDAQ and NYSE, Davis *et al.* (2014) find that price clustering is less frequent when high-frequency traders[21] involved in both sizes of transactions than when only one high-frequency trader is involved. Moreover, the probability of price clustering is higher when liquidity providing orders are submitted by non-high-frequency traders. The empirical results show positive effects of price and volatility on the probability of price clustering that are consistent with those found by Harris (1991).

The results of Davis *et al.* (2014) also imply that price clustering is driven by human trading behaviours. Goodhart and Curcio (1991) propose the attraction hypothesis that uses the psychological preference for specific prices to explain price clustering. Mitchell (2001) and Ikenberry and Weston (2008) confirm this particular preference and find that investors tend to trade prices ending in even numbers or around the most gravitational points (0 and 5). Furthermore, Christie and Schultz (1994) provide the price collusion hypothesis by investigating NASDAQ market makers. Their empirical results show that market makers tend to avoid odd-eighth quotes. With the collusion of dealers, even-number quotes are more likely to be provided in order to maximise market-making revenues. Christie and Schultz (1999), Booth *et al.* (2000), and Ni *et al.* (2005) provide further evidence on the collusive hypothesis. However, this hypothesis is challenged by later studies. For example, Godek (1996) argues that the collusion hypothesis understates the importance of preference trading.

[21] Davis et al. (2014) define high-frequency traders as those who execute orders via computer programs.

7.2.3 Literature on size clustering

Trade size is another dimension of liquidity. Traders are more likely to trade their desired quantities by sacrificing desired price or execution speed (Meng *et al.*, 2013). However, on a practical level, the financial market is unable to maintain the equilibrium between quantity demand and supply because uninformed traders do not always want to trade exactly their desired quantities (Hodrick and Moulton, 2009). Moulton (2005) uses a data set of foreign exchange transactions to explore the trade-offs between quantity, time, and price. The results uncover size clustering and propose a quarter-end hypothesis to further explain the size clustering. Moulton finds that size clustering decreases at the quarter end because some stock traders tend to rebalance their shareholding in order to reach their desired position at the end of a quarter. Verousis and ap Gwilym (2013b) consider this hypothesis as an extension of the negotiation hypothesis and provide further evidence on the London Stock Exchange (LSE). At the intraday level, Garvey and Wu (2014) find that the probability of round-size trades decreases during the end of the trading day. This reduced probability is driven by the increasing desire to reach a specific inventory level. Hence, traders sacrifice precise trade prices and sizes to obtain liquidity.

With stock trading on the NASDAQ and the NYSE from 1990 to 2001, Alexander and Peterson (2007) find the stealth hypothesis is another explanation of size clustering. This hypothesis indicates that informed traders are likely to split their order to hide information and minimise price impact (Barclay and Warner, 1993; Chakravarty, 2001). Alexander and Peterson (2007) detect more price information in medium-sized rounded trades than large rounded trades. In addition, their results also support the negotiations hypothesis as finding that price and size clustering occur simultaneously. In particular, size clustering can reduce reporting risk and thereby contribute to negotiations. This relationship between price and size clustering is inconsistent with the quantity-price trade-off relationship developed by Moulton (2005) who argues that traders are unable to optimise all dimensions of liquidity. In particular, Moulton (2005) shows a price-size trade-off relationship and observes a negative relationship between trading costs and size clustering.

ap Gwilym and Meng (2010) support the price-size substitution effect by investigating the FTSE100 index futures traded on the NYSE Euronext-LIFFE. In particular, the empirical result shows that size clustering is negatively related to price clustering and trade frequency. Furthermore, Meng *et al.* (2013) confirm the substitution effect in the credit default swap market. Verousis and ap Gwilym (2013b) further investigate the relationship between price and size clustering on the LSE but show that price and size clustering are complements rather than substitutes. Although this relationship is found in the upstairs market and downstairs market which has different market structures[22], price clustering in this relationship follows different patterns across market structures. In the upstairs market, direct negotiations lead to a clear complementary effect between price and size clustering. In the downstairs market, trading at exact sizes is also positively related to trading at exact prices, but with a wider set.

7.2.4 Clustering on options markets

A small group of studies have investigated price clustering in options markets and recognised the important role the characteristics of options play in clustering. ap Gwilym *et al.* (1998b) provide the first empirical evidence of the price clustering of

[22] The downstairs market and the upstairs market are independent at the LST. The upstairs market has market makers who have no obligation to offer quotes on the downstairs market. Unlike the downstairs market, the upstairs market also has no minimum tick restrictions and allows traders to privately negotiate.

options and connect option characteristics with the price resolution hypothesis. In particular, price clustering is less likely to occur in low priced contracts which are out-of-the-money and closing expiration dates. Ni *et al.* (2005) do not directly examine option price clustering but focus on stock price clustering on option expiration dates. The result shows a connection between stock price clustering and options trading, finding that stock prices cluster around strike prices on expiration dates. They find evidence that the price clustering is related to hedge rebalancing and stock price manipulation.

Capelle-Blancard and Chaudhury (2007) contribute to the attraction hypothesis in the price clustering of the CAC40 index options by identifying that the frequencies of certain digits are greater than others. Although they confirm the positive relationship between price clustering and the price level, their results related to moneyness and maturity differ from previous studies. ap Gwilym *et al.* (1998b) show that price clustering is negatively related to maturity and out-of-the-money contracts, while Capelle-Blancard and Chaudhury (2007) find that these effects are statistically insignificant. To further explore the effects of options contract characteristics, ap Gwilym and Verousis (2013) investigate 28 equity options at the NYSE Euronext-LIFFE. They show a negative relationship between price clustering and maturity that supports the negotiation hypothesis. In addition, price clustering is also negatively influenced by the moneyness which is driven by the intrinsic value inherent in options. Higher moneyness leads to a greater proportion of intrinsic value in an option price that implies a greater certainty in option value and results in lower price clustering.

Although existing studies observe the effects of options characteristics on price clustering, no studies are extended to option size clustering and none investigate the connection between options price and size clustering. Since traders may sacrifice more than one dimension of liquidity to execute a transaction, the investigation of size clustering and its connection to price clustering are particularly important. Additionally, the characteristics of options contracts differ from equity and other financial assets as having specific impacts on clustering. This makes the investigation into size clustering in the options markets more important. In this context, we study clustering in the options market via these two aspects.

7.3 Sample of data

The data used in this chapter are obtained from the CBOE and include all trades and quotes records of options written on the components of the Dow Jones Industrial Average Index (DJI) from 3 January 2012 to 30 June 2014. The detailed procedures of data clearing are reported in Chapter 4 of Data Cleaning. In this empirical chapter, we include all the remaining contracts left, after data cleaning. In particular, the observations of calls and puts in this chapter are 879 million and 872 million, respectively.

---------------Insert Table 6-1----------------

Table 6-1 shows the distribution of quoted and traded prices and sizes. In particular, we ranked the frequencies of the last two digits of the price and the number of sizes. Since we only focus on clustering at the market level, the results are equally weighted across the options contracts. In Table 6-1, we report the top 20 frequencies of the last two digits of prices and the size of the number of contracts. The table demonstrates the striking feature that the most reported results of prices end with the digit 0 and the digit 5. These results of quoted and traded prices are consistent with previous literature (for example, Capelle-Blancard and Chaudhury, 2007; Meng et al., 2013). Hence, we define price clustering as prices ending with 0 and 5.

Regarding results of sizes, the top 20 frequencies of size clustering have very different distributions between traded and quoted sizes. In call options, the most frequent quoted and traded size are 10 contracts (1.23%) and 1 contract (25.5%), respectively. In put options, 21 contracts (1.2%) and 1 contract (24.75%) are the most frequently quoted and traded sizes. The top 20 frequencies of quoted sizes are generally higher than 10 contracts, while traded sizes are more likely to concentrate on small numbers of contracts (for example, 1 contract and 2 contracts). The percentage of traded sizes continually declines until four-contract trades, followed by a jump in the frequency of five-contract trades. In addition, the jump in the frequency is also observed in ten-contract trades, fifteen-contract trades, and twenty-contract trades. The trend of these frequencies is inconsistent with Alexander and Peterson (2007) and Verousis and ap Gwilym (2013b) who show size clustering to be in the multiples of 500, 1000, and 5000 shares. The inconsistent results are driven by the different financial assets that are used to estimate size clustering. In particular, Alexander and Peterson (2007) and Verousis and ap Gwilym (2013b) investigate size clustering of stock while this chapter study size clustering of options. Hence, their definition cannot be employed by this chapter. Regarding the distribution of trade size clustering, we define size clustering as multiples of five contracts.

7.4 Methodology

This section shows the research design that is used to examine the determinants of price and size clustering, especially for the interaction between them. Most investors face the challenge of optimising their trade prices and sizes when the market is not in their favour. According to previous studies, the investors may optimise prices and sizes simultaneously (Verousis and ap Gwilym, 2013b) or substitute optimum prices with optimum sizes (ap Gwilym and Meng, 2010). In order to investigate the relationship between price and size clustering, we follow the method of Meng *et al.* (2013) that uses a Three Stage Least Squares Regression (3SLS) model to investigate the relationship between price and size clustering. The adoption of this method allows us to address the endogeneity between price and size clustering which may lead to biases within the estimation results. Regarding the effects of options characteristics on clustering, we include maturity and moneyness variables. Hence, the following models are estimated:

$$PC_{x,t}^{y} = \alpha_1 + \beta_1 SC_{x,t}^{y} + \beta_2 SQRTFrq_{x,t}^{y} + \beta_3 Volm_{x,t}^{y} + \beta_4 AR_{x,t}^{y} + \beta_5 PL_{x,t}^{y} + \beta_6 MON_{x,t} + \beta_7 EXP_{x,t} + \varepsilon_{x,y}^{y} \quad (1)$$

$$SC_{x,t}^{y} = \alpha_1 + \beta_1 PC_{x,t}^{y} + \beta_2 Frq_{x,t}^{y} + \beta_3 AR_{x,t}^{y} + \beta_4 EOQ_{x,t} + \beta_5 MON_{x,t} + \beta_6 EXP_{x,t} + \varepsilon_{x,y}^{y} \quad (2)$$

where t = the days in the sample, x = call/put, y = trade/quote. Hence, we estimate different regressions for call and put options, as well as for quotes and trades that provide eight sets of estimation results in total. *PC* refers to the daily percentage of traded (quoted) price which has final digits of 0 and 5 and *SC* refers to the daily proportion of traded (quoted) with multiple of 5. *SORTFreq* refers to the inverse square root of the daily number of trades (quotes). *Freq* refers to the daily number of trades (quotes). *Volm* refers to the natural logarithm of the daily average of trade size (quoted depth). *AR* refers to the daily average absolute return that is calculated over hourly intervals. In particular, the trade (quote) return is the logarithmic change in the price (quote midpoint[23]). The return over the first hour of the trading day is excluded in order to avoid the effects of overnight information arrival during the opening period (see Ahn et al., 2005; Verousis and ap Gwilym, 2013). *PL* refers to the daily average trade price (quote midpoint). *EOQ* refers to a dummy variable that equals one when the trading day is in the last two weeks of each quarter end. *MON* refers to the daily average moneyness of the options contracts. The moneyness is calculated as S/K[24]. *EXP* refers to the daily average time-to-maturity of the option contracts. All measures are weighted equally across options contracts in the sample. The expected signs of independent variables are shown in Table 6-2.

--------------Insert Table 6-2---------------

7.5 Empirical results

This section includes three sections. The first section shows the descriptive statistics of independent variables, followed by the evidence of clustering. The last section reports the results related to the relationship between price and size clustering.

7.5.1 Summary statistics

--------------Insert Table 6-3---------------

Table 6-3 represents descriptive statistics. The results show different performances between call and put options. Call options have more trades and quotes than put options on average. Additionally, the number of trades is lower than the number of quotes. The trading volume is also significantly lower than the depth which is considered as the quoted volume. ap Gwilym and Meng (2010) explain that the small

[23] Quote midpoint is the mean of the quoted ask and bid prices.
[24] S refers to the contemporaneous price of the underlying stock price, and K refers to the option strike price.

trading volume allows investors to avoid losses in trading large sizes with a better-informed counterparty. Regarding the daily average absolute return, it reflects the volatility and has no significance between calls and puts, as well as trade and quote. Regarding the price level, calls have higher prices than puts and quoted prices are higher than traded prices. Furthermore, the daily average moneyness and day-to-maturity for calls (puts) are 1.23 (1.12) and 173 (174) days, respectively.

7.5.2 Evidence of clustering

---------------Insert Table 6-4----------------

Table 6-4 shows the percentages of prices with tail numbers of 0 and 5 and of sizes with a multiple of 5 that represents the evidence of price and size clustering. It reports that quoted price clustering (65% for calls and 55% for puts) is stronger than traded price clustering (35% for calls and 32% for puts). In terms of size clustering, the extent of size clustering in trades (27% for calls and 29% for puts) is greater than in quotes (22% for calls and 22% for puts). In particular, the level of size clustering in the CBOE market is lower than in other markets, such as the equity market (Lien *et al.*, 2019) and the CDS market (Meng *et al.*, 2013).

---------------Insert Table 6-5----------------

Table 6-5 represents the distribution of clustering across different moneyness types and maturity levels. The results clearly show that the level of price and size clustering vary significantly across moneyness and maturity as suggested by ap Gwilym and Verousis (2013). Regarding moneyness, we find that at-the-money contracts have the highest level of the price and size clustering. Additionally, the extent of price clustering in all cases increase with the level of moneyness except puts' trade price clustering. However, traded and quoted sizes of put options are more clustered when the moneyness level is low. Regarding maturity, in all cases, price and size clustering is higher when the day-to-maturity is low. In particular, the contracts with day-to-maturity below 120 days report the highest price and size clustering. However, it should be noted that quote price clustering has relatively higher levels regarding long-maturity contracts as compared with trade price clustering.

7.5.3 Relationship between price and size clustering

--------------Insert Figure 6-1---------------

In this section, we will investigate the determinant of clustering and further explore the relationship between price and size clustering. Figure 6-1 represents the time series distribution of price and size clustering during the sample period. The plots roughly display the relationship between them. Regarding traded prices and sizes, their relationship is different between calls and puts. The correlation results confirm that the coefficient for calls is 0.3 and for puts is -0.018. Regarding quoted prices and sizes, they are smoother than traded prices and sizes. In addition, the level of quoted price clustering for calls decreases over the sample period. The correlation coefficient between quoted price and size clustering is 0.28 for calls and 0.11 for puts.

--------------Insert Table 6-6---------------

Before estimating the 3LSL models, we conduct the Durbin-Wu-Hausman test for examining the endogeneity of price and seize clustering. In particular, we use the first stage to generate the residuals of endogenous variables that are applied in the second stage for testing endogeneity. As shown in Table 6-6, apart from the results of the residuals of quoted size clustering, all the results are highly significant and suggest that the OLS estimation results are biased and inconsistent. Hence, the 3SLS regression model should be employed. Regarding the insignificant results, they suggest that there is no endogeneity bias in the OLS estimation. However, in order to ensure consistency, the 3SLS regression model is also employed for the quoted size clustering.

---------------Insert Table 6-7----------------

Table 6-7 shows the 3SLS regression estimates. Unlike previous studies (e.g. Meng et al., 2013; Verousis and ap Gwilym, 2013), we find both positive and negative relationship between price clustering and size clustering. In other words, we report not only the complementary effect but also the substitution effect between price and size clustering. For call options, we find positive effects of price clustering on size clustering for trades and quotes. This complementary effect of price clustering is consistent with Alexander and Peterson (2007) and Verousis and ap Gwilym (2013b), namely that the increased level of price clustering could increase the level of size clustering. The negotiations hypothesis is behind this positive relationship that price and size rounding could reduce the possibility of misreporting and thereby speed up negotiations (Alexander and Peterson, 2007). This implies that investors may sacrifice price and size dimensions of liquidity and thereby pursue the time dimension when trading call options. However, Table 6-7 also reports a negative effect of quoted size clustering on quoted price clustering. This implies that investors are less likely to use rounded prices when trading at exact sizes of put options. The adoption of a wider set of prices could compensate for the loss of optimality (Verousis and ap Gwilym, 2013b). In addition, we did not find that trade size clustering has a significant effect on trade price clustering.

The relationship between price and size clustering is reversed for put options. Size clustering significantly reduces price clustering in terms of trades and quotes, but this is not a mutual effect; price clustering has no significant impact on size clustering.

The negative relationship is consistent with Meng *et al.* (2013) who find a substitute effect of price clustering on size clustering. Since investors cannot optimise all three dimensions of liquidity, they should consider the trade-off between price and size (Moulton, 2005). Our results indicate that investors are willing to sacrifice price to get desired quantities and trade more quickly. This finding corresponds to the results in Table 6-4 that price clustering is higher than size clustering. The frequency variables (*Frq* and *SQRTFrq*) are related to the time dimension of liquidity. We find that *Frq* is significantly and negatively related to size clustering in all circumstances, while *SQRTFrq* significantly increases calls' trade price clustering and puts' quote price clustering. The findings in frequency variables are consistent with the negotiation hypothesis that high frequency implies high liquidity and thereby reduces the negotiation costs (Harris, 1991; Lien *et al.*, 2019). However, we also find a significant negative relationship between *SQRTFrq* and calls' quote price clustering. According to ap Gwilym *et al.* (1998b) and Alexander and Peterson (2007), a smaller price set would be adopted by traders in busy trading periods.

In Table 6-7, we also investigate other determinants of clustering beyond the dimensions of liquidity. Traded size (*Volm*) shows reversed impacts on calls' and puts' trade price clustering. In particular, it is positively related to trade price clustering in call options but negatively related to that in put options. The positive effect of traded size suggests that the volume is explained as a proxy of uncertainty that supports the negotiation hypothesis of Harris (1991). Besides, the quoted size has no significant impacts on clustering. The volatility (*AR*) has positive impacts on price and size clustering in all circumstance expect for puts' trade price and size clustering. Capelle-Blancard and Chaudhury (2007) confirm the positive effects on price clustering and explain the relationship using the price resolution hypothesis. The high volatility implies price uncertainty that encourages investors to use a smaller set of prices. Besides, ap Gwilym and Meng (2010) support the notion that the greater volatility is associated with greater size clustering. Price level (*PL*) has a positive impact on price clustering in all regression models that is consistent with Harris (1991) and Meng *et al.* (2013). In particular, the lower negotiation benefits can cause clustering because of the low negotiation benefits. The results for end-of-quarter dummy (*EOQ*) also follow our explanations that the dummy variable has negative impacts on size clustering. Moulton (2005) explains that investors usually desire to reach their target inventory position at the end of a quarter. Hence, they tend to sacrifice liquidity and trade at exact size.

Regarding the effects of moneyness (*MON*) and maturity (*EXP*), these do not show similar impacts on clustering in all regression models. The effects of moneyness are positive in six models and statistically significant in four of these results. The results are in contrast to the findings of ap Gwilym and Verousis (2013). Our results indicate that in-the-money contracts are more clustered than out-of-the money contracts after controlling for other factors. In particular, in-the-money contracts have a higher price level as compared with other contracts when other aspects are equal. Hence, these contracts have greater price clustering (Cheng *et al.*, 2005; ap Gwilym and Verousis, 2013). The maturity variable is significantly associated with price clustering in six models and four of these results show positive relationship. This finding is also opposite with ap Gwilym and Verousis (2013) who argue that investors have an increased urgency to trade short-maturity contracts and thereby exhibit more clustering.

7.6 Conclusion

This chapter explores price and size clustering in the CBOE. Previous literature provides limited results for the clustering in equity options markets, especially for size clustering. Hence, we initially study the extent of price and size clustering. There is clearly price clustering among trades and quotes with tail numbers of 0 and 5. In particular, 35% (32%) of traded prices in call (put) options end with 0 or 5. The percentage for quoted prices shows higher results: 65% and 55%, respectively. With the top 20 frequencies of traded and quoted size, we find that the traded and quoted size represent different distributions as compared with price clustering. Since previous studies have no investigations on size clustering in options markets, we define the size clustering as a multiple of 5 and report that the extent of traded (quoted) size clustering is 27% (22%) for calls and 29% (22%) for puts.

We subsequently investigate the relationship between price and size clustering and the effects of other determinants. The results do not show a mutual relationship between price and size clustering in most regression models. Besides, the relationship is also not consistent across call and put options. We find a positive effect of price clustering on size clustering from call options and the negative effects of size clustering on price clustering from put options. For other determinants, the results also lack consistency among all cases. In general, the estimation results of frequency variables show lower price and size clustering during the market liquid period. The higher negotiation costs and lower negotiation benefits can also lead to price and size clustering as finding the positive effects of volatility and price level. The negative relationships between end-of-quarter dummy and size clustering indicate that investors are willing to sacrifice liquidity in order to achieve the desired inventory position at the end of a quarter. Finally, we also find that moneyness and maturity can influence clustering. In particular, the out-of-the money contracts and near-to-maturity contracts show less price and size clustering.

This chapter explores the trade-off between dimensions of liquidity through investigating the price and size clustering in CBOE. In call options, investors are more likely to sacrifice price and desire to trade at exact price. In put options, they tend to sacrifice both price and size in order to speed up negotiations and pursue the time dimension of liquidity. Additionally, the characteristics of options (moneyness and maturity) can influence both price and size clustering. The further study could investigate how price and size clustering influence investors' asset selection in active portfolio management. In this context, the further study could explore the trading strategy under clustering.

Table 7-1 Distribution of price and size

Panel A: Call options

Quote price	Percentage	Trade price	Percentage	Quote size	Percentage	Trade size	Percentage
0	3.44	0	1.97	1	0.95	1	25.50
0.05	3.82	0.05	2.18	10	1.23	2	11.75
0.1	3.85	0.1	2.27	20	1.05	3	7.22
0.15	3.53	0.15	2.01	21	1.19	4	5.46
0.2	3.64	0.2	2.12	22	0.61	5	7.66
0.25	3.38	0.25	1.99	30	0.68	6	2.87
0.3	3.45	0.3	2.04	31	0.84	7	1.91
0.35	3.17	0.35	1.75	32	0.79	8	2.00
0.4	3.37	0.4	1.86	33	0.50	9	1.45
0.45	3.11	0.45	1.67	36	0.51	10	9.61
0.5	3.32	0.5	2.14	40	0.68	11	2.85
0.55	3.05	0.55	1.54	41	0.63	12	1.20
0.6	3.23	0.6	1.71	42	0.71	13	0.70
0.65	2.97	0.65	1.48	43	0.49	14	0.70
0.7	3.15	0.7	1.64	44	0.48	15	1.68
0.75	2.93	0.75	1.50	50	0.61	16	0.73
0.8	3.13	0.8	1.58	51	0.55	20	2.54
0.85	2.89	0.85	1.32	52	0.54	25	0.91
0.9	3.08	0.9	1.50	60	0.50	30	0.91
0.95	2.84	0.95	1.24	61	0.48	50	0.87

Panel B: Put options

Quote price	Percentage	Trade price	Percentage	Quote size	Percentage	Trade size	Percentage
0	2.73	0	1.57	1	0.96	1	24.75
0.05	3.62	0.01	1.35	10	1.04	2	10.31
0.1	3.45	0.02	1.49	20	0.89	3	6.62
0.15	3.22	0.03	1.55	21	1.21	4	5.08
0.2	3.02	0.04	1.50	22	0.62	5	7.40
0.25	3.07	0.05	2.40	30	0.63	6	2.82
0.3	2.91	0.1	2.23	31	0.87	7	1.91
0.35	2.87	0.15	2.00	32	0.81	8	2.02
0.4	2.76	0.2	2.03	33	0.49	9	1.50
0.45	2.67	0.25	1.86	36	0.49	10	10.75
0.5	2.66	0.3	1.89	40	0.63	11	3.34
0.55	2.59	0.35	1.65	41	0.64	12	1.29
0.6	2.56	0.4	1.77	42	0.72	13	0.75
0.65	2.41	0.45	1.53	43	0.48	14	0.76
0.7	2.39	0.5	1.83	44	0.47	15	1.86
0.75	2.45	0.55	1.45	50	0.60	16	0.79
0.8	2.33	0.6	1.58	51	0.55	20	2.72
0.85	2.29	0.65	1.35	52	0.54	25	0.95
0.9	2.27	0.7	1.45	60	0.48	30	0.99
0.95	2.24	0.8	1.36	61	0.48	50	0.91

Note: This table shows the frequencies of the last two digits of price and the numbers of sizes. Panel A shows the distribution of call options and Panel B shows the distribution of put options.

Table 7-2 Expected effects of determinant variables

Variable	Hypothesis	Expected signs
PC	Substitution effects (ap Gwilym and Meng, 2010)	-
SC	Substitution effects (ap Gwilym and Meng, 2010)	-
Freq	Negotiation hypothesis (Harris, 1991)	-
SQRTFreq	Negotiation hypothesis (Harris, 1991)	+
Volm	Negotiation hypothesis (Harris, 1991)	-
AR	Resolution hypothesis (Capelle-Blancard and Chaudhury, 2007)	+
PL	Negotiation hypothesis (Meng et al., 2013)	+
EOQ	Quarter-end hypothesis (Moulton, 2005)	-
MON	Intrinsic value (ap Gwilym and Verousis, 2013)	-
EXP	Negotiation hypothesis (ap Gwilym and Verousis, 2013)	-

Table 7-3 Descriptive statistics

	Call				Put			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Frq^T	19.04	4.87	8.89	39.41	13.93	3.33	6.57	26.25
Frq^Q	569.39	99.02	256.85	850.20	556.39	103.07	249.55	873.11
$Volm^T$	17.99	12.35	10.41	146.77	15.73	2.83	10.71	40.63
$Volm^Q$	327.47	74.23	217.51	589.14	357.82	71.09	208.28	605.21
AR^T	0.017	0.003	0.010	0.028	0.014	0.003	0.009	0.030
AR^Q	0.019	0.005	0.011	0.039	0.021	0.005	0.012	0.047
PL^T	4.76	0.71	3.37	7.44	2.53	0.32	1.78	3.86
PL^Q	12.51	1.38	8.98	15.75	7.77	0.80	6.40	10.24
MON	1.23	0.06	1.10	1.34	1.12	0.05	1.02	1.22
EXP	173.46	31.30	106.62	233.22	174.78	33.94	103.15	237.21

Note: This table shows the descriptive statistics of variables. Frq^T (Frq^Q) refers to the daily average number of trades (quotes). $Volm^T$ ($Volm^Q$) refers to the daily average of trading volume (depth). AR^T (AR^Q) refers to the daily mean absolute return of trades (quotes). PL^T (PL^Q) refers to the daily average price level (mid-quote price). MON refers to the daily average moneyness of the option contracts which are calculated as S/K. EXP refers to the daily average day-to-maturity of the option contracts.

Table 7-4 Price and size clustering

	Call				Put			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
PC^T	35.49	1.65	30.83	40.18	32.51	1.91	27.43	38.31
PC^Q	65.37	4.39	53.71	73.70	54.51	2.84	47.55	63.77
SC^T	26.99	1.15	23.94	30.05	28.98	1.65	24.41	34.12
SC^Q	22.00	0.64	20.34	24.00	21.71	0.56	20.11	23.59

Note: PC^P (PC^Q) refers to the percentage of price clustering in trades (quotes). $SC^T$$(SC^Q)$ refers to the percentage of size clustering in trades (quotes).

Table 7-5 Clustering across moneyness and maturity

Panel A: Trade price clustering

	Call options						Put options					
Maturity	DOTM	OTM	ATM	ITM	DITM	Total	DOTM	OTM	ATM	ITM	DITM	Total
1	0.39	1.95	20.14	2.47	1.68	26.64	1.99	2.96	17.51	1.21	0.37	24.04
2	0.40	0.80	2.40	0.37	0.55	4.52	1.24	0.69	1.86	0.27	0.13	4.19
3	0.69	0.61	1.46	0.43	1.14	4.33	2.00	0.59	1.15	0.25	0.29	4.28
Total	1.48	3.36	24.00	3.28	3.38	35.49	5.22	4.25	20.52	1.73	0.78	32.51

Panel B: Quote price clustering

	Call options						Put options					
Maturity	DOTM	OTM	ATM	ITM	DITM	Total	DOTM	OTM	ATM	ITM	DITM	Total
1	0.57	1.26	9.38	7.08	18.60	36.89	2.42	1.82	9.73	6.34	7.96	28.26
2	0.98	0.92	3.10	1.88	6.90	13.78	2.42	1.02	3.50	2.03	3.32	12.29
3	1.77	1.18	2.91	1.54	7.28	14.69	3.69	1.24	2.98	1.66	4.38	13.96
Total	3.32	3.36	15.39	10.50	32.78	65.37	8.53	4.07	16.21	10.03	15.66	54.51

Panel C: Trade size clustering

	Call options						Put options					
Maturity	DOTM	OTM	ATM	ITM	DITM	Total	DOTM	OTM	ATM	ITM	DITM	Total
1	0.47	2.09	17.74	1.14	0.68	22.12	2.69	3.36	16.59	0.71	0.24	23.60
2	0.36	0.66	1.37	0.19	0.22	2.79	1.31	0.57	0.95	0.14	0.10	3.06
3	0.48	0.34	0.65	0.18	0.45	2.09	1.41	0.28	0.44	0.09	0.10	2.32
Total	1.31	3.08	19.76	1.50	1.35	26.99	5.41	4.21	17.98	0.94	0.44	28.99

Panel D: Quote size clustering

	Call options						Put options					
Maturity	DOTM	OTM	ATM	ITM	DITM	Total	DOTM	OTM	ATM	ITM	DITM	Total
1	0.54	1.12	5.69	1.81	4.52	13.68	2.43	1.64	5.72	1.70	2.07	13.56
2	0.75	0.48	0.99	0.42	1.53	4.17	1.81	0.43	0.97	0.44	0.70	4.35
3	0.95	0.35	0.80	0.43	1.60	4.15	1.63	0.29	0.66	0.33	0.89	3.81
Total	2.25	1.96	7.48	2.67	7.65	22.00	5.87	2.36	7.35	2.47	3.66	21.71

Note: This table shows the average percentages of clustering across moneyness and maturity. Panel A, B, C, and D show the distribution of trade price clustering, quote price clustering, trade size clustering, and price size clustering. Price clustering is defined as the price with ending number of 0 and 5 and size clustering is defined as the multiples of five contracts. Maturity 1, 2, and 3 refer to the contacts with day-to-maturity below 120 days, between 121 to 240 days, and over 241 days. DOTM, OTM, ATM, ITM, and DITM refer to deep out-of-the-money, out-of-the-money, at-the-money, in-the-money, and deep in-the-money contracts.

Table 7-6 Tests for endogenous

Panel A: Call options

Type	Trade		Quote	
Dependent	Price clustering	Size clustering	Price clustering	Size clustering
Durbin	0.0084	0.00001	0.0011	0.0571
Wu-Hausman	0.0087	0.00001	0.0012	0.0584

Panel B: Put options

Type	Trade		Quote	
Dependent	Price clustering	Size clustering	Price clustering	Size clustering
Durbin	0.00001	0.035	0.00001	0.4888
Wu-Hausman	0.00001	0.036	0.00001	0.4917

Note: This table shows the p-values of the Durbin-Wu-Hausman test.

Table 7-7 3SLS regression estimates

Panel A: Call options

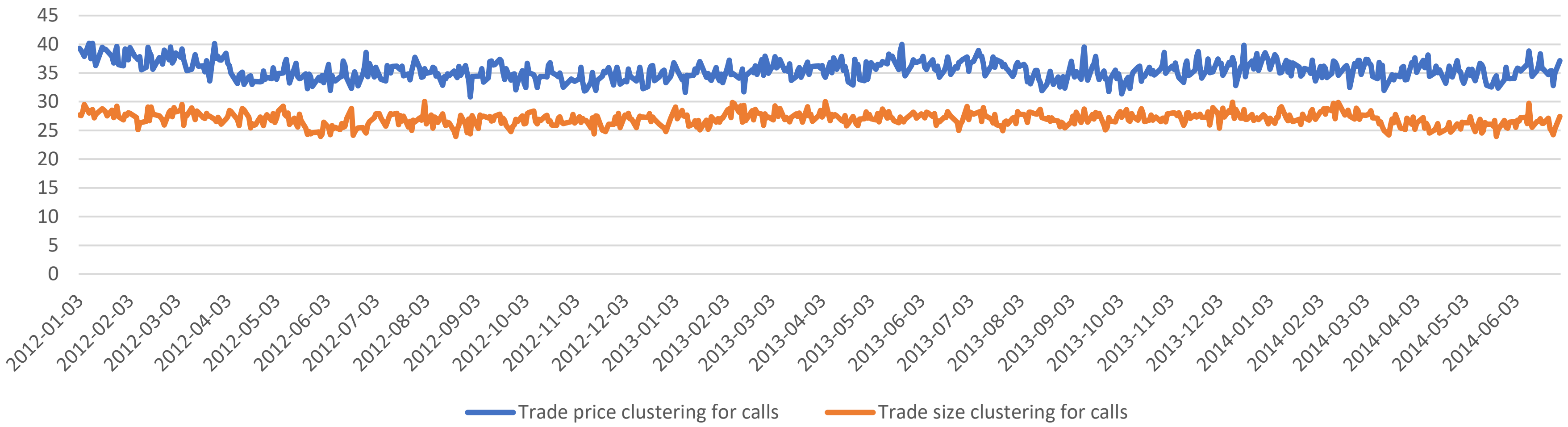
Type	Trade				Quote			
Dependent	Price clustering		Size clustering		Price clustering		Size clustering	
SC	-0.36299				-2.23337	*		
PC			0.404426	***			0.256874	***
Frq			-0.0752	***			-0.00209	***
SQRTFrq	11.30819	***			-90.8184	**		
Volm	0.456112	**			-0.93139			
AR	62.63186	**	49.4516	***	146.4478	***	18.93952	**
PL	1.746258	***			0.981225	***		
EOQ			-0.79745	***			-0.25155	***
MON	3.873817		4.06754	**	59.28483	***	-15.6406	***
EXP	-0.01477	***	0.004124	*	0.006305		-0.0032	**
Constant	29.8008	***	7.612792	***	34.58022		25.86797	***
R-square	0.3374		0.1896		0.7425		0.6511	

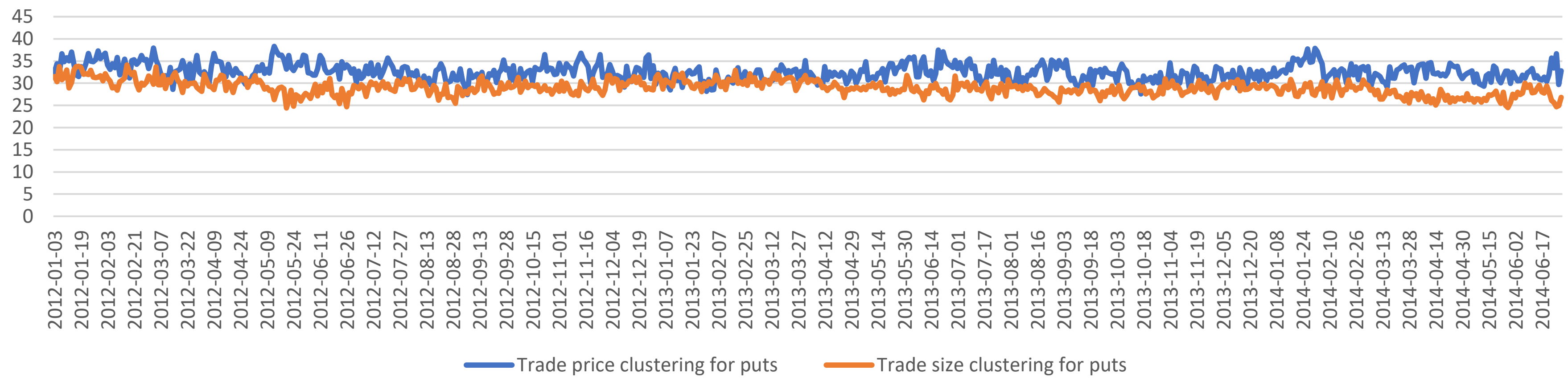
Panel B: Put options

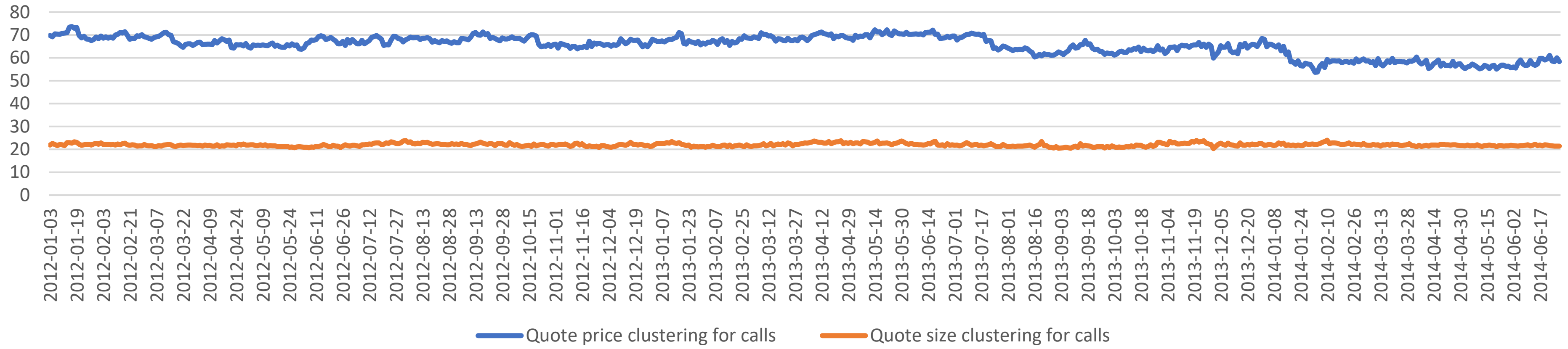
Type	Trade				Quote			
Dependent	Price clustering		Size clustering		Price clustering		Size clustering	
SC	-0.83538	***			-7.65071	***		
PC			-0.08576				0.005367	
Frq			-0.11162	***			-0.00275	***
SQRTFrq	8.1854				355.8148	***		
Volm	-0.86943	*			0.474526			
AR	-46.6662		27.31237		290.3135	***	35.69778	***
PL	3.640306	***			2.45259	***		
EOQ			-0.94134	***			-0.01664	
MON	15.87729	***	12.02196	***	4.193118		-1.97636	***
EXP	0.001598		0.017563	***	0.08389	***	0.005674	***
Constant	30.29502	***	16.52608	***	158.0702	***	23.41963	***
R-square	0.3372		0.3381		0.1887		0.1868	

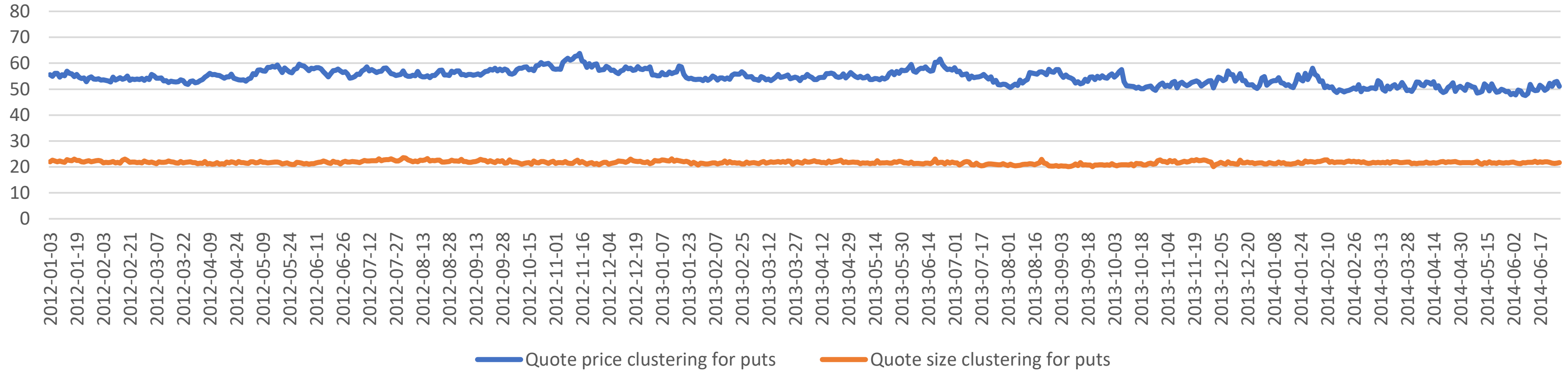
Note: ***, **, and * refer to the significant level at 1%, 5%, and 10%. The number of observations is 626. These tables show the 3SLS results of price-size clustering. Panel A shows the results of trade price-size clustering and quote price-size clustering based on call options. Panel B shows the results of trade price-size clustering and quote price-size clustering based on put options. SC refers to the trade/quote size clustering. PC refers to the trade/quote price clustering. Frq refers to the daily number of trades/quotes. SQRTFre refers to the inverse square root of the daily number of trades/quotes. Volm refers to the natural logarithm of the daily average trade volume/depth. AR refers to the daily average absolute return of trades/quotes. EOQ refers to a dummy variable that equals one when the trading day is in the last two weeks of each quarter end. PL refers to the daily average trade price/mid-quote price. MON refers to the daily average moneyness of the option contracts which are calculated as S/K. EXP refers to the daily average time-to-maturity of the option contracts.

Figure 7-1 Daily distribution of price and size clustering









Note: These plots represent the daily distribution of price and size clustering during the sample period. The first two plots represent the trade price and trade size clustering for calls and puts, respectively. The last two plots represents the quote price and quote size clustering for calls and puts, respectively.

Chapter 8: Conclusion

This thesis investigates commonality in options liquidity, predictability of information conveyed in options trades, and options price and size clustering through a high-frequency dataset obtained from the Chicago Board Options Exchange (CBOE) that includes all options contracts written on 30 components of the Dow Jones Industrial Average (DJIA) from January 2012 to June 2014.

Chapter 5 (1st empirical chapter) examines the determinants of equity options liquidity commonality in CBOE. Using a principal component analysis, we find supportive evidence that there are liquidity co-movements in CBOE. Additionally, we also show the intraday co-movement pattern of equity options liquidity in this quote-driven market. During the trading day, a stronger commonality in liquidity is shown at the beginning of the trading day. This pattern may be driven by the overnight information arrival (Verousis *et al.*, 2015). We have further investigated the fundamental sources of commonality in liquidity and find that the co-movements across options contracts are driven by inventory risks. In particular, the estimation results show that liquidity commonality increases with market volatility and market declines. Besides, we also find that trading volume, volume-weighted duration, and the number of transactions can significant influence liquidity commonality. These variables are considered as proxies for information asymmetry. Since their impacts are not following our expectation, we do not conclude information asymmetry as the source of commonality in liquidity.

Chapter 6 (2nd empirical chapter) investigates the informational role of equity options in underlying stock markets. By employing the method of Kalaitzoglou and Ibrahim (2015), we use the Smooth Transition Autoregressive Conditional Weighted Duration (STM-ACWD) model and the volume-weighted duration to capture predictive information. With the captured information, we created three sets of information measures that were used to study the predictability of options order flow. We find supportive evidence that options order flows contain information because these information measures are significantly related contemporaneous and future stock price movements. In particular, all information measures can have impact on the stock volatility and return on the same day and the next day. However, each information measure contains a different level of predictive information. Additionally, there are differences in predictability of calls and puts. We show that the predictive horizon of puts information measures is longer for the stock return but shorter for the stock volatility when comparing with calls information measures.

Chapter 7 (3rd empirical chapter) provides a study that focuses on price clustering and size clustering. In this study, we find that both call and put options show price and size clustering. There are small differences in the levels of clustering observed in calls and puts. However, quote price clustering is clearly higher than trade price clustering. Furthermore, we use Three Stage Least Squares Regression models to examine the determinants of price and size clustering. The estimation results show the relationship between price and size clustering. Unlike previous studies (e.g. Verousis and ap Gwilym, 2013), we find that price clustering can positively influence size clustering in call options, and size clustering can negatively influence price clustering in put options. Besides, the estimation results show that the characteristics of options contracts are related to price and size clustering. In particular, the out-of-the money contracts and near-to-maturity contracts show less price and size clustering.

Although this thesis contributes to literature through each empirical chapters, there are opportunities for further research. For example, further research could investigate the effects of firm-specific information on liquidity commonality. The former may influence commonality via the trading behaviours of investors. Moshirian *et al.* (2017) showed that the firm with great information transparency can reduce the level of private information and thereby increase commonality in liquidity. Moreover, further studies could extend the predictive horizon in the Chapter 6 (2nd empirical chapter) to a longer horizon (e.g. 30 days) and thereby examine the forecasting ability of information contained in options order flow. Finally, it would be interesting to investigate the intraday distributions of price and size clustering, which should provide implications for options traders.

Chapter 9: References

Ackert, L.F. and Tian, Y.S. (2001) 'Efficiency in index options markets and trading in stock baskets', *Journal of banking & finance*, 25(9), pp. 1607-1634.

Aggarwal, R. and Gruca, E. (1993) 'Intraday trading patterns in the equity options markets', *Journal of Financial Research*, 16(4), pp. 285-297.

Agyei-Ampomah, S. and Mazouz, K. (2011) 'The comovement of option listed stocks', *Journal of Banking & Finance*, 35(8), pp. 2056-2069.

Aït-Sahalia, Y. and Jacod, J. (2014) *High-frequency financial econometrics*. Princeton University Press.

Ajinkya, B.B. and Gift, M.J. (1985) 'Dispersion of financial analysts' earnings forecasts and the (option model) implied standard deviations of stock returns', *The Journal of Finance*, 40(5), pp. 1353-1365.

Alexander, C. (2008) *Market Risk Analysis: Pricing, Hedging and Trading Financial Instruments*. John Wiley & Sons.

Alexander, G.J. and Peterson, M.A. (2007) 'An analysis of trade-size clustering and its relation to stealth trading', *Journal of Financial Economics*, 84(2), pp. 435-471.

Amihud, Y. and Mendelson, H. (1982) 'Asset price behavior in a dealership market', *Financial Analysts Journal*, 38(3), pp. 50-59.

Amin, K.I. and Lee, C.M. (1997) 'Option trading, price discovery, and earnings news dissemination', *Contemporary Accounting Research*, 14(2), pp. 153-192.

Anand, A. and Chakravarty, S. (2007) 'Stealth trading in options markets', *Journal of Financial and Quantitative Analysis*, 42(1), pp. 167-187.

Anand, A., Hua, J. and McCormick, T. (2016) 'Make-take structure and market quality: Evidence from the US options markets', *Management Science*, 62(11), pp. 3271-3290.

Anand, A. and Weaver, D.G. (2006) 'The value of the specialist: Empirical evidence from the CBOE', *Journal of Financial Markets*, 9(2), pp. 100-118.

Anthony, J.H. (1988) 'The Interrelation of Stock and Options Market Trading ‐ Volume Data', *The Journal of Finance*, 43(4), pp. 949-964.

ap Gwilym, O. and Alibo, E. (2003) 'Decreased price clustering in FTSE100 futures contracts following a transfer from floor to electronic trading', *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 23(7), pp. 647-659.

ap Gwilym, O., Buckle, M. and Thomas, S. (1997) 'The intraday behavior of bid-ask spreads, returns, and volatility for FTSE-100 stock index options', *Journal of Derivatives* 4, pp. 20-32.

ap Gwilym, O., Clare, A. and Thomas, S. (1998a) 'The bid‐ask spread on stock index options: An ordered probit analysis', *Journal of Futures Markets*, 18(4), pp. 467-485.

ap Gwilym, O., Clare, A. and Thomas, S. (1998b) 'Extreme price clustering in the London equity index futures and options markets', *Journal of Banking & Finance*, 22(9), pp. 1193-1206.

ap Gwilym, O. and Meng, L. (2010) 'Size clustering in the FTSE100 index futures market', *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 30(5), pp. 432-443.

ap Gwilym, O. and Verousis, T. (2013) 'Price Clustering in Individual Equity Options: Moneyness, Maturity, and Price Level', *Journal of Futures Markets*, 33(1), pp. 55-76.

Arnold, T., Erwin, G., Nail, L. and Nixon, T. (2006) 'Do option markets substitute for stock markets? Evidence from trading on anticipated tender offer announcements', *International Review of Financial Analysis*, 15(3), pp. 247-255.

Aşçıoğlu, A., Comerton‐Forde, C. and McInish, T.H. (2007) 'Price clustering on the Tokyo stock exchange', *Financial Review*, 42(2), pp. 289-301.

Atilgan, Y., Bali, T.G. and Demirtas, K.O. (2015) 'Implied volatility spreads and expected market returns', *Journal of Business & Economic Statistics*, 33(1), pp. 87-101.

Back, K. (1993) 'Asymmetric information and options', *The Review of Financial Studies*, 6(3), pp. 435-472.

Bai, M. and Qin, Y. (2015) 'Commonality in liquidity in emerging markets: Another supply-side explanation', *International Review of Economics & Finance*, 39, pp. 90-106.

Bakshi, G., Kapadia, N. and Madan, D. (2003) 'Stock return characteristics, skew laws, and the differential pricing of individual equity options', *The Review of Financial Studies*, 16(1), pp. 101-143.

Bakshi, G. and Madan, D. (2000) 'Spanning and derivative-security valuation', *Journal of Financial Economics*, 55(2), pp. 205-238.

Bali, T.G. and Hovakimian, A. (2009) 'Volatility spreads and expected stock returns', *Management Science*, 55(11), pp. 1797-1812.

Ball, C.A., Torous, W.N. and Tschoegl, A.E. (1985) 'The degree of price resolution: The case of the gold market', *Journal of Futures Markets*, 5(1), pp. 29-43.

Barclay, M.J. and Warner, J.B. (1993) 'Stealth trading and volatility: Which trades move prices?', *Journal of Financial Economics*, 34(3), pp. 281-305.

Battalio, R., Hatch, B. and Jennings, R. (2001) *Does a national market system exist for US exchange-listed equity options? An analysis of multiple-traded equity options*. Working Paper.

Battalio, R. and Schultz, P. (2006) 'Options and the bubble', *The Journal of Finance*, 61(5), pp. 2071-2102.

Battalio, R. and Schultz, P. (2011) 'Regulatory uncertainty and market liquidity: The 2008 short sale ban's impact on equity option markets', *The Journal of Finance*, 66(6), pp. 2013-2053.

Beltran-Lopez, H., Grammig, J. and Menkveld, A.J. (2012) 'Limit order books and trade informativeness', *The European Journal of Finance*, 18(9), pp. 737-759.

Benkert, C. (2004) 'Explaining credit default swap premia', *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 24(1), pp. 71-92.

Bergsma, K., Fodor, A., Singal, V. and Tayal, J. (2019) 'Option trading after the opening bell and intraday stock return predictability', *Financial Management*, forthcoming.

Berkman, H. (1993) 'The market spread, limit orders, and options', in *Journal of Financial Services Research*. 6, pp. 399-415.

Bernales, A. (2017) 'The success of option listings', *Journal of Empirical Finance*, 40, pp. 139-161.

Bernales, A., Cañón, C. and Verousis, T. (2018a) 'Bid–ask spread and liquidity searching behaviour of informed investors in option markets', *Finance Research Letters*, 25, pp. 96-102.

Bernales, A., Chen, L. and Valenzuela, M. (2017) 'Learning and forecasts about option returns through the volatility risk premium', *Journal of Economic Dynamics and Control*, 82, pp. 312-330.

Bernales, A., Cortazar, G., Salamunic, L. and Skiadopoulos, G. (2018b) 'Learning and Index Option Returns', *Journal of Business & Economic Statistics*, pp. 1-13.

Bernales, A. and Guidolin, M. (2014) 'Can we forecast the implied volatility surface dynamics of equity options? Predictability and economic value tests', *Journal of Banking & Finance*, 46, pp. 326-342.

Bernales, A. and Guidolin, M. (2015) 'Learning to smile: Can rational learning explain predictable dynamics in the implied volatility surface?', *Journal of Financial Markets*, 26, pp. 1-37.

Bernales, A. and Valenzuela, M. (2016) *Implied Correlation and Market Returns*. Working paper.

Bernales, A., Verousis, T. and Voukelatos, N. (2016) 'Do investors follow the herd in option markets?', *Journal of Banking & Finance*, Forthcoming in Journal of Banking & Finance.

Bhattacharya, M. (1987) 'Price changes of related securities: The case of call options and stocks', *Journal of Financial and Quantitative Analysis*, 22(01), pp. 1-15.

Biais, B., Glosten, L. and Spatt, C. (2005) 'Market microstructure: A survey of microfoundations, empirical results, and policy implications', *Journal of Financial Markets*, 8(2), pp. 217-264.

Billings, M.B. and Jennings, R. (2011) 'The option market's anticipation of information content in earnings announcements', *Review of Accounting Studies*, 16(3), pp. 587-619.

Black, F. (1971) 'Toward a fully automated stock exchange, part I', *Financial Analysts Journal*, 27(4), pp. 28-35.

Black, F. (1975) 'Fact and fantasy in the use of options', *Financial Analysts Journal*, 31(4), pp. 36-41.

Black, F. and Scholes, M. (1973) 'The pricing of options and corporate liabilities', *Journal of political economy*, 81(3), pp. 637-654.

Blasco, N., Corredor, P. and Ferreruela, S. (2012) 'Does herding affect volatility? Implications for the Spanish stock market', *Quantitative Finance*, 12(2), pp. 311-327.

Blasco, N., Corredor, P. and Santamaría, R. (2010) 'Does informed trading occur in the options market? Some revealing clues', *Accounting & Finance*, 50(3), pp. 555-579.

Blau, B.M., Nguyen, N. and Whitby, R.J. (2014) 'The information content of option ratios', *Journal of Banking & Finance*, 43, pp. 179-187.

Blau, B.M., Van Ness, B.F. and Van Ness, R.A. (2009) 'Intraday stealth trading: which trades move prices during periods of high volume?', *Journal of Financial Research*, 32(1), pp. 1-21.

Blau, B.M. and Wade, C. (2013) 'Comparing the information in short sales and put options', *Review of Quantitative Finance and Accounting*, 41(3), pp. 567-583.

Bollen, N.P. (1998) 'A note on the impact of options on stock return volatility', *Journal of banking & Finance*, 22(9), pp. 1181-1191.

Bondarenko, O. (2014) 'Why are put options so expensive?', *The Quarterly Journal of Finance*, 4(3), pp. 1-50.

Booth, G.G., Kallunki, J.-P., Lin, J.-C. and Martikainen, T. (2000) 'Internalization and stock price clustering: Finnish evidence', *Journal of International Money and Finance*, 19(5), pp. 737-751.

Borochin, P. and Yang, J. (2017) 'Options, equity risks, and the value of capital structure adjustments', *Journal of Corporate Finance*, 42, pp. 150-178.

Boyer, B.H. and Vorkink, K. (2014) 'Stock options as lotteries', *The Journal of Finance*, 69(4), pp. 1485-1527.

Breymann, W., Dias, A. and Embrechts, P. (2003) 'Dependence structures for multivariate high-frequency data in finance', *Quantitative Finance*, 3(1), pp. 1-14.

Britten-Jones, M. and Neuberger, A. (2000) 'Option prices, implied price processes, and stochastic volatility', *The Journal of Finance,* 55(2), pp. 839-866.

Broadie, M., Chernov, M. and Johannes, M. (2009) 'Understanding index option returns', *The Review of Financial Studies*, 22(11), pp. 4493-4529.

Brockman, P. and Chung, D.Y. (2002) 'Commonality in Liquidity: Evidence from an Order‐Driven Market Structure', *Journal of Financial Research*, 25(4), pp. 521-539.

Brockman, P., Chung, D.Y. and Pérignon, C. (2009) 'Commonality in liquidity: A global perspective', *Journal of Financial and Quantitative Analysis*, 44(04), pp. 851-882.

Brownlees, C.T. and Gallo, G.M. (2006) 'Financial econometric analysis at ultra-high frequency: Data handling concerns', *Computational Statistics & Data Analysis*, 51(4), pp. 2232-2245.

Câmara, A., Popova, I. and Simkins, B. (2012) 'A comparative study of the probability of default for global financial firms', *Journal of Banking & Finance*, 36(3), pp. 717-732.

Cao, C., Chen, Z. and Griffin, J.M. (2005) 'Informational content of option volume prior to takeovers', *The Journal of Business*, 78(3), pp. 1073-1109.

Cao, C., Yu, F. and Zhong, Z. (2006) *How Important is Option-Implied Volatility for Pricing Credit Default Swaps?* Working Paper, UC Irvine.

Cao, J. and Han, B. (2013) 'Cross section of option returns and idiosyncratic stock volatility', *Journal of Financial Economics*, 108(1), pp. 231-249.

Cao, M. and Wei, J. (2010) 'Option market liquidity: Commonality and other characteristics', *Journal of Financial Markets*, 13(1), pp. 20-48.

Capelle-Blancard, G. and Chaudhury, M. (2007) 'Price clustering in the CAC 40 index options market', *Applied Financial Economics*, 17(15), pp. 1201-1210.

Capuano, C. (2008) *The Option-Ipod. the Probability of Default Implied by Option Prices Basedon Entropy*. IMF Working Paper No. 08/194.

Carr, P. and Wu, L. (2011) 'A simple robust link between American puts and credit protection', *The Review of Financial Studies*, 24(2), pp. 473-505.

Cartea, A. and Meyer-Brandis, T. (2010) 'How duration between trades of underlying securities affects option prices', *Review of Finance*, 14(4), pp. 749-785.

Castagna, A. and Matolcsy, Z. (1982) 'A two stage experimental design to test the efficiency of the market for traded stock options and the Australian evidence', *Journal of Banking & Finance*, 6(4), pp. 521-532.

CBOE (2016) *Equity Options Product Specifications*. Available at: http://www.cboe.com/products/equityoptionspecs.aspx (Accessed: 28/10/2017).

CBOE (2017a) *2016 Cboe Market Statistics*. Available at: http://www.cboe.com/data/market-statistics-2016.pdf (Accessed:28/10/2017)

CBOE (2017b) *Options Quick Facts - Marketplace*. Available at: http://www.cboe.com/education/getting-started/quick-facts/options-marketplace (Accessed: 23/04/2017).

Chakravarty, S. (2001) 'Stealth-trading: Which traders’ trades move stock prices?', *Journal of Financial Economics*, 61(2), pp. 289-307.

Chakravarty, S., Gulen, H. and Mayhew, S. (2004) 'Informed trading in stock and option markets', *The Journal of Finance*, 59(3), pp. 1235-1257.

Chakravarty, S., Jain, P., Upson, J. and Wood, R. (2012) 'Clean sweep: Informed trading through intermarket sweep orders', *Journal of Financial and Quantitative Analysis*, 47(2), pp. 415-435.

Chan, K., Chung, Y.P. and Fong, W.-M. (2002a) 'The informational role of stock and option volume', *The Review of Financial Studies*, 15(4), pp. 1049-1075.

Chan, K., Chung, Y.P. and Johnson, H. (1995) 'The intraday behavior of bid-ask spreads for NYSE stocks and CBOE options', *Journal of Financial and Quantitative Analysis*, 30(03), pp. 329-346.

Chan, K.C., Chang, Y. and Lung, P.P. (2009) 'Informed trading under different market conditions and moneyness: Evidence from TXO options', *Pacific-Basin Finance Journal*, 17(2), pp. 189-208.

Chang, B.Y. and Orosi, G. (2017) 'Equity Option Implied Probability of Default and Equity Recovery Rate', *Journal of Futures Markets*, 37(6), pp. 599-613.

Chang, C.-C., Hsieh, P.-F. and Lai, H.-N. (2009) 'Do informed option investors predict stock returns? Evidence from the Taiwan stock exchange', *Journal of Banking & Finance*, 33(4), pp. 757-764.

Chang, T.-H. (2011) 'Risk preference and trading motivation measurement due to moneyness: evidence from the S&P 500 Index option market', *Applied Financial Economics*, 21(14), pp. 1049-1057.

Chen, C.R., Lung, P.P. and Tay, N.S. (2005) 'Information flow between the stock and option markets: Where do informed traders trade?', *Review of Financial Economics*, 14(1), pp. 1-23.

Chen, N. and Kou, S.G. (2009) 'Credit spreads, optimal capital structure, and implied volatility with endogenous default and jump risk', *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 19(3), pp. 343-378.

Chen, Y.L. and Gau, Y.F. (2009) 'Tick sizes and relative rates of price discovery in stock, futures, and options markets: Evidence from the Taiwan stock exchange', *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 29(1), pp. 74-93.

Cheng, K.H., Fung, J.K. and Tse, Y. (2005) 'How electronic trading affects bid‐ask spreads and arbitrage efficiency between index futures and options', *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 25(4), pp. 375-398.

Chiang, C.-H. (2014) 'Stock returns on option expiration dates: Price impact of liquidity trading', *Journal of Empirical Finance*, 28, pp. 273-290.

Chiang, R. and Fong, W.-M. (2001) 'Relative informational efficiency of cash, futures, and options markets: The case of an emerging market', *Journal of Banking & Finance*, 25(2), pp. 355-375.

Cho, Y.-H. and Engle, R.F. (1999) *Modeling the impacts of market activity on bid-ask spreads in the option market*. National Bureau of Economic Research.

Chordia, T., Roll, R. and Subrahmanyam, A. (2000) 'Commonality in liquidity', *Journal of Financial Economics*, 56(1), pp. 3-28.

Chordia, T., Roll, R. and Subrahmanyam, A. (2001) 'Market liquidity and trading activity', *The Journal of Finance*, 56(2), pp. 501-530.

Chordia, T., Roll, R. and Subrahmanyam, A. (2002) 'Order imbalance, liquidity, and market returns', *Journal of Financial economics*, 65(1), pp. 111-130.

Chordia, T., Sarkar, A. and Subrahmanyam, A. (2005) 'An Empirical Analysis of Stock and Bond Market Liquidity', *Review of Financial Studies*, 18(1), pp. 85-129.

Christensen, B.J. and Prabhala, N.R. (1998) 'The relation between implied and realized volatility', *Journal of financial economics*, 50(2), pp. 125-150.

Christie, W.G. and Schultz, P.H. (1994) 'Why do NASDAQ market makers avoid odd‐eighth quotes?', *The Journal of Finance*, 49(5), pp. 1813-1840.

Christie, W.G. and Schultz, P.H. (1999) 'The initiation and withdrawal of odd-eighth quotes among Nasdaq stocks: An empirical analysis', *Journal of Financial Economics*, 52(3), pp. 409-442.

Christoffersen, P., Goyenko, R., Jacobs, K. and Karoui, M. (2017) 'Illiquidity premia in the equity options market', *The Review of Financial Studies*, 31(3), pp. 811-851.

Collver, C. (2009) 'Measuring the impact of option market activity on the stock market: Bivariate point process models of stock and option transactions', *Journal of Financial Markets*, 12(1), pp. 87-106.

Conrad, J. (1989) 'The price effect of option introduction', *The Journal of Finance*, 44(2), pp. 487-498.

Conrad, J., Dittmar, R.F. and Ghysels, E. (2013) 'Ex ante skewness and expected stock returns', *The Journal of Finance*, 68(1), pp. 85-124.

Conrad, J., Dittmar, R.F. and Hameed, A. (2017) *Cross-market and cross-firm effects in implied default probabilities and recovery values*. Working Paper.

Conrad, J., Wahal, S. and Xiang, J. (2015) 'High-frequency quoting, trading, and the efficiency of prices', *Journal of Financial Economics*, 116(2), pp. 271-291.

Constantinides, G.M., Jackwerth, J.C. and Savov, A. (2013) 'The puzzle of index option returns', *Review of Asset Pricing Studies*, 3(2), pp. 229-257.

Corredor, P., Lechon, P. and Santamaria, R. (2001) 'Option‐Expiration Effects in Small Markets: The Spanish Stock Exchange', *Journal of Futures Markets*, 21(10), pp. 905-928.

Corwin, S.A. and Lipson, M.L. (2011) 'Order characteristics and the sources of commonality in prices and liquidity', *Journal of Financial Markets*, 14(1), pp. 47-81.

Coughenour, J.F. and Saad, M.M. (2004) 'Common market makers and commonality in liquidity', *Journal of Financial Economics*, 73(1), pp. 37-69.

Coval, J.D. and Shumway, T. (2001) 'Expected option returns', *The journal of Finance*, 56(3), pp. 983-1009.

Cremers, M., Driessen, J., Maenhout, P. and Weinbaum, D. (2008) 'Individual stock-option prices and credit spreads', *Journal of Banking & Finance*, 32(12), pp. 2706-2715.

Cremers, M. and Weinbaum, D. (2010) 'Deviations from Put-Call Parity and Stock Return Predictability', *Journal of Financial and Quantitative Analysis*, 45(02), pp. 335-367.

Da Fonseca, J. and Gottschalk, K. (2014) 'Cross-hedging strategies between CDS spreads and option volatility during crises', *Journal of International Money and Finance*, 49, pp. 386-400.

Damodaran, A. and Lim, J. (1991) 'The effects of option listing on the underlying stocks' return processes', *Journal of Banking & Finance*, 15(3), pp. 647-664.

Danielsen, B.R. and Sorescu, S.M. (2001) 'Why do option introductions depress stock prices? A study of diminishing short sale constraints', *Journal of Financial and Quantitative Analysis*, 36(4), pp. 451-484.

Danielsen, B.R., Van Ness, B.F. and Warr, R.S. (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(4), pp. 1041-1062.

Dasu, T. and Johnson, T. (2003) *Exploratory data mining and data cleaning*. John Wiley & Sons.

Davis, R.L., Van Ness, B.F. and Van Ness, R.A. (2014) 'Clustering of Trade Prices by High‐Frequency and Non‐High‐Frequency Trading Firms', *Financial Review*, 49(2), pp. 421-433.

De Jong, F. and Rindi, B. (2009) *The microstructure of financial markets*. Cambridge University Press.

De Long, J.B., Shleifer, A., Summers, L.H. and Waldmann, R.J. (1990) 'Noise trader risk in financial markets', *Journal of political Economy*, 98(4), pp. 703-738.

Dennis, P., Mayhew, S. and Stivers, C. (2006) 'Stock returns, implied volatility innovations, and the asymmetric volatility phenomenon', *Journal of Financial and Quantitative Analysis*, 41(02), pp. 381-406.

Detemple, J. and Jorion, P. (1990) 'Option listing and stock returns: An empirical analysis', *Journal of Banking & Finance*, 14(4), pp. 781-801.

Detemple, J. and Selden, L. (1991) 'A general equilibrium analysis of option and stock market interactions', *International Economic Review*, pp. 279-303.

Diltz, J.D. and Kim, S. (1996) 'The relationship between stock and option price changes', *Financial Review*, 31(3), pp. 499-519.

Diz, F. and Finucane, T.J. (1993) 'The rationality of early exercise decisions: Evidence from the S&P 100 index options market', *The Review of Financial Studies*, 6(4), pp. 765-797.

Donders, M., WM, Kouwenberg, R. and Vorst, T., CF (2000) 'Options and earnings announcements: an empirical study of volatility, trading volume, open interest and liquidity', *European Financial Management*, 6(2), pp. 149-171.

Dong, W. and Sinha, N.R. (2011) *Where do informed traders trade? trading around news on dow 30 options*. Working Paper.

Driessen, J., Maenhout, P.J. and Vilkov, G. (2009) 'The price of correlation risk: Evidence from equity options', *The Journal of Finance*, 64(3), pp. 1377-1406.

Duarte, J. and Young, L. (2009) 'Why is PIN priced?', *Journal of Financial Economics*, 91(2), pp. 119-138.

Dufour, A. and Engle, R.F. (2000) 'Time and the price impact of a trade', *The Journal of Finance*, 55(6), pp. 2467-2498.

Easley, D., Kiefer, N.M. and O'Hara, M. (1997a) 'The information content of the trading process', *Journal of Empirical Finance*, 4(2-3), pp. 159-186.

Easley, D., Kiefer, N.M. and O'Hara, M. (1997b) 'One day in the life of a very common stock', *The Review of Financial Studies*, 10(3), pp. 805-835.

Easley, D., Kiefer, N.M., O'hara, M. and Paperman, J.B. (1996) 'Liquidity, information, and infrequently traded stocks', *The Journal of Finance*, 51(4), pp. 1405-1436.

Easley, D. and O'hara, M. (1987) 'Price, trade size, and information in securities markets', *Journal of Financial economics*, 19(1), pp. 69-90.

Easley, D. and O'hara, M. (1992) 'Time and the process of security price adjustment', *The Journal of finance*, 47(2), pp. 577-605.

Easley, D., O'Hara, M. and Paperman, J. (1998a) 'Financial analysts and information-based trade', *Journal of Financial Markets*, 1(2), pp. 175-201.

Easley, D., O'hara, M. and Srinivas, P.S. (1998b) 'Option volume and stock prices: Evidence on where informed traders trade', *The Journal of Finance*, 53(2), pp. 431-465.

Engle, R.F. (2000) 'The econometrics of ultra‐high‐frequency data', *Econometrica*, 68(1), pp. 1-22.

Engle, R.F. and Russell, J.R. (1998) 'Autoregressive conditional duration: a new model for irregularly spaced transaction data', *Econometrica*, pp. 1127-1162.

Eom, K.S. and Hahn, S.B. (2005) 'Traders' strategic behavior in an index options market', *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 25(2), pp. 105-133.

Evnine, J. and Rudd, A. (1985) 'Index options: The early evidence', *The Journal of Finance*, 40(3), pp. 743-756.

Fabozzi, F., Focardi, S.M. and Jonas, C. (2011) 'High-frequency trading: methodologies and market impact', *Review of Futures Markets*, 9(Special Issue), pp. 7-38.

Faccini, R., Konstantinidi, E., Skiadopoulos, G. and Sarantopoulou-Chiourea, S. (2018) 'A New Predictor of US Real Economic Activity: The S&P 500 Option Implied Risk Aversion', *Management Science,* Forthcoming.

Faff, R. and Hillier, D. (2005) 'Complete markets, informed trading and equity option introductions', *Journal of Banking & Finance*, 29(6), pp. 1359-1384.

Falkenberry, T.N. (2002) 'High frequency data filtering', *White paper, Tick Data Inc*.

Fan, R., Taylor, S.J. and Sandri, M. (2018) 'Density forecast comparisons for stock prices, obtained from high‐frequency returns and daily option prices', *Journal of Futures Markets*, 38(1), pp. 83-103.

Farhangfar, A., Kurgan, L. and Dy, J. (2008) 'Impact of imputation of missing values on classification error for discrete data', *Pattern Recognition*, 41(12), pp. 3692-3705.

Fedenia, M. and Grammatikos, T. (1992) 'Options trading and the bid-ask spread of the underlying stocks', *Journal of Business,* pp. 335-351.

Flint, A., Lepone, A. and Yang, J.Y. (2014) 'Do Option Strategy Traders Have a Disadvantage? Evidence from the Australian Options Market', *Journal of Futures Markets*, 34(9), pp. 838-852.

Florescu, I. (2016) *Handbook of High-frequency Trading and Modeling in Finance*. John Wiley & Sons.

French, K.R. and Roll, R. (1986) 'Stock return variances: The arrival of information and the reaction of traders', *Journal of financial economics*, 17(1), pp. 5-26.

Freund, S., McCann, P.D. and Webb, G.P. (1994) 'A Regression Analysis of the Effects of Options Introduction on Stock Variances', *The Journal of Derivatives*, 1(3), pp. 25-38.

Frino, A. and Fabre, J. (2004) 'Commonality in liquidity: evidence from the Australian stock exchange', *Accounting and Finance,* 44, pp. 357-368.

Frino, A., Lepone, A. and Wearin, G. (2008) 'Intraday behavior of market depth in a competitive dealer market: A note', *Journal of Futures Markets*, 28(3), pp. 294-307.

Fu, X., Arisoy, Y.E., Shackleton, M.B. and Umutlu, M. (2016) 'Option-Implied Volatility Measures and Stock Return Predictability', *The Journal of Derivatives*, 24(1), pp. 58-78.

Furfine, C. (2007) 'When is inter-transaction time informative?', *Journal of Empirical Finance*, 14(3), pp. 310-332.

Galai, D. (1978) 'Empirical tests of boundary conditions for CBOE options', *Journal of Financial Economics*, 6(2-3), pp. 187-211.

Galariotis, E.C. and Giouvris, E. (2007) 'Liquidity commonality in the London stock exchange', *Journal of Business Finance & Accounting*, 34(1‐2), pp. 374-388.

Garman, M.B. (1976) 'Market microstructure', *Journal of financial Economics*, 3(3), pp. 257-275.

Garvey, R. and Wu, F. (2014) 'Clustering of intraday order-sizes by uninformed versus informed traders', *Journal of Banking & Finance*, 41, pp. 222-235.

Ge, L., Lin, T.-C. and Pearson, N.D. (2016) 'Why does the option to stock volume ratio predict stock returns?', *Journal of Financial Economics*, 120(3), pp. 601-622.

Gençay, R., Dacorogna, M., Muller, U.A., Pictet, O. and Olsen, R. (2001) *An introduction to high-frequency finance*. San Diego: Academic press.

Gjerde, O. and Saettem, F. (1995) 'Option initiation and underlying market behavior: Evidence from Norway', *The Journal of Futures Markets (1986-1998)*, 15(18), p. 881.

Glosten, L.R. (1994) 'Is the electronic open limit order book inevitable?', *The Journal of Finance*, 49(4), pp. 1127-1161.

Godek, P.E. (1996) 'Why Nasdaq market makers avoid odd-eighth quotes', *Journal of Financial Economics*, 41(3), pp. 465-474.

Goncalves, S. and Guidolin, M. (2006) 'Predictable dynamics in the S&P 500 index options implied volatility surface', *The Journal of Business*, 79(3), pp. 1591-1635.

Goodhart, C.A. and O'Hara, M. (1997) 'High frequency data in financial markets: Issues and applications', *Journal of Empirical Finance*, 4(2), pp. 73-114.

Goodhart, C.A.E. and Curcio, R. (1991) *The clustering of bid/ask prices and the spread in the foreign exchange market*. LSE Financial Markets Group.

Govindaraj, S., Jin, W., Livnat, J. and Zhao, C. (2014) *Using Option Implied Volatilities to Predict Absolute Stock Returns-Evidence from Earnings Announcements and Annual Shareholders' Meetings*. Working Paper.

Goyal, A. and Saretto, A. (2009) 'Cross-section of option returns and volatility', *Journal of Financial Economics*, 94(2), pp. 310-326.

Goyenko, R., Ornthanalai, C. and Tang, S. (2015) *Options illiquidity: Determinants and implications for stock returns* (2492506). Working Paper.

Grossman, S.J., Miller, M.H., Cone, K.R., Fischel, D.R. and Ross, D.J. (1997) 'Clustering and competition in asset markets', *The Journal of Law and Economics*, 40(1), pp. 23-60.

Hafner, C.M. (2005) 'Durations, volume and the prediction of financial returns in transaction time', *Quantitative Finance*, 5(2), pp. 145-152.

Hameed, A., Kang, W. and Viswanathan, S. (2010) 'Stock market declines and liquidity', *The Journal of Finance*, 65(1), pp. 257-293.

Hamill, P.A., Opong, K.K. and McGregor, P. (2002) 'Equity option listing in the UK: a comparison of market-based research methodologies', *Journal of Empirical Finance*, 9(1), pp. 91-108.

Han, B. and Zhou, Y. (2012) *Variance risk premium and cross-section of stock returns*. Working Paper.

Han, J., Pei, J. and Kamber, M. (2011) *Data mining: concepts and techniques*. Elsevier.

Hao, J., Kalay, A. and Mayhew, S. (2010) 'Ex-dividend arbitrage in option markets', *The Review of Financial Studies*, 23(1), pp. 271-303.

Harris, L. (1989) 'S&P 500 cash stock price volatilities', *The Journal of Finance*, 44(5), pp. 1155-1175.

Harris, L. (1991) 'Stock price clustering and discreteness', *The Review of Financial Studies*, 4(3), pp. 389-415.

Harris, L. (2003) *Trading and exchanges: Market microstructure for practitioners*. OUP USA.

Hasbrouck, J. and Seppi, D.J. (2001) 'Common factors in prices, order flows, and liquidity', *Journal of financial Economics*, 59(3), pp. 383-411.

Hayunga, D.K. and Lung, P.P. (2014) 'Trading in the options market around financial analysts’ consensus revisions', *Journal of Financial and Quantitative Analysis*, 49(3), pp. 725-747.

Hendershott, T., Jones, C.M. and Menkveld, A.J. (2011) 'Does algorithmic trading improve liquidity?', *The Journal of Finance*, 66(1), pp. 1-33.

Hirshleifer, D., Subrahmanyam, A. and Titman, S. (1994) 'Security analysis and trading patterns when some investors receive information before others', *The Journal of Finance*, 49(5), pp. 1665-1698.

Hodrick, L.S. and Moulton, P.C. (2009) 'Liquidity: Considerations of a portfolio manager', *Financial Management*, 38(1), pp. 59-74.

Holowczak, R., Simaan, Y.E. and Wu, L. (2006) 'Price discovery in the US stock and stock options markets: A portfolio approach', *Review of Derivatives Research*, 9(1), pp. 37-65.

Hsieh, G.W.-l. and He, H.-R. (2014a) 'Informed trading, trading strategies and the information content of trading volume: Evidence from the Taiwan index options market', *Journal of International Financial Markets, Institutions and Money*, 31, pp. 187-215.

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

Huang, R.D. and Stoll, H.R. (1997) 'The components of the bid-ask spread: A general approach', *Review of Financial Studies*, 10(4), pp. 995-1034.

Huberman, G. and Halka, D. (2001) 'Systematic liquidity', *Journal of Financial Research*, 24(2), pp. 161-178.

Ikenberry, D.L. and Weston, J.P. (2008) 'Clustering in US stock prices after decimalisation', *European Financial Management*, 14(1), pp. 30-54.

Jennings, R. and Starks, L. (1986) 'Earnings announcements, stock price adjustment, and the existence of option markets', *The journal of Finance*, 41(1), pp. 107-125.

Jin, W., Livnat, J. and Zhang, Y. (2012) 'Option Prices Leading Equity Prices: Do Option Traders Have an Information Advantage?', *Journal of Accounting Research*, 50(2), pp. 401-432.

Johnson, T., Liang, M. and Liu, Y. (2016) 'What Drives Index Options Exposures?', *Review of Finance*, 22(2), pp. 561-593.

Johnson, T.L. and So, E.C. (2012) 'The option to stock volume ratio and future returns', *Journal of Financial Economics*, 106(2), pp. 262-286.

Jones, C.S. (2006) 'A nonlinear factor analysis of S&P 500 index option returns', *The Journal of Finance*, 61(5), pp. 2325-2363.

Kalaitzoglou, I. and Ibrahim, B.M. (2013) 'Does order flow in the European Carbon Futures Market reveal information?', *Journal of Financial Markets*, 16(3), pp. 604-635.

Kalaitzoglou, I.A. and Ibrahim, B.M. (2015) 'Liquidity and resolution of uncertainty in the European carbon futures market', *International Review of Financial Analysis*, 37, pp. 89-102.

Kamara, A., Lou, X. and Sadka, R. (2008) 'The divergence of liquidity commonality in the cross-section of stocks☆', *Journal of Financial Economics*, 89(3), pp. 444-466.

Kang, J. and Park, H.-J. (2008) 'The information content of net buying pressure: Evidence from the KOSPI 200 index option market', *Journal of Financial Markets*, 11(1), pp. 36-56.

Karolyi, G.A., Lee, K.-H. and van Dijk, M.A. (2012) 'Understanding commonality in liquidity around the world', *Journal of Financial Economics*, 105(1), pp. 82-112.

Kaul, G., Nimalendran, M. and Zhang, D. (2004) *Informed trading and option spreads*. Working Paper.

Kita, A. (2012) *CDS spreads explained with credit spread volatility and jump risk of individual firms*. Working Paper.

Korajczyk, R.A. and Sadka, R. (2008) 'Pricing the commonality across alternative measures of liquidity', *Journal of Financial Economics*, 87(1), pp. 45-72.

Kumar, R., Sarin, A. and Shastri, K. (1995) 'The impact of index options on the underlying stocks: The evidence from the listing of Nikkei stock average options', *Pacific-Basin Finance Journal*, 3(2-3), pp. 303-317.

Kumar, R., Sarin, A. and Shastri, K. (1998) 'The impact of options trading on the market quality of the underlying security: An empirical analysis', *The Journal of Finance*, 53(2), pp. 717-732.

Kyle, A.S. (1985) 'Continuous auctions and insider trading', *Econometrica: Journal of the Econometric Society*, pp. 1315-1335.

Lakonishok, J., Lee, I., Pearson, N.D. and Poteshman, A.M. (2007) 'Option market activity', *The Review of Financial Studies*, 20(3), pp. 813-857.

Lamoureux, C.G. and Lastrapes, W.D. (1993) 'Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities', *The Review of Financial Studies*, 6(2), pp. 293-326.

Lang, M. and Maffett, M. (2011) 'Transparency and liquidity uncertainty in crisis periods', *Journal of accounting and economics*, 52(2-3), pp. 101-125.

Latane, H.A. and Rendleman, R.J. (1976) 'Standard deviations of stock price ratios implied in option prices', *The Journal of Finance*, 31(2), pp. 369-381.

Lee, C.M., Mucklow, B. and Ready, M.J. (1993) 'Spreads, depths, and the impact of earnings information: An intraday analysis', *Review of Financial Studies*, 6(2), pp. 345-374.

Lee, J. and Cheong, H.Y. (2001) 'Trade size and information-motivated trading in the options and stock markets', *Journal of Financial and Quantitative Analysis*, 36(4), pp. 485-501.

Lehalle, C.-A. and Laruelle, S. (2013) *Market Microstructure in Practice*. New York: World Scientific.

Lemmon, M. and Ni, S.X. (2014) 'Differences in Trading and Pricing Between Stock and Index Options', *Management Science*, 60(8), pp. 1985-2001.

Levy, H. and Yoder, J.A. (1993) 'The behavior of option implied standard deviations around merger and acquisition announcements', *Financial Review*, 28(2), pp. 261-272.

Li, M., McCormick, T. and Zhao, X. (2005) 'Order imbalance and liquidity supply: Evidence from the bubble burst of Nasdaq stocks', *Journal of Empirical Finance*, 12(4), pp. 533-555.

Lien, D., Hung, P.-H. and Hung, I.-C. (2019) 'Order price clustering, size clustering, and stock price movements: Evidence from the Taiwan Stock Exchange', *Journal of Empirical Finance*, 52, pp. 149-177.

Lin, T.-C. and Lu, X. (2015) 'Why do options prices predict stock returns? Evidence from analyst tipping', *Journal of Banking & Finance*, 52, pp. 17-28.

Little, R.J. and Rubin, D.B. (2014) *Statistical analysis with missing data*. John Wiley & Sons.

Liu, Q. (2009) 'On portfolio optimization: How and when do we benefit from high‐frequency data?', *Journal of Applied Econometrics*, 24(4), pp. 560-582.

Liu, S. (2010) 'Equity Options and Underlying Stocks' Behavior: Further Evidence from Japan', *International Review of Finance*, 10(3), pp. 293-312.

Liu, X., Pong, E.S., Shackleton, M.B. and Zhang, Y. (2014) 'Option-implied volatilities and stock returns: evidence from industry-neutral portfolios', *The Journal of Portfolio Management*, 41(1), pp. 65-77.

Lung, P.P. and Xu, P. (2014) 'Tipping and option trading', *Financial Management*, 43(3), pp. 671-701.

Maberly, E.D., Pierce, R.M. and Catania, P. (2010) 'Threshold levels, strike price grid, and other market microstructure issues associated with exchange‐traded equity options', *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 30(2), pp. 188-201.

Madhavan, A. (2000) 'Market microstructure: A survey', *Journal of financial markets*, 3(3), pp. 205-258.

Madhavan, A. (2002) 'Market microstructure: A practitioner's guide', *Financial Analysts Journal*, 58(5), pp. 28-42.

Madhavan, A. and Smidt, S. (1991) 'A Bayesian model of intraday specialist pricing', *Journal of Financial Economics*, 30(1), pp. 99-134.

Malkiel, B.G. and Fama, E.F. (1970) 'Efficient capital markets: A review of theory and empirical work', *The journal of Finance*, 25(2), pp. 383-417.

Manaster, S. and Rendleman, R.J. (1982) 'Option prices as predictors of equilibrium stock prices', *The Journal of Finance*, 37(4), pp. 1043-1057.

Manganelli, S. (2005) 'Duration, volume and volatility impact of trades', *Journal of Financial markets*, 8(4), pp. 377-399.

Massimb, M.N. and Phelps, B.D. (1994) 'Electronic trading, market structure and liquidity', *Financial Analysts Journal*, 50(1), pp. 39-50.

Mayhew, S. (2002) 'Competition, market structure, and bid ‐ ask spreads in stock option markets', *The Journal of Finance*, 57(2), pp. 931-958.

Mayhew, S. and Mihov, V. (2000) *Another look at option listing effects*. Working Paper.

Mayhew, S. and Mihov, V. (2004) 'How do exchanges select stocks for option listing?', *The Journal of Finance*, 59(1), pp. 447-471.

Mayhew, S., Sarin, A. and Shastri, K. (1995) 'The allocation of informed trading across related markets: An analysis of the impact of changes in equity ‐ option margin requirements', *The Journal of Finance*, 50(5), pp. 1635-1653.

Mayhew, S., Sarin, A. and Shastri, K. (1999) *What drives option liquidity?* Working Paper.

Mayhew, S. and Stivers, C. (2003) 'Stock return dynamics, option volume, and the information content of implied volatility', *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 23(7), pp. 615-646.

Mazouz, K. (2004) 'The effect of CBOE option listing on the volatility of NYSE traded stocks: a time-varying variance approach', *Journal of Empirical Finance*, 11(5), pp. 695-708.

Meng, L., Verousis, T. and ap Gwilym, O. (2013) 'A substitution effect between price clustering and size clustering in credit default swaps', *Journal of International Financial Markets, Institutions and Money*, 24, pp. 139-152.

Mitchell, J. (2001) 'Clustering and psychological barriers: The importance of numbers', *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 21(5), pp. 395-428.

Moshirian, F., Qian, X., Wee, C.K.G. and Zhang, B. (2017) 'The determinants and pricing of liquidity commonality around the world', *Journal of Financial Markets*, 33, pp. 22-41.

Moulton, P.C. (2005) 'You can’t always get what you want: Trade-size clustering and quantity choice in liquidity', *Journal of Financial Economics*, 78(1), pp. 89-119.

Muravyev, D., Pearson, N.D. and Paul Broussard, J. (2013) 'Is there price discovery in equity options?', *Journal of Financial Economics*, 107(2), pp. 259-283.

Neal, R. (1987) 'Potential competition and actual competition in equity options', *The Journal of Finance*, 42(3), pp. 511-531.

Ni, S.X. (2008) *Stock option returns: A puzzle*. Working Paper.

Ni, S.X., Pan, J. and Poteshman, A.M. (2008) 'Volatility information trading in the option market', *The Journal of Finance*, 63(3), pp. 1059-1091.

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

Niederhoffer, V. (1965) 'Clustering of stock prices', *Operations Research*, 13(2), pp. 258-265.

Nordén, L. (2001) 'Hedging of American equity options: do call and put prices always move in the direction as predicted by the movement in the underlying stock price?', *Journal of Multinational Financial Management*, 11(4), pp. 321-340.

O'Connor, M.L. (1999) 'The Cross‐Sectional Relationship Between Trading Costs and Lead/Lag Effects in Stock & Option Markets', *Financial Review*, 34(4), pp. 95-117.

O'hara, M. (1995) *Market microstructure theory*. Blackwell Cambridge, MA.

O’Hara, M. (2015) 'High frequency market microstructure', *Journal of Financial Economics*, 116(2), pp. 257-270.

Odders-White, E.R. and Ready, M.J. (2008) 'The probability and magnitude of information events', *Journal of Financial Economics*, 87(1), pp. 227-248.

Osborne, M.F. (1962) 'Periodic structure in the Brownian motion of stock prices', *Operations Research*, 10(3), pp. 345-379.

Pan, J. and Poteshman, A.M. (2006a) 'The Information in Option Volume for Future Stock Prices', *Review of Financial Studies*, 19(3), pp. 871-908.

Patell, J.M. and Wolfson, M.A. (1979) 'Anticipated information releases reflected in call option prices', *Journal of Accounting and Economics*, 1(2), pp. 117-140.

Pinder, S. (2003) 'An empirical examination of the impact of market microstructure changes on the determinants of option bid–ask spreads', *International Review of Financial Analysis*, 12(5), pp. 563-577.

Poteshman, A.M. and Serbin, V. (2003) 'Clearly irrational financial market behavior: Evidence from the early exercise of exchange traded stock options', *The Journal of Finance*, 58(1), pp. 37-70.

Rahman, S. (2001) 'The introduction of derivatives on the Dow Jones Industrial Average and their impact on the volatility of component stocks', *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 21(7), pp. 633-653.

Rehman, Z. and Vilkov, G. (2012) *Risk-neutral skewness: Return predictability and its sources*. Working Paper.

Roll, R., Schwartz, E. and Subrahmanyam, A. (2010) 'O/S: The relative trading activity in options and stock', *Journal of Financial Economics*, 96(1), pp. 1-17.

Rösch, C.G. and Kaserer, C. (2013) 'Market liquidity in the financial crisis: The role of liquidity commonality and flight-to-quality', *Journal of Banking & Finance*, 37(7), pp. 2284-2302.

Rösch, C.G. and Kaserer, C. (2014) 'Reprint of: Market liquidity in the financial crisis: The role of liquidity commonality and flight-to-quality', *Journal of Banking & Finance*, 45, pp. 152-170.

Rubinstein, M. (1985) 'Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978', *The Journal of Finance*, 40(2), pp. 455-480.

Russell, J.R. and Engle, R. (2009) 'Analysis of high-frequency data', *Handbook of financial econometrics*, 1, pp. 383-426.

Ryu, D. (2015) 'The information content of trades: An analysis of KOSPI 200 index derivatives', *Journal of Futures Markets*, 35(3), pp. 201-221.

Ryu, D. and Yang, H. (2018) 'The directional information content of options volumes', *Journal of Futures Markets*, 38(12), pp. 1533-1548.

Sahlström, P. (2001) 'Impact of stock option listings on return and risk characteristics in Finland', *International Review of Financial Analysis*, 10(1), pp. 19-36.

Sarwar, G. (2005) 'The informational role of option trading volume in equity index options markets', *Review of Quantitative Finance and Accounting*, 24(2), pp. 159-176.

Schlag, C. and Stoll, H. (2005) 'Price impacts of options volume', *Journal of Financial Markets*, 8(1), pp. 69-87.

Segara, L. and Sagara, R. (2007) 'Intraday trading patterns in the equity warrants and equity options markets: Australian evidence', *Australasian Accounting, Business and Finance Journal*, 1(2), pp. 42-60.

Sheikh, A.M. and Ronn, E.I. (1994) 'A characterization of the daily and intraday behavior of returns on options', *The Journal of Finance*, 49(2), pp. 557-580.

Sim, M., Ryu, D. and Yang, H. (2016) 'Tests on the monotonicity properties of KOSPI 200 options prices', *Journal of Futures Markets*, 36(7), pp. 625-646.

Skinner, D.J. (1989) 'Options markets and stock return volatility', *Journal of Financial Economics*, 23(1), pp. 61-78.

Smidt, S. (1971) 'Which road to an efficient stock market: free competition or regulated monopoly?', *Financial Analysts Journal*, 27(5), pp. 18-20.

Spierdijk, L. (2004) 'An empirical analysis of the role of the trading intensity in information dissemination on the NYSE', *Journal of empirical finance*, 11(2), pp. 163-184.

Stephan, J.A. and Whaley, R.E. (1990) 'Intraday price change and trading volume relations in the stock and stock option markets', *The Journal of Finance*, 45(1), pp. 191-220.

Stoll, H.R. (1978) 'The supply of dealer services in securities markets', *The Journal of Finance*, 33(4), pp. 1133-1151.

Stoll, H.R. (2003) 'Market microstructure', in Constantinides, G., Harris, M. and Stulz, R. (eds.) *Handbook of the Economics of Finance*. Amsterdam: Elsevier, pp. 553-604.

Taylor, S.J., Tzeng, C.-F. and Widdicks, M. (2014) 'Bankruptcy probabilities inferred from option prices', *The Journal of Derivatives*, 22(2), pp. 8-31.

Taylor, S.J., Yadav, P.K. and Zhang, Y. (2010) 'The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks', *Journal of Banking & Finance*, 34(4), pp. 871-881.

Tsai, W.C., Chiu, Y.T. and Wang, Y.H. (2015) 'The information content of trading activity and quote changes: Evidence from VIX options', *Journal of Futures Markets*, 35(8), pp. 715-737.

Van Buskirk, A. (2011) *Volatility skew, earnings announcements, and the predictability of crashes*. Working Paper.

Vasquez, A. (2017) 'Equity volatility term structures and the cross section of option returns', *Journal of Financial and Quantitative Analysis*, 52(6), pp. 2727-2754.

Verousis, T. and ap Gwilym, O. (2010) 'An improved algorithm for cleaning Ultra High-Frequency data', *Journal of Derivatives & Hedge Funds*, 15(4), pp. 323-340.

Verousis, T. and ap Gwilym, O. (2013a) *The Microstructure of Individual Equity Options: Firm-Level and Common Effects on Liquidity*. Working Paper.

Verousis, T. and ap Gwilym, O. (2013b) 'Trade size clustering and the cost of trading at the London Stock Exchange', *International Review of Financial Analysis*, 27, pp. 91-102.

Verousis, T., ap Gwilym, O. and Chen, X. (2015) 'The intraday determination of liquidity in the NYSE LIFFE equity option markets', *The European Journal of Finance*, 22(12), pp. 1164-1188.

Verousis, T., ap Gwilym, O. and Voukelatos, N. (2016a) 'Commonality in equity options liquidity: evidence from European Markets', *The European Journal of Finance*, 22(12), pp. 1204-1223.

Verousis, T., ap Gwilym, O. and Voukelatos, N. (2016b) 'The Impact of a Premium - Based Tick Size on Equity Option Liquidity', *Journal of Futures Markets*, 36(4), pp. 397-417.

Vijh, A.M. (1990) 'Liquidity of the CBOE equity options', *The Journal of Finance*, 45(4), pp. 1157-1179.

Vilsmeier, J. (2016) 'Updating the option implied probability of default methodology', *Journal of Computational Finance*, 19(3).

Wang, H., Zhou, H. and Zhou, Y. (2013) 'Credit default swap spreads and variance risk premia', *Journal of Banking & Finance*, 37(10), pp. 3733-3746.

Wang, J. (1994) 'A model of competitive stock trading volume', *Journal of political Economy*, 102(1), pp. 127-168.

Watt, W.H., Yadav, P.K. and Draper, P. (1992) 'The impact of option listing on underlying stock returns: The UK evidence', *Journal of Business Finance & Accounting*, 19(4), pp. 485-503.

Wei, J. and Zheng, J. (2010) 'Trading activity and bid–ask spreads of individual equity options', *Journal of Banking & Finance*, 34(12), pp. 2897-2916.

Wong, W.K., Tan, D. and Tian, Y. (2009) 'Informed trading and liquidity in the Shanghai Stock Exchange', *International Review of Financial Analysis*, 18(1-2), pp. 66-73.

Wu, W.-S., Liu, Y.-J., Lee, Y.-T. and Fok, R.C. (2014) 'Hedging costs, liquidity, and inventory management: The evidence from option market makers', *Journal of Financial Markets*, 18, pp. 25-48.

Xing, Y., Zhang, X. and Zhao, R. (2010) 'What does the individual option volatility smirk tell us about future equity returns?', *Journal of Financial and Quantitative Analysis*, 45(3), pp. 641-662.

Xu, C. (2014) 'Expiration‐Day Effects of Stock and Index Futures and Options in Sweden: The Return of the Witches', *Journal of futures markets*, 34(9), pp. 868-882.

Xu, X.E., Chen, P. and Wu, C. (2006) 'Time and dynamic volume–volatility relation', *Journal of Banking & Finance*, 30(5), pp. 1535-1558.