A Descriptive Study of High-Frequency Trade and Quote Option Data

Torben Andersen[a], **Ilya Archakov**[b*], **Leon Grund**[b],
Nikolaus Hautsch[b], **Sergey Nasekin**[c], **Ingmar Nolte**[c],
Manh Cuong Pham[c†], **Stephen Taylor**[c] and **Viktor Todorov**[a]

[a] *Northwestern University, Kellogg School of Management*

[b] *University of Vienna, Department of Statistics and Operations Research*

[c] *Lancaster University, Department of Accounting and Finance*

August 30, 2019

Abstract

This paper provides a guide to high frequency option trade and quote data disseminated by the Options Price Reporting Authority (OPRA). First, we present a comprehensive overview of the fragmented U.S. option market, including details on market regulation and the trading processes for all 15 constituent option exchanges. Then, we review the general structure of the OPRA dataset and present a thorough empirical description of the observed option trades and quotes for a selected sample of underlying assets that contains more than 25 billion records. We outline several types of irregular observations and provide recommendations for data filtering and cleaning. Finally, we illustrate the usefulness of the high frequency option data with two empirical applications: option-implied variance estimation and risk-neutral density estimation. Both applications highlight the richer information content of the high frequency OPRA data relative to the widely used end-of-day OptionMetrics data.

Keywords: Options Data, High Frequency Data, Market Microstructure.
JEL Classification: C55, G10.

[*] Corresponding author. Email: ilya.archakov@univie.ac.at. Tel: +43 1 4277 38686.
[†] Corresponding author. Email: m.c.pham@lancaster.ac.uk. Tel: +44 1524 594816.

We would like to acknowledge financial support from the ESRC-FWF bilateral grant titled "Bilateral Austria: Order Book Foundations of Price Risks and Liquidity: An Integrated Equity and Derivatives Markets Perspective", Grant Ref: ES/N014588/1 and the Austrian Science Fund (FWF): Research project: I-2762-G27.

1 Introduction

The econometrics of option data has become a rapidly developing research area in recent years. However, both empirical and methodological studies mainly focus on cross-sections or panels of close (end-of-day) option prices. At the same time, intraday option trades and quotes, observed at higher frequency, are considered in the literature only occasionally. This tendency is unlikely to persist much longer, since high frequency option data has the potential to convey more timely and more precise information regarding investors' expectations about a company, a sector of the economy, or even the entire market, than daily data. Through the inherent connection between option and stock prices, intraday option data provides a more complete view and a better understanding of realized and expected asset price dynamics at a high frequency. It allows us to shed some light on the joint high-frequency dynamics of options and the underlying, providing insights for short-term asset return predictability, intraday risk management, price discovery, information processing, as well as the role of liquidity in both markets. Last but not least, intraday option data allows for a more precise construction of implied model-based or model-free variances and risk-neutral densities than just using end-of-day prices. Likewise, analyzing the intraday variability of high-frequency based risk-neutral densities provide insights into how strongly and rapidly investors' expectations change in response to the intraday order flow or news arrival.[1]

So far, the potential of high-frequency option markets is far from fully exploited. Only a few studies employ high-frequency option data, while mainly investigating the dynamic properties of option prices and targeting the validity of option pricing models under stochastic volatility and jumps, see, e.g., Kokoszczynski et al. (2010a,b), Abergel and Zaatour (2012), Kapetanios et al. (2014), Andersen et al. (2015b), Audrino and Fengler (2015), Amaya et al. (2018), and Taylor et al. (2018).[2] This research predominantly employs data on index options, while there is very little research on high-frequency dynamics of individual stock options (see, e.g., Abergel and Zaatour, 2012).

A possible reason for the lack of research on the microstructure and high-frequency dynamics of option markets is, on the one hand, the lack of common reliable and accessible resources on precise data at high granularity for a sufficiently large cross section of options and, on the other hand, the significantly higher computational and operational burden due to the size and high dimensionality of such data. This burden is, firstly, due to the fact that the quoting activity in options is much higher than for the underlying and, secondly, due to a higher dimensionality, as for each underlying asset, there are different contracts with different maturities and strike prices. Moreover, the chain of available options changes everyday. Consequently, in terms of size and heterogeneity, option datasets exceed those of equities by an order of magnitude. This is particularly true for actively traded equity indices and stocks. For example, during the period from January to August 2015, for Apple, we observe on average 271,641 trades and 3,283,067 quotes per day, implying an average quote-to-trade ratio of 12.1. This activity in the spot market is accompanied by an activity in the option market, amounting, on average, to 77,849 trades and 226,822,053 quotes per day across all 1,169 option classes written on Apple, suggesting an average quote-to-trade ratio of 2,913.6

[1] Intraday option prices can be used to enhance answers to many of the interesting questions expertly explored by our friend and colleague Peter Christoffersen, including the valuation of options when volatility has multiple components (Christoffersen et al., 2008), the risk-neutral dynamics of volatility (Christoffersen et al., 2010), the dynamic behaviour of the implied volatility smirk (Christoffersen et al., 2009) and estimates of systematic equity risk provided by option prices (Chang et al., 2012).

[2] Another study by Beber and Brandt (2006) utilizes high-frequency options data written on bond futures to examine how macroeconomics news influences investors' beliefs and preferences.

for Apple options.

The major contribution of this paper is to provide a comprehensive overview of the structure and features as well as pitfalls and opportunities of high-frequency option data. We rely on the most detailed and comprehensive database of the U.S. option markets, provided by the Options Price Reporting Authority (OPRA), which collects and disseminates intraday option trades and quotes data at a millisecond precision from all option exchanges operating in the U.S. The database covers all option classes written on more than 3,500 underlying equities, more than 500 exchange traded products and about 50 index-driven assets. Though this database allows for a thorough analysis of high-frequency activities across the entire market, it is only scarcely used in recent research.[3] Hence, our aim is to provide scholars with guidance on how to process and utilize such data, and provide an explorative overview and initial evidence regarding the data's potential and benefits.

Therefore, we contribute to the literature by providing (i) an overview of the institutional and regulatory settings of the fragmented option market in the U.S., based on all 15 U.S. security exchanges eligible for options listing and trading, (ii) an extensive descriptive analysis of intraday trades and quotes for options written not only on equity indices but also on individual stocks and other exchange traded products, and (iii) two applications demonstrating the advantages of high-frequency option data relative to the standard end-of-day data.

A sound understanding of the regulatory environment is necessary for properly exploiting high-frequency information on option markets. Therefore, we review the most important regulatory requirements determining the transparancy of trading, price formation, and market maker quoting obligations. The latter rules are especially important for understanding price formation since option markets are driven by the quotes of market makers, rather than from the limit order flow, as is the case for most stock markets.

Moreover, we provide an assessment of the quality of the raw data and its informativeness. We analyze the presence of possible falsely reporting misrecordings, other data errors, non-informative or irregular quotations, outliers, and the occurrence of records with identical time stamps. We present several examples to highlight data specifics and pinpoint pecularities. Finally, we provide guidance on how to identify erroneous or irrelevant records and develop a filtering algorithm in the spirit of the cleaning algorithm proposed by Barndorff-Nielsen et al. (2009) for high-frequency equity data from the trades and quotes (TAQ) database released by the New York Stock Exchange.

Focusing on data originating from 12 underlying equities and 5 exchange traded funds during the trading days in January 2015 leaves us with approximately 27 billion trades and quotes. To provide an explorative overview, we present evidence on trading and quoting activities per underlying asset, side of the market, time to maturity and moneyness and inspect liquidity characteristics based on several spread-based liquidity measures. We show that the fraction of possible data errors, strong outliers and irregular records is negligible. However, the fractions of records with identical millisecond timestamps and potentially non-informative quotes of minimal size can reach several percent for particular option classes.

The OPRA database allows us to study the quoting and trading of an instrument across all market places. This enables us to compare the cross-exchange trade and quote flows and to provide an assessment of price leadership. The latter analysis rests on a comparison of the relative frequency of participation

[3] Ae rare exception is Amaya et al. (2018) who, however, concentrate exclusively on S&P 500 index options.

in the National Best Bid and Offer (NBBO) quoting pair. Our analysis reveals that CBOE, AMEX and ARCA most often participate in the NBBO for the considered option classes.

Finally, we provide two empirical applications using OPRA data to highlight the information content of high frequency option prices. The first application focuses on the estimation of the risk-neutral variance of an asset return, which can be approximated by the value of an appropriately designed portfolio of option prices (Carr and Madan, 1999). Employing a method similar to the one in Andersen et al. (2015a), we construct a replicating option portfolio from second-by-second cross-sections of option prices to produce intraday series of the risk-neutral variance of SPY (an exchange traded fund) and GOOG (Google stock). We find that the option implied variances of SPY and GOOG returns vary significantly during the trading day and are negatively related to changes in the corresponding underlying price. Such time-varying intraday dynamics of the option implied variance cannot be extracted from OptionMetrics data, a standard dataset used in the literature only providing a single cross-section of end-of-day option prices.

In the second application, we construct estimates of risk-neutral return densities (RND) from intraday panels of option prices using the method of lognormal mixtures advocated by Ritchey (1990), Melick and Thomas (1997) and Liu et al. (2007). We compare them with density curves estimated from a single cross-section of end-of-day option prices from the OptionMetrics database, and find that the estimated RND curves using high-frequency data during the last 15 minutes of a trading day in general coincide with the end-of-day benchmark, confirming the reliability of the intraday OPRA data. Moreover, utilizing high-frequency data allows us to examine the variability of estimated RND curves at different times during a trading day. Our study shows that the RND curves estimated from medium-term options written on GOOG become more negatively skewed right after the release on the FOMC announcement at 2pm Eastern time on January 28, 2015, suggesting a higher probability of a negative outlook for the price of GOOG in the medium term. In addition, the estimated RND curves for GOOG at the beginning and the end of the same trading day exhibit noticeable differences. Such observations cannot be obtained from the widely used end-of-day OptionMetrics data.

The rest of the paper is organized as follows. In Section 2 we provide an overview of the U.S. option exchange trading industry. Section 3 contains an overview of the OPRA dataset, the characterization of its special records, and some suggested data filtering rules. A thorough description of the observed trade and quote record for the selected data sample along with detailed summary statistics are presented in Section 4. Section 5 provides two empirical examples of using intraday option prices to estimate the risk-neutral variance and risk-neutral density of the underlying asset returns. Section 6 concludes the paper. All additional materials are provided in the Appendix.

2 Options Trading in the U.S.

2.1 Overview of the U.S. Options Market

As of 2018, 15 exchanges operate in the U.S. on which equity options are traded and there are five holdings that own one or more of these exchanges. These include Nasdaq (PHLX, NOM, BX Options, ISE, GEMX, MRX), CBOE Holdings (Chicago Board Options Exchange, C2, BATS BZX, BATS EDGX), Intercontinental Exchange (NYSE AMEX, NYSE Arca), Miami International Holdings (MIAX, MIAX

Exchange	Symbol	2015 Volume	2015 Market share	2016 Volume	2016 Market share	Comments
Nasdaq PHLX	X	623,514,666	15.05%	582,093,570	14.33%	
Nasdaq NOM	Q	286,845,485	6.92%	283,822,637	6.99%	
Nasdaq BX	T	31,496,313	0.76%	29,872,954	0.74%	
Nasdaq ISE	I	482,171,761	11.64%	457,206,056	11.25%	
Nasdaq GEMX	H	117,616,072	2.84%	89,189,491	2.20%	
Nasdaq MRX	J			6,432,661	0.16%	Started in February, 2016
CBOE	C	1,043,031,630	25.17%	1,033,349,820	25.43%	
C2	W	79,226,943	1.91%	91,025,788	2.24%	
BATS BZX	Z	396,415,424	9.57%	412,034,701	10.14%	
BATS EDGX	E	1,465,760	0.04%	36,408,407	0.90%	Started in November, 2015
NYSE AMEX	A	344,463,910	8.31%	296,493,779	7.30%	
NYSE ARCA	N	381,515,438	9.21%	388,979,129	9.57%	
MIAX	M	252,605,427	6.10%	247,112,479	6.08%	
MIAX Pearl	P					Started in February, 2017
BOX	B	103,268,442	2.49%	109,202,724	2.69%	

Table 1: Trading volumes and market shares of U.S. options exchanges in 2015 and 2016. Source: Option Clearing Corporation (OCC).

Pearl) and TMX Group (BOX).[4]

All of them are security exchanges that have been approved by the U.S. Securities and Exchange Commission (SEC) for options listing and trading. The clearing services for these exchanges are provided by the Options Clearing Corporation (OCC), regulated by SEC, which is the largest equity derivative clearing organization in the U.S. and is jointly owned by the exchanges. OCC manages market participants' risks, including counterparty credit risks and risks of trading errors. To control the counterparty risk, OCC imposes daily margin requirements to each account maintained at OCC.

Table 1 reports trading volumes and corresponding market shares for each exchange in 2015 and 2016. In Table A.1 in Appendix A, we provide a brief overview of all U.S. option exchanges. Although the main regulatory requirements are identical across exchanges, there is considerable heterogeneity in trading environments, price improvement mechanisms, pricing models, as well as matching and allocation algorithms.

2.2 Market Regulation

All of the approved option exchanges are authorized by SEC to act jointly as parties to a number of National Market System (NMS) plans. These plans centralize requirements across exchange markets in order to ensure cross-exchange protection, transparency, surveillance, standardization and audit trails.

The "Plan for Reporting of Consolidated Options Last Sale Reports and Quotation Information"[5] is one of the NMS plans aimed at processing and reporting trading information from all U.S. option exchanges. The Options Price Reporting Authority (OPRA) has been registered as a securities information processor to be responsible for the implementation of this plan. OPRA is regulated by a committee of representatives from all the participating exchanges.

[4]Note that our current sample period (January - August 2015) does not contain information from Nasdaq MRX, BATS EDGX and MIAX Pearl because these exchanges were opened later.

[5]In accordance with Section 11A of the Securities Exchange Act of 1934.

OPRA currently processes option trading data from all 15 U.S. option exchange markets. According to this plan, OPRA requires its participants to report last sale information and information about current option quotes[6] (including prices, quotation sizes, and some regulatory auditing information). The Securities Industry Automation Corporation (SIAC) provides technological infrastructure (including communication lines and hardware) for collection, consolidation and dissemination of this information in real time. OPRA provides market data to professional (directly or through the vendors) and non-professional subscribers (only through the vendors) for a fee.

In 2005, SEC adopted the Regulation NMS - a set of rules aimed to improve fairness and efficiency in price formation across securities eschanges, by means of boosting information dissemenation and competition among exchange markets as well as among individual orders. The most important parts of this regulation are the following. First, the so-called "Access Rule" ensures fair access to the market data (Rule 610). Second, the "Sub-Penny Rule" implies the presence of a minimal price increment (Rule 612). Third, the "Market Data Rules" are related to the market data revenue allocation. Finally, the "Order Protection Rule" (or "Trade-Through Rule") prevents an exchange from executing transactions at prices that are inferior to publicly displayed protected quotes at another exchange (Rule 611).

Although the Regulation NMS was supposed to make markets more liquid, efficient and transparent, many of its rules are subjects of debates and critique, see, for example, Chung and Chuwonganant (2012). One of the most controversial is the "Trade Through Rule" adopted in 1975.[7] It protects investors' orders from execution at prices inferior to the National Best Bid and Offer (NBBO).[8] However, it enforces routing fractions of an order to other exchanges in case they display a better quotation, thus deteriorating both the speed and certainty of execution.

The Regulation NMS was originally passed for trading NMS stocks, not for derivatives trading. In 2009, SEC approved the Decentralized Options Linkage Plan[9] that incorporates several specific amendments and modifications reflecting option trading specifics. According to the Decentralized Options Linkage Plan, the Order Protection Rule prevents trade-throughs during normal market conditions, but allows for a number of exceptions. The latter include, for instance, intermarket sweep orders, system glitches, and transactions which occured during a crossed market. In addition, the Decentralized Options Linkage Plan requires exchanges to avoid locked and crossed markets, i.e., markets where a protected bid is equal to (higher than) a protected offer.

There are three more plans which are mandatory for all exchanges. The Options Regulatory Surveillance Authority (ORSA) Plan has been adopted in 2006 and is intended to prevent insider option trading.

[6]Each national security exchange is obligated to collect, process, and make available this data in accordance with Rule 602 of the Regulation NMS.

[7]See a detailed dissent report at https://www.sec.gov/rules/final/34-51808-dissent.pdf, or brief discussions, for example, at www.etf.com/docs/Nasdaq_Primer.pdf, https://www.cboe.com/publish/ComLet/20150511.pdf or https://www.greenwich.com/account/time-fresh-look-trade-through-rule. See also Stoll (2006) for a discussion of the Order Protection Rule from the perspective of electronic trading and Hendershott and Jones (2005) who investigate the effect of the Trade-Through Rule exemption on market quality.

[8]The NBBO represents the highest bid and the lowest offer among best bids and offers (BBO) from all exchanges at each instant. It is calculated from the flow of protected quotes, i.e., the best bid or offer of all U.S. option exchanges processed through the OPRA system that are publicly displayed and disseminated in accordance with the OPRA plan.

[9]The Decentralized Options Linkage Plan is the revised version of the original option market linkage plan that was previously adopted in 2000. The revised plan has been introduced partially in a response to the growth of option trading and the evolution of the trading environment on option markets. The key difference is that the new plan replaced a centralized market linkage structure with a decentralized one.

It is aimed to maintain "the regulation, surveillance, investigation and detection of the unlawful use of undisclosed, material information in the trading of securities options". The Options Listing Procedures (OLP) Plan was introduced in 2006 "to facilitate the listing and trading of standardized option contracts on each of the exchanges". The Consolidated Audit Trail (CAT) was initially filed in 2014 with the objective to collect all orders and identify them as cancellations, modifications and excecutions for all exchange listed equities and options across all U.S. markets.

2.3 Options Trading and Market Maker Obligations

The core trading session in the U.S. option markets begins at 8:30 and lasts until 15:00 Central Time (CT) every business day. Index and exchange traded product (ETP) options have an extended trading session until 15:15 CT. In addition, some exchanges (BATS BZX and BATS EDGX) provide a pre-market trading session up to two hours in advance.

The OCC summarizes the most important product-specific information, sets daily position limits (250,000 contracts for most liquid stocks), and requires minimum customer margins (up to 120% of the aggregate contract volume for writers of uncovered options).

A standard equity or ETF/ETP option contract covers 100 shares of the underlying and an exercise of a contract results in physical delivery of the underlying shares. In case of stock splits or dividend payments, the number of shares and the excercise price are adjusted accordingly. Options usually expire at the close of the trading session on the third Friday of each month. Depending on the demand and the liquidity of the underlying, further weekly expiring options can be issued. In fact, the issuance of such weekly options has exploded in recent years.

The option market is a *quote driven market*, where bid and offer quotes are posted by market makers. Often, there are different types of market makers that differ in privileges and responsibilities. More senior categories (e.g., lead or primary market makers), which are assumed to have more trading experience, are granted with an allocation priority in the relevant option classes, but they are subject to stricter capital requirements and quoting obligations and bear other responsibilities beyond ordinary market makers. Regular market makers may also be registered as preferred or directed market makers who have certain privileges in executing preferenced and directed orders.

There are several market-wide obligations for option market makers. First, the SEC requires market makers to post firm quotes, i.e., to guarantee that the option quotes are valid for order executions for at least one contract.[10] Second, in 2010, the SEC proposed an amendment to the local exchange rules that prohibit market maker stub quotes, i.e., quotes that are far away from the prevailing market.[11] According to the new rule, quotes are required to be within a certain percentage band around the NBBO (or around the consolidated last sale, if the NBBO is not available). These requirements are supposed to make options trading less risky for investors and prevent transactions from being executed at irrational prices.

By the Quote Rule, market makers have an obligation to provide continuously updated two-sided

[10] In February 2001, SEC introduced a market-wide firm quote obligation that was an amendment to the Quote Rule (Securities Exchange Act Rule 11Ac1-1) that previously was applied to the equity market. Before that, option exchanges had imposed different firm quote requirements on market makers independently of each other.

[11] Stub quotes might be posted when market makers attempt to fulfill quoting obligations with no actual intention to trade. Stub quotes are considered among the main factors leading to the Flash Crash on May 6, 2010.

quotes throughout a trading day. Each option exchange (or a self-regulatory organization) imposes specific quoting obligations on its market makers. In general, these quoting obligations are independent from market conditions. Therefore, during the episodes of market stress, market makers are supposed to keep providing liquidity, thus absorbing the impact of shocks on individual investors.[12] The requirement of continuous quoting is especially important for the option markets where an appreciable fraction of securities is highly illiquid.

Most exchanges require market makers to quote in at least 90% the time during the trading day, with compliance being assessed on a monthly basis. The quote size should exceed a minimum number of contracts, usually determined by an exchange on a class-by-class basis, and may vary with the type of the market maker. Moreover, market makers are supposed to continuously quote in some minimum fraction of the option classes and series to which they are assigned. Such a minimum fraction ranges from 60% to 100% across exchanges and market makers' types. More detailed information about the quotation requirements in U.S. option markets is collected in Table A.2 in Appendix A.

The minimum tick size – minimum possible price increment – depends on the price level of a given option. Thus, for options that are traded below \$3, the minimum tick constitutes \$0.01 for the option classes participating in the option penny pilot program and \$0.05 for the other classes. For options traded above \$3, the minimum tick is \$0.05 for the classes from the penny pilot program and \$0.10 for the other series. The options written on the several market-wide ETFs (namely, QQQQ, IWM, and SPY) and option-related products (XSP and VIXW) represent an exception with a minimal increment of \$0.01 for all corresponding option series.

3 Data Overview, Special Records and Data Filtering

In this section, we provide a general overview of the OPRA data, select a working data sample and inspect the quality of trade and quote records within it. First, we categorize trades and quotes which might be considered as irrelevant for certain types of analyses. Then, we suggest a filtering algorithm that aims to remove such records from the raw OPRA data. Finally, we inspect the selected data sample and explore how often records from the specified categories actually occur in the raw data.

3.1 Data Overview

The OPRA Data (Bulk) package of CBOE Livevol[13] covers all transactions and top-level quotes disseminated from all U.S. option exchanges in accordance with the OPRA Plan on a millisecond basis. Each OPRA record corresponds to either a quote or trade event realization for one of the available option contracts (which are identified with the underlying, expiration date (tenor), strike price and contract type – put or call) on one of the U.S. option exchange markets.

Each transaction record displays the transaction price and the corresponding trading volume. Each quote record contains top-level bid and ask prices together with the corresponding quoted amounts. Each new quote record reflects an update of a bid-ask pair (either a change in prices, or a change in quoted

[12] See, for example, https://www.sec.gov/comments/s7-05-15/s70515-34.pdf and Nagel (2012).

[13] CBOE Livevol is a leading data provider in equity and index options technology and services, see https://datashop.cboe.com.

	Asset classes			Specific examples		
	Equity	ETF/ETN	Index	AAPL	SPY	SPX
Underlyings	3,686	566	47			
Option symbols	3,814	585	53			
Option classes	620,123	156,162	33,690	1,169	3,685	2,544
Quotes	5,021,787,050	1,965,243,131	321,813,772	226,822,053	520,188,594	10,183,398
Trades	678,120	222,143	50,896	77,849	75,509	10,994
Volume	8,871,502	6,024,324	1,672,754	1,029,155	2,521,188	493,663
Notional	$19,523,973	$9,016,319	$19,080,347	$3,436,161	$4,534,401	$13,133,924
MPS	214,635	83,994	13,755	9,697	22,234	436
QPC	8,098	12,585	9,552	194,031	141,164	4,003
QPT	7,405	8,847	6,323	2,914	6,889	926
QPV	566	326	192	220	206	21
QPN	257	218	17	66	115	0.78

Table 2: OPRA average daily statistics based on data from January until August 2015. MPS refers to messages per second, while QPC, QPT, QPV and QPN are the numbers of quote messages divided by option classes, trade messages, traded volume and traded notional in USD, respectively.

amounts) relative to the preceeding quote record on a given exchange market. Together with each option quote or trade record the data also provides the most recent first-level quotes of the underlying.[14] A detailed description of the content and structure of the OPRA dataset is provided in Appendix B.

Our dataset spans the first eight months in 2015, containing a total of 167 trading days. Altogether, we identify 3,686 equities, 566 exchange traded funds (ETF) or exchange traded notes (ETN) and 47 indices as underlying instruments in the data. In addition, 335 underlyings have a non-standard deliverable, resulting in multiple listed option symbols for the same underlying.[15] On average, 160 contracts are listed per option symbol per day. For the most liquid underlying assets, however, we might find up to 4,000 different contract variations. In terms of expiration dates, 45% of the contracts can be classified as standard (i.e. equity, ETF or index options). However, weekly (20%), quarterly (25%) and LEAPS (Long Term Equity Anticipation Securities - 10%) options are listed in considerable proportions as well. On the contrary, Mini Options rarely appear and can only be found for five of the most traded equities and ETFs.[16]

Between January and August 2015, OPRA recorded 1.22 trillion quotes and nearly 159 million trade records, and a traded volume of more than 2.76 billion contracts with a total notional amount of 7.95 billion USD. Table 2 summarizes average daily quotes and trades with respect to the three main asset classes. The most actively traded option classes within each group are AAPL, SPY and SPX. These underlyings have multiple option classes,[17] but for illustrative purposes we only report statistics for the standard classes.

To provide a representative overview of the OPRA data and to keep it feasible and succinct, our descriptive analysis covers only a small subset of the underlyings. We select a sample of twelve stocks and

[14]This is the case when the underlying is a tradable instrument. For example, SPX index option records do not contain such quotes since they are written on the market index which is not a traded asset by itself. In contrast, records of the SPY options, written upon the tradable exchange traded fund, contain the most recent top quotes of this underlying.

[15]Often, non-standard deliverables are Mini and Jumbo options, but can be, for example, also corporate action affected stocks.

[16]Furthermore, on the S&P 500 Index, we can also find the non-standard option types - *binary options* and *range options* (BSZ and SRO, respectively). Another *binary* exists for the S&P 500 Volatility Index (BVZ). These irregular option types are only listed and traded on the Chicago Board Options Exchange.

[17]For example, the mini options AAPL7 and SPY7 were not considered as standard belongings to any option class.

five ETFs from different sectors, mirroring the heterogeneity in terms of the (average) number of records (see Table 3). We focus on equities and ETFs to study cross-exchange patterns, which are non-existent for index options. Furthermore, in this paper we do not consider option classes such as Mini and Jumbo options, or corporate action adjusted ones.[18]

Sector	Asset	Ticker	Name	Industry	Average records (per day)
Consumer staples	Equity	PG	The Procter & Gamble Company	Personal Products	22,194,052
	Equity	WMT	Wal-Mart Stores, Inc.	Discount, Variety Stores	20,171,655
	Equity	K	Kellogg Company	Processed & Packaged Goods	5,461,349
	ETF	XLP	Consumer Staples Select Sector SPDR Fund		5,517,801
Energy	Equity	XOM	Exxon Mobil Corporation	Major Integrated Oil & Gas	31,764,057
	Equity	CVX	Chevron Corporation	Major Integrated Oil & Gas	28,095,240
	Equity	MPC	Marathon Petroleum Corporation	Oil & Gas Refining & Marketing	4,389,222
	ETF	XLE	Energy Select Sector SPDR Fund		42,901,922
Financial	Equity	JPM	JPMorgan Chase & Co.	Money Center Banks	32,209,032
	Equity	BAC	Bank of America Corp	Money Center Banks	24,726,836
	Equity	BLK	BlackRock Inc	Asset Management	1,701,449
	ETF	XLF	Financial Select Sector SPDR Fund		7,465,856
Technology	Equity	AAPL	Apple Inc.	Electronic Equipment	226,899,902
	Equity	GOOG	Alphabet Inc.	Internet Information Providers	61,812,973
	Equity	EA	Electronic Arts Inc.	Multimedia & Graphics Software	11,184,655
	ETF	XLK	Technology Select Sector SPDR Fund		6,963,341
Global	ETF	SPY	SPDR S&P 500 ETF		520,264,103

Table 3: Sample of underlying assets for descriptive analysis.

3.2 Potentially Irrelevant Observations

Depending on the analysis to be implemented, some OPRA records might be irrelevant, redundant or even introduce undesirable noise through data errors or market microstructure peculiarities. We classify such special records largely in agreement with the corresponding categorizations for the algorithms in the TAQ data cleaning suggested in the prior literature (Brownlees and Gallo (2006) and Barndorff-Nielsen et al. (2009)). Our classification contains, however, several categories specific for high-frequency data from options markets.

We suggest six categories of OPRA records that, depending on the context, may be eliminated. For expositional convenience, we provide all six classes of special records in Table 4, along with the detailed rules for authentication. For each group, we introduce a filter that applies separately for trade and quote records from a given exchange.

The first category (F1) consists of all observations recorded before the start of a regular daily trading session (8:30 CT), or after the close (15:00 CT for stock options and 15:15 CT for ETF/ETP options). Although some exchanges (e.g., BATS) accept early quotation and even run pre-market trading sessions, such records are typically more noisy and might cause time mismatching, if observations from several exchange markets are considered jointly.

The second group (F2) constitutes entries which are potentially infected by data errors. In particular, we identify trade records with a zero transaction price or zero size. Likewise, we filter quotes associated with negative spreads as well as entries with zero offer price or size. Note that, in contrast to the stock

[18]However, by including AAPL and SPY, two of the most liquid assets among all 4,562 option class symbols, we still cover more than 10% of the entire OPRA quote data (see Table 2).

Group	Description	Trades	Quotes
F1	Records outside the regular trading hours	Entries with timestamps outside the normal trading period (from 8:30 until 15:00 CT for underlying stocks and until 15:15 CT for underlying ETF/ETP)	
F2	Records with possible misrecordings and data errors	Entries with zero transaction price or size	(a) Entries with zero offer price or size (b) Entries with a negative spread
F3	Records with irregular conditions	Entries which correspond to canceled transactions (observations with condition codes 'A', 'C', 'E', 'G' and 'O')	Entries which correspond to non-firm quotes (observations with condition code 'F')
F4	Records corresponding to non-informative quotes		Entries for which both bid and offer sizes do not exceed 1 contract
F5	Possibly outlying records	Entries for which a transaction price is either lower than the current bid price minus the current spread, or higher than the current offer price plus the current spread	(a) Entries for which the spread exceeds 50 median spreads on that day (b) Entries for which the mid-quote is by more than 10 mean absolute deviations apart from the rolling centered median (that is based on a rolling window of 50 observations) (c) Entries for which the ratio of offer price to bid price exceeds 5 when the bid price is non-zero.
F6	Records with identical timestamps	At the millisecond frequency, can be replaced with a single entry with the median transaction price and the total size. For lower frequencies, the observations with the latest millisecond time stamp can be used.	At the millisecond frequency, can be replaced with a single entry with the median bid and offer prices and the total sizes. For lower frequencies, the observations with the latest millisecond time stamp can be used.

Table 4: Classification of special trade and quote OPRA records according to six groups.

market data cleaning algorithm from Barndorff-Nielsen et al. (2009), we do not eliminate quotes with zero bid prices and positive offer prices. This scenario occurs regularly for deep out-of-the-money options and should not be treated as a data error.

The next category (F3) represents trades and quotes with specific conditions which indicate a potentially irregular nature of a given entry, such as records of cancelled trades and non-firm quotes.

The fourth group (F4) contains quote records with minimum size, because such quotes might provide less reliable information about the latent equilibrium price of an option. For example, if being uncertain about the fair price, market makers may try to discover it at the low exposure risk by posting a minimal quote at a distinct price level. Alternatively, due to the obligation to continuously maintain quoting throughout a trading day (see Table A.2 in Appendix A), market makers sometimes may find it optimal to use the minimum size quotes just to meet the requirements without providing extra liquidity. More particularly, we flag quotes with the size less than or equal to a single contract at both the bid and offer sides.[19]

We note, that since only the top bid and ask quote records are available in the OPRA dataset, we can not always identify whether a quote is of minimum size, or not. For example, each time a bid or ask price changes with a new quote record, we observe a different number of contracts available at a new price level. Since this number is typically unobservable in the data prior to the price change, it makes it impossible to determine the exact size of the newly arrived quote. Hence, into category F4 we attribute only those minimum size quotes which can be identified with certainty from the data.

The fifth group (F5) contains potentially outlying observations. In particular, we focus on trades with abnormally high or low transaction prices relative to the current spread. As for quotes, we identify entries with excessive spreads, extraordinary high or low midquotes, and entries for which the ratio between offer and bid prices is unusually high. We follow detection rules similar to those of Barndorff-Nielsen et al. (2009) for trades and quotes. In many cases, the presence of such outliers introduces undesirable irregularities in the data structure, rendering the analysis difficult. Moreover, some outlying records may reflect data errors or non-informative market microstructural artefacts, such as stub quotes.

Records with identical timestamps are referred to the sixth group (F6). It is often the case that many quotes or transactions are recorded at the same millisecond. When option prices are analyzed as a time series, it is reasonable to replace them with a single record. This can be done in several ways.

First, the data provides information about the exact order of records, so the last record within a given millisecond can be used as such a single entry to replace multiple entries with the same millisecond stamp. Though this approach seems natural and is supposed to deliver the most recent observations for each time stamp, it, however, has shortcomings. In particular, the approach discards all information contained in other records, preserving only the latest one. This may lead to information losses (regarding the aggregate transaction size in a given millisecond, for example). Moreover, such an approach is not robust to the occurrence of noisy or outlying records at the latest position. Therefore, the alternative is to use the median transaction, bid and offer prices and the total size, aggregated across the "simultaneous" records within the same millisecond. This approach not only robustifies the record against the presence of noisy observations, but also aggregates information contained in all entries at a given millisecond. More

[19] Note that we do not remove quotes which have a minimal size at only one side. Such situations often appear when one side of the market (supply or demand) dominates. For example, for deep out-of-the-money options offer orders usually dominate. Conversely, in the case of in-the-money options bid orders prevail.

generally, suppose that one chooses a frequency for analysis different from a millisecond (e.g., second or minute). Then, if the data is to be treated as time series, only the record corresponding to the latest millisecond timestamp is preserved within each frequency unit. When there are several such records, the aggregation scheme described above may be used.

The overall presence of records belonging to group F2 (and to lesser extent, groups F1 and F5) might signal the general data quality by indicating the fraction of possible misrecordings and data errors.

3.3 Empirical Evidence

We inspect trade and quote records from the raw OPRA data for all option contracts (among selected underlyings) expiring on February 20, 2015, throughout all the trading days in January, 2015. In total, our sample consists of 2.77 billion quote records and 730,000 trade records. Tables A.3 and A.4 in Appendix A contain detailed aggregate statistics on the presence of special records in the categories classified above across underlyings and exchange markets, respectively. Figures 1 and 2 summarize several key aspects of the results. We report the average daily fractions of records that belong to all special categories F1-F6 (white bars) as well as the fraction of records that belong to several selected categories (colored bars). Figure 1 contains the fraction of special option records for each underlying asset. Figure 2 depicts the fraction of special option records for each of the exchange markets.

Due to the quoting obligations for market makers, quote records are far more numerous than trade records. For example, in January 2015, SPY options expiring on February 20 have on average more than 66 million daily quotes across all exchanges[20] relative to less than 16,500 transactions. On average, the number of special entries for SPY options exceeds 8.4 million quotes (more than 12% of all SPY quote records) and about 2,000 trades (again more than 12% of all SPY trade records) per trading day.

We observe that, depending on the underlying asset, 5% to 25% of all available trade records have non-unique millisecond time stamps (category F5), which may be generated by a large order being executed against several smaller orders on the other side of the trade at the same millisecond. Moreover, this group constitutes on average more than 99% of all special records for trades. As evident from our data sample, groups F1, F2, F3 and F5 are almost irrelevant when trade data is inspected. This suggests that trade records are remarkably clean, valid and free of extreme outliers. As seen from Figure 2, trade records with identical millisecond stamps are especially frequent on NYSE Amex (exchange symbol A), BOX (B) and MIAX Options Exchange (M), reaching 30-35% of all trade records detected on these platforms. In contrast, on Nasdaq BX (T) and C2 (W) the fraction of such trade records is less than 1.5%.

In total, between 10% and 30% of all quote records for a given option class are identified as special records and fall into one of six groups described above. Quote records that fall outside the regular trading hours (F1) or are erroneous (F2) appear very infrequently (less than 0.1% of observations). Non-firm quotes (F3) appear rarely and never exceed 0.5% of the total quote record. Minimal size quotes (F4) appear frequently on all exchanges (sometimes exceeding 15% of the total quotes, as for the BOX exchange (B)). Potentially outlying quote records (F5) amount to 7.5% of all quotes for certain underlyings and appear with a similar frequency on all exchange markets. As can be seen from Figure 1, the fraction of quotes with identical millisecond stamps (F6) is substantial and varies from 3% to 12% of all quote entries across the underlyings. For example, quotes on Apple options with repeating timestamps reach almost

[20] On average, it implies about 240,000 quotes per traded contract for 282 available contract specifications.

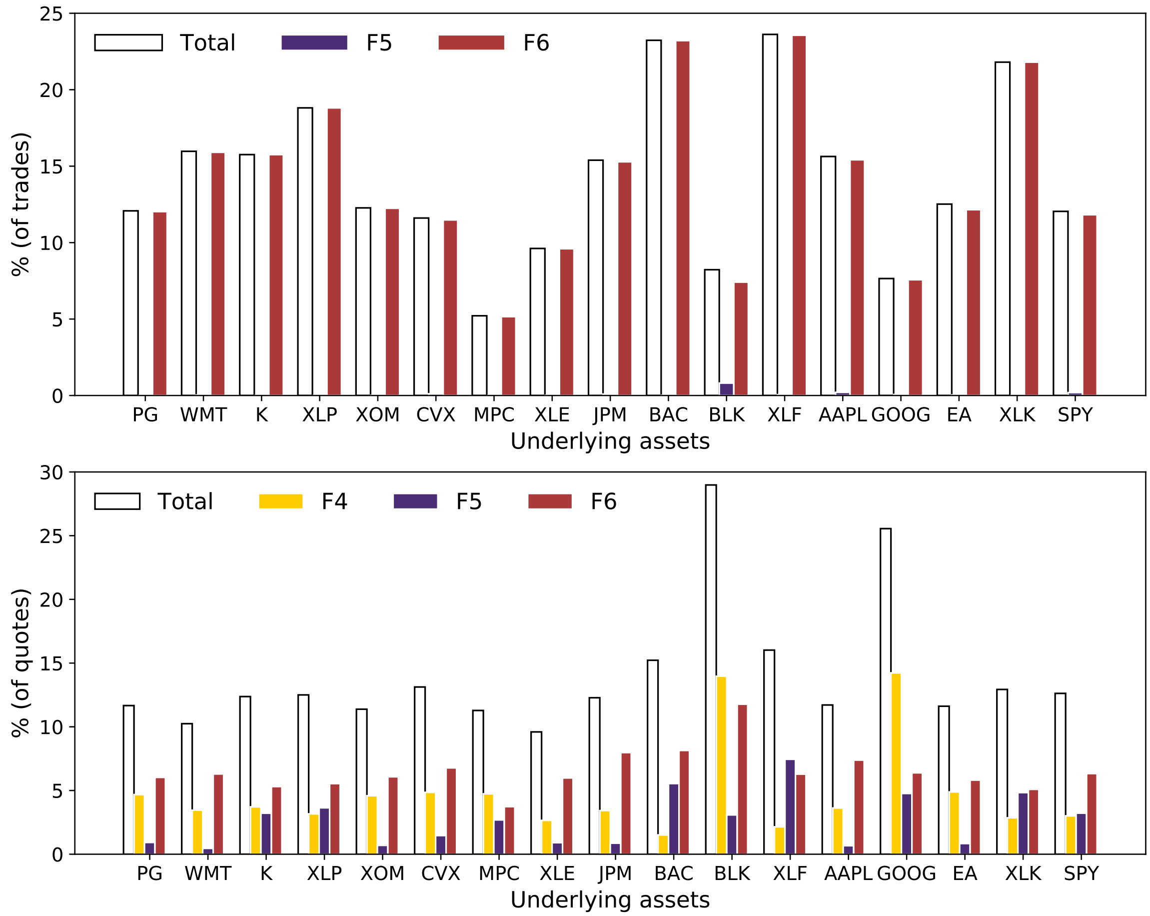

Figure 1: Average daily percentage of special trade (top panel) and quote (bottom panel) option records across selected underlying assets. White bars correspond to the total fraction of special records in the raw data (all categories, F1-F6). Yellow bars (available for quotes only) correspond to the fraction of records with the minimal quoting size (category F4). Blue bars correspond to the fraction of potentially outlying records (category F5). Red bars correspond to the fraction of records with the same millisecond time stamps (category F6). The results are based on the data for all option contracts traded in January 2015, which expire on February 20, 2015, observed on all available exchange markets.

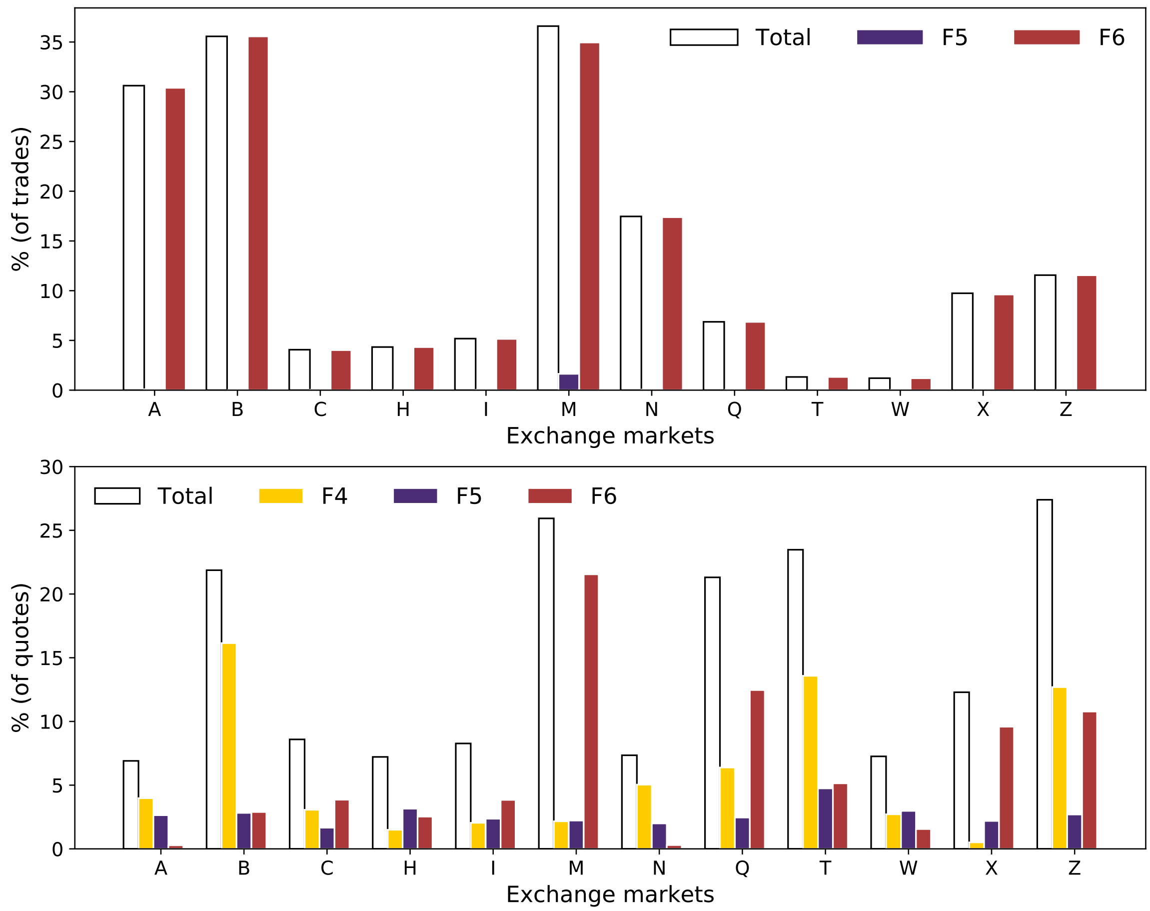

Figure 2: Average daily percentage of special trade (top panel) and quote (bottom panel) option records across 12 option exchange markets. White bars correspond to the total fraction of special records in the raw data (all categories, F1-F6). Yellow bars (available for quotes only) correspond to the fraction of records with the minimal quoting size (category F4). Blue bars correspond to the fraction of potentially outlying records (category F5). Red bars correspond to the fraction of records with the same millisecond time stamps (category F6). The results are based on the data for all option contracts traded in January 2015, which expire on February 20, 2015, observed on all available exchange markets.

2 million entries per day. Although group F6 contributes substantially to the amount of special records, its relative contribution is not as overwhelming as it is for trades. Thus, we conclude that OPRA quote records are mostly clean from data errors and invalid observations, but they contain a non-negligible amount of potential outliers.

4 An Explorative Analysis of OPRA Trades and Quotes

4.1 Basic trade and quote statistics

Table 5 reports the daily numbers of trades and quotes of option contracts written on the 17 underlying assets in January 2015, before and after applying the cleaning algorithm in Section 3.2. On average, 14.53% of the records are eliminated due to data filtering. The average daily number of option quotes and trades varies greatly across the underlyings and signals the liquidity of the latter. Amongst the 17 underlying assets, the most actively quoted (and traded) ones are the ETF SPY and Apple, whose average numbers of cleaned quotes per day exceed 628 million and 279 million, respectively. On individual day, the number of SPY quotes exceeds 870 million. Average order-to-trade ratios range between 3036.9 (for BAC) and 22433.6 (for EA), indicating that option markets rely on dramatically higher quoting than trading activities. The dominance of quotes over trades in option markets substantially exceeds that in equity markets, where average order-to-trade ratios rarely exceed 100, even for the order-to-trade ratios submitted by algorithmic and high frequency traders (see, for example, Hagströmer and Nordén, 2013, Brogaard et al., 2015).

	Underlying stocks																
	SPY	**AAPL**	**GOOG**	**EA**	**XLK**	**XOM**	**CVX**	**MPC**	**XLE**	**PG**	**WMT**	**K**	**XLP**	**JPM**	**BAC**	**BLK**	**XLF**
Number of records (before cleaning)																	
Avg	700,838,009	301,575,942	73,543,546	15,184,625	9,489,613	36,198,297	28,728,799	5,215,386	59,890,142	23,723,427	22,802,939	745,353	8,778,407	43,001,824	36,051,422	3,166,838	11,930,495
Std Dev	139,414,541	78,014,816	17,064,430	4,269,292	2,406,511	7,548,864	5,085,071	1,290,710	11,537,807	8,134,120	6,288,729	120,562	2,341,311	8,747,138	6,725,070	724,910	2,642,094
Median	694,355,725	310,310,810	71,363,687	14,629,507	8,729,688	34,021,556	27,206,324	5,066,729	58,010,337	24,989,843	23,184,688	718,706	8,477,310	41,790,218	34,030,353	3,139,884	11,100,126
Min	409,607,244	159,094,428	45,479,752	9,744,870	5,399,868	24,735,171	21,143,683	3,245,001	45,196,390	9,543,493	12,224,635	574,871	4,851,901	25,718,345	23,281,419	1,968,863	6,549,215
Max	984,010,543	417,942,352	113,190,636	23,388,291	13,105,051	49,712,308	38,714,736	8,314,240	84,261,758	38,230,503	33,970,789	981,853	13,133,224	59,398,468	49,911,031	5,340,371	17,999,368
Number of trades (after cleaning)																	
Avg	80,476	78,298	8,078	607	462	3,846	2,979	405	3,635	1,728	1,699	127	375	4,117	9,895	157	1,388
Std Dev	15,585	27,043	3,932	643	125	1,028	1,084	261	1,223	742	623	59	179	1,941	3,724	79	449
Median	80,004	72,670	6,637	424	456	3,820	2,828	301	3,277	1,513	1,515	122	337	3,443	8,869	115	1,321
Min	44,787	45,928	4,947	185	224	1,847	1,568	143	1,933	912	858	52	164	1,617	3,649	55	685
Max	104,560	162,313	20,738	2,898	706	5,809	6,149	936	6,542	3,714	2,911	243	826	9,552	19,100	350	2,299
Number of quotes (after cleaning)																	
Avg	628,611,886	279,290,504	64,487,534	13,617,224	7,987,132	32,616,857	25,518,930	3,958,422	54,090,705	21,348,654	20,543,501	508,302	7,325,981	38,322,047	30,050,269	2,532,815	9,503,469
Std Dev	125,112,503	72,431,205	14,818,791	3,903,640	2,323,915	7,434,977	4,757,349	1,572,582	11,359,835	7,476,082	5,899,002	136,216	2,196,516	7,838,284	5,886,635	699,342	2,421,343
Median	622,338,806	286,858,000	61,416,002	13,364,134	7,488,202	29,583,046	24,582,347	4,098,120	51,518,109	22,057,529	20,700,550	465,697	7,268,446	366,82,359	28,184,448	2,534,352	9,194,951
Min	368,646,694	147,174,936	40,321,726	8,347,573	3,747,030	22,156,109	19,131,646	1,293,294	37,783,596	8,222,172	10,856,256	318,438	3,309,087	22,522,639	19,449,233	1,336,222	4,186,745
Max	872,278,526	389,048,310	97,324,007	20,947,630	11,382,152	46,440,632	34,812,831	7,343,634	77,700,703	34,932,632	31,134,228	838,467	11,548,186	51,724,816	43,635,724	4,508,035	14,354,507

1st month data are used. "Trades-only" contracts are not considered for the calculation.

Table 5: Descriptive statistics for daily OPRA data records in January 2015.

4.2 Trade and Quote Records by Time to Maturity and Moneyness

The standard expiration day of an option contract is the third Friday of a month. Given the different economic incentives for options trading, a monthly investment horizon, however, seems insufficient. Therefore, for most underlyings, we also observe quarterly or yearly listed option contracts. In addition, in the past decade, so-called weekly options, or weeklies, have increased in importance. For instance, the trading volume of S&P 500 weekly options (SPXW) increased from about 12% in 2010 to 25% in 2014.[21]

According to the OCC contract specifications, regular options expire on the third Friday each month, weeklies on any other near-term Friday (up to five consecutive weeks), quarterlies on the financial quarter-end (not necessarily Fridays), and LEAPS are characterized by tenors greater than twelve months.

In the first eight months of 2015, most contracts in our sample (about 45%) can be classified as standard (i.e., equity options, ETF options or index options), but weekly (20%), quarterly (25%) and LEAPS (10%) options are listed in considerable proportions as well. In order to examine how quotes and volumes are distributed with respect to tenor, we classify the time to expiry into ultra short-term, short-term, mid-term and long-term according to the following classification,

$$\textit{Time to expiry} := \begin{cases} \textit{ultra short-term} & \text{for } T - t \leq 7, \\ \textit{short-term} & \text{for } T - t \in (7, 30], \\ \textit{mid-term} & \text{for } T - t \in (30, 90], \\ \textit{long-term} & \text{for } T - t > 90, \end{cases}$$

where T and t denote the expiration date and the current date, respectively, measured in calendar days.

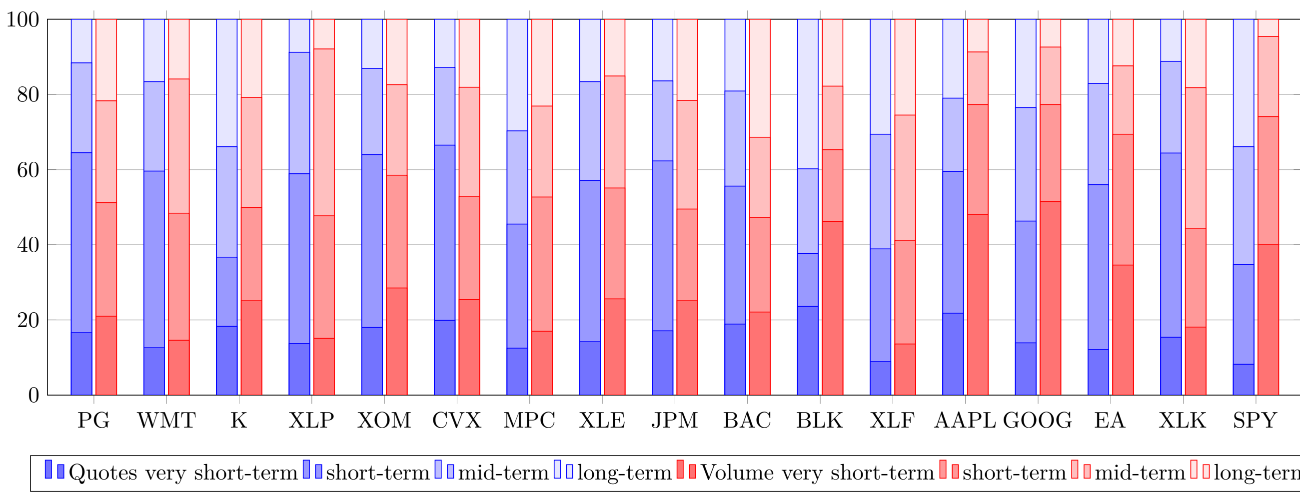

Figure 3: Average daily proportions of quotes and traded volumes by time to expiration (in %). The two stacked bar charts represent the proportions of quotes (left) and traded volumes (right) for each underlying between January and August 2015. Bars are partitioned by the *time to expiry* classified as ultra short-term (bottom), short-term, mid-term and long-term (top).

Figure 3 displays the average daily proportions of quotes and trading volumes by time to expiration. For most option classes, the relative trading volume is significantly higher than the relative quoting

[21]In August 2017, weekly options on S&P 500 Index exceeded the 50% mark on CBOE for the first time in terms of traded volume, see charts on www.cboe.com/products/stock-index-options-spx-rut-msci-ftse/s-p-500-index-options.

activity for ultra short-term contracts. For example, listed options on GOOG can be traded at twelve different expiration dates (14 for AAPL and 24 for SPY), but contracts with the nearest maturity alone account for 51.5% of all trades (48.1% for AAPL and 40.0% for SPY). This suggests that it is not only the weekly maturity profile which is steadily increasing in popularity and trading activity, but even less liquid classes without weeklies (e.g., BLK with 46.2%, MPC with 17.0% and K with 25.1%) are traded increasingly more in the last week before expiration.

In order to categorize trading and quoting activities according to moneyness, we define three moneyness groups according to the following scheme:

$$\textit{Moneyness types} := \begin{cases} \textit{Out of the money (OTM)} & \text{for } (S_t/K-1)\cdot CP < -5\%, \\ \textit{At the money (ATM)} & \text{for } |S_t/K-1| \leq 5\%, \\ \textit{In the money (ITM)} & \text{for } (S_t/K-1)\cdot CP > 5\%, \end{cases}$$

where S_t refers to the underlying mid-quote price at the time of the quote or trade, K is the exercise price and $CP = 1$ for calls and $CP = -1$ for puts. The classifying scheme implies that we have to consider intraday data for both options and the underlying assets in order to correctly determine the moneyness category of the observation at hand. Figure 4 depicts the corresponding proportions in quotes and trading volumes depending on these moneyness categories.

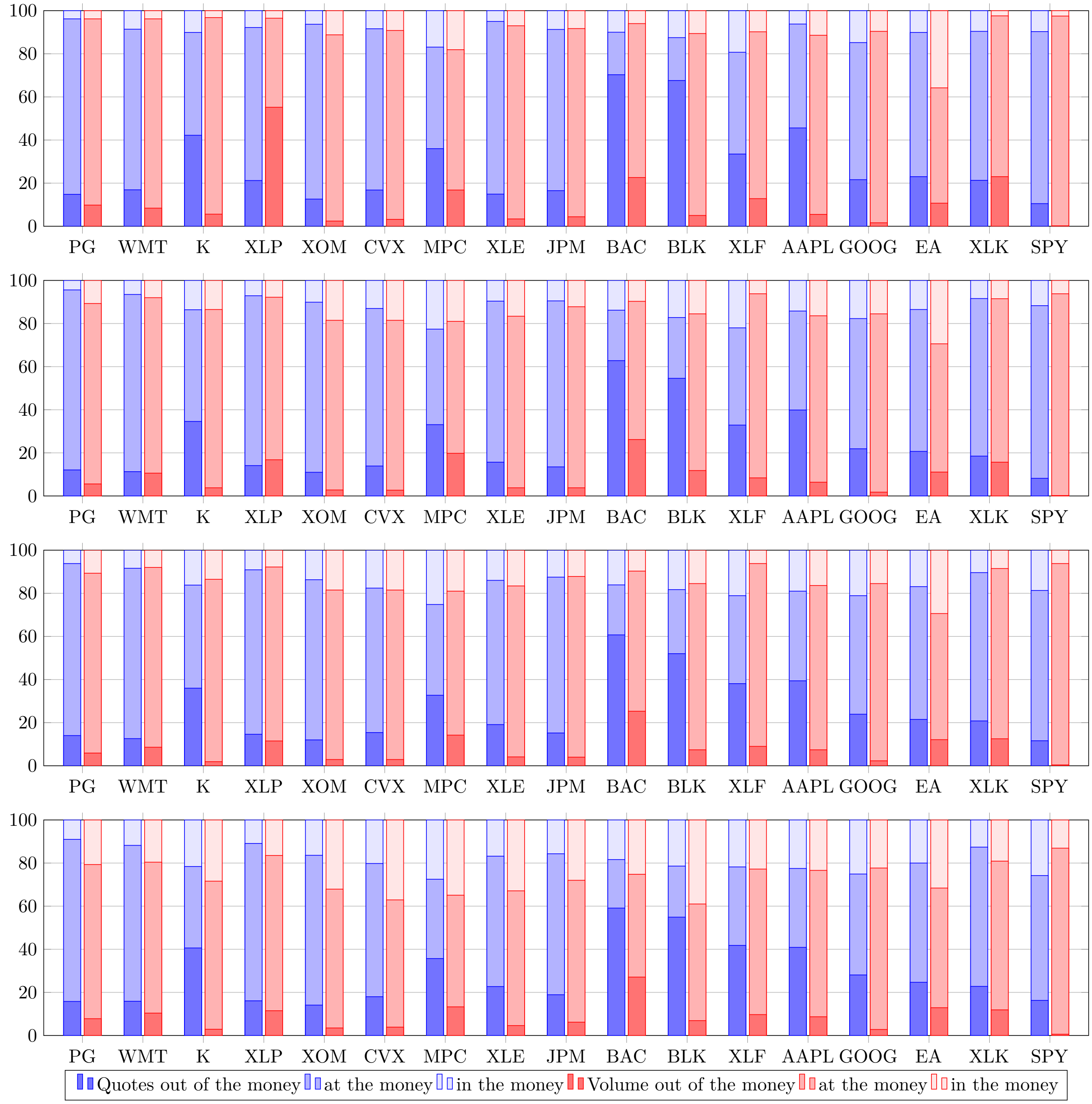

Figure 4: Average daily proportions of quotes and traded volumes by moneyness types (in %) for 4 maturity categories (ultra short-term (top plot), short-term, mid-term and long-term (bottom plot)). The two stacked bar charts for each underlying represent the proportions of quotes (left) and traded volumes (right) between January and August 2015, grouped by the *moneyness types* OTM (bottom), ATM (middle) and ITM (top).

In general, quotes for at-the-money options are updated most frequently for the majority of underlyings. Notable deviations, however, are representatives of the financial sector, such as BAC or BLK, where the quote activity of out-the-money options dominates. For the trading volumes, we note that out-of-the-money options are traded approximately twice as much as in-the-money options for short and ultra short-term (top two plots). For mid and long-term options (bottom two plots), the trend is reversed, but at-the-money trading still dominates across all underlyings and maturity terms.

4.3 Relative Activity across Exchange Platforms

A distinct feature of the OPRA data is the availability of trade and quote records disseminated by all U.S. option exchanges. Figure 5 shows the relative proportion of the quotation activity and trading volume across exchange holding groups and individual exchanges. In general, the pattern of relative trading activity across exchanges agrees with the evidence from Table 1.[22]

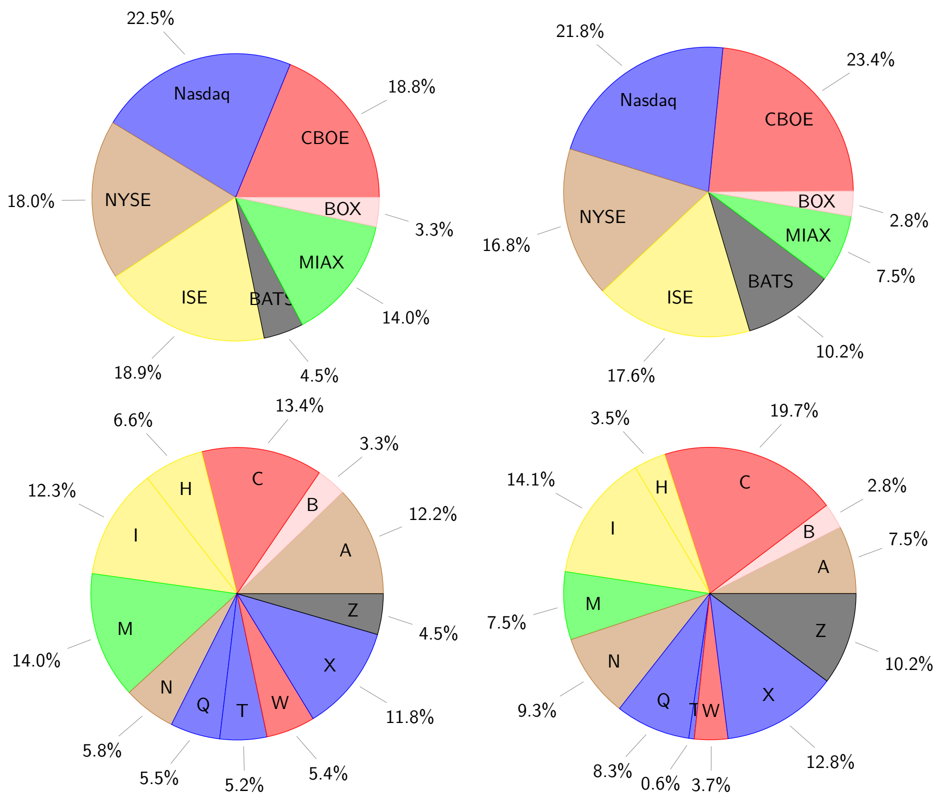

Figure 5: Quotation and traded volume market shares per exchange. Pie charts on the left-hand side reflect the quotation distribution among holding groups (top) and individual exchanges (bottom) based on the sample assets selection. Analogously, traded volumes are reflected on the right-hand side.

The oldest marketplaces within each holding group account for the majority of the trading (see bottom right pie chart in Figure 5). The same is generally true for the distribution of quotes, which are more than twice as frequent as for the secondary exchanges within each group. The only exception is NYSE with its exchange AMEX (A) that was launched three years later than ARCA (N), and has more than twice as many quotation messages in 2015. Both markets offer floor and complex trading, but they differ in their pricing and allocation structures (see Table A.1 in Appendix A). AMEX utilizes the classical *customer-priority* model with a *pro-rata* allocation, which encourages deeper liquidity. On the other hand, ARCA focuses on price efficiency exploiting *maker-taker* and *price-time* approaches. In this case, the liquidity seeking approach is more quote-expensive, but yields lower trading volume relative to the alternatives.

[22] This refers to the quoting and trading behavior in January 2015, so bear in mind that ISE and BATS were still independent from NASDAQ and CBOE, respectively. Furthermore, the new options exchanges MCRY, EDGX and MPRL had not been launched yet.

However, the Miax exchange, which also follows the *maker-taker* approach, had the highest number of quote updates in 2015 among all option marketplaces in the U.S. on the same set of underlyings. Hence, the pricing structure is not the only factor accounting for the heterogeneity in quotation and trading activity across the markets.

Based on the short time frame of eight months and the high volatility of daily market shares in options trading, it is difficult to detect a clear trend on exchange competitiveness. From Figure 6, the leading groups Nasdaq and Cboe started with approximately 25% market share each, ahead of Ise and Nyse with approximately 15% each. The remaining three exchange holdings, Bats, Miax and Box, started with significantly lower and less volatile daily traded volumes in 2015. In August, however, the four largest exchange holdings went head-to-head with market shares around 20%, reflecting a higher competitiveness, but also a shift of trading interest towards smaller exchanges. A remarkable growth in traded volumes can be observed for Bats in August 2015. They gained an additional 5% of market share, perhaps due to market stress that generated incentives for trade in the *pre-market trading session* offered solely by Bats.[23]

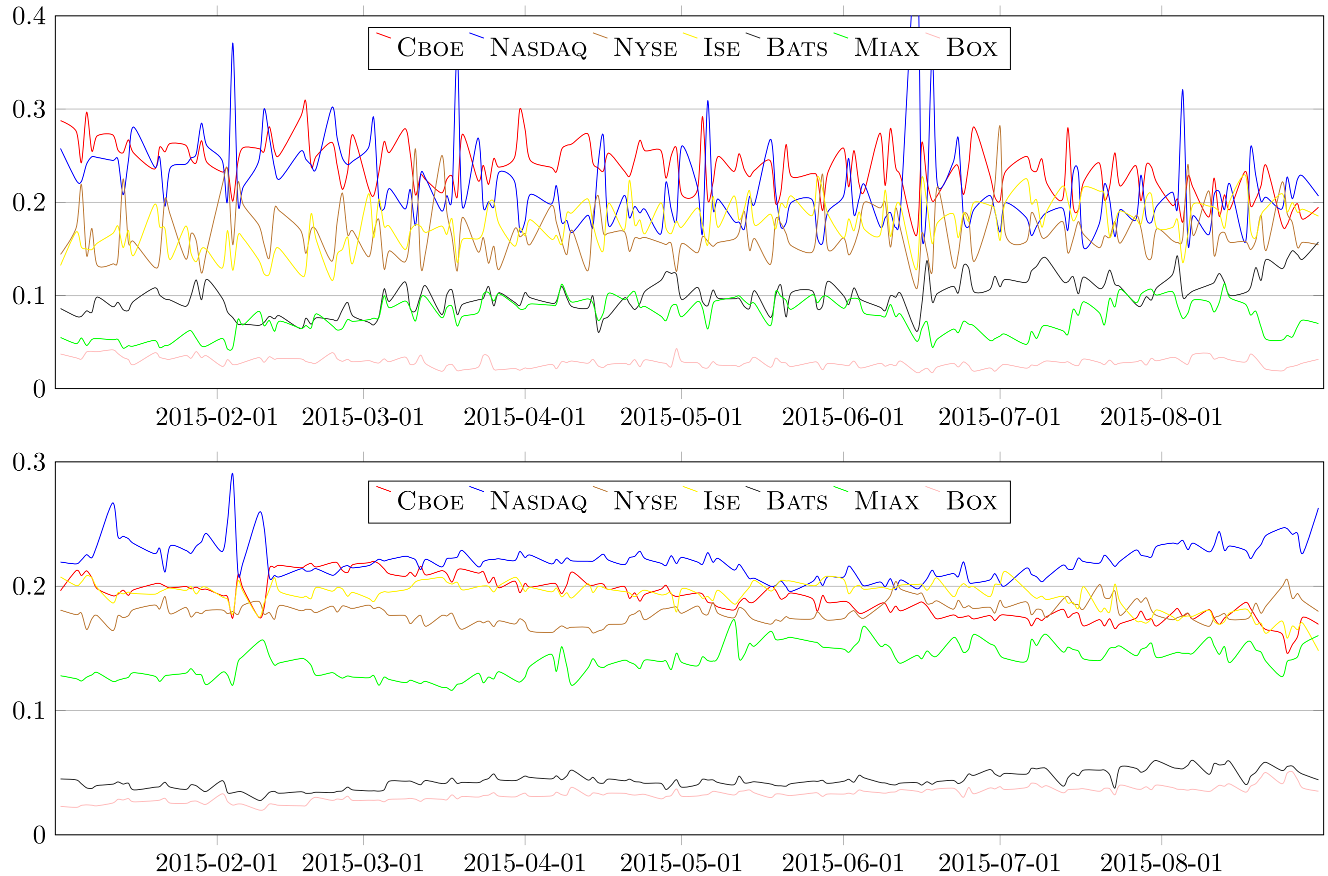

Figure 6: Daily average traded volume (top plot) and quotation (bottom plot) market shares per exchange holding group.

On the contrary, market shares of quotations are relatively flat. If the underlying market is more

[23] For example, Greek default on June 30, 2015 or China's stock market crash (Shanghai Composite shed 38% between June 12 and August 24.)

volatile, option markets proportionally increase their quote updates, while the overall market share remains relatively stable. Nasdaq was the only group operating three individual exchanges in this period (Cboe, Ise and Nyse run two exchanges). Therefore, it is surprising that Miax, with only one options exchange, increased its quotation frequency in line with the multi-exchange groups. However, with less than one third of the quote updates on the same set of underlyings, Bats still traded a higher contract volume on a daily basis.

Code	A	B	C	H	I	M	N	Q	T	W	X	Z
Exchange	AMEX	BOX	CBOE	GEMX	ISE	MIAX	ARCA	NOM	NOBO	C2	PHLX	BATS
PG	0.08	0.07	0.18	0.03	0.04	0.10	0.13	0.10	0.01	0.02	0.12	0.13
WMT	0.06	0.07	0.14	0.03	0.12	0.11	0.09	0.09	0.01	0.03	0.12	0.13
K	0.08	0.08	0.12	0.04	0.10	0.11	0.16	0.08	0.02	-	0.10	0.11
XLP	0.07	0.06	0.12	0.03	0.07	0.13	0.11	0.08	0.02	0.02	0.09	0.19
XOM	0.06	0.06	0.18	0.02	0.05	0.09	0.10	0.14	0.01	0.02	0.09	0.17
CVX	0.07	0.05	0.15	0.02	0.07	0.14	0.09	0.15	0.01	0.02	0.10	0.13
MPC	0.08	0.06	0.15	0.04	0.11	0.08	0.12	0.10	0.01	-	0.10	0.16
XLE	0.08	0.06	0.20	0.03	0.09	0.10	0.09	0.10	0.01	0.03	0.09	0.12
JPM	0.06	0.08	0.17	0.02	0.06	0.09	0.10	0.14	0.01	0.01	0.15	0.11
BAC	0.06	0.07	0.15	0.03	0.08	0.22	0.10	0.08	0.01	0.01	0.08	0.11
BLK	0.05	0.03	0.19	0.04	0.12	0.06	0.10	0.06	0.02	-	0.13	0.20
XLF	0.10	0.08	0.13	0.04	0.06	0.18	0.07	0.06	0.01	0.04	0.12	0.11
AAPL	0.07	0.07	0.18	0.03	0.08	0.13	0.09	0.11	0.01	0.02	0.08	0.15
GOOG	0.06	0.04	0.14	0.06	0.09	0.06	0.12	0.09	0.01	0.02	0.14	0.17
EA	0.06	0.05	0.12	0.02	0.06	0.05	0.12	0.15	0.01	0.04	0.19	0.14
XLK	0.07	0.06	0.12	0.03	0.07	0.16	0.08	0.07	0.02	0.02	0.13	0.16
SPY	0.07	0.07	0.18	0.03	0.15	0.16	0.07	0.07	0.01	0.03	0.07	0.10

Table 6: Shares of trading volumes per exchange for selected underlyings. Highlighting: $x \geq 0.15$, $0.15 > x \geq 0.1$, and $0.1 > x \geq 0.05$.

Tables 6 and 7 demonstrate trading and quoting activity across exchange platforms for option contracts related to individual underlyings from our sample. We can observe that CBOE consistently exhibits high options trading and quoting activity for all considered assets relative to other markets. In contrast, NOBO and C2 reveal relatively low activity compared to the other exchanges.

Code Exchange	A AMEX	B BOX	C CBOE	H GEMX	I ISE	M MIAX	N ARCA	Q NOM	T NOBO	W C2	X PHLX	Z BATS
PG	0.16	0.04	0.15	0.09	0.12	0.14	0.07	0.03	0.03	0.03	0.10	0.04
WMT	0.15	0.04	0.13	0.06	0.13	0.14	0.05	0.04	0.05	0.05	0.13	0.04
K	0.16	0.03	0.16	0.08	0.13	0.11	0.05	0.04	0.05	-	0.13	0.05
XLP	0.13	0.03	0.14	0.05	0.19	0.14	0.05	0.04	0.05	0.05	0.10	0.03
XOM	0.18	0.03	0.15	0.06	0.13	0.14	0.05	0.04	0.05	0.03	0.11	0.04
CVX	0.14	0.04	0.15	0.04	0.12	0.15	0.06	0.05	0.05	0.04	0.13	0.04
MPC	0.25	0.02	0.17	0.07	0.10	0.12	0.04	0.03	0.03	-	0.15	0.03
XLE	0.15	0.03	0.14	0.07	0.15	0.13	0.05	0.04	0.05	0.05	0.10	0.03
JPM	0.14	0.03	0.14	0.07	0.15	0.14	0.05	0.04	0.04	0.04	0.12	0.04
BAC	0.10	0.03	0.13	0.11	0.13	0.12	0.05	0.05	0.07	0.03	0.13	0.04
BLK	0.14	0.04	0.15	0.06	0.09	0.15	0.06	0.05	0.07	-	0.15	0.04
XLF	0.12	0.03	0.13	0.10	0.15	0.14	0.04	0.04	0.05	0.06	0.10	0.03
AAPL	0.13	0.03	0.16	0.07	0.11	0.15	0.04	0.04	0.05	0.06	0.11	0.04
GOOG	0.15	0.04	0.11	0.08	0.09	0.13	0.08	0.06	0.07	0.03	0.12	0.06
EA	0.16	0.03	0.16	0.05	0.13	0.10	0.06	0.05	0.04	0.03	0.13	0.05
XLK	0.11	0.02	0.13	0.14	0.15	0.13	0.04	0.04	0.04	0.05	0.10	0.03
SPY	0.11	0.04	0.12	0.06	0.13	0.14	0.06	0.06	0.05	0.06	0.12	0.05

Table 7: Shares of aggregate quotation amounts per exchange for selected underlyings. Highlighting: $x \geq 0.15$, $0.15 > x \geq 0.1$, and $0.1 > x \geq 0.05$.

Another important point to address is the relative participation of exchange markets in the NBBO. This analysis sheds additional light on the relative exchange efficiency. For each exchange, underlying asset, and second-to-second stamp, we compare the BBO of a given exchange against the National BBO. We only consider the core-trading session, lasting from 8:30 a.m. until 3:00 p.m. CT, to avoid the results being shifted in favor of Bats, which is the only exchange group offering a pre-trading session.[24]

From Table 8, C2 and GEMX are the least competitive exchanges with respect to NBBO participation for almost all underlyings. Therefore, it is not surprising that these exchanges had less than half the quote updates compared to the market leaders. However, two other "small" exchanges, BATS and BOX, with even less quotation messages (see Figure 5) exhibit significantly higher rates of NBBO participations. In general, however, Table 8 suggests that the overall price quality for first-level quotations in the U.S. is remarkably similar across all twelve options exchanges.

[24]For example, the value '0.50' in Table 8 for exchange CBOE and underlying SPY means that CBOE's market makers quote the best bid and offer prices for option classes on SPY among all market makers on other U.S. option exchanges for 3.25 hours per day on average.

Code Exchange	A AMEX	B BOX	C CBOE	H GEMX	I ISE	M MIAX	N ARCA	Q NOM	T NOBO	W C2	X PHLX	Z BATS
PG	0.56	0.62	0.64	0.47	0.55	0.65	0.69	0.63	0.63	0.43	0.54	0.64
WMT	0.57	0.56	0.64	0.59	0.62	0.62	0.65	0.58	0.61	0.50	0.61	0.55
K	0.52	0.44	0.43	0.37	0.40	0.47	0.52	0.45	0.47	0.00	0.44	0.47
XLP	0.53	0.47	0.47	0.42	0.41	0.47	0.55	0.47	0.47	0.33	0.45	0.47
XOM	0.61	0.50	0.57	0.60	0.56	0.58	0.57	0.56	0.54	0.44	0.64	0.53
CVX	0.63	0.43	0.54	0.48	0.49	0.55	0.55	0.51	0.55	0.30	0.62	0.50
MPC	0.58	0.58	0.56	0.49	0.54	0.61	0.62	0.59	0.59	0.00	0.55	0.59
XLE	0.68	0.67	0.69	0.66	0.64	0.69	0.71	0.69	0.68	0.48	0.70	0.62
JPM	0.67	0.68	0.70	0.69	0.62	0.71	0.71	0.72	0.70	0.57	0.66	0.71
BAC	0.38	0.36	0.41	0.29	0.36	0.40	0.45	0.44	0.40	0.31	0.41	0.45
BLK	0.61	0.49	0.57	0.38	0.37	0.61	0.66	0.53	0.49	0.00	0.64	0.47
XLF	0.63	0.51	0.53	0.51	0.46	0.52	0.63	0.52	0.52	0.42	0.49	0.52
AAPL	0.76	0.58	0.74	0.62	0.69	0.72	0.73	0.71	0.72	0.66	0.69	0.70
GOOG	0.66	0.46	0.62	0.67	0.42	0.54	0.57	0.58	0.69	0.20	0.69	0.53
EA	0.51	0.53	0.54	0.38	0.51	0.51	0.60	0.50	0.53	0.45	0.47	0.52
XLK	0.57	0.51	0.54	0.51	0.53	0.54	0.57	0.51	0.55	0.46	0.53	0.52
SPY	0.47	0.36	0.50	0.40	0.45	0.44	0.47	0.42	0.41	0.39	0.49	0.41

Table 8: Heatmap on NBBO participation per exchange and underlying. Percentages of average time each exchange displays the two-sided NBBO prices, calculated from all option contracts within a given class. For multiple messages per minute, only the last quoted levels per exchange and asset were considered. Quotation before 8:30 a.m., after 3:00 p.m. and those having either zero bid or ask sizes were neglected. Highlighting: green ≥ 0.6, $0.6 >$ light green ≥ 0.5 and yellow ≤ 0.4.

4.4 Relative Spreads

We analyze the spreads observed in option quotes across different markets, stocks and moneyness levels. For this purpose we focus on the relative spread statistics. The relative spread is given as

$$RS_t \stackrel{\text{def}}{=} \frac{Q_t^A - Q_t^B}{MQ_t},$$

where MQ_t is a mid-quote price at time t. To keep the analysis manageable, we consider only those option contracts traded in January 2015 and expiring on February 20, 2015. Thus, the tenor of the options ranges from three to seven weeks. We remove entries belonging to the filtering groups F1, F2, F3 and F5 from our quote records to focus on regular quotes and avoid the impact of outliers.

We primarily focus on out-of-the-money option contracts. The mid-quote prices for these options are typically much lower than the underlying, so the corresponding relative spreads are less impacted by the latter. The situation differs for in-the-money options, for which a mid-quote price is of a similar order as the underlying price, so both the price level and spread of the underlying substantially affect the relative spreads for such options.

Figure 7 illustrates the average relative spread of high freqency quote records. Statistics are provided for twelve different option exchanges and separately for call and put option contracts (on the top and bottom panels, respectively). Each dot indicates the average relative spread for quotes corresponding to a particular underlying asset and given moneyness range (out-of-the-money, at-the-money or in-the-money).

First, as expected, the relative spreads for OTM options substantially exceed those for ATM and ITM options for both put and call contracts and for all exchange markets, partly reflecting the mechanical effect that OTM option prices are lower than those of ATM and ITM options. Howefer, it may also highlight a higher degree of disagreement about their market valuation. Second, we observe that although the cross-exchange pattern of relative spreads is similar for both put and call options, OTM call options have systematically larger spreads than OTM put options. This reflects the systematically higher liquidity and trading intensity of put options, which are often regarded as a hedging instrument against sudden market downturns. Finally, we notice that the scatter of average spread levels is roughly similar across the twelve options exchanges, implying that they exhibit similar spreads.

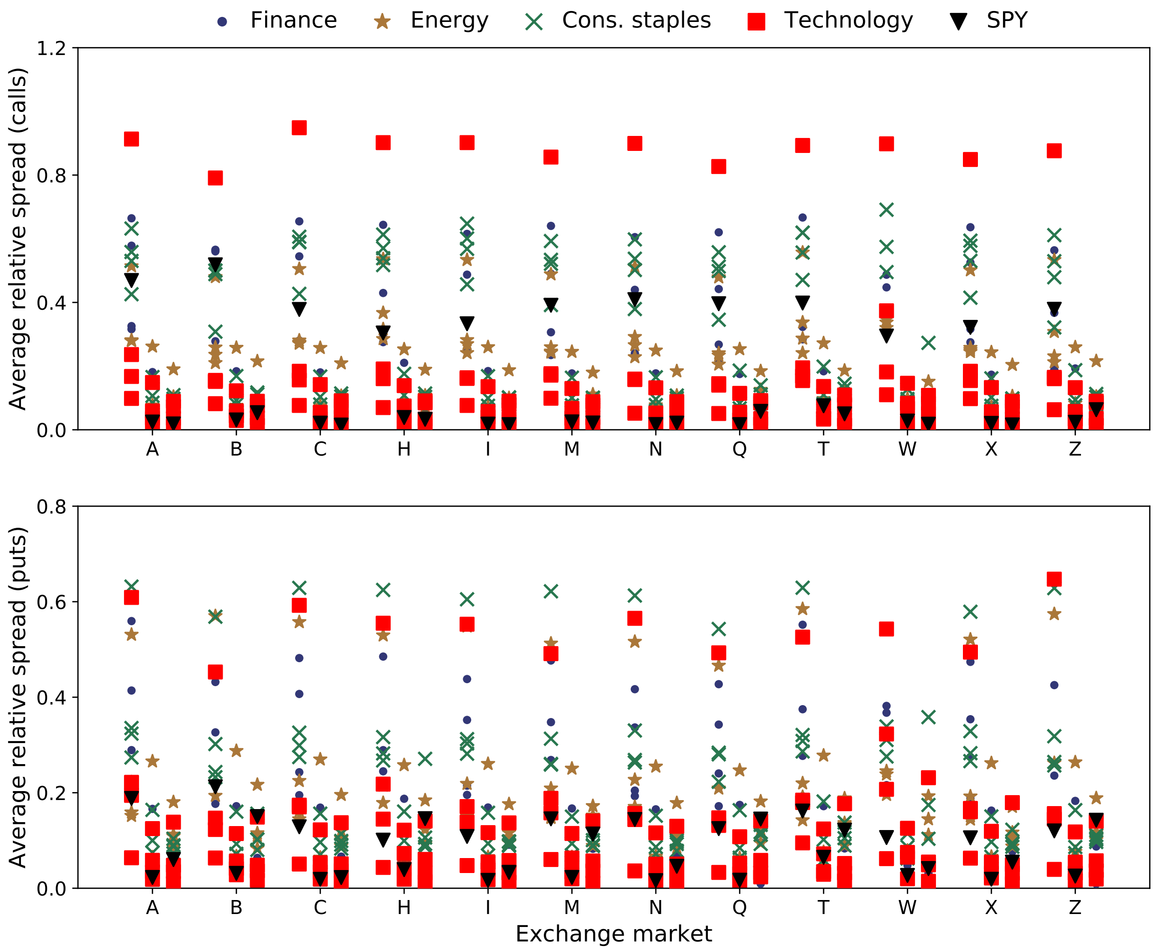

Figure 7: Average relative spread measures computed from the OPRA quote records for call (top panel) and put (bottom panel) option contracts for 17 selected underlying assets (different markers correspond to the underlying assets from different sectors). Average spreads are provided for 12 exchange markets (horizontal axis) for all option contracts traded in January 2015 and expiring on February 20, 2015. For each exchange market we plot three columns with average relative spreads corresponding to the out-of-the-money (left), at-the-money (center) and in-the-money (right) options.

Figure 7 also reveals that the relative spreads of option contracts vary substantially from one underlying to another, even within the same sector. Options written on more actively traded assets, such as SPY and AAPL, have narrower relative spreads irrespective of the exchange market, as demonstrated in Tables A.5 and A.6 of Appendix A, which provide detailed statistics on average relative spreads for the options of all 17 assets under consideration. We also notice from these two tables that options written on different underlyings exhibit distinct cross-exchange relative spread patterns. For example, while the lowest relative spreads for OTM AAPL options are found on NYSE Arca and Nasdaq, those for OTM CVX options are detected on the BATS and BOX exchanges.

5 Empirical Applications

In this section, we illustrate the usefulness of high frequency option prices by means of two empirical applications. In the first application, we obtain second-by-second model-free estimates of the risk-neutral variance based on intraday cross-sections of option prices. In the second application, we focus on the estimation of a risk-neutral asset return density from high-frequency option prices.

5.1 Real Time Measures of the Risk-Neutral Variance

An important piece of information that can be extracted from option prices is a market expectation of volatility of an underlying asset return. In particular, a cross-section of option prices for a range of strikes allows us to estimate the expected variance of an asset return under the risk-neutral probability measure, or simply the risk-neutral variance. Such an option implied variance can be especially useful for volatility trading, hedging and portfolio management purposes as it contains a predictive component. Furthermore, it can serve as a signal about economic uncertainty expected by the market, providing thus an important indicator of macro-financial linkages.

A risk-neutral variance can be computed with a cross-section of option prices, but without specific modeling assumptions. The method exploits the fact that the expectation of the average variance of an asset return from the present until a given future time can be replicated by an appropriate portfolio of European-style options with the corresponding expiration period (see, e.g., Carr and Madan, 1999). In the absence of arbitrage, the market value of such an option portfolio approximates[25] the risk-neutral asset return variance. This measure is referred to as the model-free implied variance, or MFIV.[26]

The typical option data used in the academic literature consists of end-of-day cross-sections of prices (e.g., OptionMetrics data). This format allows for a single estimate of a risk-neutral variance per day only. This is restrictive, as it rules out an investigation of the dynamics of implied variances in real time, the reaction of volatility expectations to specific events and announcements that occur during a day, intraday co-dynamics of implied variances across multiple assets, etc. Furthermore, the quality of the MFIV based on a single cross-section of prices might suffer from noise or data errors in the observed prices, whereas a more robust measure can be constructed when several cross-sections are available for a given day.[27]

Given high-frequency observations of option prices, intraday series of the implied variance can be extracted from multiple intraday cross-sections. Two distinctive features of the OPRA data support the accuracy of the extracted information. First, a major resolution of timestamps for recorded trades and quotes enables us to retrieve cross-sections of option strikes at very high frequencies. For sufficiently

[25] A perfect replication is possible only when a cross-section of options with a continuum of strike prices is used. In practice, a finite set of strikes provides an approximation to the risk-neutral variance.

[26] We use terms *option implied variance* and *MFIV* interchangeably when we refer to the risk-neutral variance extracted from option prices in this manner.

[27] For example, assuming that all the relevant state variables remain approximately constant over a short time interval, a more precise implied variance measure can be constructed by averaging over multiple implied variances estimated within this window.

liquid assets, intraday cross-sections can be extracted at frequencies of one second and higher. Second, the information flow from multiple exchange markets allows us to reduce the number of missing observations in the cross-sections of option prices. This is especially important for illiquid option contracts.

A celebrated example of an intraday approximation to the implied variance is the VIX index disseminated by the CBOE. The VIX reflects the expected risk-neutral volatility of the S&P 500 index over a 30-days period and is calculated from intraday mid-quotes of standard and weekly SPX options. The VIX is released at a 15-second frequency and provides a benchmark indicator of volatility expectations for the U.S. stock market in real time. It is worth mentioning, however, that Andersen et al. (2015a) documented a non-trivial delay in the high frequency VIX series when they compared it to a replication index computed from actual option prices. They found that the distribution time of the CBOE VIX lagged the time stamp of the option prices used in the computation of the replication index by 15-45 seconds and, furthermore, such a time gap varied unpredictably over time. On the contrary, the OPRA data allows us to construct MFIV series that will timely reflect the dynamics of the market variance implied by option prices at even higher frequencies than those at which the VIX is disseminated.

The OPRA data enables us to construct real-time implied variance series not only for the market index, but for a variety of stocks and ETFs traded on the U.S. equity market, for a range of time horizons (determined by options expiration dates) and for higher frequencies than each 15 seconds. In our empirical application, we extract an MFIV index in real time using high frequency option prices from the OPRA dataset. In particular, we obtain cross-sections of out-of-the-money option prices on a second-by-second basis. Then, from each cross-section we construct a replication portfolio that provides us with an instanteneous risk-neutral variance measure at every second of the trading day.

The method we use to construct a replication option portfolio from a cross-section of option prices resembles the one used in Andersen et al. (2015a), though different in several aspects. For instance, we use a log-linear extrapolation of option prices in the strike domain for the tails of a return distribution, rather than relying on a possibly noisy tail truncation via the specific cut-off rule for the strike range (as it is applied for the VIX computation), or a robust corridor-based measure (as it has been originally suggested in Andersen and Bondarenko (2007)). For a detailed description of our option portfolio design, we refer to Appendix C.

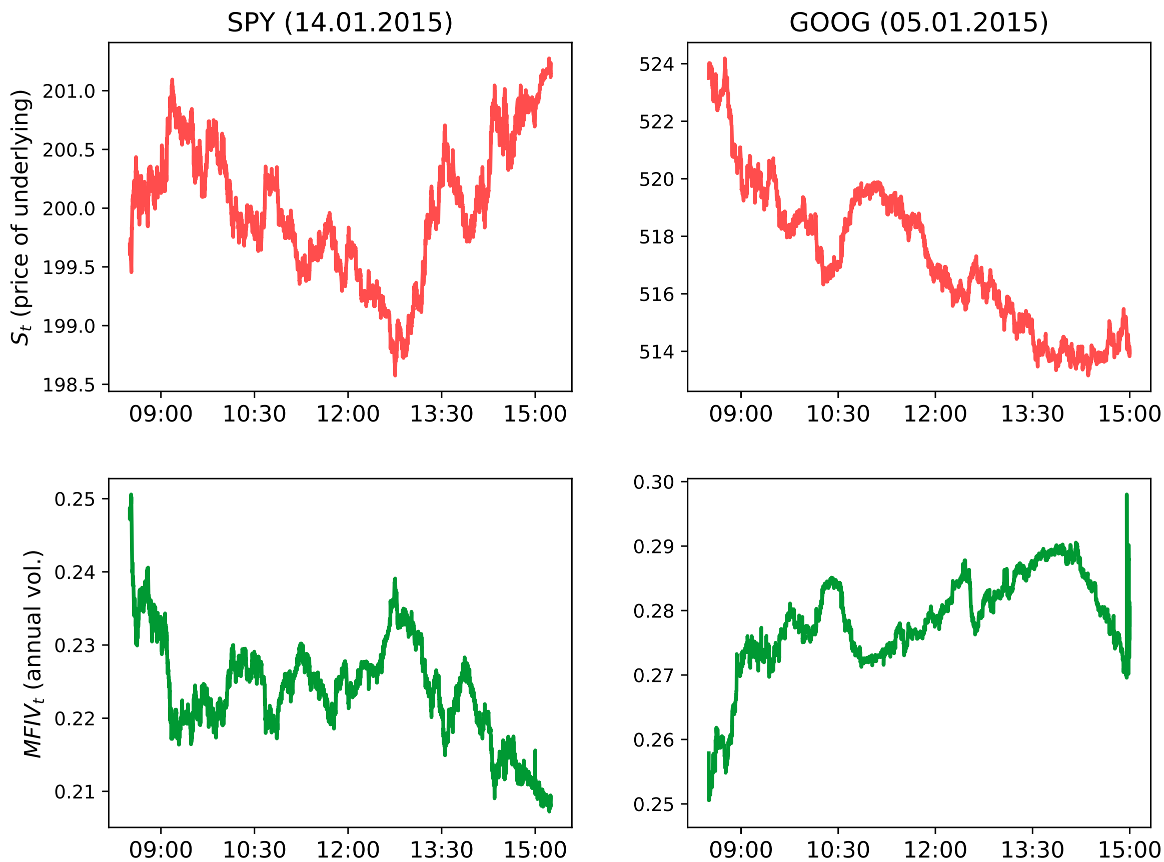

Figure 8: On the top plots, intraday prices of SPY (left side) and GOOG (right side) observed on January 14 and January 5, 2015, respectively. On the bottom plots, intraday MFIV series constructed with SPY options on January 14, 2015 (expired on January 17, 2015) and with GOOG options on January 5, 2015 (expired on January 9, 2015). Real time MFIV is calculated on a second-by-second basis and normalized to annual volatility units.

Here, we provide a representative example of MFIV estimation in real time. Figure 8 depicts a series of risk-neutral variances extracted at a second frequency using intraday option quotes on SPY and GOOG. To calculate the corresponding MFIV series for SPY and GOOG, we use options contracts with 3 and 4 days until expiration, respectively. We choose such close-to-expiry contracts for two reasons. First, for ultra-short maturities, the difference between the values of American and European-style options is negligible, so we can avoid the use of price corrections and proceed with American-style option prices. Second, a possible variation of a risk-free interest rate and dividend yields can be ignored for such a short time interval.

Figure 8 plots the intraday prices on underlyings SPY and GOOG (top panel) and the second-by-second series of the annualized model-free implied volatility estimates (bottom panel). The intraday dynamics of the option implied variances reveal several representative features. First, the level of the risk-neutral variance exhibits an appreciable amount of variation within a trading day. This fact is not surprising as the option prices are tightly related to the price of the underlying asset which might be highly volatile. However, this fact directly motivates us to take into account multiple MFIV measures based on several intraday option cross-sections in order to obtain a more reliable and robust signal about the future volatility expected by the market. This is contrary to the use of a single measure based on end-of-day prices which can be distorted by the dynamics of the underlying near the market closure.

Second, for both SPY and GOOG, the extracted implied variance is very volatile during the first minutes after the market open. This might be an artefact not only of a highly volatile underlying price due to intense price discovery right after the market opening, but also of disagreement and uncertainty about future volatility among market participants. Finally, we notice that the dynamics of the option implied variance is negatively related to the dynamics of the underlying price (realized correlations computed on 2-minute increments are -0.86 for SPY and -0.47 for GOOG). Thus, an implied variance tends to rise when the underlying price falls and, conversely, it falls in case of a rise of the underlying price. This negative correlation is in line with the natural intuition that a downward movement of an asset price is associated with an elevated pricing of uncertainty regarding future returns.

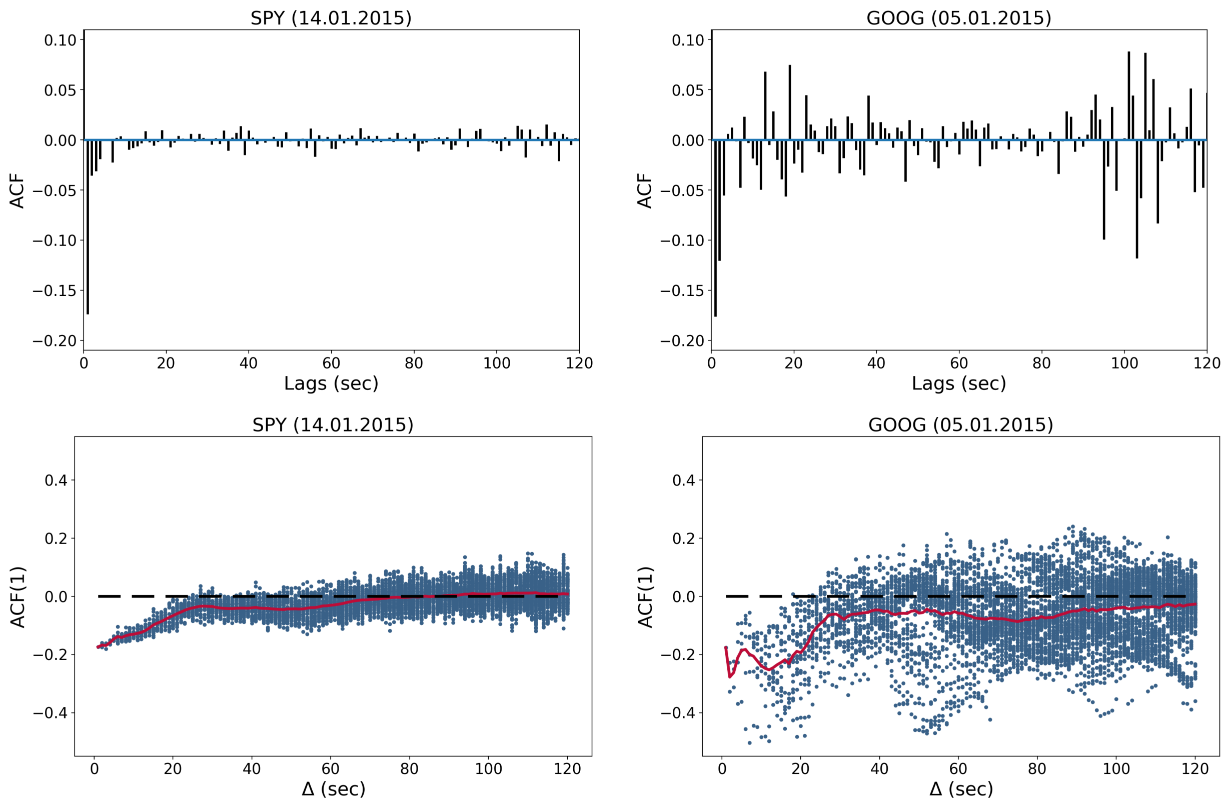

Figure 9: Autocorrelation functions for MFIV series constructed with SPY options (left side) on January 14, 2015 (expired on January 17, 2015) and with GOOG options (right side) on January 5, 2015 (expired on January 9, 2015). Top plots show the autocorrelations as functions of lags constructed for the second-by-second increments of intraday MFIV (with the maximum lag of 120 seconds). Bottom plots picture the first-order serial correlations as functions of a sampling frequency (Δ) calculated for the increments of intraday MFIV obtained at the corresponding frequency. We consider Δ ranging from 1 sec to 120 sec with a second step. Blue dots represent the first-order autocorrelations computed for a given Δ on multiple sampling "grids" achieved by shifting the initial MFIV observation by one second (thus, for Δ=1 sec we have one "grid" and for Δ=120 sec we have 120 "grids"). Solid red line is an average autocorrelation across all "grids" for a given Δ.

Figure 9 illustrates dynamic properties of the constructed high-frequency MFIV indices. In particular, we consider the serial correlations of MFIV increments for different lags and different sampling frequencies. In the top panels of Figure 9, we show the empirical autocorrelations for the second-by-second MFIV increments with the lags ranging from 1 to 120 seconds. In the bottom panels, we depict the first-order autocorrelations of MFIV increments for a range of sampling frequencies Δ (from 1 to 120 seconds). For each $\Delta = 1, ..., 120$, we compute the serial correlations of MFIV increments on Δ different sampling grids.

The first observation of the initial grid is taken at 8:35:00 CT and consecutive observations follow with a step of Δ seconds.[28] Each next grid is obtained by shifting the previous grid one second forward. The initial grid and $\Delta - 1$ grids produced by the one-second shifts constitute Δ different sampling grids. Thus, for each sampling frequency Δ, we have Δ serial correlation estimates (blue dots) obtained for Δ distinct sampling grids. The red solid line corresponds to the average first-order autocorrelations calculated across all grids for each sampling frequency Δ.

The autocorrelation functions of both SPY and GOOG share several qualitatively similar features. The autocorrelation functions of the second-by-second MFIV increments exhibit substantial negative dependence for ultra-short lags (3-4 seconds) and then, for higher lags, change signs randomly and become smaller in magnitude. From the bottom plots, it is evident that the first-order serial correlations of the MFIV increments are strongly negative for high sampling frequencies. As the sampling frequency gets lower, the serial correlations tend to concentrate around zero.

Interestingly, the autocorrelation properties of the SPY and GOOG intraday implied variances reveal several important differences as well. First, apart from the first-order, the autocorrelations of the SPY's MFIV second-by-second increments are substantially smaller in magnitude. Second, the first-order serial correlations of the SPY's MFIV increments are much more robust to the choice of the sampling grid, indicated in the bottom plots of Figure 9 by a less dispersed locations of the autocorrelation estimates (blue dots) around the average. Finally, the negative first-order serial dependence of the MFIV increments persists over different ranges of sampling frequencies for both SPY and GOOG. For SPY, the negative serial dependence in the MFIV increments on average disappears when sampling every 70 seconds and longer, whereas, for GOOG, it remains even for Δ equal to 120 seconds.

The evidence indicates that the high frequency MFIV series and, in particular, the corresponding increments for SPY demonstrate much more regular and, possibly, less noisy dynamics than the ones obtained for GOOG. This effect might be partialy attributable to a substantially different number of strikes in the OTM options cross-sections (182 for SPY against 57 for GOOG). Hence, both the Riemann approximation of the theoretical implied variance (which is the core of the MFIV estimation) and the cross-section reduction of microstructural noise in the observed options mid-quotes are much more effective for the SPY options.

5.2 High-frequency-based Estimation of the Risk-Neutral Density

An important source of information to be extracted from a cross-section of option prices is the risk-neutral density (RND) of the underlying asset price. Under no-arbitrage assumptions, an option price is the discounted expectation of the future payoff under the risk-neutral distribution. Hence, the measurement of this distribution is key for option pricing. Apart from this, the estimation of the option-implied return density has implications for a broad range of applications in macroeconomics and finance as it conveys important forward-looking signals about the economic uncertainty, investors' sentiments and their attitudes to risk.

Most related studies consider the extraction of a risk-neutral density from a single cross-section of

[28]We exclude the MFIV estimates calculated for the first 5 minutes of trading to avoid distortions related to market opening effects, including a lack of observable mid-quotes for some strikes and excessively volatile mid-quote revisions due to active price discovery.

daily (end-of-day) option prices. Intraday options prices significantly increase the amount of available information. Specifically, several intraday cross-sections can be used, in lieu of a single end-of-day cross-section. The purpose of our empirical illustration is to investigate whether the use of intraday data can improve the precision of the estimated risk-neutral densities and whether it provides any additional information beyond what can be extracted from end-of-day option prices.

Below, we describe our methodology of RND estimation from high frequency option prices based on mixtures of log-normal densities, as advocated by Ritchey (1990), Melick and Thomas (1997) and Liu et al. (2007). The results are compared to the benchmark RND computed from end-of-day OptionMetrics prices.

5.2.1 Estimation methodology

As documented by Breeden and Litzenberger (1978), the RND of an asset price can be computed as the second partial derivative of the call option price C at current time t with respect to strike price K, when the strike price is evaluated at the price of the underlying asset at expiry S_T,

$$f_Q(S_T) \stackrel{\text{def}}{=} e^{r_f \tau} \left. \frac{\partial^2 C}{\partial K^2} \right|_{K=S_T}, \tag{1}$$

where $\tau = T - t$ is the time to maturity of the call option, and r_f is the risk-free interest rate.

This study employs the log-normal mixture method to estimate the RND. Following Ritchey (1990), Melick and Thomas (1997) and Liu et al. (2007), the RND $f_Q(x)$ associated with a given time to maturity τ can be modeled with a mixture of M log-normal distributions as

$$f_Q(x|\boldsymbol{\theta}) = \sum_{i=1}^{M} w_i f_{LN}(x|F_i, \sigma_i, \tau), \tag{2}$$

with

$$f_{LN}(x|F, \sigma, \tau) = \frac{1}{x\sigma\sqrt{2\pi\tau}} \left(-\frac{1}{2\sigma^2\tau} \left(\log(x) - [\log(F) - 0.5\sigma^2\tau] \right)^2 \right), \tag{3}$$

where $\boldsymbol{\theta} = (F_1, \cdots, F_M, \sigma_1, \cdots, \sigma_M, w_1, \cdots, w_M)'$ is the parameter vector of the forward prices of the underlying, the implied volatility of the underlying, and the weights of the M densities in the mixture. The density weights w_i are non-negative and sum up to unity. The density $f_Q(x|\boldsymbol{\theta})$ is risk-neutral only when the expectation of the expiration price of the underlying asset equals the current forward price of the asset F_t, i.e. when $\sum_{i=1}^{M} w_i F_i = F_t$. The RND $f_Q(x|\boldsymbol{\theta})$ in (2) has a flexible shape determined by $(3M - 2)$ free parameters, as opposed to only one free parameter in the case where $f_Q(x|\boldsymbol{\theta})$ is modeled with a single log-normal distribution as in the Black (1976) model. Meanwhile, the theoretical price of a call option with a strike price K under the risk-neutral density $f_Q(x|\boldsymbol{\theta})$ is simply the weighted average of the prices of call options with the same strike given by the Black (1976) option pricing formula. That is,

$$C(K|\boldsymbol{\theta}, r_f, \tau) = \sum_{i=1}^{M} w_i C_B(K, F_i, \sigma_i, r_f, \tau), \tag{4}$$

where $C_B(\cdot)$ denotes the Black (1976) option pricing formula for call options on forward contracts.

We estimate the parameter vector $\boldsymbol{\theta}$ that underlies the RND $f_Q(x|\boldsymbol{\theta})$ in (2) by minimizing the sum of squared deviations of the observed market prices for N call options from their corresponding theoretical prices for each tenor. Thus, the estimate of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$, is defined as,

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \left(C_{\text{market}}(K_i) - C(K_i|\boldsymbol{\theta})\right)^2, \tag{5}$$

where $C(\cdot|\boldsymbol{\theta})$ is the theoretical option price given in equation (4).

The minimization in (5) is a standard nonlinear least squares estimation problem. Under standard regularity conditions and correct model specifications, the nonlinear least squares estimate $\widehat{\boldsymbol{\theta}}$ is $\sqrt{N}$-consistent and follows an asymptotic normal distribution derived in Jennrich (1969) and Wu (1981).[29] Given the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$, we can easily derive the asymptotic distribution of the estimated RND $f_Q(x|\widehat{\boldsymbol{\theta}})$ in (2) using the Delta method.

5.2.2 Empirical results

We illustrate the estimation of the RND using intraday quote data on call options written on GOOG on January 28, 2015, when the Federal Reserve Board (Fed) released its first Federal Open Market Committee (FOMC) in 2015 at 2pm Eastern time (i.e. 1pm CT). We filter out potentially erroneous quote entries using the filtering rules (i.e. F1, F2, F3 and F5c) detailed in Section 4. Following standard practice in the literature, we use the midquotes of the best bid and ask prices of the options and the underlying stock price at any time instance as the observed market prices of the options and the underlying. Consistent with no arbitrage, we assume that the forward price F_t of the underlying asset associated with a particular option contract equals the dividend-adjusted future value of the underlying spot price S_t corresponding to the tenor of the option contract, i.e. $F_t = S_t \exp\left((r_f - \delta_t) \cdot \tau\right)$, where r_f is the risk-free rate, δ_t is the dividend yield of the underlying, and τ is the time to maturity of the option contract. As GOOG is a technology stock that does not pay dividends, we set $\delta_t = 0$. We obtain the annualized risk-free rates available on January 28, 2015 from the U.S. Treasury's website, and use cubic spline interpolations to compute the risk-free rate prevailing over the life of an option contract.

Table 9 reports summary statistics about the strikes and spreads for call options, of four different maturities of 9, 23, 37 and 79 days, written on GOOG on January 28, 2015 obtained from the last 15 minute (14:45-15:00) intraday OPRA data and the end-of-day OptionMetrics data. As expected, there are many more observations for the intraday OPRA data than for the daily OptionMetrics data. While both the intraday and end-of-day data possess the same range of strike prices, the former covers a wider range of spread measures. This suggests a richer information set provided by the intraday OPRA data relative to the daily OptionMetrics data.

[29] We also follow Newey and West (1994) procedure to obtain a heteroskedasticity and autocorrelation consistent estimator $\widehat{\boldsymbol{\theta}}$.

Time to maturity (days)		Last 15min OPRA			EOD OptionMetrics		
		Strike	**Spread**	**Rel. spread**	**Strike**	**Spread**	**Rel. spread**
9	Min	395.0	0.100	0.022	395.0	0.250	0.034
	Mean	511.3	1.598	0.179	502.8	1.847	0.201
	Max	600.0	7.700	1.333	600.0	4.200	0.909
	Nobs	116,916	116,916	116,916	58	58	58
23	Min	260.0	0.050	0.006	260.0	0.100	0.015
	Mean	501.9	1.065	0.064	454.3	2.089	0.075
	Max	650.0	11.100	1.333	650.0	4.000	1.333
	Nobs	293,852	293,852	293,852	94	94	94
37	Min	425.0	0.400	0.033	425.0	0.450	0.033
	Mean	512.1	1.616	0.149	507.9	1.717	0.156
	Max	610.0	5.050	1.294	610.0	3.300	1.000
	Nobs	90,859	90,859	90,859	47	47	47
79	Min	405.0	0.100	0.008	405.0	0.200	0.012
	Mean	527.8	0.959	0.080	527.5	1.157	0.112
	Max	650.0	7.600	0.868	650.0	3.200	0.667
	Nobs	158,459	158,459	158,459	50	50	50

Table 9: Summary statistics about the strikes ($), spreads ($) and relative spreads (Rel. spreads) for call options, of various times to maturity, written on GOOG on January 28, 2015 obtained from last 15 minute (14:45-15:00) intraday OPRA data and daily OptionMetrics data.

One big advantage of high-frequency data is that it offers many intraday cross-sections of option prices, rather than just one single end-of-day cross-section. This feature of OPRA data facilitates a study of the differences, due to, for example, news releases, of the estimated RND curves at different times during a trading day, which cannot be examined using end-of-day OptionMetrics data. As an illustration, Figure 10 represents the estimated RNDs from the intraday OPRA data on call options written on GOOG during two 15-minute trading intervals right before and after the release of the FOMC announcement at 2pm Eastern time (i.e. 1pm CT) on January 28, 2015, and for four different times to maturity (i.e. 9, 23, 37 and 79 days). Each plot contains the RND curve estimated from a mixture of $M = 2$ log-normal densities.[30] To ensure the risk-neutrality requirement that the risk-neutral expectation of the expiration price of the underlying asset is equal to "the current forward price" of the underlying in cases where the option data are not observed at the same time but during some time interval (e.g. 12:45-13:00 CT), we use the median of the forward prices over that time interval as "the current forward price" of the underlying, and the instance at which the corresponding forward price of the underlying is closest to the median forward price is assumed to be "the current time" of that interval.[31]

At 1pm CT on January 28, 2015, the Fed announced that it would keep the federal funds rate at the 0 to 1/4 percent target range and await further market indicators for the reassessement of its stance in the future. Not surprisingly, the Fed's decision to maintain the funds rate at the "exceptionally" low level had little effects on the RND curves for GOOG estimated from options with a short time to maturity (i.e.

[30] We obtain similar estimated RND curves, in terms of shape and scale, using a mixture of $M = 3$ log-normal densities.

[31] This assumption is not needed for end-of-day OptionMetrics data, since the prices of all options from OptionMetrics that are quoted on the same day are assumed to be observed contemporaneously.

9 days). However, it changed the RND curves estimated from longer-term options slightly in which the estimated RNDs during 15 minutes after the announcement became more negatively skewed than those during 15 minutes prior to the announcement, suggesting that market participants might anticipate a higher probability of more unfavorable events for the price of GOOG in the near or medium term, which might be triggered if the funds rate was to be raised. Clearly, this observation cannot be extracted from daily OptionMetrics data.

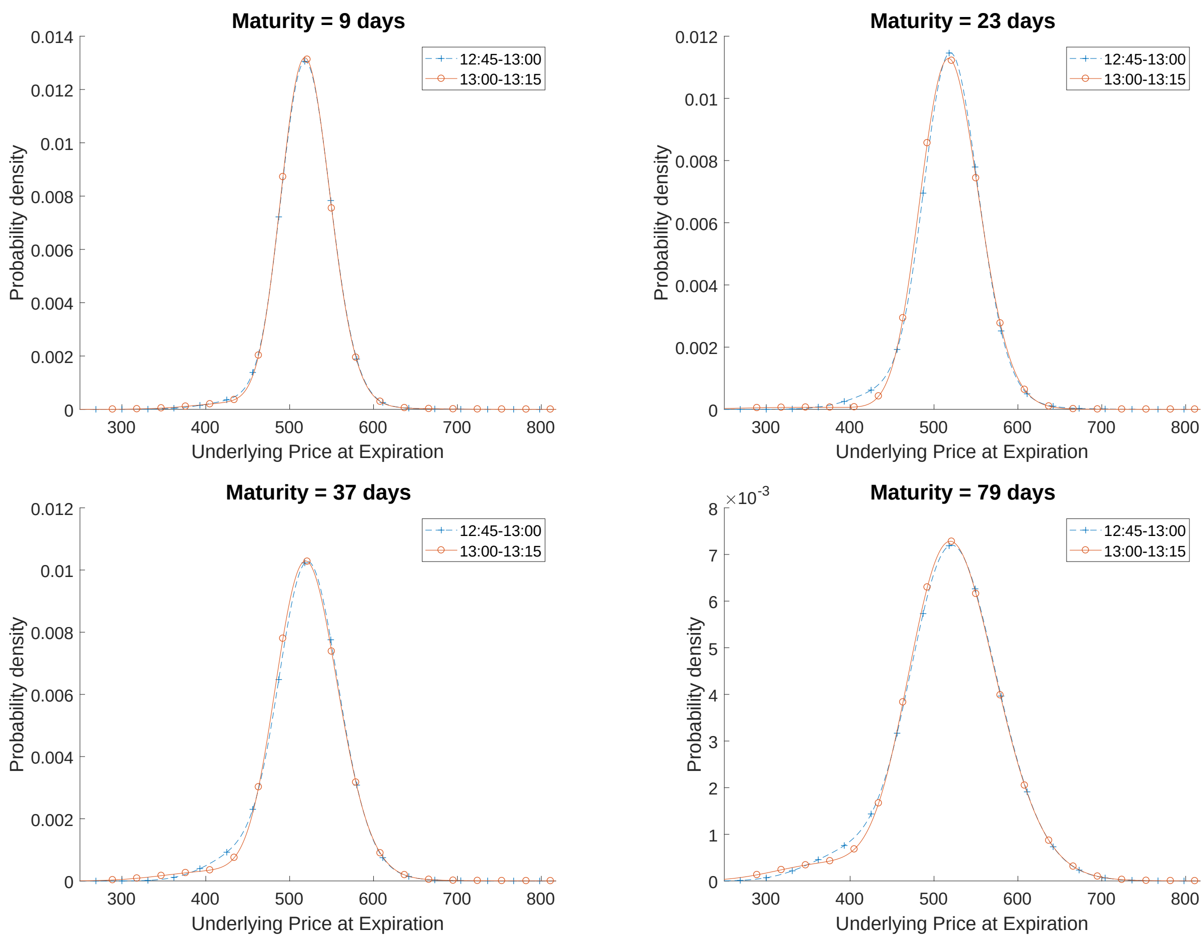

Figure 10: Estimated Risk-neutral densities (RNDs), as a function of the underlying price at expiry S_T, for various times to maturity obtained from intraday OPRA data for call options written on GOOG during two 15 minute intervals, before and after the FOMC announcement at 2pm Eastern time (i.e. 1pm CT) on January 28, 2015. Each RND curve is estimated from a mixture of $M = 2$ log-normal distributions. Time in each figure is CT.

Figure 11 depicts the RND curves estimated from intraday call options on GOOG during 15 minutes at the market open and close on January 28, 2015, with the estimated RND curve produced by the end-of-day OptionMetrics data acting as a benchmark. While the estimated RND during the last 15 minutes of the trading day generally coincides with the end-of-day benchmark, especially for options with 37 and 79 days to maturity, which confirms the reliablity of the high-frequency OPRA data, there are noticeable differences between the estimated RNDs at the market open and close in terms of both peak location and left tail properties, which are likely tied to the price dynamics of the underlying during the course of the trading day and which may be missed if one only considers the end-of-day option data.

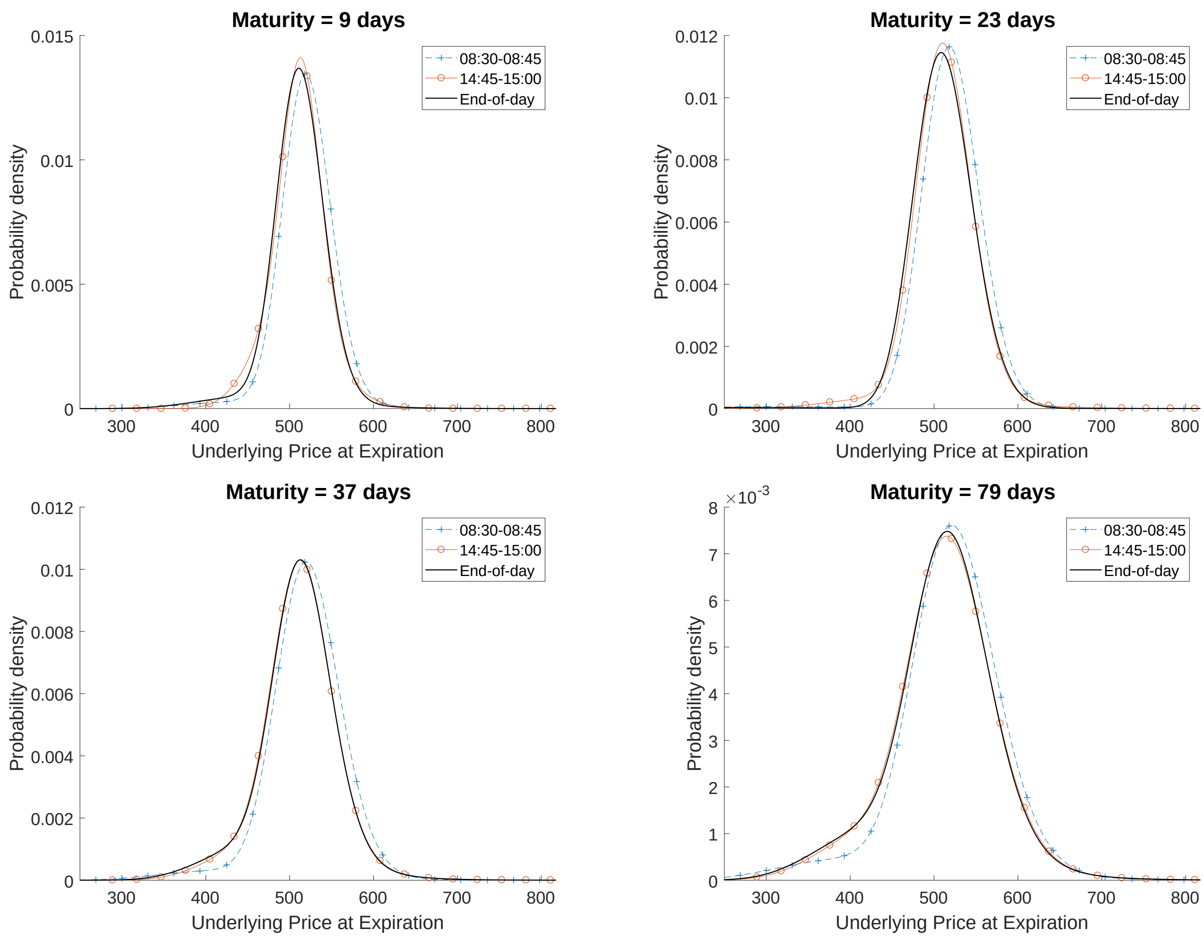

Figure 11: Estimated Risk-neutral densities (RNDs), as a function of the underlying price at expiry S_T, for various times to maturity obtained from intraday OPRA data for call options written on GOOG during two 15 minute intervals, at the market open and close on January 28, 2015. Each graph also plots the corresponding RND curve estimated from the End-of-day OptionMetrics data. Each RND curve is estimated from a mixture of $M = 2$ log-normal distributions.

Overall, the high-frequency OPRA data provides significantly richer information about option prices than does the end-of-day OptionMetrics data. In particular, it enables a study of how the estimated RND curves evolve during a trading day, which cannot be done using daily OptionMetrics data. Indeed, we find subtle changes in the estimated RND curves for GOOG following the FOMC announcement at 2pm Eastern time on January 28, 2015, as well as noticeable differences in the estimated RNDs at the market open and close on the same day. Thus, the ability to conduct a cross-time study with high-frequency OPRA data is beneficial to researchers who are interested in examining the effects of various types of news events on option prices and other option-based measures.

6 Conclusion

This study provides a detailed description of high frequency trade and quote data on options traded in the U.S. The study starts with a comprehensive review of the U.S. option market by outlining its structure and 15 constituent option exchanges, summarizing the market regulatory plans that govern option trading, and discussing specific market maker quoting obligations that are pertinent to the quote driven U.S. option market.

Our dataset is provided by OPRA in accordance with the "Plan for Reporting of Consolidated Options Last Sale Reports and Quotation Information", and it contains more than 150 million trade and 1.2 trillion quote records at a millisecond resolution for all option classes written on individual equities, stock indices and exchange traded products that are traded in the U.S. during the first eight months of 2015. Our dataset is much more comprehensive than other high frequency option datasets that have been employed in a very few studies in the literature which focus only on index options. We provide a detailed analysis to assess the quality and informativeness of our dataset, based on which a general filtering algorithm is suggested for data cleaning in a similar spirit to the Barndorff-Nielsen et al. (2009) algorithm for tick-by-tick data on equities.

Based on a representative sample of underlyings in January 2015, we find a very small fraction of erroneous and irregular records, which suggests that OPRA data is of high quality. An analysis of various liquidity measures of the selected sample confirms our expectation that options written on more liquid underlyings are generally more liquid and have tighter spreads. In addition, a cross-exchange investigation suggests that amongst 12 option exchanges that were available in 2015, CBOE, AMEX and ARCA were the most competitive exchanges that participated most frequently in the NBBO quoting pair.

We present two applications - the estimation of firstly the risk-neutral variance of an asset's returns and secondly the risk-neutral density (RND) - using intraday OPRA data. Our results show that the option-implied risk-neutral variance of returns on SPY (an exchange traded fund) and GOOG (Google stock) varies considerably during a trading day and is inversely related to the corresponding intraday underlying price dynamics. In addition, there are marked differences in the RND curves for GOOG estimated from high-frequency OPRA data before and after an FOMC announcement, as well as at the market open and close of a trading day. Such observations cannot be revealed by the typical end-of-day OptionMetrics data.

Our work serves as the first study that thoroughly investigates and describes the high frequency option trade and quote data provided by OPRA, and it highlights several advantages of the OPRA data over the end-of-day OptionMetrics data. As has already been demonstrated in the case of high frequency stock data, we expect that the availability of high frequency option data such as OPRA data will open an active area for future research, since it will help enrich our understanding of the dynamics of option prices and option-implied measures of the underlying at a finer resolution.

References

Abergel, F. and Zaatour, R. (2012). What drives option prices? *Journal of Trading*, 7(3):12–28.

Amaya, D., Bégin, J.-F., and Gauthier, G. (2018). Extracting latent states from high frequency option prices. Available at SSRN: https://ssrn.com/abstract=2975355.

Andersen, T. G. and Bondarenko, O. (2007). Construction and interpretation of model-free implied volatility. NBER Working Papers 13449, National Bureau of Economic Research, Inc.

Andersen, T. G., Bondarenko, O., and Gonzalez-Perez, M. T. (2015a). Exploring return dynamics via corridor implied volatility. *Review of Financial Studies*, 28(10):2902–2945.

Andersen, T. G., Bondarenko, O., Todorov, V., and Tauchen, G. (2015b). The fine structure of equity-index option dynamics. *Journal of Econometrics*, 187(2):532 – 546.

Audrino, F. and Fengler, M. R. (2015). Are classical option pricing models consistent with observed option second-order moments? evidence from high-frequency data. *Journal of Banking & Finance*, 61:46–63.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: trades and quotes. *Econometrics Journal*, 12(3):C1–C32.

Beber, A. and Brandt, M. W. (2006). The effect of macroeconomic news on beliefs and preferences: Evidence from the options market. *Journal of Monetary Economics*, 53(8):1997–2039.

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, 3(1-2):167–179.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651.

Brogaard, J., Hagströmer, B., Norden, L., and Riordan, R. (2015). Trading fast and slow: Colocation and liquidity. *Review of Financial Studies*, 28(12):3407–3443.

Brownlees, C. and Gallo, G. (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis*, 51(4):2232–2245.

Carr, P. and Madan, D. B. (1999). Option valuation using the fast fourier transform. *Journal of Computational Finance*, 2:61–73.

Chang, B.-Y., Christoffersen, P., Jacobs, K., and Vainberg, G. (2012). Option-implied measures of equity risk. *Review of Finance*, 16(2):385–428.

Christoffersen, P., Heston, S., and Jacobs, K. (2009). The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science*, 55(12):1914–1932.

Christoffersen, P., Jacobs, K., and Mimouni, K. (2010). Volatility dynamics for the S&P500: Evidence from realized volatility, daily returns, and option prices. *Review of Financial Studies*, 23(8):3141–3189.

Christoffersen, P., Jacobs, K., Ornthanalai, C., and Wang, Y. (2008). Option valuation with long-run and short-run volatility components. *Journal of Financial Economics*, 90(3):272–297.

Chung, K. H. and Chuwonganant, C. (2012). Regulation NMS and market quality. *Financial Management*, 41(2):285–317.

Gencay, R., Dacorogna, M., Muller, U. A., Pictet, O., and Olsen, R. (2001). *An Introduction to High-Frequency Finance*. Elsevier.

Hagströmer, B. and Nordén, L. (2013). The diversity of high-frequency traders. *Journal of Financial Markets*, 16(4):741–770.

Hendershott, T. and Jones, C. M. (2005). Trade-through prohibitions and market quality. *Journal of Financial Markets*, 8(1):1–23.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics*, 40(2):633–643.

Kapetanios, G., Neumann, M., and Skiadopoulos, G. (2014). Jumps in option prices and their determinants: Real-time evidence from the E-mini S&P 500 option market. Available at SSRN: https://ssrn.com/abstract=2368385.

Kokoszczynski, R., Sakowski, P., and Slepaczuk, R. (2010a). Midquotes or transactional data? the comparison of Black model on HF data. University of Warsaw, Faculty of Economic Sciences, Working Papers 15/2010 (38).

Kokoszczynski, R., Sakowski, P., and Slepaczuk, R. (2010b). Which option pricing model is the best? high frequency data for Nikkei225 index options. University of Warsaw, Faculty of Economic Sciences, Working Papers 16/2010 (39).

Liu, X., Shackleton, M. B., Taylor, S. J., and Xu, X. (2007). Closed-form transformations from risk-neutral to real-world distributions. *Journal of Banking & Finance*, 31(5):1501–1520.

Melick, W. R. and Thomas, C. P. (1997). Recovering an asset's implied PDF from option prices: An application to crude oil during the Gulf crisis. *Journal of Financial and Quantitative Analysis*, 32(1):91–115.

Nagel, S. (2012). Evaporating liquidity. *Review of Financial Studies*, 25(7):2005–2039.

Newey, W. K. and West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61(4):631–653.

Ritchey, R. J. (1990). Call option valuation for discrete normal mixtures. *Journal of Financial Research*, 13(4):285–296.

Stoll, H. R. (2006). Electronic trading in stock markets. *The Journal of Economic Perspectives*, 20(1):153–174.

Taylor, S. J., Tzeng, C.-F., and Widdicks, M. (2018). Information about price and volatility jumps inferred from options prices. *Journal of Futures Markets*, 38(10):1206–1226.

Wu, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics*, 9(3):501–513.

Appendices

A Additional Tables

Group	Option exchange operator	Headquarter	Since	OPRA symbol	Type	Overview (pricing model, order allocation algorithm, etc.)
Nasdaq	OMX PHLX	Philadelphia, PA	2008	X	FE	Based on a customer priority pricing model (a traditional model where customer accounts - e.g., brokerage firms - receive marketing fees/rebates and a priority to fill orders, while non-customers are charged a fee) and a pro-rata allocation algorithm (assigns the fills across orders based on percentages of the total quantity).
	OMX Options Market (NOM)	New York, NY	2008	Q	E	Based on a maker-taker pricing model (exchange pays a rebate to suppliers of liquidity - market makers - and charges liquidity takers - traders) and a price/time priority allocation algorithm, or FIFO (fills orders at the same price level according to time priority).
	OMX BX Options (NOBO)	Boston, MA	2012	T	E(?)	Based on a maker-taker pricing model and either a price/time priority or a pro-rata allocation algorithm. Focused on retail order flow.
	International Securities Exchange (ISE)	New York, NY	2000	I	E	Based on a modified maker-taker pricing model (offers rebates and fees based on a quote performance, product, client designation and order category) and a pro-rata allocation algorithm. Founded in 2000, it became the first fully electronic US options exchange.
	GEMX (ISE Gemini)	New York, NY	2013	H	E	Based on a maker-taker pricing model and a pro-rata allocation model with a priority to preferenced orders. Offers a flat pricing structure for price improvement mechanisms.
	MRX (ISE Mercury Exchange)	New York, NY	2016	J	E	Based on a customer priority pricing model and a pro-rata allocation model with a priority to preferenced orders.
CBOE	Chicago Board Options Exchange	Chicago, IL	1973	C	FE	Based on a customer priority pricing model and a price/time priority allocation algorithm. CBOE is the oldest U.S. options exchange offering standardized option contracts. Trading is executed via CBOE Hybrid Trading System, which enables customers to submit orders either to the face-to-face open outcry or to the electronic environment. CBOE disseminates the best bid and offer from all market participants.
	C2 Options Exchange	Chicago, IL	2010	W	E	Based on a maker-taker pricing model and a pure pro-rata allocation algorithm for multiply-listed classes.
	BATS BZX Options	Kansas, MO	2010	Z	E	Based on a maker-taker pricing model (with a flat pricing structure) and a pure price/time allocation algorithm. Offers penny executions in all options regardless of premium or class.
	BATS EDGX Options	Kansas, MO	2015	E	E	Based on a customer priority pricing model and a pro-rata allocation algorithm.
Intercontinental Exchange (ICE)	NYSE AMEX Options	New York, NY	2008	A	FE	Based on a customer priority pricing model and a pro-rata allocation algorithm. Blends it with the benefit of an open outcry trading floor in New York to offer traders deep liquidity across listed option contracts.
	NYSE Arca Options	New York, NY	2006	N	FE	Based on a customer priority pricing model and price-time priority allocation algorithm. Employs unique Lead Market Maker participation (LMM) model (where LMMs have a special fee schedule and priorities in order allocation). Market participants can trade electronically through the all-electronic trading platform, or by open outcry at the NYSE Arca Options floor in San Francisco.
Miami International Holdings (MIH)	MIAX Options Exchange	Princeton, NJ	2012	M	E	Based on a maker-taker pricing model and a pro-rata allocation algorithm.
	MIAX Pearl	Princeton, NJ	2017	P	E	Based on a maker-taker pricing model and a price-time allocation algorithm.
TMX Group	BOX Options Exchange	Chicago, IL; Boston, MA	2004	B	E	Based on a maker-taker model for most penny classes and on a price/time priority allocation algorithm except during the Trade-Through Filter Exposure Period and at the end of the price improvement period (PIP) auction process. PIP auction is a patented automated trading mechanism which improves executable client orders.

Table A.1: Overview of the OPRA participant exchanges. Column 'Type' refers to the trading environment on the exchange where 'E' stands for all-electronic markets and 'FE' for the markets with a mixed (floor and electronic) trading environment. ISE, ISE Gemini and ISE Mercury were purchased by Nasdaq from Deutsche Boerse in June, 2016. BATS BZX and BATS EDGX were acquired by CBOE Group in February, 2017. Regular trading hours on most of the exchanges are between 8:30 and 15:00 CT (except for certain exchange traded products which can be traded until 15:15 CT). BATS exchanges begin order acceptance at 6:30 and run early and pre-market trading sessions before the start of regular trading day at 8:30 CT.

Exchange	Market Makers	Continuous Quoting Obligations
Nasdaq PHLX	Market Makers (MM), or Registered Option Traders (these include Streaming Quote Traders (SQT)(*), Remote SQT and non-SQT(*)), Specialists (S)(*) (including Remote Specialists), Directed Market Makers (DMM) (these include S, SQT and RSQT that receives a Directed Order)	non-SQT - no continuous quoting obligations; SQT, RSQT - 90% (60% of its assigned series); S - 90% (99% of its assigned series); DMM - 90% (99% of the series listed on the Exchange, in each case in at least 60% of the options in which such DMM is assigned)
Nasdaq NOM	Market Makers (MM)	MM - 60% (100% of a MM's registered options collectively to all appointed issues, rather than on an option-by-option basis)
Nasdaq BX	Market Makers (MM)	MM - 60% (100% of a MM's registered options collectively to all appointed issues, rather than on an option-by-option basis)
Nasdaq ISE	Competitive Market Makers (CMM), Primary Market Makers (PMM)	CMM - 60% for option classes to which it is appointed and 90% for option classes in which it receives preferenced orders (CMM are not required to enter quotations in the options classes to which it is appointed, but whenever a CMM enters a quote in some options class to which it is appointed, it must maintain continuous quotations); PMM - 90% (100% of its assigned series)
Nasdaq GEMX	—	—
Nasdaq MRX	—	—
CBOE	Market Makers (MM)(*), Designated Primary Market Makers (DPM)(*) (including Off-Floor DPM), Lead Market Makers (LMM)(*) (there is only one LLM or DPM in a given option class), Preferred Market Makers (PMM) (MM having preferences in some option classes)	MM - 90% (60% of the assigned non-adjusted option series that have a time to expiration of less than nine months); PMM, LMM, DPM (there is only one DPM in a given option class) - 90% (99%)
C2	Market Makers (MM), Designated Primary Market Makers (DPM) (there can be only one DPM in a given option class), Preferred Market Makers (PMM) (MM having preferences in some option classes)	MM - 90% (60% of the non-adjusted option series of each registered class that have a time to expiration of less than nine months); DPM - 90% (99%); PMM - 99% (90%)
BATS BZX	Market Makers (MM), Lead Market Makers (LMM) (with respect to one or more securities listed on the Exchange), Competitive Liquidity Providers (CLP) (a MM can be registered as CLP if meets certain requirements)	MM, LMM - 90% (75% of the options series in which a MM is registered); CLP - must have Winning Bid (Offer) SETs equal to at least 10% of the total Bid (Offer) SETs
BATS EDGX	Market Makers (MM)	MM - 90% (75% of the options series in which a MM is registered)
NYSE AMEX	Market Makers (MM) (these include Remote Market Makers and Floor Market Makers(*)), Specialists (S) (including e-Specialists, only one specialist can be appointed per an option class)	MM - 60% (100% of its appointed issues); S - 90% (100% of its appointed issues)
NYSE ARCA	Market Makers (MM)(*), Lead Market Makers (LMM)	MM - 60% (100% of its appointed issues); LMM - 90% (100% of its appointed issues)
MIAX	Registered Market Makers (RMM), Lead Market Makers (LMM), Primary Lead Market Makers (PLMM)	RMM - 90% (60% of the non-adjusted series that have a time to expiration of less than nine months in each of its appointed classes); LMM - 90% (90% of the non-adjusted option series in each of its appointed classes); PLMM (there is only one PLMM in a given option class) - 90% (99% of the non-adjusted option series in which a PLMM is assigned)
MIAX Pearl	Market Makers (MM)	MM - 90% (75% of the options series in which a MM is registered)
BOX	Market Makers (MM), Preferred Market Makers (PMM) (one PMM is designated for each Preferenced Order)	MM - 60% (99% of the appointed classes collectively, rather than on a class-by-class basis); PMM - 90% (99% of the non-adjusted option series of each class for which it accepts Preferenced Orders)

Table A.2: Electronic quotation requirements for market makers. Column 'Market Makers' contains distinct types of market makers defined on a given exchange which differ in quoting obligations (star indicates that a market maker also enrolled for quoting and making transactions as a dealer-specialist on the trading floor). Column 'Continuous Quoting Obligations' contains a minimum fraction of time required for quoting (as a percentage of the total number of minutes in a trading day). In parentheses there are given a minimum set of option series to which the continuous quoting requirement is applied. Information in the table reflects market maker obligations specified in the official exhange market rules (available on July 25, 2017).

	Trade records						Quote records										
Ticker	Total	F1	F2	F3	F5	F6	Total	F1	F2(a)	F2(b)	F3	F4	F5(a)	F5(b)	F5(c)	F6	#
PG	**414.85**	0.10	0.00	0.10	0.00	49.90	**3,039,685.95**	193.70	135.50	4.35	709.15	142,246.40	250.75	5,630.10	22,154.70	183,264.80	54.8
	100.00	*0.02*	*0.00*	*0.02*	*0.00*	*12.03*	***100.00***	*0.01*	*0.00*	*0.00*	*0.02*	*4.68*	*0.01*	*0.19*	*0.73*	*6.03*	
WMT	**456.45**	0.05	0.00	0.10	0.15	72.60	**3,372,531.85**	237.55	126.00	6.80	824.80	116,502.05	174.85	6,338.65	8,941.30	212,378.35	52.8
	100.00	*0.01*	*0.00*	*0.02*	*0.03*	*15.91*	***100.00***	*0.01*	*0.00*	*0.00*	*0.02*	*3.45*	*0.01*	*0.19*	*0.27*	*6.30*	
K	**46.65**	0.00	0.00	0.00	0.00	7.35	**163,854.40**	73.35	84.60	0.10	55.55	6,089.10	2.25	502.95	4,763.80	8,701.20	24.5
	100.00	*0.00*	*0.00*	*0.00*	*0.00*	*15.76*	***100.00***	*0.04*	*0.05*	*0.00*	*0.03*	*3.72*	*0.00*	*0.31*	*2.91*	*5.31*	
XLP	**148.05**	0.00	0.00	0.00	0.00	27.85	**1,131,852.35**	488.95	444.90	2.35	786.30	35,837.00	208.15	3,516.75	37,491.00	62,726.45	77.3
	100.00	*0.00*	*0.00*	*0.00*	*0.00*	*18.81*	***100.00***	*0.04*	*0.04*	*0.00*	*0.07*	*3.17*	*0.02*	*0.31*	*3.31*	*5.54*	
XOM	**764.45**	0.05	0.00	0.10	0.05	93.60	**4,885,233.55**	276.45	162.05	9.15	650.30	224,220.20	204.65	8,860.00	24,849.50	296,794.80	67.0
	100.00	*0.01*	*0.00*	*0.01*	*0.01*	*12.24*	***100.00***	*0.01*	*0.00*	*0.00*	*0.01*	*4.59*	*0.00*	*0.18*	*0.51*	*6.08*	
CVX	**650.90**	0.00	0.00	0.05	0.75	74.75	**3,875,495.75**	359.65	250.55	10.30	590.75	188,189.40	244.30	6,946.70	49,477.90	262,532.65	72.1
	100.00	*0.00*	*0.00*	*0.01*	*0.12*	*11.48*	***100.00***	*0.01*	*0.01*	*0.00*	*0.02*	*4.86*	*0.01*	*0.18*	*1.28*	*6.77*	
MPC	**87.25**	0.00	0.00	0.05	0.00	4.50	**484,733.30**	148.90	222.95	2.20	195.55	22,978.50	0.00	1,137.85	11,903.85	18,092.30	53.8
	100.00	*0.00*	*0.00*	*0.06*	*0.00*	*5.16*	***100.00***	*0.03*	*0.05*	*0.00*	*0.04*	*4.74*	*0.00*	*0.23*	*2.46*	*3.73*	
XLE	**833.40**	0.00	0.00	0.10	0.05	80.00	**6,888,951.20**	1,019.15	674.95	3.00	1,981.30	182,999.60	137.25	12,824.90	49,338.40	412,240.55	113.8
	100.00	*0.00*	*0.00*	*0.01*	*0.01*	*9.60*	***100.00***	*0.01*	*0.01*	*0.00*	*0.03*	*2.66*	*0.00*	*0.19*	*0.72*	*5.98*	
JPM	**946.75**	0.00	0.00	0.15	0.85	144.70	**7,008,967.80**	299.45	166.55	10.00	784.90	239,941.55	744.90	12,592.90	47,183.75	559,035.75	67.5
	100.00	*0.00*	*0.00*	*0.02*	*0.09*	*15.28*	***100.00***	*0.00*	*0.00*	*0.00*	*0.01*	*3.42*	*0.01*	*0.18*	*0.67*	*7.98*	
BAC	**1,683.95**	0.00	0.00	0.15	0.20	390.85	**4,594,295.80**	283.50	134.90	1.05	551.45	69,515.95	273.35	10,760.75	243,706.75	374,137.20	59.3
	100.00	*0.00*	*0.00*	*0.01*	*0.01*	*23.21*	***100.00***	*0.01*	*0.00*	*0.00*	*0.01*	*1.51*	*0.01*	*0.23*	*5.30*	*8.14*	
BLK	**45.42**	0.00	0.00	0.00	0.37	3.37	**553,401.11**	128.21	464.26	28.68	236.84	77,307.32	0.00	1,000.00	16,061.05	65,144.47	50.0
	100.00	*0.00*	*0.00*	*0.00*	*0.81*	*7.42*	***100.00***	*0.02*	*0.08*	*0.01*	*0.04*	*13.97*	*0.00*	*0.18*	*2.90*	*11.77*	
XLF	**454.35**	0.00	0.00	0.00	0.25	107.05	**1,919,046.85**	943.65	353.00	37.70	1,232.20	41,246.90	428.15	4,979.65	137,567.95	120,626.00	63.3
	100.00	*0.00*	*0.00*	*0.00*	*0.06*	*23.56*	***100.00***	*0.05*	*0.02*	*0.00*	*0.06*	*2.15*	*0.02*	*0.26*	*7.17*	*6.29*	
AAPL	**12,042.65**	0.20	0.00	0.85	24.90	1,856.60	**26,532,461.60**	3,546.45	507.30	180.15	4,854.75	962,079.85	3,007.75	71,053.10	102,615.15	1,959,568.10	77.6
	100.00	*0.00*	*0.00*	*0.01*	*0.21*	*15.42*	***100.00***	*0.01*	*0.00*	*0.00*	*0.02*	*3.63*	*0.01*	*0.27*	*0.39*	*7.39*	
GOOG	**1,158.85**	0.00	0.00	0.15	0.75	87.75	**9,783,760.85**	6,311.45	2,006.75	355.50	8,049.60	1,392,289.60	131.10	16,574.25	449,652.80	624,808.85	220.9
	100.00	*0.00*	*0.00*	*0.01*	*0.06*	*7.57*	***100.00***	*0.06*	*0.02*	*0.00*	*0.08*	*14.23*	*0.00*	*0.17*	*4.60*	*6.39*	
EA	**136.60**	0.20	0.00	0.25	0.05	16.60	**2,852,405.70**	541.05	348.05	2.20	1,185.90	139,412.20	177.15	4,916.60	18,790.50	165,937.85	64.0
	100.00	*0.15*	*0.00*	*0.18*	*0.04*	*12.15*	***100.00***	*0.02*	*0.01*	*0.00*	*0.04*	*4.89*	*0.01*	*0.17*	*0.66*	*5.82*	
XLK	**123.15**	0.00	0.00	0.00	0.00	26.85	**1,068,806.10**	502.00	333.60	0.15	690.40	30,587.05	227.60	2,577.15	48,884.05	54,415.70	65.3
	100.00	*0.00*	*0.00*	*0.00*	*0.00*	*21.80*	***100.00***	*0.05*	*0.03*	*0.00*	*0.06*	*2.86*	*0.02*	*0.24*	*4.57*	*5.09*	
SPY	**16,465.60**	2.10	0.00	3.05	31.25	1,946.55	**66,756,914.05**	17,713.50	3,560.60	959.25	19,182.05	2,008,607.85	61,442.70	244,963.95	1,843,949.20	4,226,239.30	281.7
	100.00	*0.01*	*0.00*	*0.02*	*0.19*	*11.82*	***100.00***	*0.03*	*0.01*	*0.00*	*0.03*	*3.01*	*0.09*	*0.37*	*2.76*	*6.33*	

Table A.3: Aggregate statistics for potentially irrelevant observations in January 2015. All contracts expiring on February 20, 2015 are considered. The table contains average daily numbers and percentage fractions of special option records for each selected underlying. Records for group F5 are filtered based on millisecond stamps. Column "#" provides the average daily number of available option contract specifications.

Ticker	Trade records						Quote records									
	Total	F1	F2	F3	F5	F6	Total	F1	F2(a)	F2(b)	F3	F4	F5(a)	F5(b)	F5(c)	F6
NYSE Amex	**2,011.70**	0.40	0.00	2.25	1.75	611.45	**17,793,772.65**	161.95	584.65	8.25	0.00	707,862.90	545.10	53,230.85	416,470.30	49,876.90
	100.00	*0.02*	*0.00*	*0.11*	*0.09*	*30.39*	***100.00***	*0.00*	*0.00*	*0.00*	*0.00*	*3.98*	*0.00*	*0.30*	*2.34*	*0.28*
BOX	**2,955.30**	0.00	0.00	0.05	0.05	1,051.05	**3,889,041.85**	35.15	246.95	18.15	0.00	627,994.25	2,605.45	4,441.40	102,727.45	112,489.40
	100.00	*0.00*	*0.00*	*0.00*	*0.00*	*35.56*	***100.00***	*0.00*	*0.01*	*0.00*	*0.00*	*16.15*	*0.07*	*0.11*	*2.64*	*2.89*
CBOE	**8,154.50**	0.65	0.00	0.75	2.15	328.15	**20,192,318.80**	436.50	46.25	2.55	0.00	619,486.75	1,110.35	46,914.70	287,497.05	779,180.45
	100.00	*0.01*	*0.00*	*0.01*	*0.03*	*4.02*	***100.00***	*0.00*	*0.00*	*0.00*	*0.00*	*3.07*	*0.01*	*0.23*	*1.42*	*3.86*
ISE Gemini	**899.75**	0.00	0.00	0.00	0.05	38.95	**10,940,923.95**	1,752.55	1,581.00	1.55	0.00	164,216.75	8,269.05	40,931.05	295,915.10	276,891.90
	100.00	*0.00*	*0.00*	*0.00*	*0.01*	*4.33*	***100.00***	*0.02*	*0.01*	*0.00*	*0.00*	*1.50*	*0.08*	*0.37*	*2.70*	*2.53*
ISE	**2,848.40**	0.00	0.00	0.10	0.75	146.75	**19,256,193.35**	1,852.65	1,547.55	2.25	0.00	393,963.95	1,472.85	49,745.85	404,272.30	740,140.90
	100.00	*0.00*	*0.00*	*0.00*	*0.03*	*5.15*	***100.00***	*0.01*	*0.01*	*0.00*	*0.00*	*2.05*	*0.01*	*0.26*	*2.10*	*3.84*
MIAX	**3,093.10**	0.00	0.00	0.15	50.75	1,081.15	**17,436,973.85**	23.95	152.05	215.65	0.00	377,507.30	9,628.50	62,228.45	315,595.70	3,757,241.20
	100.00	*0.00*	*0.00*	*0.00*	*1.64*	*34.95*	***100.00***	*0.00*	*0.00*	*0.00*	*0.00*	*2.16*	*0.06*	*0.36*	*1.81*	*21.55*
NYSE Arca	**3,198.70**	0.25	0.00	0.50	1.70	556.25	**8,101,861.50**	92.80	638.10	12.05	0.00	408,495.20	594.20	14,222.25	146,943.75	24,140.20
	100.00	*0.01*	*0.00*	*0.02*	*0.05*	*17.39*	***100.00***	*0.00*	*0.01*	*0.00*	*0.00*	*5.04*	*0.01*	*0.18*	*1.81*	*0.30*
Nasdaq OM	**3,881.20**	0.00	0.00	0.00	0.25	266.35	**8,096,120.35**	51.70	80.20	7.25	0.00	516,918.70	5,052.95	23,440.00	170,556.90	1,009,068.25
	100.00	*0.00*	*0.00*	*0.00*	*0.01*	*6.86*	***100.00***	*0.00*	*0.00*	*0.00*	*0.00*	*6.38*	*0.06*	*0.29*	*2.11*	*12.46*
BX Options	**286.35**	0.00	0.00	0.00	0.00	3.80	**7,034,444.70**	21.95	320.10	0.25	0.00	955,360.75	16,537.20	27,657.00	289,721.30	361,821.10
	100.00	*0.00*	*0.00*	*0.00*	*0.00*	*1.33*	***100.00***	*0.00*	*0.00*	*0.00*	*0.00*	*13.58*	*0.24*	*0.39*	*4.12*	*5.14*
C2	**771.85**	0.00	0.00	0.00	0.05	9.25	**8,258,266.50**	765.60	81.55	0.55	0.00	224,218.65	4,811.50	28,504.10	212,728.55	128,415.55
	100.00	*0.00*	*0.00*	*0.00*	*0.01*	*1.20*	***100.00***	*0.01*	*0.00*	*0.00*	*0.00*	*2.72*	*0.06*	*0.35*	*2.58*	*1.55*
Nasdaq PHLX	**3,507.25**	1.40	0.00	1.30	1.65	337.30	**17,663,467.80**	4.85	53.45	1.85	0.00	90,336.35	721.10	43,284.20	343,144.85	1,693,604.55
	100.00	*0.04*	*0.00*	*0.04*	*0.05*	*9.62*	***100.00***	*0.00*	*0.00*	*0.00*	*0.00*	*0.51*	*0.00*	*0.25*	*1.94*	*9.59*
BATS BZX	**4,848.95**	0.00	0.00	0.00	0.45	560.25	**6,221,342.85**	27,860.90	4,621.45	1,341.15	42,549.95	789,823.60	16,306.65	20,526.40	130,955.35	670,516.70
	100.00	*0.00*	*0.00*	*0.00*	*0.01*	*11.55*	***100.00***	*0.45*	*0.07*	*0.02*	*0.68*	*12.70*	*0.26*	*0.33*	*2.10*	*10.78*

Table A.4: Aggregate statistics for potentially irrelevant observations in January 2015. All contracts expiring on February 20, 2015 are considered. The table contains average daily numbers and percentage fractions of special option records for each exchange market. Records for group F5 are filtered based on millisecond stamps.

Market	PG	WMT	K	XLP	XOM	CVX	MPC	XLE	JPM	BAC	BLK	XLF	AAPL	GOOG	EA	XLK	SPY
	Out-of-the-money options																
Nasdaq PHLX	0.594	0.415	0.578	0.531	0.258	0.251	0.501	0.245	0.275	0.316	0.636	0.525	0.098	0.184	0.155	0.849	0.322
Nasdaq OM	0.499	0.346	0.513	0.558	0.233	0.205	0.479	0.241	0.213	0.268	0.620	0.442	0.051	0.140	0.145	0.827	0.396
BX Options	0.619	0.471	0.619	0.557	0.241	0.337	0.557	0.287	0.284	0.323	0.667	0.559	0.155	0.195	0.169	0.893	0.398
ISE	0.647	0.457	0.568	0.601	0.269	0.241	0.534	0.282	0.254	0.251	0.615	0.487	0.076	0.161	0.163	0.902	0.333
ISE Gemini	0.570	0.539	0.613	0.517	0.316	0.367	0.536	0.285	0.291	0.275	0.644	0.430	0.070	0.161	0.191	0.902	0.304
CBOE	0.589	0.427	0.606	0.591	0.274	0.271	0.505	0.281	0.274	0.280	0.655	0.545	0.076	0.184	0.157	0.949	0.378
C2	0.691	0.575	N/A	0.495	0.347	0.337	N/A	0.320	0.323	0.447	N/A	0.487	0.110	0.374	0.182	0.898	0.294
BATS BZX	0.479	0.321	0.529	0.611	0.231	0.213	0.533	0.308	0.186	0.188	0.564	0.367	0.063	0.160	0.165	0.876	0.379
NYSE Amex	0.632	0.426	0.558	0.530	0.281	0.280	0.514	0.279	0.316	0.326	0.664	0.578	0.098	0.236	0.167	0.913	0.469
NYSE Arca	0.598	0.379	0.502	0.537	0.292	0.228	0.512	0.266	0.240	0.239	0.605	0.440	0.052	0.158	0.159	0.899	0.409
MIAX	0.534	0.392	0.593	0.521	0.246	0.243	0.488	0.259	0.235	0.306	0.640	0.490	0.099	0.172	0.176	0.857	0.391
BOX	0.500	0.308	0.500	0.489	0.259	0.210	0.480	0.236	0.244	0.278	0.565	0.560	0.082	0.152	0.155	0.791	0.518
All markets	0.593	0.424	0.574	0.557	0.267	0.260	0.510	0.274	0.267	0.290	0.647	0.505	0.086	0.190	0.164	0.899	0.366
	At-the-money options																
Nasdaq PHLX	0.075	0.054	0.162	0.102	0.070	0.063	0.244	0.072	0.042	0.050	0.173	0.078	0.023	0.033	0.057	0.133	0.020
Nasdaq OM	0.067	0.051	0.186	0.083	0.071	0.069	0.253	0.068	0.027	0.030	0.175	0.052	0.016	0.032	0.055	0.114	0.016
BX Options	0.082	0.061	0.199	0.110	0.070	0.097	0.272	0.074	0.048	0.049	0.184	0.074	0.034	0.036	0.073	0.136	0.075
ISE	0.075	0.055	0.168	0.099	0.072	0.064	0.260	0.070	0.037	0.039	0.184	0.064	0.019	0.041	0.057	0.136	0.018
ISE Gemini	0.072	0.062	0.176	0.108	0.083	0.088	0.253	0.070	0.050	0.034	0.210	0.055	0.020	0.042	0.075	0.139	0.039
CBOE	0.080	0.054	0.167	0.103	0.075	0.061	0.257	0.069	0.040	0.052	0.180	0.073	0.020	0.036	0.055	0.142	0.021
C2	0.097	0.066	N/A	0.127	0.098	0.080	N/A	0.075	0.057	0.088	N/A	0.062	0.022	0.083	0.068	0.146	0.027
BATS BZX	0.077	0.053	0.187	0.094	0.071	0.067	0.259	0.083	0.027	0.029	0.193	0.050	0.016	0.033	0.058	0.132	0.024
NYSE Amex	0.085	0.056	0.166	0.107	0.077	0.063	0.262	0.069	0.040	0.055	0.181	0.076	0.020	0.036	0.059	0.148	0.024
NYSE Arca	0.078	0.053	0.165	0.096	0.077	0.062	0.249	0.068	0.029	0.032	0.177	0.058	0.016	0.031	0.052	0.132	0.017
MIAX	0.077	0.055	0.164	0.100	0.074	0.064	0.245	0.070	0.038	0.055	0.177	0.071	0.021	0.036	0.067	0.130	0.025
BOX	0.080	0.056	0.169	0.092	0.076	0.068	0.258	0.071	0.040	0.057	0.184	0.079	0.029	0.043	0.059	0.122	0.031
All markets	0.079	0.056	0.169	0.103	0.075	0.066	0.255	0.071	0.039	0.046	0.181	0.067	0.020	0.037	0.060	0.137	0.025
	In-the-money options																
Nasdaq PHLX	0.060	0.074	0.098	0.099	0.105	0.078	0.204	0.074	0.050	0.014	0.066	0.025	0.011	0.033	0.072	0.089	0.015
Nasdaq OM	0.073	0.089	0.140	0.111	0.126	0.092	0.184	0.077	0.064	0.010	0.065	0.030	0.019	0.035	0.084	0.092	0.058
BX Options	0.068	0.083	0.145	0.129	0.114	0.119	0.186	0.072	0.053	0.013	0.064	0.037	0.015	0.033	0.078	0.109	0.050
ISE	0.061	0.079	0.102	0.100	0.100	0.078	0.187	0.075	0.058	0.013	0.069	0.024	0.012	0.034	0.079	0.090	0.016
ISE Gemini	0.064	0.084	0.104	0.114	0.108	0.092	0.189	0.074	0.061	0.012	0.071	0.027	0.012	0.034	0.084	0.091	0.033
CBOE	0.060	0.075	0.105	0.114	0.107	0.071	0.209	0.075	0.049	0.014	0.072	0.026	0.011	0.039	0.072	0.091	0.015
C2	0.100	0.082	N/A	0.273	0.151	0.098	N/A	0.081	0.046	0.025	N/A	0.032	0.015	0.051	0.083	0.108	0.018
BATS BZX	0.070	0.079	0.113	0.103	0.104	0.078	0.216	0.086	0.059	0.010	0.072	0.028	0.014	0.036	0.085	0.090	0.063
NYSE Amex	0.058	0.073	0.098	0.109	0.104	0.075	0.190	0.073	0.050	0.014	0.074	0.025	0.011	0.036	0.073	0.090	0.018
NYSE Arca	0.060	0.076	0.100	0.108	0.108	0.069	0.184	0.070	0.048	0.010	0.061	0.025	0.012	0.037	0.080	0.088	0.021
MIAX	0.070	0.077	0.101	0.101	0.107	0.111	0.180	0.074	0.054	0.015	0.070	0.027	0.011	0.037	0.078	0.088	0.021
BOX	0.061	0.077	0.111	0.117	0.102	0.079	0.215	0.069	0.048	0.016	0.066	0.027	0.014	0.037	0.075	0.090	0.054
All markets	0.064	0.077	0.106	0.108	0.107	0.086	0.193	0.074	0.053	0.013	0.069	0.026	0.012	0.036	0.077	0.091	0.023

Table A.5: Average relative spread measures computed from the OPRA quote records for call option contracts. Average spreads are provided for 12 exchange markets (as well as aggregated across all exchanges) for all call option contracts written on a given underlying asset traded in January, 2015, which expire on February 20th, 2015. All intra-daily option quotes used for the analysis are partitioned into three groups of moneyness for which the average relative spreads are computed (out-of-the money options - top panel, at-the-money options - middle panel, in-the-money options - bottom panel).

Market	PG	WMT	K	XLP	XOM	CVX	MPC	XLE	JPM	BAC	BLK	XLF	AAPL	GOOG	EA	XLK	SPY
	Out-of-the-money options																
Nasdaq PHLX	0.283	0.266	0.579	0.329	0.151	0.144	0.521	0.192	0.188	0.275	0.474	0.354	0.063	0.167	0.160	0.494	0.106
Nasdaq OM	0.223	0.280	0.543	0.284	0.146	0.136	0.466	0.209	0.172	0.240	0.427	0.343	0.033	0.132	0.147	0.493	0.125
BX Options	0.286	0.308	0.629	0.321	0.143	0.191	0.585	0.220	0.193	0.277	0.552	0.375	0.095	0.180	0.184	0.526	0.162
ISE	0.303	0.282	0.605	0.312	0.149	0.146	0.548	0.219	0.195	0.215	0.438	0.352	0.048	0.138	0.171	0.553	0.108
ISE Gemini	0.268	0.316	0.625	0.282	0.179	0.215	0.530	0.215	0.220	0.245	0.485	0.289	0.044	0.145	0.218	0.555	0.101
CBOE	0.299	0.275	0.629	0.326	0.162	0.145	0.557	0.225	0.195	0.243	0.482	0.407	0.051	0.170	0.174	0.592	0.129
C2	0.339	0.311	N/A	0.276	0.239	0.196	N/A	0.245	0.216	0.368	N/A	0.382	0.062	0.323	0.207	0.543	0.106
BATS BZX	0.257	0.264	0.628	0.318	0.143	0.135	0.574	0.262	0.124	0.151	0.425	0.236	0.039	0.145	0.156	0.647	0.120
NYSE Amex	0.324	0.274	0.632	0.335	0.159	0.152	0.531	0.222	0.221	0.289	0.559	0.414	0.064	0.222	0.194	0.609	0.188
NYSE Arca	0.269	0.263	0.613	0.330	0.171	0.143	0.516	0.227	0.193	0.205	0.417	0.337	0.036	0.144	0.157	0.565	0.144
MIAX	0.260	0.259	0.622	0.313	0.149	0.147	0.512	0.209	0.164	0.269	0.477	0.348	0.060	0.158	0.188	0.491	0.145
BOX	0.243	0.233	0.569	0.302	0.143	0.132	0.570	0.194	0.177	0.220	0.432	0.326	0.063	0.123	0.147	0.453	0.212
All markets	0.288	0.275	0.618	0.317	0.157	0.150	0.535	0.218	0.193	0.252	0.499	0.363	0.054	0.173	0.176	0.559	0.130
	At-the-money options																
Nasdaq PHLX	0.066	0.053	0.150	0.097	0.058	0.056	0.262	0.066	0.037	0.043	0.163	0.066	0.021	0.032	0.054	0.119	0.019
Nasdaq OM	0.064	0.050	0.163	0.084	0.059	0.060	0.247	0.063	0.025	0.027	0.174	0.046	0.016	0.031	0.052	0.108	0.017
BX Options	0.072	0.059	0.182	0.102	0.057	0.076	0.278	0.068	0.043	0.042	0.172	0.063	0.029	0.035	0.072	0.124	0.065
ISE	0.069	0.053	0.158	0.094	0.063	0.057	0.260	0.065	0.033	0.033	0.169	0.055	0.018	0.039	0.055	0.116	0.017
ISE Gemini	0.067	0.058	0.161	0.099	0.069	0.071	0.258	0.066	0.043	0.029	0.187	0.049	0.019	0.041	0.073	0.122	0.039
CBOE	0.070	0.053	0.156	0.096	0.063	0.054	0.270	0.064	0.036	0.044	0.169	0.060	0.019	0.034	0.053	0.122	0.019
C2	0.086	0.062	N/A	0.114	0.079	0.066	N/A	0.071	0.048	0.071	N/A	0.055	0.020	0.075	0.065	0.126	0.027
BATS BZX	0.073	0.052	0.164	0.086	0.064	0.060	0.264	0.076	0.024	0.026	0.183	0.045	0.016	0.032	0.055	0.118	0.026
NYSE Amex	0.074	0.054	0.164	0.100	0.065	0.055	0.266	0.064	0.035	0.047	0.166	0.061	0.019	0.034	0.058	0.125	0.023
NYSE Arca	0.072	0.052	0.152	0.086	0.067	0.055	0.255	0.062	0.027	0.028	0.165	0.050	0.016	0.030	0.050	0.116	0.016
MIAX	0.069	0.054	0.150	0.094	0.062	0.055	0.251	0.065	0.034	0.047	0.167	0.062	0.020	0.034	0.063	0.115	0.023
BOX	0.072	0.054	0.160	0.093	0.064	0.060	0.288	0.067	0.036	0.049	0.172	0.067	0.028	0.041	0.057	0.114	0.031
All markets	0.071	0.054	0.159	0.096	0.063	0.058	0.263	0.065	0.035	0.040	0.169	0.058	0.019	0.036	0.058	0.120	0.024
	In-the-money options																
Nasdaq PHLX	0.087	0.094	0.093	0.123	0.117	0.107	0.171	0.088	0.077	0.014	0.060	0.029	0.010	0.032	0.049	0.179	0.055
Nasdaq OM	0.093	0.118	0.133	0.112	0.137	0.123	0.182	0.095	0.096	0.009	0.059	0.033	0.024	0.032	0.058	0.141	0.145
BX Options	0.089	0.105	0.118	0.138	0.137	0.134	0.188	0.090	0.083	0.013	0.057	0.043	0.016	0.032	0.052	0.177	0.122
ISE	0.089	0.097	0.091	0.098	0.116	0.107	0.176	0.095	0.091	0.013	0.061	0.025	0.013	0.032	0.058	0.137	0.033
ISE Gemini	0.087	0.106	0.093	0.271	0.136	0.123	0.184	0.093	0.093	0.011	0.064	0.031	0.014	0.032	0.060	0.141	0.146
CBOE	0.083	0.091	0.099	0.114	0.120	0.105	0.196	0.088	0.066	0.015	0.066	0.026	0.010	0.037	0.051	0.137	0.022
C2	0.174	0.104	N/A	0.358	0.193	0.145	N/A	0.112	0.061	0.026	N/A	0.031	0.015	0.049	0.055	0.231	0.041
BATS BZX	0.097	0.105	0.102	0.111	0.117	0.110	0.189	0.101	0.087	0.009	0.063	0.031	0.019	0.037	0.058	0.140	0.142
NYSE Amex	0.073	0.088	0.090	0.104	0.112	0.109	0.180	0.086	0.067	0.014	0.080	0.027	0.010	0.035	0.048	0.138	0.060
NYSE Arca	0.078	0.100	0.093	0.097	0.121	0.109	0.179	0.085	0.069	0.010	0.057	0.026	0.013	0.033	0.051	0.130	0.046
MIAX	0.088	0.101	0.090	0.120	0.150	0.152	0.172	0.089	0.083	0.015	0.062	0.036	0.011	0.036	0.057	0.141	0.114
BOX	0.080	0.102	0.102	0.156	0.115	0.110	0.217	0.083	0.062	0.016	0.064	0.027	0.017	0.036	0.048	0.149	0.150
All markets	0.086	0.098	0.097	0.116	0.127	0.124	0.180	0.090	0.079	0.014	0.063	0.030	0.012	0.034	0.053	0.145	0.062

Table A.6: Average relative spread measures computed from the OPRA quote records for put option contracts. Average spreads are provided for 12 exchange markets (as well as aggregated across all exchanges) for all put option contracts written on a given underlying asset traded in January 2015, which expire on February 20, 2015. All intra-daily option quotes used for the analysis are partitioned into three groups of moneyness for which the average relative spreads are computed (out-of-the money options - top panel, at-the-money options - middle panel, in-the-money options - bottom panel).

B Structure of the OPRA Dataset

The OPRA data can be informally split into two parts. The first part directly refers to the option contract, while the second part contains information about the underlying asset. Each message of data represents a line in the data file with 33 fields. Data is structured in such a way that all the messages are grouped by option classes according to a strict order (date, option symbol, expiry, type, strike and time), ensuring that adjacent quotes and trades refer to the same option contract.[32]

Below we briefly discuss the information contained in the messages. A detailed specification can be found in the official documentation by OPRA.[33]

B.1 Data Content

Table B.7 summarizes the OPRA data structure with a brief description of the data fields. The descriptive information about an event (quote or transaction) is contained in fields 1-12. In particular, these fields specify whether a quote or trade is recorded, time stamps, a marketplace, option class, contract specification (put or call, tenor and strike) and the condition code, which contains extra information about the event.

For standard options, the option class symbol (field *CLASS_SYM*) is equivalent to the underlying ticker. Additional digits ("1", "2", etc.) after the ticker indicate adjusted options after corporate actions, such as stock splits, special dividends, spin-offs or mergers, where the deliverable per contract deviates from 100 shares. Furthermore, Mini Options (written for 10 shares instead of 100 in case of a standard option) are identified by the digit "7" (that might change to "8" or "9" as a result of corporate actions). Jumbo Options (written for 1000 shares) are indicated with the additional letter "J". For example, the standard, Mini and Jumbo deliveries for options on SPY are denoted by the option symbols SPY, SPY7 and SPYJ, respectively.

The quote (or transaction) prices and sizes are stored in fields 13-16. Note that for this dataset OPRA does not report on what side (buy or sell) a transaction has been executed. Although this information can often be deduced from the preceeding quotes, trades inside the best bid and offer levels (which are mainly seen with pit-trading or multi-leg options) can not be classified with certainty.

Fields 17 and 33 specify the marketplace and ticker of the underlying instrument. Fields 18-24 contain information on the last quotes (top-of-the-book) and sales (if available) of the underlying, which precede the considered trade or quote of a given option contract. Fields 25-32 provide additional descriptive and technical information.

[32] Since the data is provided at a millisecond precision, several records can have the same time stamp. To ensure that the data are kept in the same order as received by OPRA, a special field with a sequence number, *OPRA_SEQ_NBR*, is introduced.

[33] Please refer to the "OPRA Binary Data Recipient Spec" file retrieved from https://www.opraplan.com/document-library.

No	Column names	Example	Description
1	RECORD_TYPE_CODE	Q	Either "Q" for quote or "T" for trade records
2	TRADE_DATE	20150102	Trade date in the format *yyyymmdd*
3	TRADE_TIME	083301	Trade time (in CT) in the format *hhMMss*
4	TRADE_TIME_THOU	837	Trade times millisecond, i.e. 8:33:01.*837*
5	OPRA_SEQ_NBR	16858240	Sequence number assigned by OPRA
6	EXCHANGE_CODE	X	Options exchange code
7	OPTION_CONDITION_CODE		Additional information and condition codes
8	CLASS_SYM	Y	Options class symbol
9	EXPIRY_TYPE_CODE	A	Indicates option type and expiry
10	EXPIRATION_DATE	20150117	Expiration date in the format *yyyymmdd*
11	PUT_CALL_CODE	C	Either "P" for put or "C" for call options
12	EXERCISE_PRICE	360.00	Strike price of the option
13	BID_TRADE_PRICE	105.3	Either the trade price or the best quoted bid price at the particular exchange
14	BID_TRADE_SIZE	10	Either the traded contracts or (for quote records) the number of contracts available at the best bid price at the particular exchange
15	ASK_PRICE	107.3	Either zero (for trades) or the best quoted offer price at the particular exchange
16	ASK_SIZE	10	Either zero (for trades) or the number of contracts available at the best quoted offer price
17	STOCK_PRIMARY_MARKET	N	Primary stock exchange
18	STOCK_LAST_CONDITION_CODE	R	Condition code for the stock last sale
19	STOCK_LAST_SALE_PRICE		Last sale price of the underlying (if available)
20	STOCK_LAST_SALE_SIZE		Last sale size of the underlying (if available)
21	STOCK_BID_PRICE	465.43	First-level bid price
22	STOCK_BID_SIZE	2	First-level bid size (divided by 100)
23	STOCK_ASK_PRICE	470.00	First-level ask price
24	STOCK_ASK_SIZE	2	First-level ask size (divided by 100)
25	TRADE_DATE2	02-JAN-15	Trade date in the format *dd-MON-yy*
26	ORIG_OPRA_REC_TYPE	F	Original record type code from OPRA
27	OPRA_YEAR_CODE	5	Year code as defined by OPRA
28	QUOTE_PRICE		Unused
29	QUOTE_SIZE		Unused
30	EXPIRATION_MONTH	01	Expiration month of the contract
31	EXPIRATION_YEAR	2015	Expiration year of the contract
32	THREAD_CODE	7	Internal CBOE threat code
33	UNDER_SEC_SYM	Y	Underlying Ticker Symbol

Table B.7: Information contained in the OPRA dataset.

B.2 Data Examples

Table B.8 provides an example of raw data. It displays the first five consequtive quote messages recorded on CBOE (records are taken from the raw data file that contains trades and quotes for options whose underlying tickers begin with letter "S", on January 2, 2015).

Line	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
26	Q	20150102	83018	402	58448642	C		S	A	20150102	C	0.5	3.6	36	3.75	36
42	Q	20150102	83023	674	65249664	C		S	A	20150102	C	0.5	3.6	37	3.75	32
43	Q	20150102	83023	700	65288833	C		S	A	20150102	C	0.5	3.6	55	3.75	48
55	Q	20150102	83046	173	88081921	C		S	A	20150102	C	0.5	3.6	107	3.75	100
56	Q	20150102	83046	858	88523392	C		S	A	20150102	C	0.5	3.6	114	3.75	112

Table B.8: Example of raw data. The first column reflects the number of a message in the file, while the remaining columns are defined in Table B.7.

As it can be seen from this example, the best bid and ask sizes (fields 14 and 16) are always modified simultaneously, illustrating the specifics of quote-driven markets. Often, there are several quote updates in a single message ("bulk-quoting"), having the identical timestamps for a particular exchange and multiple option contracts.

The first two trade records in the same file can be found in lines 18,030 and 19,390, respectively (Table B.9). Both are executed electronically on the Nasdaq Options Market (code Q in field 6). The first trade was executed at *10:23:40.334* and perfectly matches the previous ask quote in price and size.[34] Right after this trade, in the very same millisecond, a new quote reflects the adjusted ask side offer.

Line	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
18,018	Q	20150102	102304	23	3224704	Q		S	A	20150102	C	3.5	0.64	108	0.8	85
18,029	Q	20150102	102340	333	14308480	Q		S	A	20150102	C	3.5	0.64	108	0.74	3
18,030	T	20150102	102340	334	14308608	Q	I	S	A	20150102	C	3.5	0.74	3	0.0	0
18,031	Q	20150102	102340	334	14308736	Q		S	A	20150102	C	3.5	0.64	108	0.79	74
...																
19,388	Q	20150102	130900	226	19821824	Q		S	A	20150102	C	3.5	0.69	79	0.79	14
19,389	Q	20150102	130900	228	19837568	Q		S	A	20150102	C	3.5	0.69	79	0.81	63
19,390	T	20150102	130900	228	19837696	Q	I	S	A	20150102	C	3.5	0.79	14	0.0	0
19,391	Q	20150102	130900	229	19840640	Q		S	A	20150102	C	3.5	0.69	79	0.82	77

Table B.9: Example of raw data. The first column reflects the number of a message in the file, while the remaining columns are defined in Table B.7. Highlighted lines refer to the quotes and the trades, originated from these quotes.

The second trade, however, refers to the second last quote record (line 19,388). The quote message in between (line 19,389) is recorded at the same millisecond as a trade, thus, breaking the natural quote-trade order and making it harder to analyze the trade directions.

Figures B.1 and B.2 illustrate the data samples for two selected option contracts.

[34] Regardless of whether the trade is a buy or sell, it is always stored in columns 13 and 14.

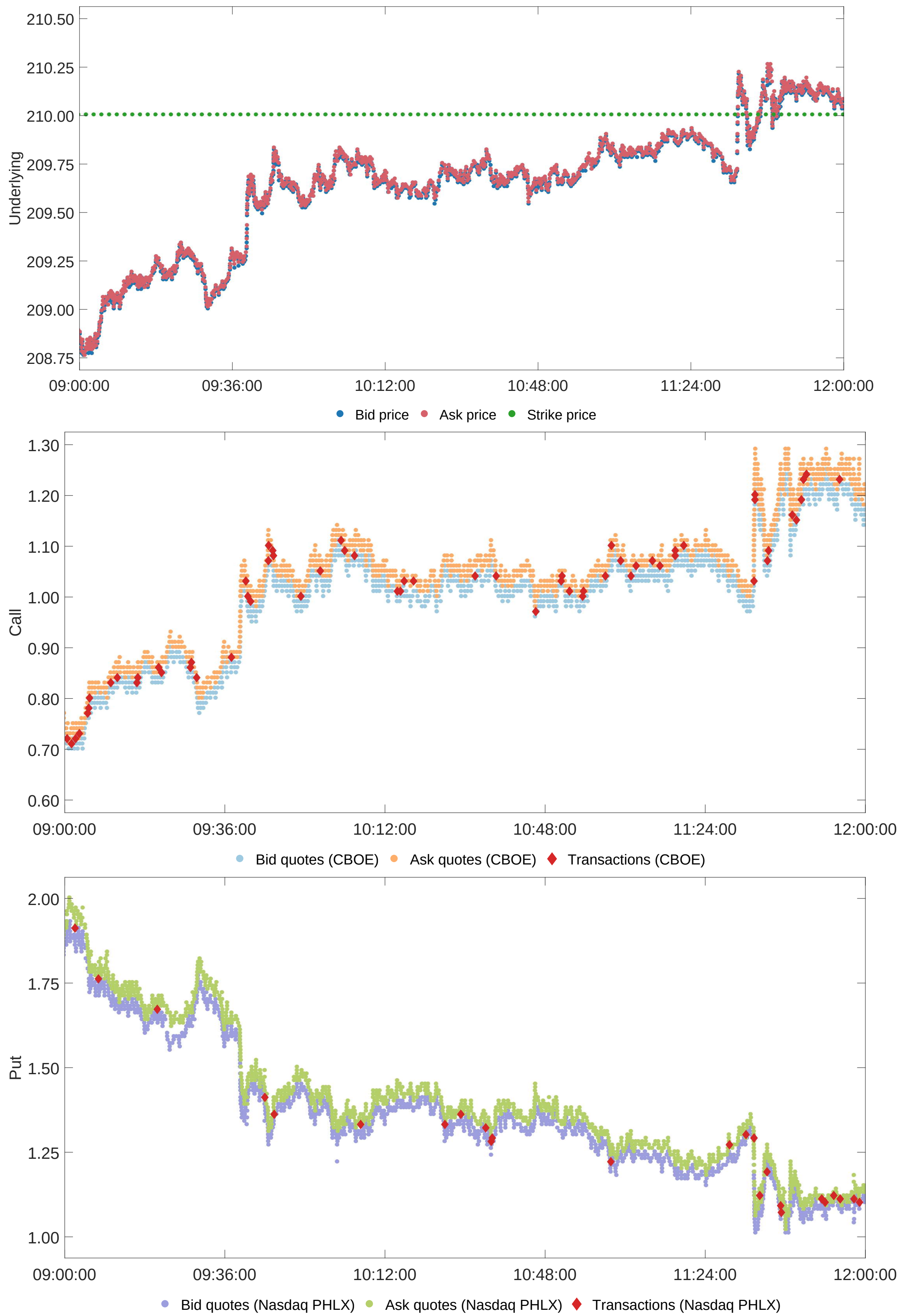

Figure B.1: An example of trade and quote data for at-the-money SPY options (strike price \$210, one week before expiration) traded on February 20, 2015 for the intra-daily interval between 9:00 and 12:00 CT. The underlying prices are depicted on the top plot. The middle plot contains 84,322 quotes and 71 transactions for the call option from CBOE. The bottom plot contains 52,654 quotes and 28 transactions for the put option from Nasdaq PHLX.

Figure B.2: An example of trade and quote data for at-the-money Electronic Arts options (strike price \$57, one week before expiration) traded on February 20, 2015 for the intra-daily interval between 9:00 and 12:00 CT. The underlying prices are depicted on the top plot. The middle plot contains 6,446 quotes and 1 transaction for the call option from MIAX. The bottom plot contains 2,376 quotes and 1 transaction for the put option from Nasdaq NOM.

C Construction of MFIV in Real Time

Here we provide a detailed description of how we construct the MFIV index in real time from intra-daily cross-sections of option prices. Namely, such intra-daily cross-sections can be used to construct a weighted portfolio of prices that replicates risk neutral expected variance of the asset return. Our real time MFIV index represents the sequence of such weighted portfolios constructed on a second-by-second basis.

To begin with, we assume that we have a chunk of raw OPRA dataset that consists of quote records for options with a selected expiration date, for a selected underlying asset and trading day. In our illustrative example in Section 5.1, we use quote data for SPY options traded on January 14, 2015 and expiring on January 17, 2015 (3 days prior to expiry), and for GOOG options traded on January 5, 2015 and expiring on January 9, 2015 (4 days prior to expiry). The construction of real time MFIV is conducted via the following steps.

1. Apply a set of selected filtering rules, introduced in Section 3.2, to the raw data (we use F1, F2, F5c). In addition, we discard the quote records with zero bid prices in order to reduce potential distortions in mid-quotes due to the bounded bid-ask corridor.

   In our empirical application, we retain only the records from the CBOE exchange market to avoid potential effects related to asynchronicity in quoting across multiple exchanges. Optionally, however, the records from several exchange platform of interest (or even from all available platforms) can be retained, while records from the other markets are filtered away.

2. Split the data into pieces which are characterized by a contract type (put or call) and a strike price ($K_0 < K_1 < ... < K_N$).

3. For each piece, bring the records to the second-by-second grid. For this, we preserve the last quotes recorded in all the calendar seconds for which OPRA records are available. For the remaining calendar seconds, we use the previous quote interpolation (see Gencay et al. (2001), for example). As a result, for each calendar second we have a cross-section of quotes for put and call options for the entire range of strike prices traded on that day.

4. Determine the futures price, F_t, by using option mid-quotes and put-call parity for each calendar second, t. Based on F_t, select only out-of-the-money options for each calendar second (put options with $K_i \leq F_t$ and call options with $K_i > F_t$). We denote mid-quote prices of such out-of-the-money options by $\mathcal{O}_t(K_i)$.

   Our task now is to approximate the following integral for each calendar second t,

   $$\int_0^\infty f_t(K)\mathcal{O}_t(K)dK,$$

   where $f_t(K) = \frac{2}{K^2}\Big(1 - \log K + \log F_t\Big)$ is a weighting function. The integral can be viewed as a portfolio of a continuum of options that replicates the expected realized variance of an asset return. The integral value provides an option implied variance that we are aiming to extract. Given that only a finite number of strikes is available, we must approximate the integral by Riemann sums. To enhance the approximation accuracy we split a strike range into three parts - a middle section, left and right tails.

5. Compute Black-Scholes implied volatility, σ_t, for each calendar second by using mid-quotes of at-the-money option contracts. We use σ_t to define tail sections on a strike range.

6. For each calendar second, approximate the option portfolio for the middle section of a strike range by the following sum

$$S_{m,t} = \sum_{i \in \mathcal{I}_{m,t}} f_t(K_i) \mathcal{O}_t(K_i) \Delta K_i$$

where $\mathcal{I}_{m,t}$ is a set of all available strikes which satisfy $-2.5\sigma_t\sqrt{\tau} \leq \log\big(K_i/F_t\big) \leq 1.5\sigma_t\sqrt{\tau}$ and τ denotes the time until expiration. The range is asymmetric and implies that we use more put than call options to approximate the middle section of the integral. This is the case because out-of-the-money put options are presumably more liquid than out-of-the-money call options and have more narrow spreads. Thus, the mid-quote prices of out-of-the-money puts are reliable for a wider moneyness range than the mid-quotes of out-of-the-money calls.

7. Since the number of strikes available for tail sections is typically very limited, we extrapolate option prices for these sections. We use a log-linear extrapolation of the tail prices for which we need to determine the corresponding slopes. We calculate such left and right tail slopes once for a specific intra-daily time interval, say, 30 minutes. Begin with the left tail. We pull all the observed mid-quote prices $\mathcal{O}_t(K_i)$ such that $-15\sigma_t\sqrt{\tau} < \log\big(K_i/F_t\big) < -2.5\sigma_t\sqrt{\tau}$ for all t from a given half-hour interval and compute the slope as an average of

$$\frac{\log \mathcal{O}_t(K_i) - \log \mathcal{O}_t(K_{i-1})}{\log K_i - \log K_{i-1}},$$

over all appropriate K_i and t.

Similarly, we determine the slope based on the log-linear extrapolation for the right tail prices. For this we use the mid-quotes prices $\mathcal{O}_t(K_i)$ with strikes $1.5\sigma_t\sqrt{\tau} < \log\big(K_i/F_t\big) < 15\sigma_t\sqrt{\tau}$ and for all t from the considered half-hour interval.

8. For each calendar second, we extrapolate mid-quotes from the middle section, $\mathcal{I}_{m,t}$, into the left and right tails of the strike range using the slopes calculated before. Price extrapolation is implemented while $\log\big(K/F_t\big) < 15\sigma_t\sqrt{\tau}$ with the strike step equals to the minimal strike increment ΔK_i for $i \in \mathcal{I}_{m,t}$. We denote the sets of strikes in the left and right tail sections, for which the option mid-quotes are extrapolated, by $\mathcal{I}_{l,t}$ and $\mathcal{I}_{r,t}$, respectively. Extrapolated mid-quote prices are denoted by $\widetilde{\mathcal{O}}_t(K_i)$.

9. For each calendar second, approximate the option portfolio for the left tail section and for the right tail section of a strike range by the following sums

$$S_{l,t} = \sum_{i \in \mathcal{I}_{l,t}} f_t(K_i)\widetilde{\mathcal{O}}_t(K_i)\Delta K_i \qquad \text{and} \qquad S_{r,t} = \sum_{i \in \mathcal{I}_{r,t}} f_t(K_i)\widetilde{\mathcal{O}}_t(K_i)\Delta K_i,$$

where $\mathcal{I}_{l,t}$ is a set of extrapolated strikes in the left tail which satisfy $-15\sigma_t\sqrt{\tau} < \log(K_i/F_t) < -2.5\sigma_t\sqrt{\tau}$ and $\mathcal{I}_{r,t}$ is a set of extrapolated strikes in the right tail which satisfy $1.5\sigma_t\sqrt{\tau} < \log(K_i/F_t) < 15\sigma_t\sqrt{\tau}$.

10. An approximation of the entire integral by Riemann sums obtained from the three sections of the strike range represents the model-free implied variance measure,

$$MFIV_t = S_{l,t} + S_{m,t} + S_{r,t},$$

and is computed for all calendar seconds within a considered time interval (e.g., a trading day).