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

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

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

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

[c] *University of Manchester, Alliance Manchester Business School*

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

August 21, 2020

Abstract

This paper provides a guide to high frequency option trade and quote data disseminated by the Options Price Reporting Authority (OPRA). We present a comprehensive overview of the U.S. option market, including details on market regulation and the trading processes for all 16 constituent option exchanges. We review the existing literature that utilizes high-frequency options data, summarize 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 rich information content of the high frequency OPRA data.

Keywords: Options, 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 thank the guest editor Kris Jacobs and two anonymous referees for their invaluable comments which greatly improved the quality of this paper. 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 been a rapidly developing research area in recent years. Nonetheless, the full range of available data is underexploited, as empirical studies typically rely on closing (end-of-day) option prices. We advocate for a change in this regard. High-frequency option data has the potential to convey accurate real-time information regarding investors' expectations about a company, a sector, or even the entire market. Through the tight connection between option and underlying stock prices, intraday data provides a more comprehensive view of the realized and expected asset price dynamics, offering potential insights for short-term asset return predictability, intraday risk management, price discovery, information processing, and the role of liquidity. Finally, intraday option data allows for a more precise construction of popular model-free option-implied risk-neutral measures concerning the future return distribution and volatility. As such, they convey information regarding how strongly investors' expectations and risk appetites change in response to the intraday order flow and news arrivals.[1]

Among the vast literature on the U.S. options market, the use of high-frequency option prices is relatively rare.[2] There is a stream of literature employing high-frequency data to explore the intraday option price dynamics and provide inference on key quantities, including jumps, implied volatility surfaces and risk-neutral densities, e.g. Birru and Figlewski (2012), Andersen et al. (2015b), Audrino and Fengler (2015), Amaya et al. (2018), Taylor et al. (2018), Kapetanios et al. (2019), Dalderop (2020). However, these studies predominantly focus on options written on indices or index futures, while there is little work on the high-frequency dynamics of individual equity options.

Another severely under-researched area is the microstructure of equity option markets. The structure and organization of the U.S. market for exchange-traded options has undergone dramatic changes over recent decades. The number of exchange holding groups and trading venues has increased rapidly, the regulatory oversight has strengthened, infusing the markets with a higher degree of transparency and competition, the trading technology has developed at a rapid pace, experimentation with diverse incentives for market making and provision of order flow is rampant, and product innovation has been strong, with entirely new option categories gaining market share quickly. The implications of these developments for trading costs, price efficiency, liquidity and market depth across exchanges and product categories are largely unexplored. A few recent studies are Muravyev and Pearson (2020) on trading costs, exploiting minute-by-minute option trades and quotes for stocks in the S&P 500 index, as well as Anand et al. (2016) and Battalio et al. (2016), who study the impact of the fee structure on the effective option bid-ask spread and total trading costs.

The paucity of research on the microstructure and high-frequency dynamics for the full cross-section of securities and option venues is due to a variety of factors. One primary reason is the large amount of data storage and processing resources required to analyze even a short sample comprising all options for a limited number of underlying assets. This is particularly striking for actively traded securities. For example, from January to August 2015, for the Apple stock, we observe an average of 271,641 trades and

[1] In summary, intraday option prices are useful in improving our understanding of many of the issues explored by our friend, collaborator, 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 behavior of the implied volatility smirk (Christoffersen et al., 2009), estimates of systematic equity risk provided by option prices (Chang et al., 2012), and the existence of liquidity premiums in option prices and quotes (Christoffersen et al., 2018).

[2] See Section 3 for a comprehensive review.

3,283,067 quotes per day, implying a quote-to-trade ratio of 12.1. This activity is greatly surpassed by the option market featuring, on average, 77,849 trades and 226,822,053 quotes *per day* across all 1,169 option classes written on Apple, suggesting an option quote-to-trade ratio of almost 3,000. Second, the maturity structure evolves every day, with maturities dropping and new option contracts entering the sample, as others expire. Hence, the panel is unbalanced, and it contains many thinly traded options. Third, the market environment is constantly shifting, with new venues appearing, exchange mergers eliminating existing ones, and trading protocols undergoing frequent modification. It can be difficult to generate a sufficiently large sample under stable conditions to obtain good empirical estimates of relevant structural quantities in this setting, but high-frequency data should help. Fourth, no single source exists for option trades and quotes, reported in a homogenous manner, over the full period of exchange-based options trading in the U.S. Fifth, obtaining a complete record of the market activity at the highest frequency has, at least until very recently, come at a high financial cost. As a result, scholars have exploited a plethora of ad hoc arrangements to obtain partial coverage of the market activity over limited sample periods.

Our main objective is to provide an overview of the current U.S. exchange-based option market, with an emphasis on the pitfalls and opportunities associated with high-frequency data. We rely on the most comprehensive database – the CBOE OPRA Data (Bulk) package – provided by the Options Price Reporting Authority (OPRA), which collects and disseminates intraday trade and quote data at a millisecond precision from all option exchanges operating in the U.S. It covers all option classes written on more than 3,500 equities, more than 500 exchange traded products and about 50 index-driven assets. Hence, we seek to provide scholars with guidance on how to process and utilize such data, provide an explorative overview, and present some initial illustrations 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 16 U.S. security exchanges eligible for options listing and trading, (ii) a selective, yet fairly extensive, descriptive analysis of the intraday trades and quotes for options written not only on equity indices, but also individual stocks and other exchange traded products, and (iii) a few illustrative applications, implemented to demonstrate the feasibility of constructing high-frequency option-based measures through standard procedures, and then assess their advantages and drawbacks relative to end-of-day option data.

As a consequence, we do not pursue any specific research question in depth, but review relevant market institutions and features. Along the way, we identify opportunities for new research, that are opening up with the increasing availability of databases covering the option market activity at the tick-by-tick level, coupled with the rapid advances in processing power and the declining data storage costs.

We first briefly review the history and competitive forces that have shaped the market evolution since the start of organized option trading in the early 1970's. We then outline the most critical regulatory initiatives determining the transparency of trading, price formation, and market maker quoting obligations. The latter are especially important given the dominance of quotes relative to trades in option markets.

We next explore the quality of the raw (bulk) data. We identify the extent of potential faulty recordings, other data errors, non-informative or irregular quotations, outliers, and the occurrence of records with identical time stamps. To provide practical guidance, we develop a filtering algorithm in the spirit of the cleaning algorithm proposed by Barndorff-Nielsen et al. (2009) for high-frequency equity data.

To keep the analysis manageable, we focus on data originating from 12 representative underlying equities and 5 exchange traded funds, with parts of the study concentrating on the trading during January 2015. The latter still leaves us with approximately 27 billion trades and quotes. We summarize the trading and quoting activities for each underlying asset, with further categorization based on option maturity and moneyness. We also explore the liquidity characteristics through bid-ask spread measures. The latter are critical for assessing the signal-to-noise ratio associated with the option quotes. Finally, we check for arbitrage violations associated with a basic put-call parity relation. These occur at non-trivial frequencies and are observed from quotes at all exchange venues, albeit at somewhat different intensities.

A unique feature of the new OPRA database is that it allows for a study of the quoting and trading of an instrument across all exchange venues. This enables us to compare the cross-exchange trade and quote flows and to assess potential price leadership, determined by the degree to which the exchange participates in the National Best Bid and Offer (NBBO) quoting pair. We find that CBOE, AMEX and ARCA most often match the NBBO, but participation is otherwise quite evenly distributed across all active exchanges, lending credence to the hypothesis of an integrated national options market.

Finally, we provide a pair of illustrations using OPRA data to gauge the feasibility and reliability of standard techniques for constructing popular option-implied measures at very high frequencies. The first application involves estimation of the risk-neutral return variance. We construct these measures second-by-second from option prices written on SPY (an exchange traded fund) and GOOG (Google stock). We verify that they display substantial, genuine intraday variation, so they add significantly to the information content provided by typical end-of-day measures extracted from, say, OptionMetrics data. However, we also find that the series suffer from significant serial correlation, indicating a non-trivial impact of noise. We find that such effects only vanish for sampling frequencies of about one minute or lower.

In the second application, we estimate the risk-neutral return density (RND) from intraday option prices. We compare the estimated RND curves for all underlyings 30 minutes before and after the news release from the FOMC meeting in March 2015. We document a uniform reduction in implied volatility along with a dampening of the left tail of the distribution after the FOMC statement. The point is that such analyses are perfectly feasible across a wide cross-section of equity options, allowing for future studies investigating heterogeneity in the response across stocks in different sectors or with distinct characteristics.

The rest of the paper is organized as follows. Section 2 provides an overview of the U.S. option exchange trading industry. Section 3 reviews the relevant literature that uses high-frequency options data. Section 4 contains an overview of the OPRA dataset, the characterization of its special records, and develops suggestive data filtering rules. A thorough description of the observed trade and quote record for the selected data sample are presented in Section 5. Section 6 provides our empirical illustrations using intraday option prices to estimate the risk-neutral variance and density of the underlying asset returns. Section 7 concludes the paper. All additional materials are relegated to the Web Appendix.

2 Option Trading in the U.S.

2.1 The Evolution of the U.S. Options Market

Exchange-based options trading in the U.S. began in April 1973 with the foundation of the Chicago Board Options Exchange (CBOE). During 1973-1999, three additional exchanges opened and continued operating as independent trading venues for options till 1999, namely the American Stock Exchange (AMEX), the Philadelphia Stock Exchange (PHLX) and the Pacific Exchange (PCX). These four exchanges were all floor-based with either an open outcry or a specialist structure. Over this period, the market was highly fragmented, as actively traded options, almost exclusively, were listed on a single exchange, so the trading activities were governed by the listing options exchange only, see, e.g., Battalio et al. (2004).

This fragmentation was targeted by an options listing campaign in August 1999, when prior exchange-exclusive options began to be listed at competing exchanges. This campaign sharpened the competition among exchanges so that, in short order, 37% of all equity option volume were for contracts traded on multiple exchanges (De Fontnouvelle et al., 2003). Mayhew (2002) and De Fontnouvelle et al. (2003) find the enhanced competition to improve market quality in terms of smaller quoted or effective spreads.

Inspired by the heightened exchange competition, in 2000, the Security and Exchange Commission (SEC) approved the Plan for the Purpose of Creating and Operating an Intermarket Options Linkage[3] (the 'Linkage Plan') and Firm Quote and Trade-Through Disclosure Rules for Options[4] to facilitate the creation of a national market. The Linkage Plan is essentially a central routing system operated by the Options Clearing Corporation (OCC) for the participating exchanges to route order flows with the aim to limit trade-throughs and execute at the National Best Bid and Offer (NBBO) price. Comparing market quality before and after the Linkage Plan, Battalio et al. (2004) find the quality of the options market to be substantially improved, with the number of crossed quotes reduced by 85%, a 7% decrease in the trade-through rates, and an overall reduction in the effective spread of over 60%, suggesting the Linkage Plan was a major driver in transforming the fragmented options market into a national market system.

The year 2000 also witnessed the foundation of the International Securities Exchange (ISE) – a fully electronic options venue without a physical trading floor. This innovative design presented a challenge to the traditional floor-based options exchanges. According to Simaan and Wu (2007), ISE generated more informative and executable option quotes with a smaller bid-ask spread compared to its rivals. These advantages catapulted ISE into the leading exchange in terms of trading volume by 2003, and the electronic-based market structure was adopted by CBOE in 2001, PCX in 2003 and the new Boston Options Exchange (BOX) in 2004.

The development of electronic option trading platforms and the associated boom in market activity in the mid-2000s created a problem for the Linkage Plan, as the centralized routing system was not designed to handle the elevation in order flow. This problem was compounded by the introduction of the Penny Quoting Pilot Program[5] (the 'Penny Plot') in 2007, which increased the possibility of trade-throughs and locked/crossed markets due to a smaller tick size. Based on the Regulation National Market System (NMS) for the equity market in 2005, in 2009, the SEC approved the Options Order Protection and Locked/Crossed Market Plan[6] (the 'Decentralized Linkage Plan'). This plan augments the former central routing channel with the Intermarket Sweep Order, which allows market makers to route their orders in an efficient and decentralized manner. The intention was to eliminate locked/crossed markets and allow for more efficient price protection following the Regulation NMS.

In 2007, the newly formed NYSE Arca exchange introduced a make-take fee structure to the option classes affected by the Penny Pilot. This led to direct competition with the traditional payment for the

[3]See e.g. https://www.sec.gov/rules/sro/34-43086.htm.

[4]https://www.sec.gov/rules/final/34-43591.htm.

[5]https://www.sec.gov/news/press/2007/2007-10.htm.

[6]https://www.sec.gov/rules/sro/nms/2009/34-60405.pdf.

order flow (PFOF) model used by other exchanges. Anand et al. (2016) find that NYSE Arca's shift to a make-take fee structure reduced the execution costs for liquidity demanders and improved the quotes posted by market makers. This is partially confirmed by Battalio et al. (2016), although they note that the PFOF may lead to lower effective transaction costs for low-priced options. Subsequently, the make-take model was adopted by various exchanges, including Nasdaq, BOX, BATS, ISE and PHLX in 2010.

The number of option venues has doubled over the past decade in parallel with multiple exchange mergers. As of January 2020, 16 option exchanges operate in the U.S., with five holding companies owning one or more of these. They 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 Pearl, MIAX Emerald) and TMX Group (BOX). Table 1 summarizes major events that influenced the evolution of option trading venues in the U.S. A detailed description of all exchanges is provided in Table A.1 of Web-Appendix A.

Date	Event	No. of Venues
Apr 1973	The **Chicago Board Options Exchange (CBOE)** was launched.	1
Jan 1975	Options trading initiated at the American Stock Exchange (AMEX).	2
Jun 1975	Options trading initiated at the Philadelphia Stock Exchange (PHLX).	3
Apr 1976	Options trading initiated at the Pacific Exchange (PCX).	4
Dec 1976	Options trading initiated at the Midwest Stock Exchange (MSE).	5
Jun 1980	Options business at MSE was consolidated with CBOE.	4
Jun 1985	Options trading initiated at the New York Stock Exchange (NYSE).	5
Apr 1997	Options business at the NYSE was consolidated with CBOE.	4
May 2000	The International Securities Exchange (ISE) was launched.	5
Feb 2004	The **Boston Options Exchange (BOX)** was launched.	6
Sep 2005	The Archipelago Exchange (ArcaEx) acquired PCX.	6
Feb 2006	The NYSE Group acquired ArcaEx to form **NYSE Arca Options**.	6
Jul 2007	Nasdaq acquired PHLX to form **Nasdaq PHLX**.	6
Mar 2008	**Nasdaq Options Market (NOM)** was launched.	7
Oct 2008	The NYSE Group acquired AMEX to form NYSE AMEX Options.	7
Feb 2010	BATS Options was launched.	8
Oct 2010	**CBOE C2** was launched.	9
Jun 2012	Options trading initiated at **Nasdaq BX Options (NOBO)**.	10
Jul 2013	ISE Gemini was launched.	11
Dec 2013	The **MIAX Options Exchange** was launched.	12
Nov 2015	BATS EDGX Options was launched.	13
Feb 2016	ISE Mercury was launched.	14
Jun 2016	Nasdaq acquired ISE to form **Nasdaq ISE**, **Nasdaq GEMX** (former ISE Gemini) and **Nasdaq MRX** (former ISE Mercury).	14
Feb 2017	CBOE acquired BATS to form **CBOE BZX** (former BATS Options) and **CBOE EDGX**.	14
Feb 2017	**MIAX Pearl** was launched.	15
Aug 2017	NYSE AMEX Options was renamed to **NYSE American Options**.	15
Mar 2019	**MIAX Emerald** was launched.	16

Table 1: The Evolution of Options Exchanges in the U.S. The active options exchanges as of February 2020 are in **bold**. The information in the table is partially based on Mayhew (2002).

Figure 1 depicts the total annual number of option contracts traded on various exchanges along with the corresponding market shares for 1973-2016. In the two decades leading up to 2000, the trading activity was fairly stable, with CBOE being the dominant venue. The foundation of ISE in 2000 and the introduction of electronic trading triggered unprecedented growth, tripling the overall volume by 2010 and leading to a sharply increased market share for ISE. Finally, after 2010, the trading activity has stabilized, while also becoming more evenly distributed across exchanges.

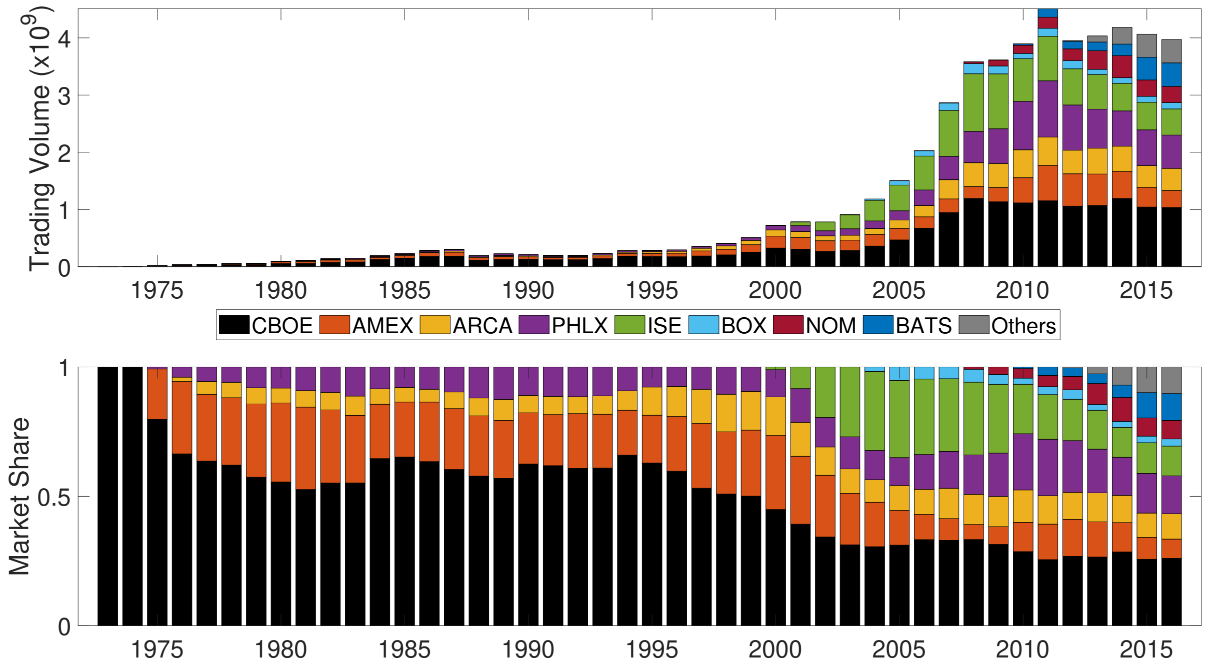


Figure 1: Toal annual option contract volume traded on different exchanges. The 'Others' category contains C2, EDGX, MIAX, MRX, NOBO and GEMX. We exclude NYSE, MSE and NASDAQ, as their respective option trading activities amassed only minor market shares during brief periods prior to 2000. The data stem from the CBOE Annual Market Statistics.

2.2 Market Regulation

The option exchanges are authorized by SEC to act jointly, as parties to a number of NMS plans, including the aforementioned Decentralized Linkage Plan. These plans centralize requirements across exchanges to ensure cross-exchange protection, transparency, surveillance, standardization and audit trails.

The 'Plan for Reporting of Consolidated Options Last Sale Reports and Quotation Information'[7] is one of the NMS plans aimed at reporting trading information from all U.S. option exchanges. The Options Price Reporting Authority (OPRA) is registered as a securities information processor, responsible for the implementation of this plan, and is regulated by a committee comprising all participating exchanges.

OPRA currently processes option trading data from all 16 U.S. option exchange markets. OPRA requires participants to report information on last sale and current quotes in accordance with Rule 602 of Regulation NMS (including prices, quotation sizes, and some regulatory auditing information). The Securities Industry Automation Corporation (SIAC) provides technological infrastructure for collection, consolidation and dissemination of this real-time information. OPRA provides market data to professional (directly or through vendors) and non-professional subscribers (only through vendors) for a fee.

There are three additional mandatory plans for the exchanges. The Options Regulatory Surveillance Authority (ORSA) Plan was adopted in 2006. It seeks to deter insider trading. 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 filed in 2014 to collect all orders and identify them as cancellations, modifications, or executions for the exchange-listed equities and options across all U.S. markets.

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

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 session ending at 15:15 CT. In addition, some exchanges (BATS BZX and BATS EDGX) provide a pre-market trading session, initiated up to two hours prior to the regular market open.

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.

The most heavily traded options are written on single stocks, cash indices and ETF/ETPs, but many other types of underlyings are available, e.g., commodities, interest rates, and foreign exchange. A standard equity or ETF/ETP option contract is an American-type option covering 100 shares of the underlying, and exercise of the contract results in physical delivery of the underlying shares. In case of stock splits or dividend payments, the number of shares and exercise prices are adjusted accordingly. Index options are, in contrast, mostly European-style and settled in cash. Regular options usually expire at the close of trading on the third Friday of each month. In 2005, CBOE issued weekly options that expire every Friday[8] for some indices and ETFs, and expanded it to cover individual securities in 2012. Monday and Wednesday-expiring weekly options were launched in 2016 for selected indices and ETFs.[9]

The option market is a *hybrid quote-driven* market, where market makers are responsible for providing continuous bid and offer quotes. Often, there are multiple types of market makers that differ in privileges and responsibilities. More senior categories (e.g., lead or primary market makers) are granted allocation priority in the relevant option classes, but are subject to stricter capital requirements, quoting obligations, and other responsibilities. Regular market makers may also be registered as preferred or directed market makers with certain privileges in executing preferenced and directed orders. Since the type and role of market makers differ across exchanges and over time, we only discuss some of their general obligations. We refer interested readers to Mayhew (2002) and Simaan and Wu (2007) for a more detailed account.

There are several market-wide obligations for option market makers. First, in February 2001, the SEC introduced a market-wide firm quote obligation through an amendment of the Quote Rule (Securities Exchange Act Rule 11Ac1-1), which was previously applied only to the equity market. This rule requires the market makers to post firm quotes, that are valid for order executions of at least one contract. Before that, option exchanges imposed their own firm quote requirements on market makers independently of each other. 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. Stub quotes might be posted, when market makers attempt to fulfill quoting obligations without an actual intent to trade. Stub quotes were viewed as a contributing factor to the Flash Crash on May 6, 2010. The new rule requires quotes to be within a certain percentage band around the NBBO (or the consolidated last sale, if the NBBO is not available). These requirements seek to make options trading less risky for investors and prevent transactions from being executed at irrational prices.

By the Quote Rule, market makers must provide continuously updated two-sided quotes throughout

[8] Weekly options are not issued when an existing option of the same specification also expires on the same Friday.

[9] Underlyings with weekly options available can be found in http://www.cboe.com/products/weeklys-options/available-weeklys.

the trading day. Each option exchange (or a self-regulatory organization) imposes additional obligations on its market makers. In general, these quoting obligations are in force irrespective of the prevailing market conditions. Therefore, during episodes of stress, market makers are supposed to maintain liquidity, absorbing the impact of shocks on individual investors.[10] The requirement of continuous quoting is especially important for option markets, because an appreciable fraction of the securities is thinly traded.

Most exchanges require market makers to quote at least 90% of the time during the trading day, with compliance assessed on a monthly basis. Moreover, the quote size should exceed a minimum number of contracts, usually determined on a class-by-class basis, and may vary with market maker type. Moreover, market makers must quote continuously in some minimum fraction of the option classes and series to which they are assigned. These fractions range from 60% to 100% across exchanges and market maker type. More detailed information on quotation requirements is collected in Table A.2 of Web-Appendix A.

The minimum tick size - the smallest possible price increment - depends on the price level of an option. For those traded below $3, the minimum tick constitutes $0.01 for the option classes participating in the option penny pilot program and $0.05 for 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 many market-wide ETFs (namely, QQQQ, IWM, and SPY) and option-related products (XSP and VIXW) represent exceptions with a minimal increment of $0.01 for all corresponding option series.

3 Literature Review on High-Frequency Options Data

This section provides a brief and selective review of the existing literature using U.S. high-frequency options data and discusses the differences across the data sources with emphasis on the distinction between the OPRA bulk dataset and those used in prior studies. Towards that purpose, Table 2 references a number of papers along with their associated datasets, sampling periods and option classes used.

The first widely-adopted intraday U.S. stock options data stem from the Berkeley Options Data Base (BODB), which collected each transaction and bid/ask update for every option series on the CBOE from the Market Data Report, time-stamped to the second. The BODB only covers CBOE trades and quotes until 1997, as the systematic collection of BODB data seems to end by December 1996.[11] However, some authors may still, subsequently, have acquired this type of data directly from the CBOE.[12]

After the termination of BODB, scholars focused on an older version of the OPRA dataset for intraday stock options data, stemming from the early 2000s, labeled OPRA (old) in Table 2. This edition of OPRA provides a complete record of option trades and best available quotes for all exchanges, time-stamped to-the-second, but depth information is unavailable. This dataset appears to have been discontinued later in the 2000s. We have been unable to establish the exact sample period covered.

The old OPRA is considerably less granular than the up-to-date, to-the-millisecond OPRA dataset exploited in this paper, and labeled OPRA (new) in Table 2. New OPRA also includes depth information at the best quotes for each exchange. The new OPRA dates back to July 2004, but is discontinued after September 2019. Moreover, not all papers using new OPRA acquire data from CBOE, e.g., Battalio and Schultz (2011) cites a market maker as their source, Hu (2014) uses data from Trade Alert LLC,

[10] See, for example, https://www.sec.gov/comments/s7-05-15/s70515-34.pdf and Nagel (2012).
[11] See e.g. http://libguides.stanford.edu/az.php and https://catalog.princeton.edu/catalog/2593696.
[12] Bollen and Whaley (2004) use such data covering 01/1995 - 12/2000 for the 20 most heavily traded stock options.

Authors	Dataset	Sampling Period	Sampled Options Classes
Bhattacharya (1987)	BODB	06/1977 - 08/1978	Options on the 32 top stocks by volume
Stephan and Whaley (1990)	BODB	01/1986 - 03/1986	All call options
Vijh (1990)	BODB	03/1985 - 04/1985	Options on NYSE-listed stocks
Chan et al. (1993)	BODB	01/1986 - 03/1986	All options
Sheikh and Ronn (1994)	BODB	01/1986 - 09/1987	30 most heavily traded stock options
Chan et al. (1995)	BODB	01/1986 - 03/1986	Options on NYSE-listed stocks
Mayhew et al. (1995)	BODB	09/1985 - 06/1986, 01/1988 - 10/1988	Short-dated options on approx. 110 stocks
Bakshi et al. (1997)	BODB	06/1988 - 05/1991	Options on the S&P 500 index
Easley et al. (1998)	BODB	10/1990 - 11/1990	Options on the 50 top stocks by volume
Lee and Yi (2001)	BODB	01/1980 - 12/1990	Call options on NYSE-listed stocks
Chan et al. (2002)	BODB	01/1995 - 03/1995	Options on the 60 top stocks by volume
Mayhew (2002)	BODB	01/1986 - 08/1997	All stock options traded on CBOE
Pan (2002)	BODB	01/1989 - 12/1996	Options on the S&P 500 index
Chakravarty et al. (2004)	BODB	01/1988 - 12/1992	60 most heavily traded stock options
Cao et al. (2005)	BODB	01/1986 - 12/1994	Options on firms involved in merger and acquisition activities
George and Longstaff (1993)	CBOE	01/1989 - 12/1989	Options on the S&P 100 Index
Bollen and Whaley (2004)	CBOE	06/1988 - 12/2000	Options on the S&P 500 index and 20 individual stocks
De Fontnouvelle et al. (2003)	OPRA (old)	08/1999, 08/2000	28 Multi-listed option classes
Battalio et al. (2004)	OPRA (old)	06/2000, 01/2002	71 (615) option classes in 2000 (2002)
Harris and Mayhew (2005)	OPRA (old)	01/2003	Options on 451 stocks
Battalio and Schultz (2006)	OPRA (old)	01/2000 - 06/2000	Options on up to 49 stocks
Holowczak et al. (2006)	OPRA (old)	05/2002 - 07/2002	Options on the 40 most actively traded stocks
Anand and Chakravarty (2007)	OPRA (old)	07/1999 - 08/1999	Options on 100 sample firms
Simaan and Wu (2007)	OPRA (old)	01/2002	Options on the 50 top stocks by volume
Battalio and Schultz (2011)	OPRA (new)	08/2006 - 10/2008	Options on stocks subject to the short sale ban and a matched sample
Birru and Figlewski (2012)	OPRA (new)	09/2006 - 10/2006, 09/2007 - 10/2007, 09/2008 - 11/2008	December-expiry options on S&P 500 Index
Muravyev et al. (2013)	OPRA (new)	04/2003 - 10/2006	Options on 36 liquid U.S. stocks and 3 ETFs
Cakici et al. (2014)	OPRA (new)	05/2010	Options on S&P 500 and S&P 100 constituents
Holowczak et al. (2014)	OPRA (new?)	02/2006 - 12/2006	Options on QQQQ
Hu (2014)	OPRA (new?)	04/2008 - 08/2010	Options on all individual stocks
Mishra and Daigler (2014)	OPRA (new?)	10/2008 - 12/2008, 10/2009 - 12/2009	Options on SPX and SPY
Chatrath et al. (2015)	OPRA (new)	01/2011 - 05/2012	Options on S&P 500 Index
Anand et al. (2016)	OPRA (new)	01/2007 - 12/2010, 11/2012 - 01/2013	Options traded on NYSE Arca and a matched sample
Muravyev (2016)	OPRA (new)	04/2003 - 10/2006	Options on 39 most actively traded stocks (including 4 ETFs)
Amaya et al. (2018)	OPRA (new?)	07/2004 - 12/2012	Options on S&P 500 Index
Zhang (2018)	OPRA (new)	01/1996 - 01/2015	Options on S&P 500 Index and its constituents, and sector ETFs
Muravyev and Pearson (2020)	OPRA (new) LiveVol (?)	04/2003 - 10/2006 01/2004 - 04/2013	Options on 39 most actively traded stocks (including 2 ETFs)
Simon (2013)	LiveVol	05/2005 - 04/2010	Options on SPY ETF
Battalio et al. (2016)	LiveVol	05/2010 - 06/2010	3233 Options classes on all stocks and ETFs
Christoffersen et al. (2018)	LiveVol	01/2004 - 12/2012	Options on S&P 500 constituents
Battalio et al. (2020)	LiveVol	03/2010	Options on 2945 stocks

Table 2: List of papers using high-frequency options data from the U.S. market. We report the number of option classes before any filtering. BODB stands for the Berkeley Options Data Base. For datasets with unidentifiable data source, we insert a question mark.

while Muravyev et al. (2013) and Muravyev (2016) rely on data from NANEX, time-stamped every 25 milliseconds. The OPRA data in Anand et al. (2016) is preprocessed by the Baruch College Options Data Warehouse, and the source for Zhang (2018) is Thomson Reuters Tick History (TRTH). These third-party vendors receive feeds directly from OPRA, but apply various data cleaning and aggregation procedures which, inevitably, imply some (unknown) loss of information. For example, TRTH processes the raw quote messages from OPRA and provides time-stamped NBBO quotes, but exchange identifiers for the NBBO quotes and exchange-level best quotes are not provided by TRTH.

As recent alternatives to new OPRA, LiveVol,[13] also based at the CBOE, offers two separate intraday datasets. One contains to-the-minute NBBO quotes and trades for all exchanges, obtained by aggregating all underlying individual OPRA entries. A second dataset provides all (tick-by-tick) option transaction prices and volume along with the concurrent best bid-ask quotes for the option and the underlying (time and sales data). These datasets have dramatically reduced granularity relative to OPRA new, but have the advantage of a more manageable size, and the former includes the disseminated NBBO quotes.

In summary, OPRA new is more granular than all of the alternative datasets referenced in Table 2. Moreover, along with LiveVol and TRTH, it is the main source of information regarding recent intraday option market activity. Hence, our study of the OPRA new sample, covering 01/2015 - 08/2015, provides a detailed look at the most granular option dataset available to scholars.

4 Data Overview, Special Records and Data Filtering

In this section, we provide a broad overview of the bulk OPRA data, select a working sample and inspect the associated trade and quote records. Next, we categorize trades and quotes, which might be irrelevant, or even detrimental, for certain types of analyses. Furthermore, we explore the frequency with which such potentially problematic records appear in the bulk data. Finally, we explore the extent to which intraday option quotes recorded by OPRA imply violations of a basic no-arbitrage put-call parity relation.

4.1 Data Overview

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

Each transaction record displays the price and corresponding trading volume. Each quote record contains top-level bid and ask prices along with the quoted amounts, implying that each such record reflects an update of a bid-ask pair (a change in the quoted prices or amounts) relative to the preceding quote record for a given exchange market. In addition, each option quote or trade record contains the most recent first-level quotes for the underlying.[14] A more detailed description of the content and structure of the OPRA dataset is provided in Web-Appendix B.

[13] `https://datashop.CBOE.com/option-quotes-intervals`.

[14] This applies only when the underlying is a tradable instrument. For example, for SPX options, such quotes are not available, while records for SPY options, written on a tradable ETF, contain the most recent top quotes of the underlying. For an in-depth comparison between SPX and SPY options, see Mishra and Daigler (2014).

	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 3: OPRA average daily statistics from January through August 2015. MPS refers to messages per second, while QPC, QPT, QPV and QPN are the number of quote messages divided by option classes, trade messages, traded number of contracts, and notional trading volume in dollar terms, respectively.

Our dataset spans the first eight months of 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 addition, 335 underlyings have a non-standard deliverable, resulting in multiple listed option symbols for the same underlying entity.[15] On average, 160 contracts are listed per option symbol on a daily basis but, for the most liquid underlyings, there may be up to 4,000 different contract variations. In terms of tenor, 45% of the contracts may be classified as standard (equity, ETF or index options), while weekly (20%), quarterly (25%) and LEAPS (Long Term Equity Anticipation Securities, 10%) options are listed in sizable proportions as well. In contrast, Mini Options rarely appear and are only 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 trades, as the trading volume exceeded 2.76 billion contracts with a total notional value of $7.95 billion. Table 3 summarizes the average daily quote and trade activity for the three main asset classes, with the most active constituents within each class being AAPL, SPY and SPX. These underlyings have multiple option classes, but for illustrative purposes we only report statistics for the standard categories.[17]

To provide a representative overview of the OPRA data, while keeping it reasonably succinct, we select a small subset of the underlyings. This sample of twelve stocks and five ETFs from different sectors mirrors the general heterogeneity in terms of the (average) number of records (see Table 4). We focus on equities and ETFs to study the cross-exchange patterns, which are absent for index options. Furthermore, we exclude option classes such as Mini and Jumbo options, or corporate-action adjusted ones.[18]

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

[16] There is trading in some non-standard S&P 500 index contracts, such as *binary* (BSZ) and *range* (SRO) options. Another *binary* option exists for the S&P 500 Volatility Index (BVZ). These contracts are only listed and traded on the CBOE.

[17] For example, the mini options AAPL7 and SPY7 were excluded.

[18] However, by including AAPL and SPY, two of the most liquid assets among the 4,562 option class symbols, we still cover more than 10% of the entire OPRA quote data, see Table 3.

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 4: Sample of underlying assets for the descriptive analysis.

4.2 Potentially Irrelevant or Faulty Observations

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

4.2.1 Classification of Special Records

We identify six categories of OPRA records that, depending on the context, may be suppressed. We characterize these special records in Table 5, 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 F1 category consists of all observations recorded before the start of the regular 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, such records are often noisy and may induce misalignments in terms of timing, if observations from multiple exchange markets are considered jointly.

The F2 category includes entries that likely involve data error. In particular, we identify trade records with a zero transaction price or zero size. Likewise, we filter quotes associated with negative spreads and entries with zero offer price or size. Note that, in contrast to the stock market cleaning algorithm from Barndorff-Nielsen et al. (2009), we do not eliminate quotes with zero bid prices (and positive offer prices). This occurs regularly for deep out-of-the-money options and is typically not associated with an error.

The F3 category represents trades and quotes with specific conditions, indicative of potentially irregular features of the given entry. This occurs for records of cancelled trades and non-firm quotes.

The F4 group contains quote records of minimum size. Such quotes might be less informative about the latent equilibrium option price. The presumption is that market makers experiment, when uncertain

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 median/mean/volume-weighted 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 median/mean/volume-weighted bid and offer prices and the total sizes. For lower frequencies, the observations with the latest millisecond time stamp can be used.

Table 5: Classification of special trade and quote OPRA records according to six groups.

about the fair price, by posting minimally-sized quotes at a distinct price level – simply to uncover latent demand with minimal exposure. Alternatively, due to the obligation to continuously maintain quotes throughout the trading day (see Table A.2 in Web-Appendix A), market makers may at times post minimum sized quotes to satisfy requirements without providing genuine liquidity. Thus, we flag quotes when the size is less than or equal to a single contract at both the bid and offer side.[19]

We caution that, since only the top bid and ask quotes are available in (new) OPRA, we cannot always identify whether a quote is of minimal size. For example, whenever a bid or ask quote changes, we observe a different set of contracts available at the new level. Since this figure is unobserved prior to the quote shift, it is impossible to determine the exact size of the new quote. As a result, we allocate only those minimally sized quotes, that can be identified with certainty, to category F4.

The F5 group contains outliers. For trades, we focus on abnormally high or low transaction prices relative to the current spread. For quotes, we identify entries with excessive spreads, extraordinarily high or low mid-quotes, and entries for which the ratio between ask and bid is unusually high. Our detection rules are similar to those of Barndorff-Nielsen et al. (2009) for trades and quotes. Such outliers may introduce unwarranted irregularities into the data structure, complicating the analysis. In particular, outliers may reflect data errors or market microstructure artifacts, such as stub quotes.

Records with identical time-stamps are collected in category F6. Often, multiple quotes or transactions are recorded at the same millisecond. When treating the observed option prices as a time series, it is convenient to replace such entries with a single record. This can be done in several ways.

First, OPRA captures the exact execution time for each record, so the final entry within a given millisecond may be used as the single price observation in lieu of the multiple records with the same time stamp. Obviously, this approach has drawbacks, as it discards the information contained in the other records and, in particular, fails to take advantage of the possibility of constructing a more robust observation from the available records. An alternative is to develop a procedure that aggregates the "simultaneous" records within a given millisecond by computing the median, mean or volume-weighted average of the trade/quote prices and summing up individual trade/quote sizes. Specifically, for time series analysis, where the frequency is considerably lower than a millisecond, e.g., a second or minute, it is sensible to consider an approach that mitigates the distortions induced by potential outliers or other microstructure noise effects associated with the reliance on a single end-of-interval observation.

We note that the frequency of records in group F2 and, to a lesser extent, F5 might provide a signal regarding the overall data quality, as such observations are more likely to be associated with data errors.

Finally, we stress that the filters are intended solely as a tool for option selection in the context of a given research objective. They can be ignored or imposed independently of each other. Moreover, they apply separately to the trade and quote records from a given exchange. If a specific study needs to guard against data errors, deal with multiple records at a single timestamp, or remove noisy and outlying observations, the filters in Table 5 provide a guide for identifying the relevant records in the dataset.[20]

[19] We do not remove quotes, which feature minimal size at only one side. Such scenarios often materialize when one side of the market dominates the quoting activity. Specifically, for deep out-of-the-money options ask orders tend to dominate, while, conversely, bid orders prevail for in-the-money options.

[20] Studies focusing on specific topics will often apply additional filters. For example, Christoffersen et al. (2018) impose a positive daily volume requirement and check for violations of minimal tick size rules implied by the option quotes.

4.2.2 Summary Statistics of Special Records

We inspect trade and quote records from our selected raw OPRA option sample throughout all trading days in January 2015 for contracts expiring on February 20, 2015. This sample consists of 2.77 billion quote and 730,000 trade records. Tables A.3 and A.4 in Web-Appendix A provide detailed aggregated statistics on the presence of records in the special categories identified above across underlyings and exchange markets, respectively. Figures 2 and 3 summarize key aspects of the findings. They report the average fraction of daily records belonging to one of the categories F1-F6 (white bars) and the fraction belonging to multiple categories (colored bars). Figure 2 depicts the fraction of special option records for each underlying asset. Figure 3 displays the fraction of special option records for each exchange venue.

Quote records are far more numerous than trade records. For example, in January 2015, SPY options expiring on February 20 feature, on average, more than 66 million daily quotes across the exchanges[21] compared to less than 16,500 transactions. The average 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.

Depending on the underlying asset, 5% to 25% of the trade records have non-unique millisecond time stamps (category F6), which may reflect the execution of larger aggressive orders hitting several smaller limit orders in the book simultaneously. In fact, this group constitutes more than 99% of the special trade records. Importantly, the categories F1, F2, F3 and F5 are largely absent for trade data. This suggests that the trade records are remarkably clean and free of extreme outliers. As seen from Figure 3, trades with identical timestamps are especially prevalent on the NYSE Amex (exchange symbol A), BOX (B) and MIAX Options Exchanges (M), reaching 30-35% of all trade records on these platforms. In contrast, for Nasdaq BX (T) and C2 (W), the fraction of such trade records is less than 1.5%.

In total, between 10% and 30% of the quote records for a given option class are deemed special and fall into one of six categories. Quote records outside regular trading hours (F1) or appearing erroneous (F2) are rare (less than 0.1%). Non-firm quotes (F3) are also infrequent, never exceeding 0.5% of the total quote records. Minimal two-sized quotes (F4) are common on all exchanges, exceeding 15% of all quotes for the BOX exchange (B). Quote records representing potential outliers (F5) amount for up to 7.5% of the total for certain underlyings, and are about equally common across all markets. Figure 2 shows that the fraction of quotes with identical millisecond stamps (F6) is substantial, varying between 3% to 12% of all entries across underlyings. For example, for Apple options, such quotes amount to almost 2 million entries per day. Although group F6 contributes substantially to the amount of special records, its relative contribution is less important than for trades. Thus, we conclude that OPRA quote records are mostly void of obvious data errors and invalid observations, but contain a non-trivial amount of potential outliers, which may require some attention, depending on the research objective.

4.3 Intraday Deviations from Put-Call Parity

Identifying records violating no-arbitrage principles is important for some applications, as it suggests the presence of frictions, market failures, faulty data entries, or non-synchronous recording of option and underlying asset prices. Consequently, the frequency of such violations at a given venue may be indicative

[21] It implies an average of about 240,000 quotes for the 282 distinct contract specifications.

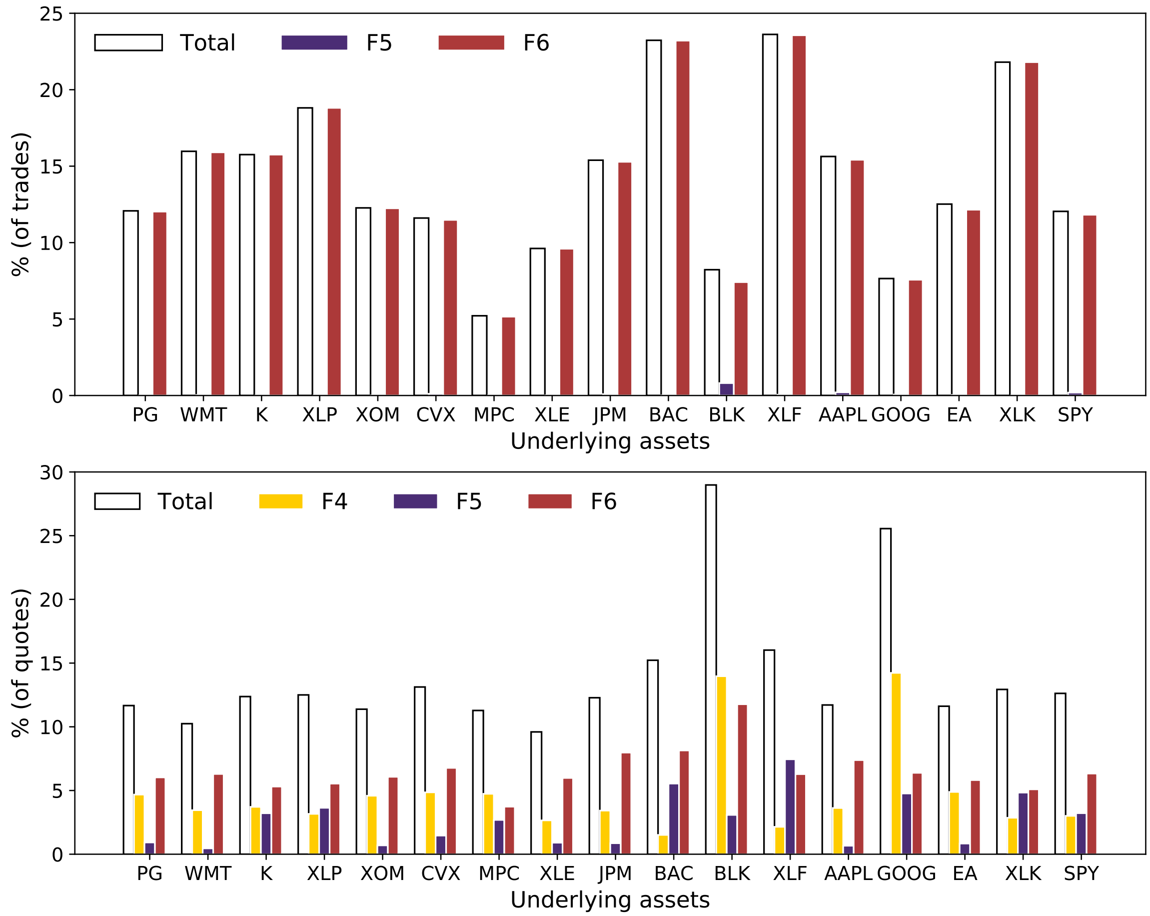


Figure 2: Average daily percentage of special option trade (top panel) and quote (bottom panel) records across selected underlying assets. White bars correspond to the total fraction of special records in the raw data (all categories, F1-F6). Yellow bars (available for quotes only) correspond to the fraction of records with minimal quote 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 cover data for all options traded in January 2015, expiring on February 20, 2015, observed across all available exchange markets.

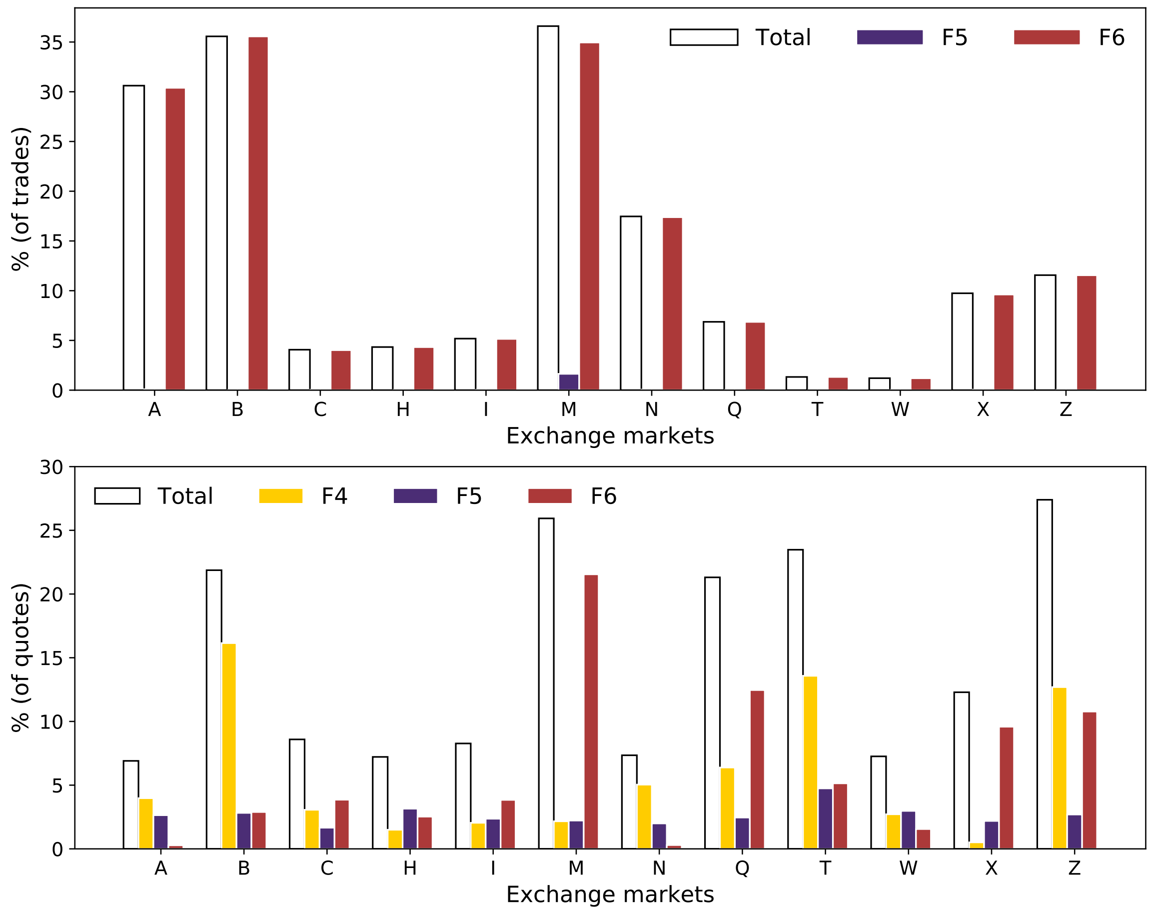


Figure 3: Average daily percentage of special option trade (top panel) and quote (bottom panel) records across 12 option exchange markets. White bars correspond to the total fraction of special records in the raw data (all categories, F1-F6). Yellow bars (available for quotes only) correspond to the 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.

of the relative exchange efficiency and the reliability of the associated option price and quote record.

We hasten to add that minor deviations from the put-call parity may not offer actual arbitrage opportunities. Whether the violation is economically relevant hinges on the proxy for the risk-free interest rate, representing a cost of capital for the option trader, and certain shadow costs associated, e.g., with margin requirements and inventory control.[22] As such, one may think in terms of "apparent" violations instead. Nonetheless, near arbitrage violations may also be informative, as they reflect scenarios with high call option quotes relative to the corresponding put options. In particular, the measures capturing the put-call parity deviations potentially can be relevant for asset pricing more broadly. For example, Cremers and Weinbaum (2010) find that such end-of-day metrics are indicative of price pressures, which provide significant predictability for future asset returns.

Theory predicts that the price of European options must be convex and monotone functions – increasing for put and decreasing for calls – of the strike price, see, e.g., Breeden and Litzenberger (1978). For brevity, we focus on one specific relation – the put-call parity – which establishes a no-arbitrage pricing relationship between put and call options at a given strike jointly with the price of the underlying asset.[23]

For American options, put-call parity cannot be represented as an equality. Instead, certain inequalities apply, including the following,[24]

$$C_t + Ke^{-r_f \tau} \;\leq\; P_t + S_t,$$

where C_t and P_t are call and put option prices at time t, respectively, K is the strike price, S_t is the underlying asset price, r_f denotes the risk-free rate, and τ is the tenor of the option pair. Consequently, put-call parity is violated if,

$$C_t^B \;>\; P_t^A + S_t^A - Ke^{-r_f \tau},$$

where the superscripts A and B refer to ask and bid prices, respectively.

For all underlying assets in our sample, we identify violations of this inequality using intraday option quote records separately for each of the 12 exchanges. We restrict our attention to data from 20 trading days in January 2015 and consider only options expiring on February 20, 2015. Consequently, the tenor ranges from seven to three weeks. For each trading day, underlying security and strike price, we inspect all put-call quote pairs, checking the put-call inequality record-by-record.

Table A.5 in Web-Appendix A presents aggregate results from this analysis, spanning all 17 underlying assets for each of 12 option venues that were active at the time. The results reveal strong heterogeneity in the (apparent) violations across options for different underlyings. Quotes for JPM, BAC and SPY options display the highest rate of violations, whereas we detect none at all for XOM, CVX and MPC in January 2015. The variation of put-call parity violations across exchanges is much less pronounced, reinforcing the point that such violations are linked closely to features of options for specific underlying asset.

[22]For example, when using the 3-month Libor instead of the Treasury rate in our analysis, we detect significantly lower numbers of such violations.

[23]The filters described in the previous section might also identify records that violate general no-arbitrage principles. To identify all records violating a given restriction, prices over a cross-section of options contracts (possibly along with the underlying prices) may need to be analyzed jointly. This type of comprehensive analysis necessitates additional assumptions for the construction of option cross-sections from non-synchronous intraday prices and specific criteria for identifying records that violate the specific arbitrage condition in question. In general, such studies differ substantively from the analysis of special records in the previous section, based only on the data from a single contract.

[24]There is a closely related upper bound on the put plus asset price. However, given the illustrative nature of this exercise, we ignore that constraint, which involves data on the dividend yield as well.

We illustrate how the (apparent) violations vary over time and depend on both moneyness and exchange venue by providing a more granular visualization in Figure 4. Each small square reflects a particular trading day and strike price in January 2015. The color is red, if put-call parity was violated for a combined period exceeding one minute; otherwise, the color is blue. To conserve space, we display results only for the BAC option quotes recorded on two exchanges, NYSE Arca and CBOE. This choice is useful in exemplifying the type of economic events and exchange-specific features that can be explored through high-frequency options data. To put the results in context, we note that the financial sector performed poorly over the first half of the month, with the BAC stock displaying heightened volatility and suffering a cumulative loss of approximately 14% over that period.

Several patterns stand out in Figure 4. First, there is a clustering of the violations in the first half of the month, when volatility was elevated. Second, they center on the ATM strike price, which was close to \$18 at the start of the month and between \$14 and \$16 over the last two weeks. Third, they are asymmetric across the strikes, as there are more violations for lower strikes. Fourth, they are almost absent for the deep ITM and deep OTM contract pairs. Fifth, for BAC during this month, the violations are much more prevalent on NYSE Arca than CBOE.

For a different, but related perspective, Figure 5 displays the frequency with which we observe a violation of put-call parity for any one of the quoted BAC or JPM options within a given 10-minute interval for a specific exchange in January 2015 with February 20 expiry. Here, two observations are striking. First, the violations are concentrated on a few venues, NYSE Arca, Nasdaq Options Market (NOM), and BATS plus, in the case of BAC, also ISE Gemini. Nonetheless, violations are observed occasionally for almost every venue. Second, the violations seem to be less frequent around the opening of the trading day, when the underlying assets typically display high volatility and wide bid-ask spreads.

The above observations raise numerous research questions that fall outside the scope of the current paper. One, is the volatility of the underlying a primary driver in the put-call parity violations? Two, does the direction of the violations possess short-term predictive power for the asset return? Three, do effective bid-ask spread measures (using trade prices relative to mid-quotes) also display arbitrage violations? Four, do the violations cluster around the ATM strike because the posted spreads are particularly narrow, in an economically meaningful sense, for this (liquid) part of the strike range? Five, why do we observe large discrepancies in violations across exchange venues? Six, does the option bid-ask spread display strong intraday patterns, possibly explaining why there may be fewer violations during morning trading? In Section 5, we present some findings related to the three latter points, so, at that stage, we shed a bit more light on a few of these issues, even if we do not provide comprehensive evidence.

In summary, instances of put-call parity violation are observed across all exchanges, they do at times cluster across consecutive trading days, they are more frequent on certain exchanges, they are more prevalent for some underlyings at a given point in time, and they vary in intensity across the trading day. Although such observations often are filtered out, they may potentially bring forth new and interesting market microstructure issues, and even broader asset pricing questions. However, we also recall the opposite argument: the violations may be economically inconsequential, as other (shadow) costs may render exploitation unprofitable, or they arise solely because the cost of capital is understated. Irrespective, a thorough analysis, expanding on our approach above, requires a comprehensive dataset like the bulk OPRA package, as one needs to monitor the top-level quotes across all active exchanges.

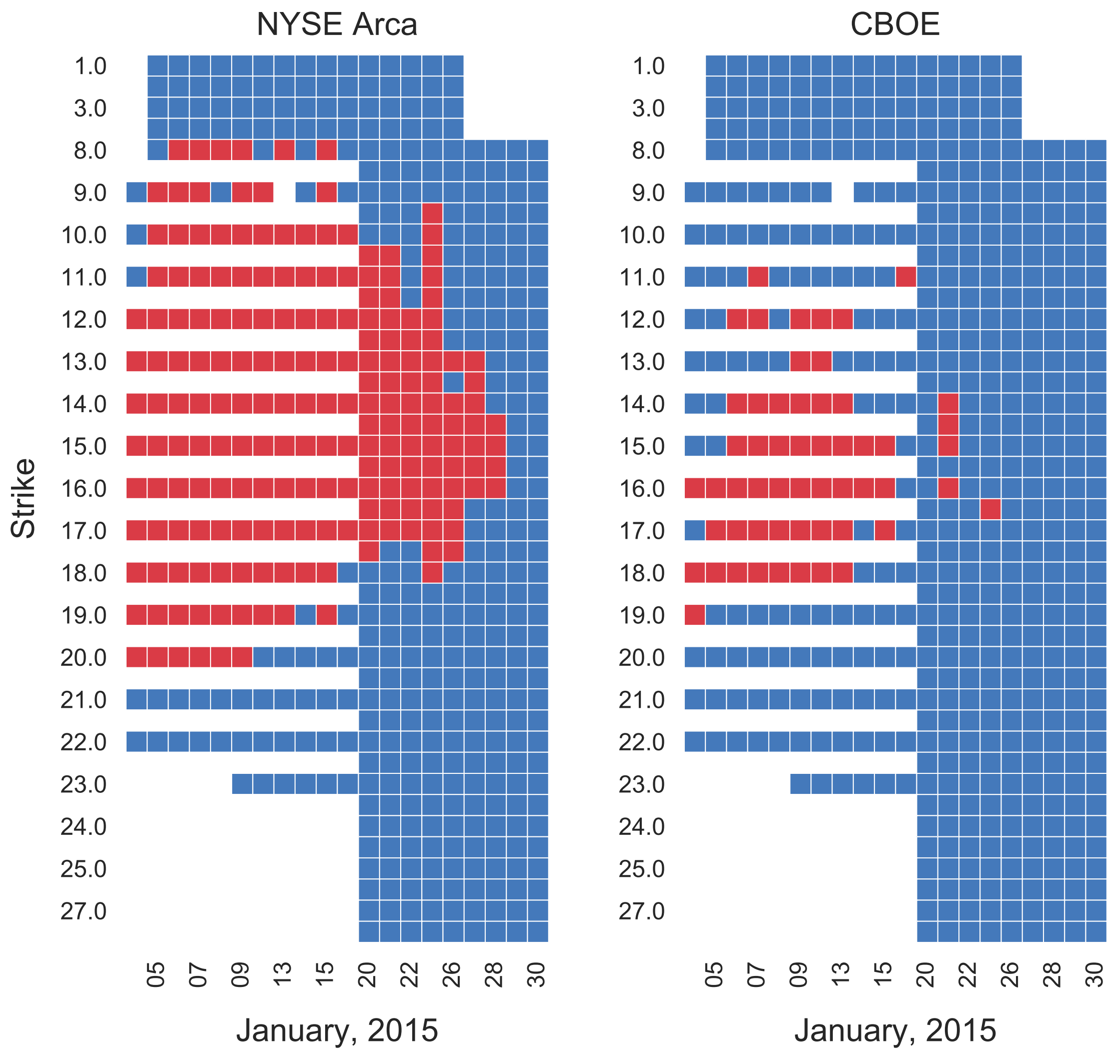


Figure 4: An illustration of the put-call parity violations for BAC option quotes posted at NYSE Arca (left panel) and CBOE (right panel) in January 2015. A square represents BAC options expiring on February 20, 2015, if the corresponding strike price (vertical axis) were available for trade on the specified date (horizontal axis). Red color implies that the put-call parity has been violated for more than one minute for a given strike price (in dollars) on a given date. Otherwise, the blue color is used.

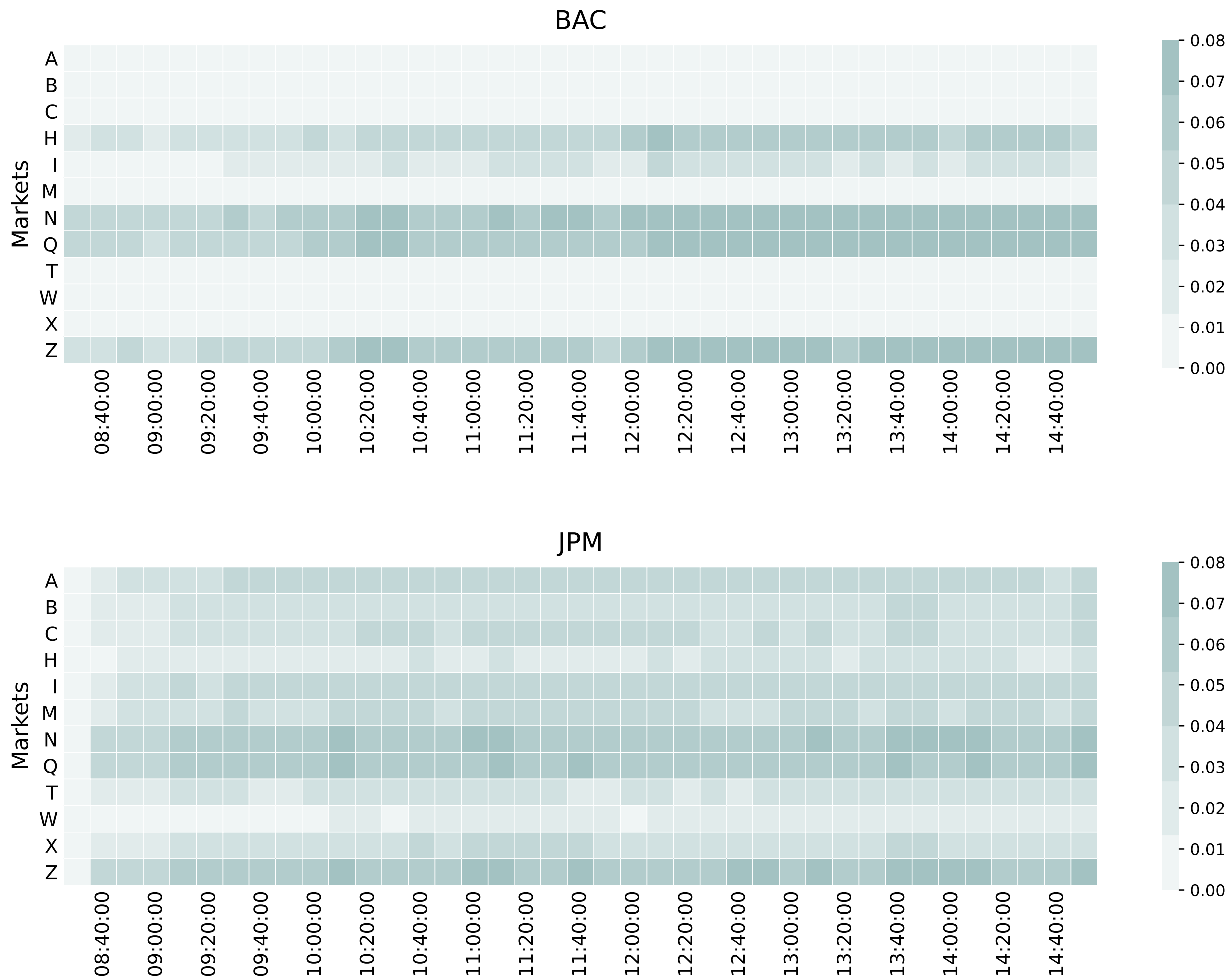


Figure 5: Intraday frequency of put-call parity violations for BAC and JPM options. The figure displays the frequency of put-call parity violations for BAC and JPM option quotes across all exchange venues in January 2015, for options which expire on February 20th, 2015. The violations are compiled for 10-minute intervals across the trading day. The color coding reflects the frequency of observed violations within the given interval and exchange across the full month.

5 An Explorative Analysis of OPRA Trades and Quotes

5.1 Basic Trade and Quote Statistics

Table 6 reports the daily number of trades and quotes for option contracts written on the 17 underlying assets in January 2015, before and after applying all the filtering algorithms in Section 4.2. On average, 14.53% of the records are eliminated due to this (aggressive) data cleaning. The number of option quotes and trades varies greatly across underlyings, reflecting largely the liquidity of the latter. The most actively quoted (and traded) option contracts are those written on the ETF SPY and Apple, with on average, respectively, more than 628 million and 279 million clean quotes daily. On the most active trading day, the SPY quotes exceed 872 million after filtering and close to 1 billion prior to cleaning.

The average order-to-trade ratio ranges between 3,036.9 (for BAC) and 22,433.6 (for EA), implying a dramatic excess of quotes relative to trades in option markets. This is an order of magnitude larger than for equity markets, where order-to-trade ratios rarely exceed 100, even for algorithmic and high-frequency traders, see, e.g., Hagströmer and Nordén (2013), and Brogaard et al. (2015).

5.2 Trade and Quote Records by Tenor and Moneyness

The usual expiration day for a standard option contract – following conventions adopted when exchange-based option markets were initially established – is the third Friday of the month. Given the diverse economic incentives for option trading, the subsequent successful introduction of quarterly and yearly option contracts for most underlyings is not surprising. In addition, over the past decade, weekly options, or weeklies, have increased dramatically in importance. For instance, the trading volume of S&P 500 weeklies (SPXW) grew from about 12% in 2010 to 25% in 2014.[25]

According to current 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. For the first eight months of 2015, we classify about 45% of the contracts in our sample as regular (i.e., equity, ETF, or index options), but weekly (20%), quarterly (25%) and LEAPS (10%) options are also listed in considerable proportions. To convey how quotes and volume are distributed across maturities, we categorize options as ultra short-, short-, medium-, or long-term, depending on tenor,

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

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

[25] In August 2017, weekly options on the S&P 500 index passed 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.

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

Table 6: Descriptive statistics for daily OPRA data records in January 2015. “Trades-only” contracts are not considered for the calculation.

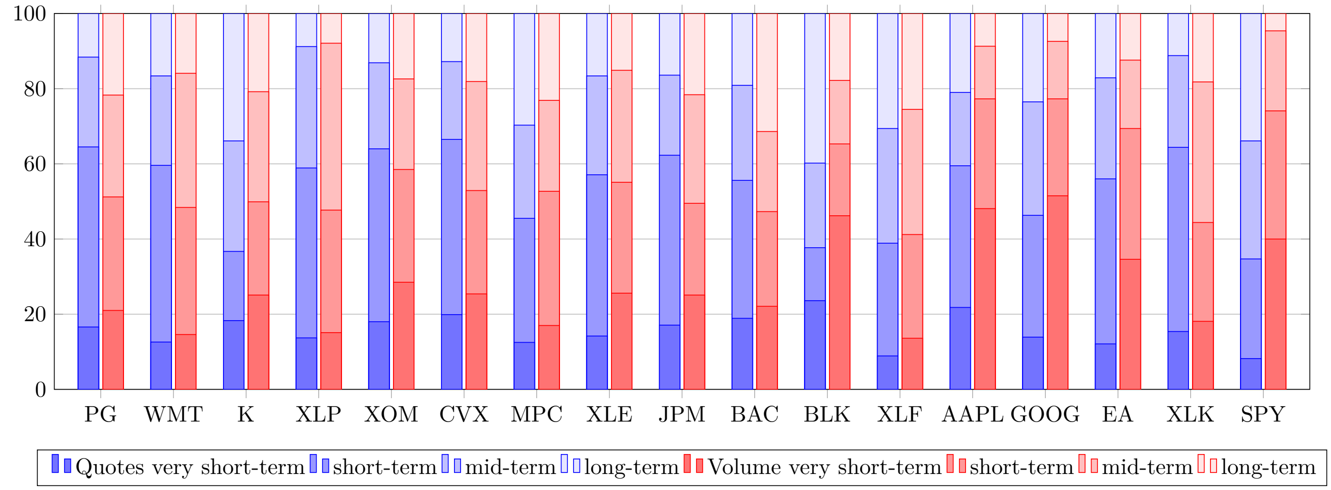


Figure 6: Average daily proportions of quote and trading volume by tenor (in percent). The two stacked bar charts represent the proportion of quotes (left) and volume of traded option contracts (right) for each underlying between January and August 2015. The bars are partitioned in segments by *time-to-expiry*, as ultra short-term (bottom), short-term, medium-term and long-term (top).

Figure 6 displays the average daily proportion of quotes and trading volume by tenor. For most option classes, the trading volume is relatively more concentrated among the ultra short-term contracts compared to the corresponding quoting activity. For example, listed options on GOOG can be traded for 12 separate expiration dates (14 for AAPL and 24 for SPY), but contracts for the nearest maturity account for 51.5% of all trades (48.1% for AAPL and 40.0% for SPY) relative to about 15% of the quotes (22% for AAPL and 8% for SPY). Hence, not only is the weekly maturity profile increasing in popularity and trading activity, but less liquid option classes without weeklies, e.g., BLK, MPC, and K with 46.2%, 17.0%, and 25.1%, respectively, are also traded more intensively in the week prior to expiration.

We now turn towards the option trade and quote activity as a function of moneyness. Figure 7 tabulates the percentage of option contracts traded across different degrees of moneyness between January 2 and February 18, with expiration February 20, 2015. Across the securities, the vast majority of the option trading is in close-to-the-money options. For the individual stocks, the trading volume for options with positive moneyness tends to exceed that of options with negative moneyness, when the options are within two (Black-Scholes) "sigmas" of the ATM strike. However, the reverse is true, when considering moneyness beyond two "sigmas", where the options with "downside" strikes command the highest volume, almost uniformly across all the underlyings in our sample, and often by a wide margin. The equity indices are clearly different, as we observe a very strong asymmetry towards excess trading in options with below ATM strikes, irrespective of the degree of moneyness. This is, of course, consistent with a strong motive for downside hedging by fund managers who hold net long positions in equities.

Figure C.3 in Web-Appendix C confirms that these findings apply qualitatively for the total option trading volume as well, while the quote activity is much more balanced for "up" and "down" strikes, as seen in Figure C.4, even though a slight tilt towards the downside strikes remains.

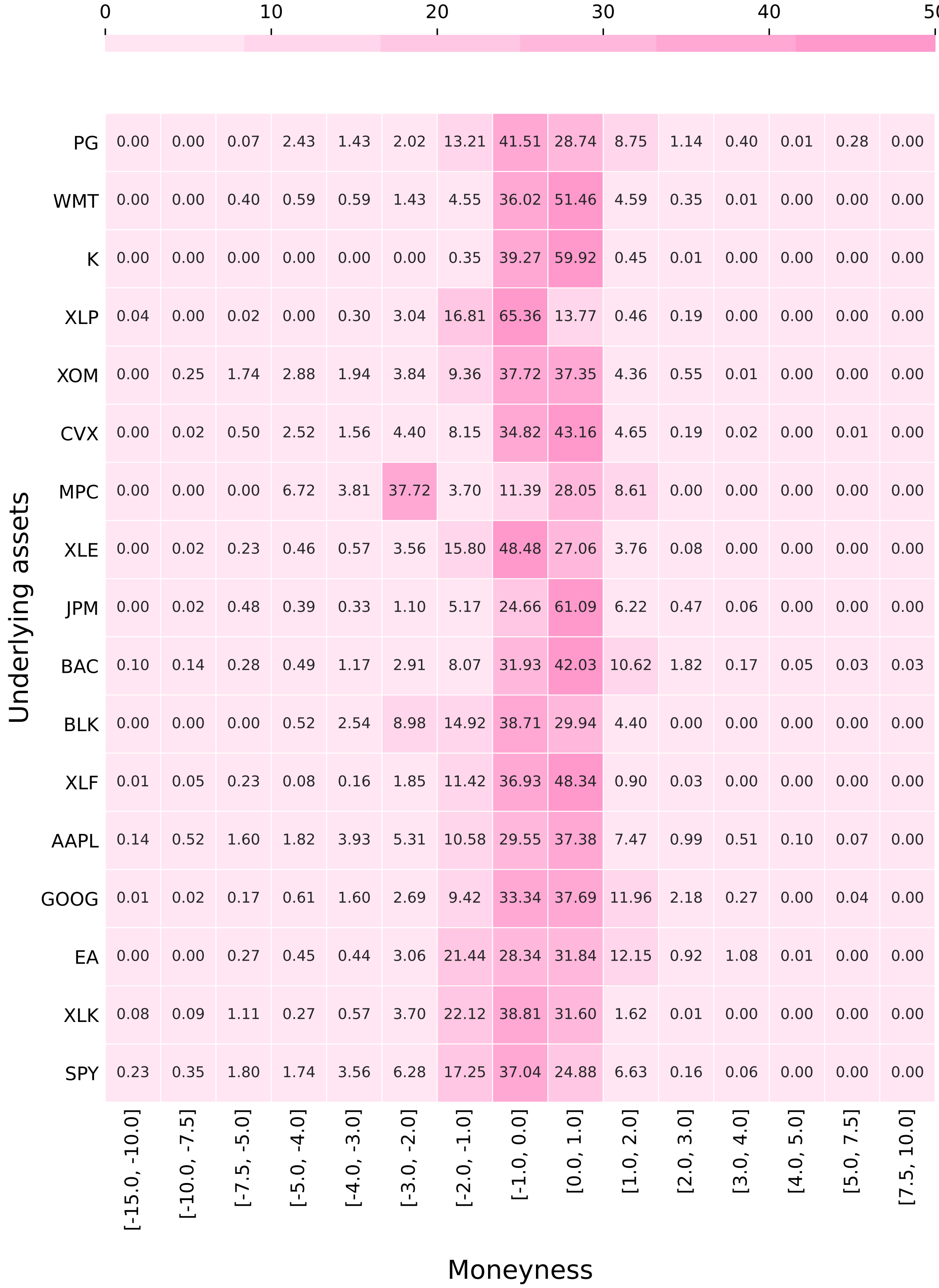


Figure 7: Proportions of contracts traded (in %) by moneyness for the selected option classes. The results are obtained using the OPRA quote records for put and call option contracts traded between January 2nd and February 18th, 2015, which expire on February 20th, 2015. The moneyness is defined as $m = log(K/F_t)/(\sigma_t\sqrt{\tau})$.

5.3 Activity across Exchange Platforms

A distinct feature of the OPRA data is the availability of trade and quote records disseminated by all of the individual U.S. option exchanges. Figure 8 shows the relative proportion of quoting and trading volume across exchange holding groups and individual exchanges. In general, the pattern for the trading activity across exchanges is in line with the evidence from Figure 1.[26]

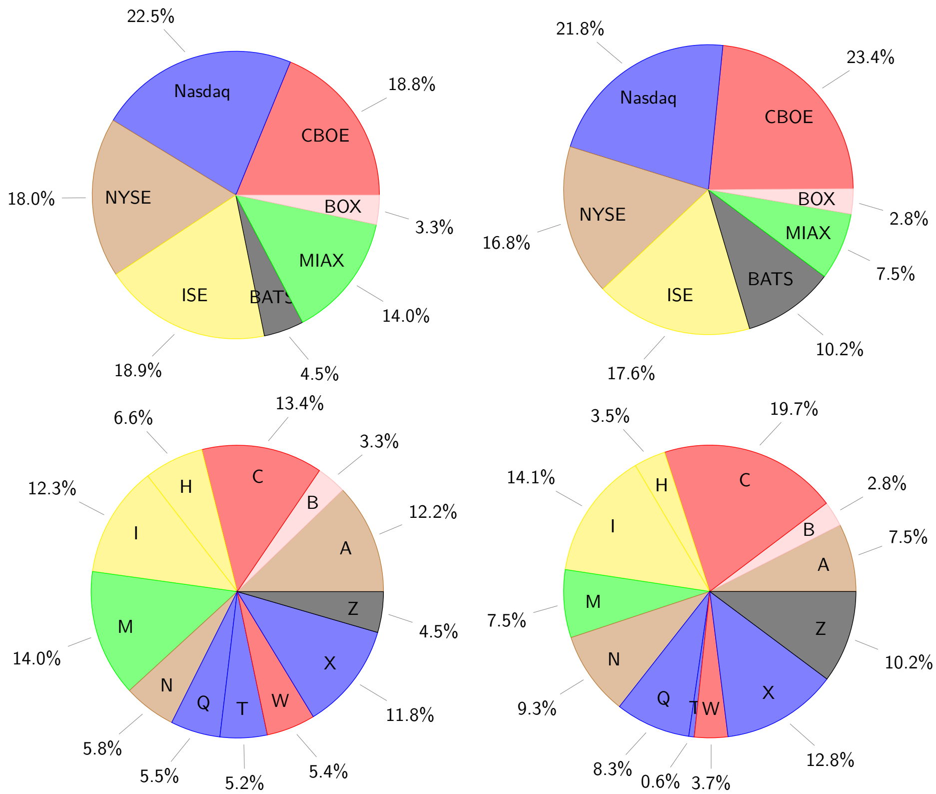


Figure 8: Quotation and trading volume market shares per exchange.Pie charts on the left-hand side reflect the quotation distribution among holding groups (top) and individual exchanges (bottom) for our sample of underlying assets. Analogously, the trading volumes are indicated on the right-hand side. The identity of the individual venues associated with letter codes can be gleaned from the header in Table 7.

Within each holding group, the longest-serving venues account for the majority of the trading, as seen from the bottom right chart in Figure 8. The same is generally true for quotes, which typically are more than twice as frequent for these exchanges than for the secondary venues within each group. However, NYSE is an exception, as the AMEX (A) exchange was launched three years after ARCA (N), but has double the number of quotation messages by 2015. Both markets offer floor and complex trading, but they differ in their pricing and allocation structures, as documented in Table A.1 of Web-Appendix A. AMEX utilizes the classical *customer-priority* model with a *pro-rata* allocation, which encourages deeper liquidity. In contrast, ARCA focuses on price efficiency, exploiting *maker-taker* and *price-time* procedures. In this case, the liquidity-inducing approach is more quote-rich, but generates a lower trading volume relative to the alternative. It is evident, however, that the pricing structure is not the only factor accounting for the

[26] This figure is based on market activity in January 2015, when ISE and BATS were still independent from NASDAQ and CBOE, respectively. Furthermore, the option exchanges MCRY, EDGX, MPRL and EMLD had not yet been launched.

heterogeneity in quotation and trading activity across markets. The MIAX exchange, which also follows the *maker-taker* model, had the highest number of quote updates in 2015 among all option marketplaces in the U.S. for this set of underlyings.

Given our relatively short sample and the high variability of market shares in options trading, we cannot identify a clear trend in exchange competitiveness. From Figure 9, the leading groups NASDAQ and CBOE started out with a market share of approximately 25% each, ahead of ISE and NYSE with about 15%. The remaining exchange holdings, BATS, MIAX and BOX, sported significantly lower daily volumes in 2015. By August, however, the four largest exchange holdings went head-to-head with market shares of around 20% each, suggesting a stronger degree of competition, but also a shift in trading interest towards the smaller exchanges. In addition, there was remarkable growth in the volume at BATS in August 2015, when it gained an additional 5% market share, possibly in response to a bout of market stress.[27] By posting limit orders in the pre-market session, offered solely by BATS, traders obtain time priority for the opening call auction and, if not executed at that stage, within the following regular trading session.

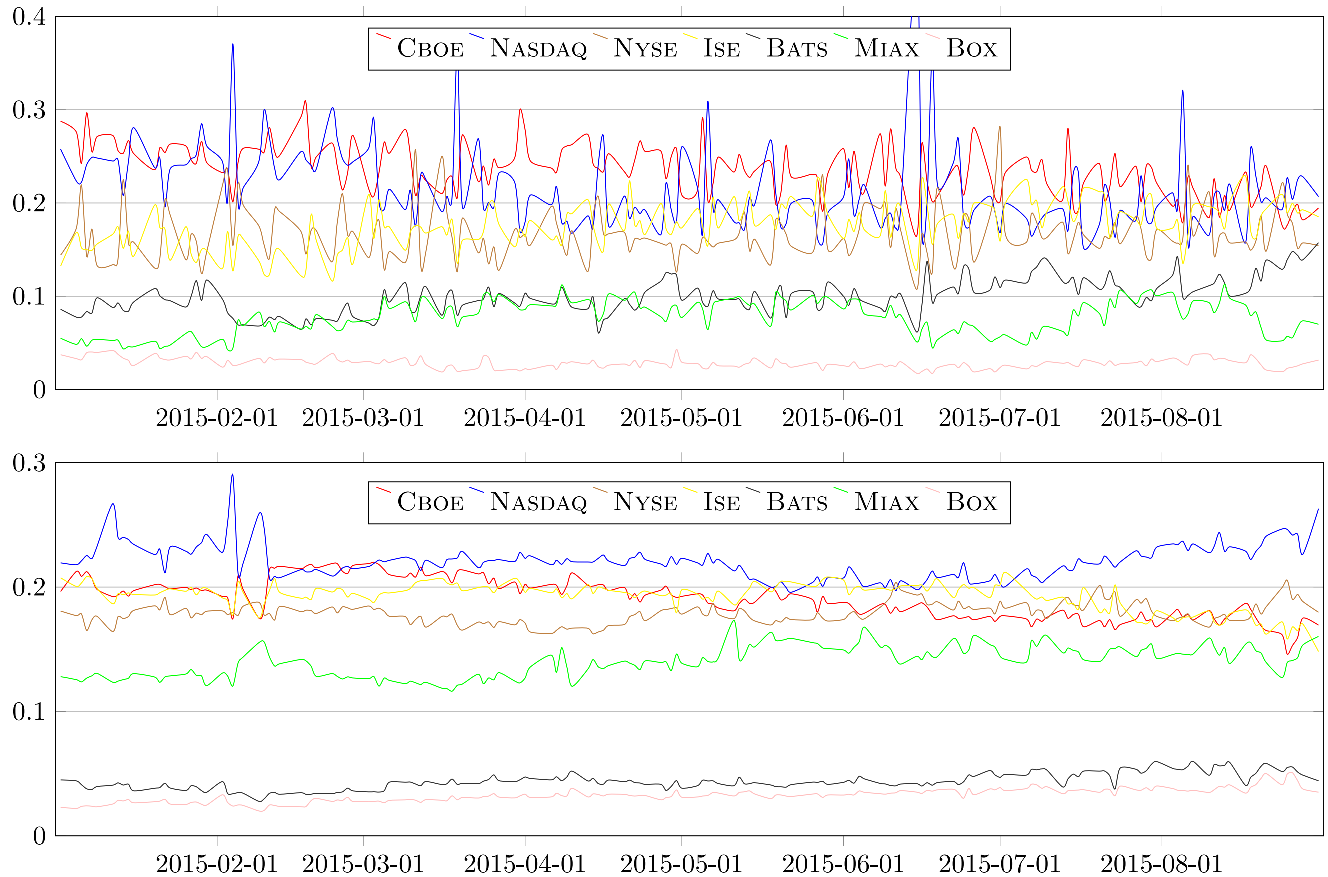


Figure 9: Market Shares for the daily contract trading volume (top plot) and quotation activity (bottom plot) per exchange holding group for our option sample.

In contrast, the relative number of quotations are fairly stable across exchanges. This is not unexpected. As the price on the underlying asset fluctuates, option prices change as well, necessitating

[27] Notable events include the Greek default on June 30, 2015, and a market crash in China, where the Shanghai Composite shed 38% between June 12 and August 24.

adjustments to a wide range of quotes. Hence, if the exchanges maintain a menu of options with similar characteristics across this period, the frequency of their quote updates is naturally aligned with the volatility of the underlying assets, irrespective of the actual transactions they facilitate. Consequently, the relative quotation shares may well remain stable, unless there is an underlying shift in the market structure. Such changes are often associated with an innovation on one or more venues in the type of marketed options, in the incentive structure for giving and taking liquidity, and in technology-driven changes in trading costs, along with affiliated shifts in trading strategies. In the absence of major developments in these areas, the quote frequency for a given type of option is likely relatively stable.

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	0.08	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 7: 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 7 and 8 provide the market shares of trading volume and quoting activity across the exchange platforms for the option contracts written on our sample of underlying securities. The most striking feature is the wide dispersion in trading across the venues, with the largest market share being the 22% obtained by MIAX for BAC options. Overall, the dominant venue is CBOE, but BATS and MIAX are not very far behind, followed by PHLX, NOM, ARCA and ISE. In fact, all exchanges display non-trivial participation, except for the marginal activity on OBO, C2 and, to some extent, GEMX. The quoting activity on the individual venues is correlated with the trading volume, but AMEX, ISE, and PHLX are relatively more active in this dimension, while the quoting activity on BATS, NOM, and ARCA is very subdued compared to their share of trading volume. Clearly, there is a strong heterogeneity in the mode of operation across exchanges, altering the relative quoting versus trading frequency substantially in some cases. As an aside, we note that the venue-specific trade and quote frequencies are not reliable indicators of the put-call parity violations identified in Section 4.3.

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 8: Shares of aggregate quotation activity per exchange for selected underlying securities. Highlighting: $x \geq 0.15$, $0.15 > x \geq 0.1$, and $0.1 > x \geq 0.05$.

The participation of the different markets in the NBBO is of substantial interest, as this figure may be suggestive of the relative price efficiency of the venues. For each exchange, underlying asset, and second-to-second stamp, we compare the BBO of a given venue against the NBBO. We only consider the core trading session, from 8:30 am until 3:00 pm CT, to avoid the results being trivially tilted in favor of Bats – the only exchange group offering a form of pre-trading session.[28]

Table 9 indicates that C2 and GEMX are the least competitive venues with respect to NBBO participation for most underlyings and, not surprisingly, these exchanges have less than half the quote updates of the market leaders. However, two other "small" exchanges, BATS and BOX, with even less quotation messages (Figure 8) exhibit significantly higher rates of NBBO participation. Nonetheless, the general impression from Table 9 is that the overall price quality for first-level quotations is remarkably similar across all twelve options exchanges. Undoubtedly, this is an important factor behind the highly dispersed option trading volumes, and it is consistent with the notion of a competitive national options market.

These tentative conclusions do not, however, clarify why we see large differences in the rate of apparent arbitrage violations across venues, as illustrated for BAC options quoted at CBOE and NYSE Arca in Section 4.3. From Tables 7-9, we find NYSE Arca to be an active venue for BAC option trading, but more marginal in its quoting intensity. Moreover, it is one of the most competitive venues in terms of matching the NBBO quotes for BAC options. Hence, while CBOE is more actively engaged in BAC option trading and quotation, NYSE Arca is, on average, providing equal or better top-level quotes. This suggests that, during turbulent periods, a slow quote update frequency may be an important contributor to the arbitrage opportunities identified in Section 4.3.[29]

[28] For example, the value '0.50' in Table 9 for CBOE and underlying SPY means that CBOE market makers match or improve on the best bid-offer quotes for SPY options across all other venues for 3.25 hours per day (i.e. 50% of the daily trading period) on average.

[29] A robust exploration of this conjecture requires an elaborate empirical analysis and is outside the scope of this paper.

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 9: 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. Highlights: green ≥ 0.6, $0.6 >$ light green ≥ 0.5 and yellow ≤ 0.4.

5.4 Quoted Spreads

We explore the size of the spreads in option quotes across different markets, stocks, tenor and moneyness. We focus on the relative spread, defined as,

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

where MQ_t is a mid-quote price at time t. To keep the analysis manageable, we restrict our attention to option contracts traded from January 2 to February 18 with expiry February 20, 2015. Thus, the option tenor ranges from a couple of days to seven weeks. We remove entries belonging to the filtering groups F1, F2, F3 and F5, so our results reflect only regular quotes and mitigate the impact of outliers.

Numerous factors may help rationalize the size and variability of the option bid-ask spread. We convey our main findings through a few illustrative figures, while deferring extensive tabulations to Web-Appendix A and additional illustrative evidence to Web-Appendix D.

First, for equity indices, it is known that put options tend to be more liquid and have lower spreads than call options, but corresponding stylized facts for individual equity options are less well established. Figure 10 displays the relative quoted spread differential between calls and puts at the identical degree of moneyness. The figure renders the asymmetry in the option spread across the full set of underlying assets transparent. The discrepancy is particularly dramatic for OTM spreads, with only positive, and often very large, entries appearing for moneyness in excess of unity. In contrast, for ITM options, the spreads are quite closely aligned and, if anything, they are slightly smaller for call options, as indicated by the negative entries on the left-hand side of the figure. Consequently, the evidence on spreads is qualitatively quite similar for the individual equity and equity index options.

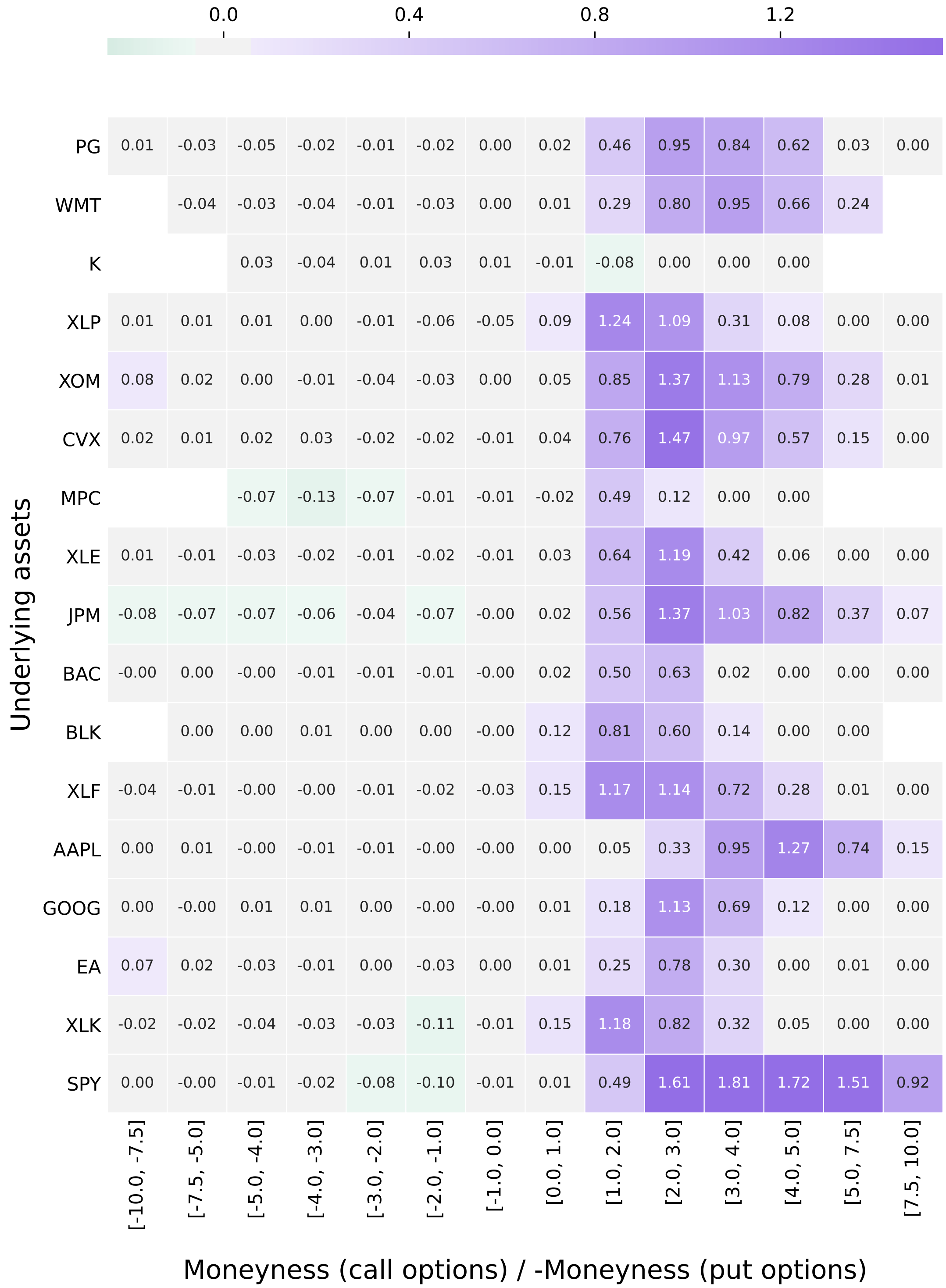


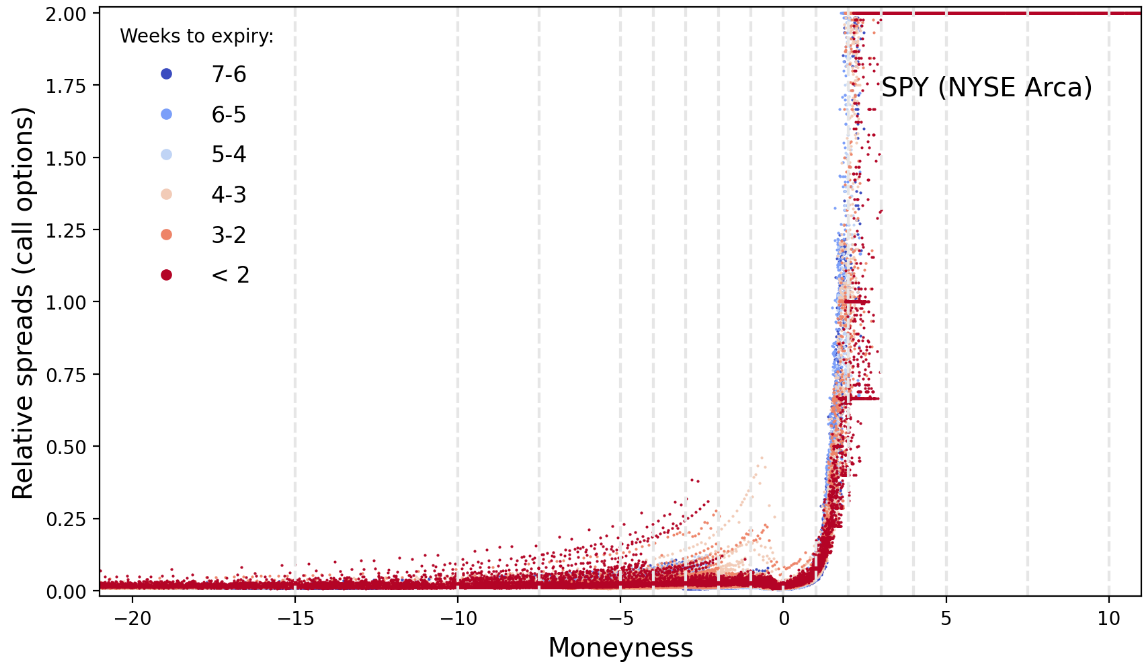
Figure 10: Difference in average relative spreads between call and put options across moneyness for each underlying asset in our sample. The results are based on OPRA quote records for put and call options between January 2nd and February 18th, 2015, that expire February 20th, 2015. Moneyness is defined as $m = log(K/F_t)/(\sigma_t \sqrt{\tau})$ with m multiplied by -1 for put options.

Next, Figure 11 displays the average intraday spreads for put options over three distinct tenor categories, covering the same period as above.[30] The left column depicts an "L-shaped" pattern in the spread over the trading day for three separate equity options. These are representative of the results for our individual stock options. In contrast, the option spreads for the two equity indices in the right column are approximately "U-shaped," whilst the Apple spread configuration in the middle panel looks like a mixture of the two patterns. Apple is also noteworthy by having the smallest relative option spread, by far, among all the underlying securities in our sample, followed by the SPY ETF. Interestingly, the energy ETF, XLE, has the largest uptick in relative spread towards the close of trading, while XOM, a large constituent of this energy ETF, shows no signs of an elevation at the end of the session. This divergence in trading costs mimics the discrepancies often observed in the volatility of the underlying equity indices and ETFs relative to individual stocks towards the market close. Finally, we note that the relative spread is almost monotonically increasing as time-to-maturity shrinks, which aligns well with the fact that option implied volatilities increase dramatically, as the tenor approaches zero.

[30]The corresponding figure for call options convey the identical qualitative results, see Web-Appendix D.

Figure 11: Relative spread measures (average over 15 min intra-daily intervals) computed from the OPRA quote records for put option contracts traded in January and February, 2015, which expire on February 20th, 2015. The reported results are obtained using only those local 15-minute intervals, where moneyness is in the range $-1 \leq m \leq 1$. The moneyness is defined as $m = log(K/F_t)/(\sigma_t \sqrt{\tau})$.

The actual relationship between tenor and spreads is a bit more involved than conveyed above. Figure 12 displays the relative spreads for a set of SPY options observed on NYSE Arca exchange in January 2015.



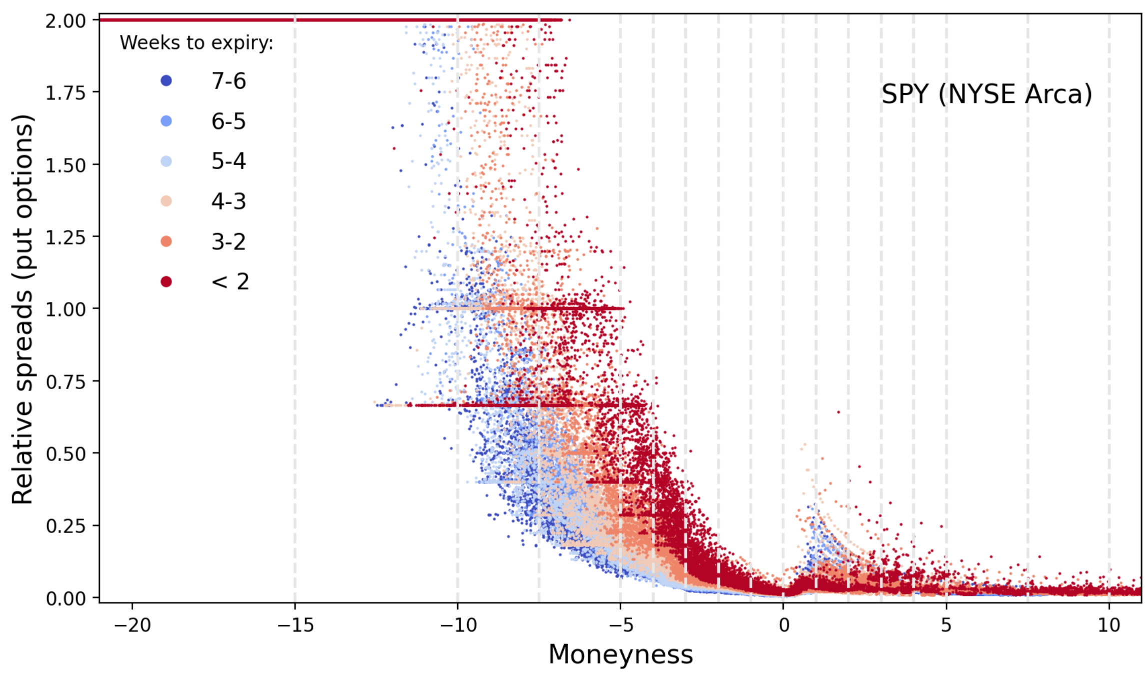


Figure 12: Relative spread measures (average within 15-minute intraday intervals observed between 9:00 and 10:00 CT) of SPY options across maturity groups computed from the OPRA quote records for call (upper plot) and put (bottom plot) option contracts traded on NYSE Arca exchange between January 2nd and February 18th, 2015, which expire on February 20th, 2015. The moneyness is defined as $m = \log(K/F_t)/(\sigma_t\sqrt{\tau})$.

Figure 12 exemplifies the strong asymmetry in spreads between puts and calls. In the top panel, OTM call options hit the maximum relative spread of 2 for moneyness 2-3. In the bottom panel, a large fraction of the OTM put options feature much lower relative spreads for moneyness well beyond -5. It is also evident that certain spread values dominate for the short-maturity options, visible as straight horizontal red lines for both OTM put and calls. This stems from identical posted bid and ask quotes for a string of options across adjacent strikes, which is indicative of clustering at certain ticks.

Further instances of non-monotonicity in spreads as a function of tenor may be gleaned from Figure 12, where moderately ITM put options with long maturities attain the largest relative spreads (the blue dots lie above the red for low positive m values in the bottom panel). This feature is corroborated more broadly for our option sample in Tables A.6-A.7 (Web-Appendix A), which capture the relation between moneyness and spreads along with Figures 7, C.3 and C.4 (Web-Appendix C), that relate moneyness to trade and quote activity. We conjecture this stems from opposing effects of liquidity and tenor on the spread.

In summary, the relative option trading costs, as represented by the quoted bid-ask spread, are highly heterogeneous, varying substantially from one underlying to another, even within the same sector, and also showing non-trivial deviations across exchanges. A second factor is the option tenor, since relative spreads tend to increase, as the time value of the options shrinks in line with maturity. Nonetheless, the most important determinant is the identity of the underlying and the associated liquidity of the options market. Specifically, the actively traded options written on SPY and AAPL have dramatically narrower spreads than the less liquid options, irrespective of the exchange market and controlling for option tenor. Moreover, as short-dated options often are quite liquid, we do observe a non-monotonic tenor-spread or volume-spread relation across a subset of the maturity spectrum.

6 Empirical Applications

This section explores whether a couple of commonly-used option-implied measures may be generated in a meaningful manner from high-frequency option prices exploiting standard techniques. If shortcomings become evident, it serves as motivation for future work on generating more robust intraday measurement procedures. It is beyond the scope of the present paper to pursue any such comprehensive remedies.

The first application involves second-by-second model-free estimation of the risk-neutral variance for an underlying security based on the intraday cross-section of option prices. The second focuses on estimation of the risk-neutral asset return distribution from high-frequency option prices.

6.1 Intraday Risk-Neutral Return Variation Measures

The expected future return volatility is a critical input to numerous financial and economic applications. As is well known, the cross-section of European-style option prices, covering the full range of strikes for a given expiration date, enables model-free computation of the expected return variation for the underlying security return under the risk-neutral probability measure, see, e.g., Carr and Madan (1999) and Britten-Jones and Neuberger (2000). This model-free implied-volatility (MFIV) measure contains both a predictive and a risk premium component, as it reflects the expected future variance as well as the risk pricing for a security with a payoff equal to the future *realized* return variation.

In practice, the MFIV can only be approximated, as we do not observe "true spot" option prices, but rather bid-ask quotes. Likewise, the set of strikes is discrete, finite, and does not cover the entire positive real axis. As such, there are legitimate questions about the reliability of high-frequency MFIV measures.

The typical option data used in the academic literature consists of end-of-day cross-sections of prices or quotes (e.g., OptionMetrics data). This precludes investigation of the implied variance dynamics in real time, the reaction of volatility expectations to specific events within the trading day, the intraday co-dynamics of implied variances across multiple assets, etc. Furthermore, the quality of a MFIV measure, based on a single cross-section of option prices, depends on the degree of noise or data error at the observation time. Intuitively, we obtain a more robust measure using multiple cross-sections from adjacent seconds, much in the spirit of pre-averaging for the measurement of the return variation from intraday log asset prices, see, e.g., Jacod et al. (2009).

For the broader U.S. market, we already have a popular high-frequency return variation measure. The VIX index disseminated by the CBOE, capturing the expected risk-neutral volatility of the S&P 500 index over a 30-day period, is computed on a continuous basis throughout the trading day and released at a 15-second frequency, providing a real-time benchmark volatility indicator. Unfortunately, the intraday VIX series is not a reliable real-time MFIV measure. Andersen et al. (2015a) document significant spurious outliers in the high-frequency VIX index, due largely to the random tail truncation of the OTM option prices. In addition, it is subject to a non-trivial delay of 15-45 seconds stemming from random variation in the processing and dissemination speed. As a consequence, direct use of the high-frequency VIX series can result in severely distorted inference, especially during periods of market stress, when accurate real-time measures, arguably, are most needed. A number of studies recognize the potential distortion arising from the tail truncation inherent in the VIX computation and provide alternative recipes that amend the tails in different ways using the observed cross-section of option quotes. However, these procedures are almost invariably implemented at the daily frequency, and not on an intraday basis.

The highly accurate time-stamps in the OPRA data allow for the construction of MFIV series for a large number of stocks and ETFs traded on the U.S. equity market across a range of time horizons and at almost arbitrarily high frequencies. Of course, it remains an open question whether lack of liquidity or other distinct market microstructure features will render such measures excessively noisy or biased.

For our empirical illustration, following Carr and Wu (2009), we rely on a log-linear extrapolation of option prices in the strike domain for the tails of the return distribution rather than using tail truncation, as applied in the official VIX computation, or a robust corridor-based measure, as suggested in Andersen and Bondarenko (2007). Web-Appendix E.1 provides a detailed description of our option portfolio design.

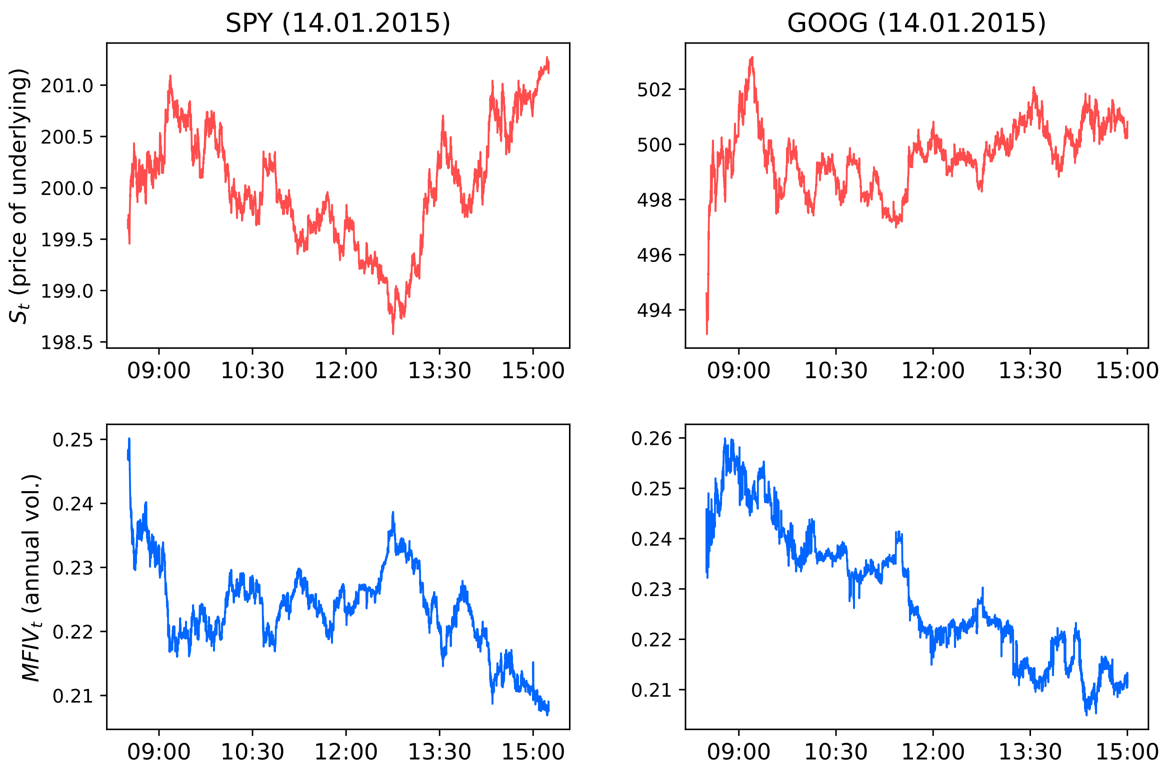


Figure 13: On the top plots, intraday prices of SPY (left side) and GOOG (right side) observed on January 14, 2015. On the bottom plots, intraday MFIV series constructed from SPY and GOOG options on January 14, 2015 (expiring January 17, 2015). Real-Time MFIV is calculated on a second-by-second basis and normalized to annual volatility units.

Our illustration focuses on MFIV's extracted at the one second frequency using intraday quotes for American-style SPY and GOOG options. The former provide an alternative way to construct a high-frequency S&P 500 volatility measure, while the GOOG options speak to the possibility of generating reliable intraday MFIV estimates for individual stocks. The choice of very liquid options provides us with a near best-case scenario to assess the properties of such measures at extremely high frequencies. We use options with only 3 days till expiration. This generates a return variation measure that is closely related to the spot volatility of the underlying series, given the very short tenor.[31] Moreover, it ensures that the price differential between American- and European-style options truly is negligible.

Figure 13 plots the intraday prices for SPY and GOOG (top panel) and corresponding second-by-second annualized MFIV estimates (bottom panel). Several features stand out. First, the MFIV exhibits an appreciable amount of variation within the trading day, allowing for direct identification of shifts in the value across 10-20 minute intervals. Nonetheless, the series are quite choppy, indicating a fair degree of measurement error, inducing negative serial correlation over short horizons. This suggests it may be worthwhile to construct more robust measures by smoothing suitably across adjacent second-by second MFIV measures. Second, for both SPY and GOOG, the extracted MFIV is very volatile immediately after the market open. This is consistent with the elevated return volatility and high bid-ask spreads for the underlying securities in the morning. It likely reflects (option) price discovery, when overnight

[31] There is a gap between spot (diffusive) volatility and the above option-based measure due to the risk-neutral jump variation, which can differ significantly from its statistical counterpart. This component of the option-based variation measure can be further removed in a nonparametric way using the approach of Todorov (2019).

information and newly arriving orders are absorbed into the market. Finally, we note that the leverage effect, defined as a negative return-volatility relation, is evident, but clearly more pronounced for the SPY index than GOOG. The realized correlation between MFIV and the underlying price, computed from 2-minute increments, equals -0.85 for SPY and -0.36 for GOOG. Existing studies of the high-frequency leverage effect from option data has focused exclusively on equity indices, see, e.g., Andersen et al. (2015a) and Kalnina and Xiu (2017), but Figure 13 points to the feasibility of extending such studies to individual equity series.

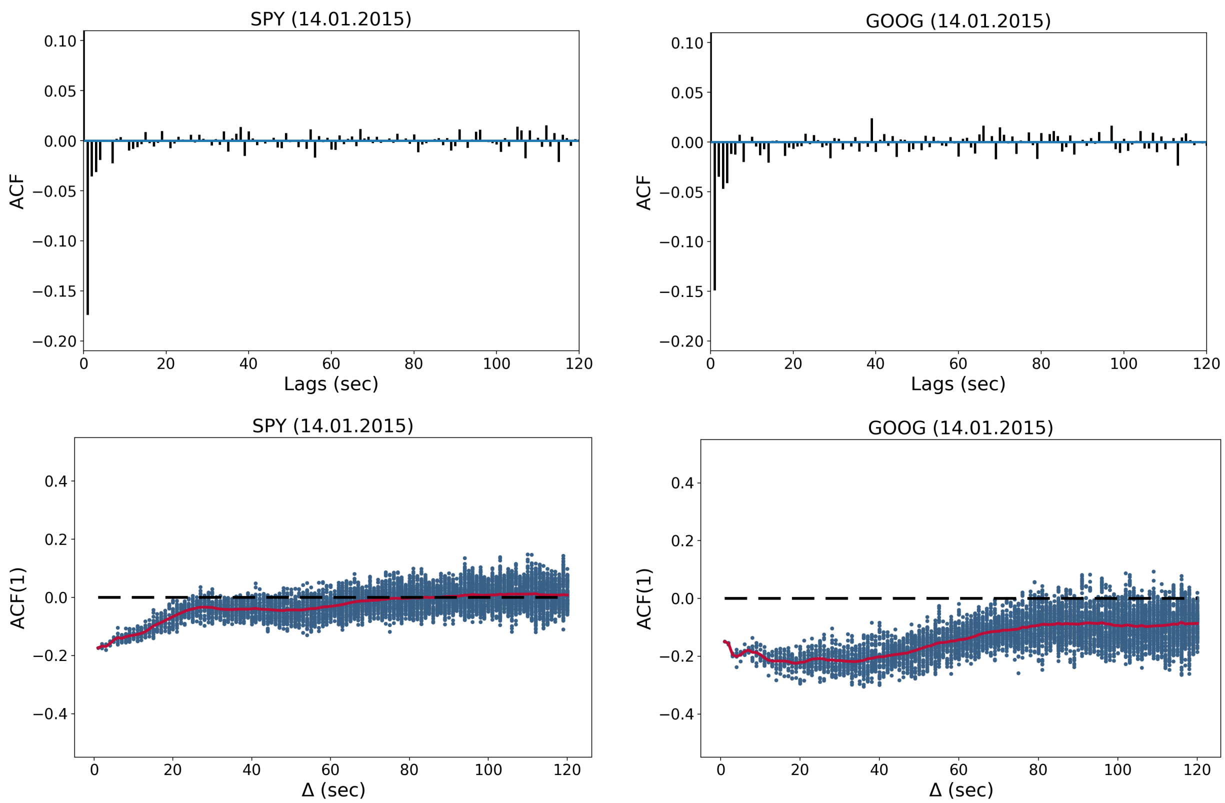


Figure 14: Autocorrelation functions for MFIV series constructed with SPY options (left side) and with GOOG options (right side) on January 14, 2015 (expired on January 17, 2015). Top plots show the autocorrelations as functions of lags constructed for the second-by-second increments of intraday MFIV (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.

To gauge the degree of noise in our intraday series, Figure 14 displays the serial correlation pattern for MFIV increments at different lags and sampling frequencies. The top panels present the empirical autocorrelations across lags ranging from 1 to 120 seconds. The bottom panels depict the first-order autocorrelation of MFIV increments for sampling frequencies $\Delta = 1 \ldots 120$ seconds across different sampling grids, obtained by initiating the computation of the autocorrelations at each possible one-second grid point $\Delta = 1, \ldots, 120$. The first observation of the initial grid is set to 8:35:00 CT and consecutive observations follow with a step of Δ seconds.[32] Hence, for each sampling frequency Δ, we have Δ serial

[32]We exclude MFIV estimates for the first 5 minutes of trading to avoid distortions related to market opening effects,

correlation estimates (blue dots) generated from 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 MFIV increments for both SPY and GOOG exhibit substantial negative autocorrelations for ultra-short lags (3-4 seconds) while, for higher lags, the autocorrelations change signs randomly and become smaller in magnitude. However, the bottom plots of Figure 14 also suggest important discrepancies. The negative first-order serial dependence in the SPY MFIV increments vanishes, on average, for sampling frequencies of 70 seconds or lower, whereas, for GOOG, they remain sizable, even as Δ reaches 120 seconds. This is consistent with the SPY MFIV series being less impacted by measurement error, which may be due to the lower spreads and moderately richer set of observations available for the OTM SPY relative to GOOG options, with cross-sections of 182 and 155 strikes for SPY and GOOG, respectively.

The illustration highlights the feasibility of generating, but also the need for additional scrutiny of, high-frequency option-implied return variation measures. In Figures E.9-E.14 in Web-Appendix E.2, we provide the corresponding January 14, 2015, MFIV measures for another six securities in our sample. The negative autocorrelations vanish fairly rapidly, within 50 seconds, for AAPL, but for the remaining ETFs and stocks, we observe non-zero correlations throughout the 120-second horizon. Our tentative conclusion is that MFIV measures obtained at the 1-2 minute frequency using standard procedures may be reliable, when based on quotes from the most active option markets. In general, however, caution is warranted, and sampling at distinctly lower frequencies is advised for less liquid markets.[33]

6.2 High-Frequency Risk-Neutral Density Extraction

Another popular source of information extracted from the cross-section of option prices is the risk-neutral density (RND) of the underlying asset price over the period until option expiry. Since this measure seeks to provide a detailed picture of the risk-neutral distribution, while the corresponding MFIV only reflects the return variation, the RND is likely to be estimated with a greater degree of imprecision. However, they may still provide useful insights across lower intraday sampling intervals, and they may be particularly informative regarding the response to specific events across the cross-section of underlying assets.

In this section, we examine the intraday changes in the RND of both the market index ETF and several individual stocks before and after a Federal Open Market Committee (FOMC) meeting. The FOMC meetings are closely related to the equity risk premium and can have a significant impact on the prices of both stocks (e.g. Bernanke and Kuttner (2005), Savor and Wilson (2013), Lucca and Moench (2015) among others) and the underlying options (Andersen et al., 2017). By exploiting high-frequency options data, we document immediate market-wide and idiosyncratic RND responses to scheduled macroeconomic news announcements, which is otherwise not feasible using only end-of-day options data.

Firstly, we introduce some notations. The price of a call option written on an underlying asset at time t with a strike price K and expiration date T is given by,

$$C(K) \;=\; e^{-r_f\,\tau}\,\mathbb{E}^{\mathbb{Q}}\left(\max(S_T-K,0)\right) \;=\; e^{-r_f\,\tau}\int_{S_T=K}^{\infty}(S_T-K)\,f_{\mathbb{Q}}(S_T)\,dS_T,$$

where S_T is the price of the underlying asset at expiry, $\tau = T-t$ is the time to maturity of the call option,

including a lack of mid-quotes for some strikes and excessively volatile mid-quote revisions due to active price discovery.

[33] We leave for future work the formal analysis of microstructure noise in option panels observed at high frequency.

r_f is the risk-free interest rate, and $f_{\mathbb{Q}}(S)$ denotes the RND of the asset price.

The object of interest is the RND $f_{\mathbb{Q}}(x)$. We adopt the well-established mixture of lognormals (MLN) approach, advocated by Ritchey (1990), Melick and Thomas (1997) and Liu et al. (2007). It implies that the call option price equals the weighted average of the option prices implied by the individual lognormal densities, which may be calculated in closed form via the Black (1976) formula. Improved estimates are obtained, asymptotically, by employing a larger number of distinct lognormal densities. A detailed description of the MLN methodology is provided in Web-Appendix F.1.

For all securities with sufficiently liquid option markets, we estimate the RND before and after the FOMC announcement, released at 13:00 CT on March 18, 2015. To this end, for each underlying, we construct cross-sections containing the last observed option quote for each strike 30 minutes before and after the announcement. We use the shortest-dated options with tenor exceeding 3 calendar days, leading to cross-sections with 9 days to expiry for this trading day.[34]

We follow standard procedures (e.g., Andersen et al., 2015a, Song and Xiu, 2016) to remove potentially erroneous quote entries.[35] Furthermore, to avoid excessively noisy estimates, we proceed only if the option cross-section has at least 7 strikes after the data cleaning.[36] For brevity, we only discuss results for the most liquid stock and ETF in each sector, relegating the remaining results to Web-Appendix F.3.[37]

Figure 15 depicts the price series for seven underlyings over the one-hour window straddling the FOMC announcement. At 13:00 CT, it was announced that the federal funds rate would remain at the 0 to 1/4 percent target range and the stance of monetary policy would be reassessed only after further market indicators became available. This decision triggered an immediate price jump for all 7 underlyings, ultimately generating one-hour returns ranging from 1.20% to 1.51%.

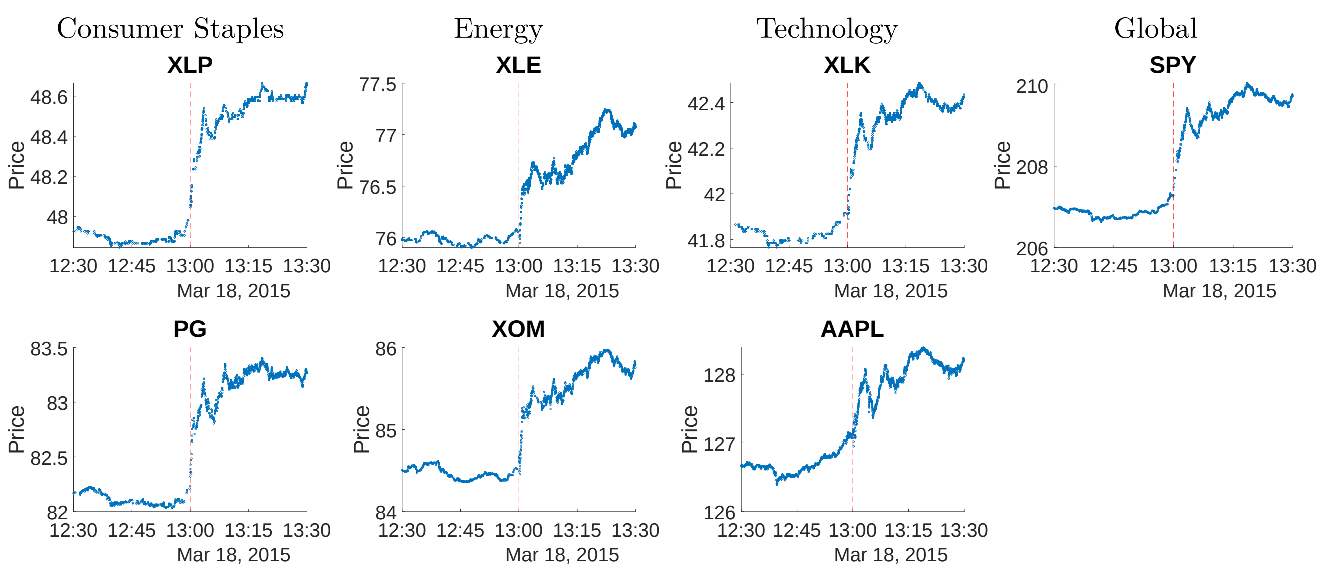


Figure 15: Time series plots of prices of different underlyings during 30 minutes before and after the FOMC announcement at 14:00 Eastern time (i.e. 13:00 CT) on 18 March 2015. Time in each plot is CT.

[34] We exclude K (Kellogg Company) and BLK (BlackRock), because their shortest-dated options have a 30-day tenor – much longer than the available tenor for the remaining underlyings.

[35] A detailed description of our data preparation steps and estimation procedure are available in Web-Appendix F.2.

[36] This leads to the removal of BAC (Bank of America) and XLF (Financial Sector ETF) from this analysis.

[37] We only report results for three sectors, Consumer Staples, Energy, and Technology, as all financial sector assets, except for JPM (JPMorgan & Chase), are excluded by our liquidity filter.

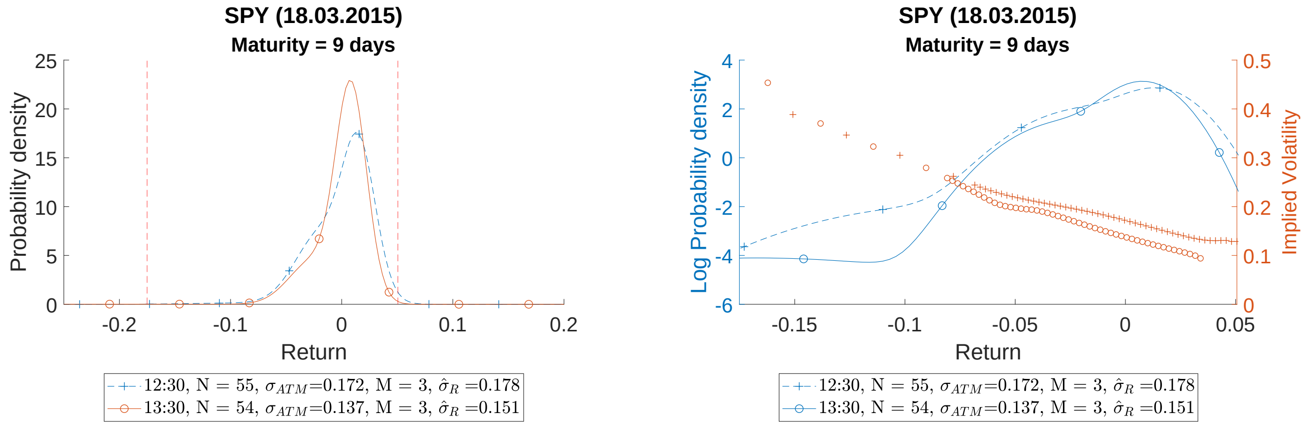


Figure 16: The left plots depict the estimated RNDs, as a function of return $S_T/F_t - 1$, for the shortest time to maturity obtained from intraday OPRA data for options written on SPY in the **Global** sector at **30 minutes** before and after the FOMC announcement at 14:00 Eastern time (i.e. 13:00 CT) on 18 March 2015. Each RND curve is estimated from a mixture of M lognormal distributions. In each subplot, the number of strikes (N), the ATM Black-Scholes implied volatility (σ_{ATM}), the number of lognormal densities in each mixture (M), and the estimated annualized standard deviation ($\hat{\sigma}_R$) of each RND curve are reported. Vertical red dashed lines indicate the observed return ($K/F_t - 1$) range. Right plots show the logarithm of the RNDs (left axis) and the Black-Scholes implied volatility (right axis) over the observed return range. Time in each plot is CT.

Figure 16 depicts the implied RND for the return, $r_T = S_T/F_t - 1$, obtained from a mixture of $M = 3$ lognormal densities, using SPY options with 9 days to maturity at 12:30 and 13:30 CT on March 18, 2015.[38] Both RND curves exhibit pronounced negative skewness, consistent with the extant literature. This notwithstanding, the pertinent question is whether the RND changes shape in response to the FOMC announcement. In fact, from the log-density plot, we do identify a thinning of the left SPY tail and we find both the ATM Black-Scholes implied volatility and the implied risk-neutral return dispersion, $\hat{\sigma}_R$, to have dropped.[39] This suggests that market participants, in response to the policy statement, have lowered their expectations regarding near-term unfavorable events impacting the SPY.

Of course, the statistical significance of the above shift in the RND is hard to assess. The ability to estimate concurrent shifts in the RND curves for a number of diverse assets is one way to gauge the robustness of this finding.

[38] These RNDs are readily derived from $f_{\mathbb{Q}}(S_T)$ using a Jacobian transformation.

[39] We define the dispersion of the RND curve (σ_R) as the annualized risk-neutral standard deviation of return $\frac{1}{\sqrt{\tau}}\sigma_{r_T}$, which equals $\frac{1}{\sqrt{\tau}F_t}\sigma_{S_T} = \frac{1}{\sqrt{\tau}F_t}\sqrt{\left[\sum_{i=1}^{M} w_i F_i^2 \exp(\sigma_i^2\tau)\right] - F_t^2}$, as implied by Equations (F.1) and (F.2) in Web-Appendix F.1.

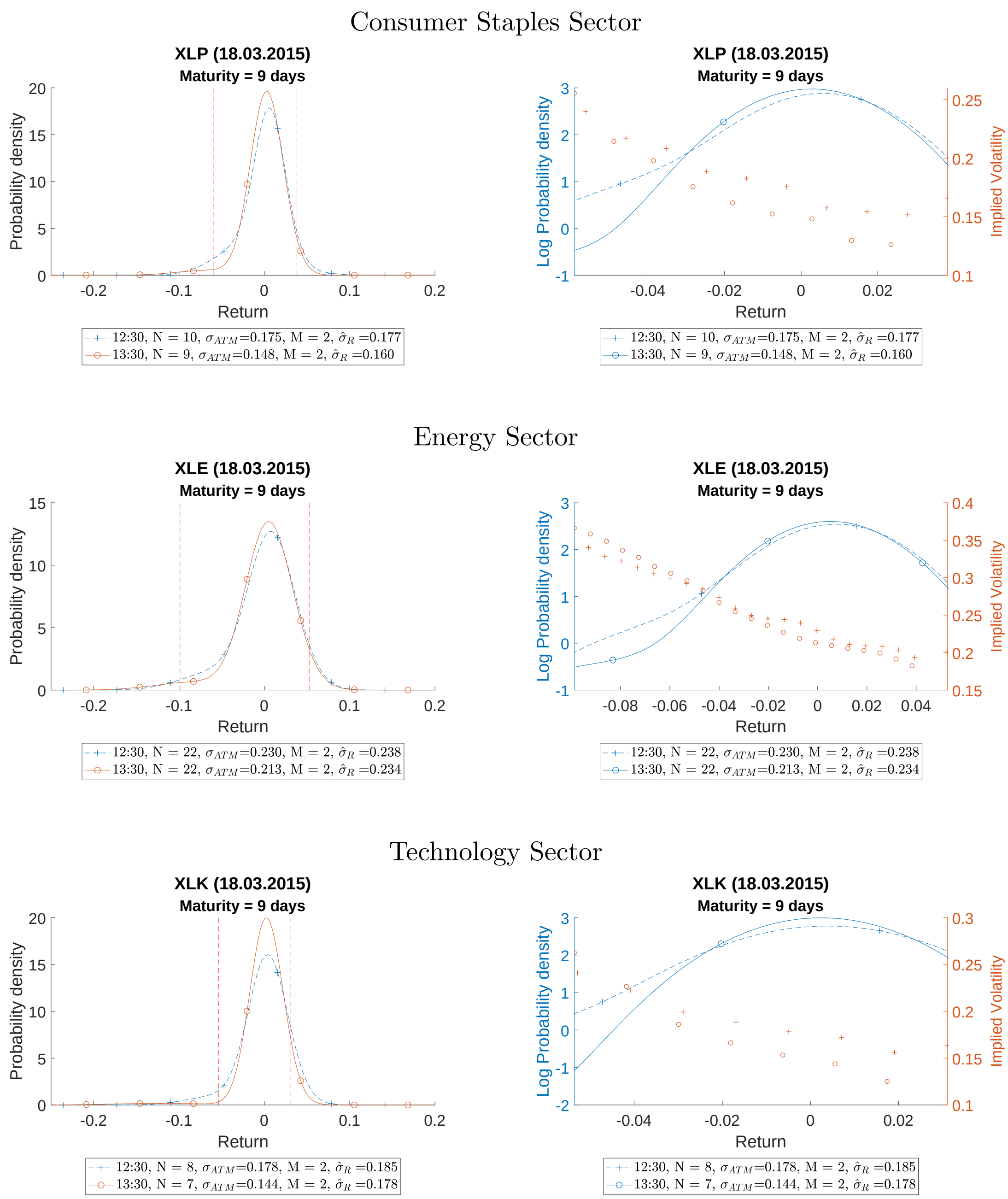


Figure 17: The left plots depict the estimated RNDs, as a function of return $S_T/F_t - 1$, for the shortest time to maturity obtained from intraday OPRA data for options written on ETFs of different sectors at **30 minutes** before and after the FOMC announcement at 14:00 Eastern time (i.e. 13:00 CT) on 18 March 2015. Each RND curve is estimated from a mixture of M lognormal distributions. In each subplot, the number of strikes (N), the ATM Black-Scholes implied volatility (σ_{ATM}), the number of lognormal densities in each mixture (M), and the estimated annualized standard deviation ($\hat{\sigma}_R$) of each RND curve are reported. Vertical red dashed lines indicate the observed return ($K/F_t - 1$) range. Right plots show the logarithm of the RNDs (left axis) and the Black-Scholes implied volatility (right axis) over the observed return range. Time in each plot is CT.

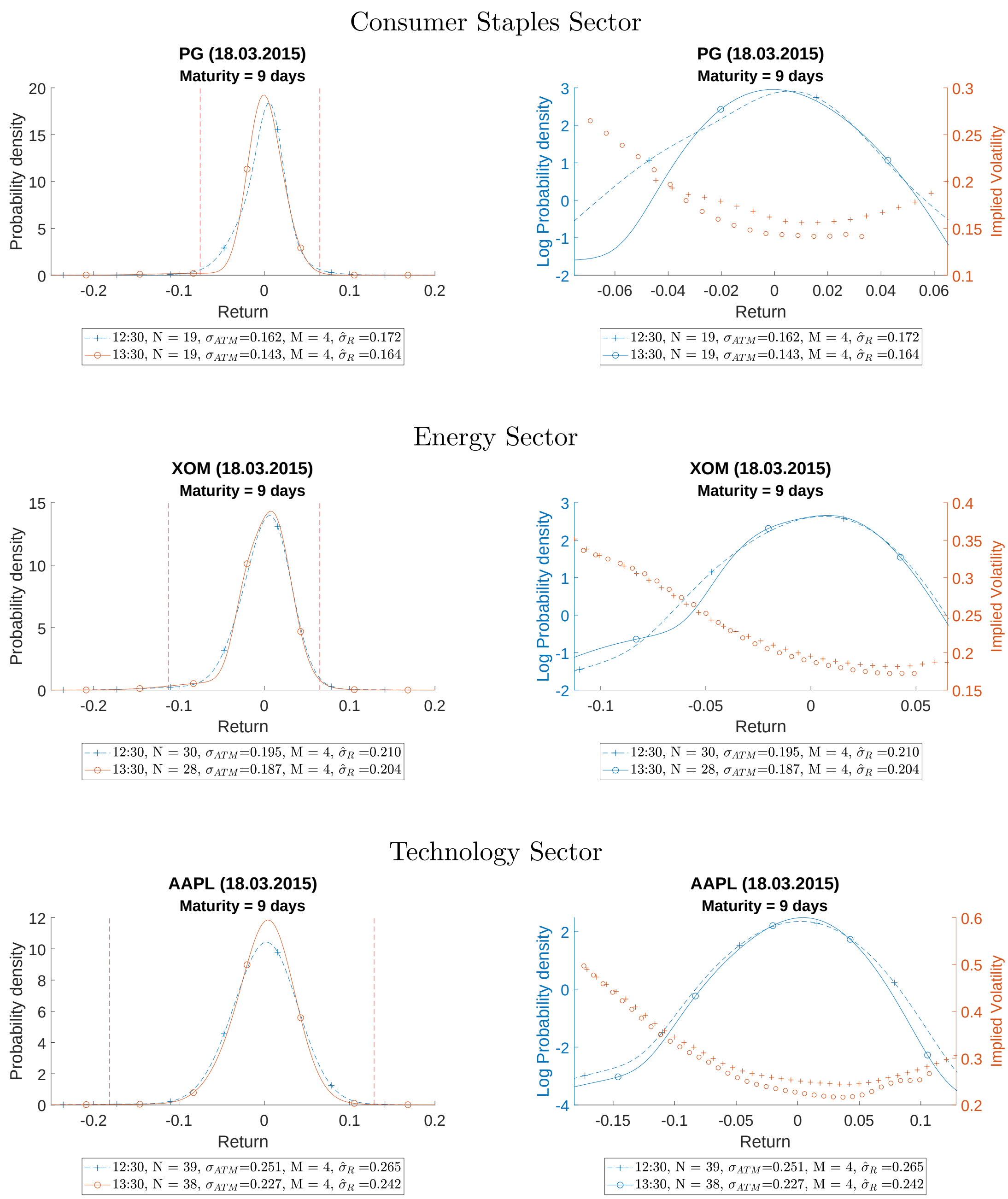


Figure 18: The left plots depict the estimated RNDs, as a function of return $S_T/F_t - 1$, for the shortest time to maturity obtained from intraday OPRA data for options written on individual stocks in different sectors at **30 minutes** before and after the FOMC announcement at 14:00 Eastern time (i.e. 13:00 CT) on 18 March 2015. Each RND curve is estimated from a mixture of M lognormal distributions. In each subplot, the number of strikes (N), the ATM Black-Scholes implied volatility (σ_{ATM}), the number of lognormal densities in each mixture (M), and the estimated annualized standard deviation ($\hat{\sigma}_R$) of each RND curve are reported. Vertical red dashed lines indicate the observed return ($K/F_t - 1$) range. Right plots show the logarithm of the RNDs (left axis) and the Black-Scholes implied volatility (right axis) over the observed return range. Time in each plot is CT.

Figures 17 and 18 reveal how the RND for three ETFs and the most liquid stock within each of these sectors react to the announcement. In analogy to the SPY, the estimated RND (left plots) for all these securities become less dispersed (have lower standard deviation $\widehat{\sigma}_R$) and experience a downward shift in ATM Black-Scholes implied volatilities after the FOMC statement. This is consistent with the policy decision reducing the uncertainty about the future price and volatility of not only the equity index, but also each of the sector ETFs and individual stocks. In addition, consistent with SPY, the left tail of the estimated RNDs of the sector ETFs shrinks after the FOMC announcement. This is generally also true for the individual stocks, but the estimation uncertainty concerning the tail shape is occasionally quite severe due to the limited number of OTM strikes, so the evidence is at times ambiguous.

Figures F.15 and F.16 in Web-Appendix F.3 provide the corresponding evidence for the remaining securities in our sample. The findings are qualitatively similar, although we once again encounter a few cases where the shift in the RND is ambiguous. Nonetheless, across all the underlying assets explored, we find the left RND tail to either shrink or remain largely unaltered, while the measures of dispersion and volatility almost uniformly drop. In other words, the high-frequency options data indicate a reduction in the risk-neutral risk measures both for the aggregate equity index and across the various sectors and individual stocks. For other types of economic news, the reaction across different sectors and individual stocks is likely to be more heterogenous, which should facilitate interpretation regarding the shift in the perception of underlying risks and associated risk premiums. Since many other factors impact the RND across the full trading day, inference based solely on end-of-day option implied measures is much less efficient.

We conclude that the construction of meaningful intraday RND estimates is feasible for the assets with liquid option trading, as long as the sampling frequency is moderate. At the same time, it is evident that further research into the construction of robust and reliable extraction procedures is warranted for less liquid option markets. These issues notwithstanding, it is clear that the increasing availability and richness of high-frequency, short-dated option quotes enhances our ability to explore the economic and financial effects of various types of news events through their differential impact on a large option cross-section.

7 Conclusion

This study provides a detailed description of high-frequency trade and quote data for options traded in the U.S. It reviews the current structure of the U.S. market by characterizing the 16 constituent option exchanges, summarizing the market regulatory plans governing the option trading, and discussing specific market maker quoting obligations that are pertinent to the functioning of the markets.

Our data is provided by OPRA in accordance with the "Plan for Reporting of Consolidated Options Last Sale Reports and Quotation Information." 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 traded in the U.S. during the first eight months of 2015. Our dataset is more comprehensive than the alternative high-frequency option datasets employed in the limited number of existing studies, which, typically focus strictly on index options. We provide a detailed assessment of the quality of our dataset, and develop a general filtering algorithm for data cleaning the spirit of the Barndorff-Nielsen et al. (2009) algorithm for tick-by-tick data on equities.

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

We characterize the trade and quote intensities across the securities and exchange venues as a function of option tenor and moneyness. Likewise, we explore the association between the size of the quoted option bid-ask spreads and the moneyness, tenor and type of option (call or put).

Finally, we present two illustrative applications - the estimation of the risk-neutral return variance and the risk-neutral density (RND) - using intraday OPRA data. We confirm that such measures can be constructed and yield information beyond what can be gleaned from end-of-day option data, but we also point towards limitations associated with the measures at the very highest frequencies.

The various findings serve as inspiration for new studies addressing either market microstructure or asset pricing questions related to the option markets. The literature on high-frequency equity data is voluminous. We expect the much richer, but also less manageable, option data to provide a stimulating basis for novel work in the future, both through exploration within the option space itself, but also in terms of the interaction between the option and equity markets at a fine resolution. In order for this to materialize, however, the rather complex market structure and regulatory environment should be recognized, so that appropriate hypotheses and empirical procedures can be designed.

Overall, we hope this overview will serve as inspiration to explore the OPRA dataset to its fullest. We are convinced it holds the key to progress on many market microstructure and high-frequency asset pricing questions.

References

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.

Anand, A. and Chakravarty, S. (2007). Stealth Trading in Options Markets. *Journal of Financial and Quantitative Analysis*, 42(1):167–187.

Anand, A., Hua, J., and McCormick, T. (2016). Make-take structure and market quality: Evidence from the U.S. options markets. *Management Science*, 62(11):3271–3290.

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.

Andersen, T. G., Fusari, N., and Todorov, V. (2017). Short-Term Market Risks Implied by Weekly Options. *Journal of Finance*, 72(3):1335–1386.

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.

Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, 52(5):2003–2049.

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.

Battalio, R., Figlewski, S., and Neal, R. (2020). Option Investor Rationality Revisited: The Role of Exercise Boundary Violations. *Financial Analysts Journal*, 76(1):82–99.

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

Battalio, R. and Schultz, P. (2006). Options and the bubble. *Journal of Finance*, 61(5):2071–2102.

Battalio, R. and Schultz, P. (2011). Regulatory Uncertainty and Market Liquidity: The 2008 Short Sale Ban's Impact on Equity Option Markets. *Journal of Finance*, 66(6):2013–2053.

Battalio, R., Shkilko, A., and Van Ness, R. (2016). To Pay or Be Paid? The Impact of Taker Fees and Order Flow Inducements on Trading Costs in U.S. Options Markets. *Journal of Financial and Quantitative Analysis*, 51(5):1637–1662.

Bernanke, B. S. and Kuttner, K. N. (2005). What Explains the Stock Market's Reaction to Federal Reserve Policy? *The Journal of Finance*, 60(3):1221–1257.

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

Birru, J. and Figlewski, S. (2012). Anatomy of a meltdown: The risk neutral density for the S&P 500 in the fall of 2008. *Journal of Financial Markets*, 15(2):151–180.

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

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

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

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

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.

Cakici, N., Goswami, G., and Tan, S. (2014). Options resilience during extreme volatility: Evidence from the market events of May 2010. *Journal of Banking and Finance*, 49:262–274.

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

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

Carr, P. and Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, 2:1311–1341.

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

Chan, K., Chung, Y. P., and Fong, W.-M. (2002). The informational role of stock and option volume. *Review of Financial Studies*, 15(6):1049–1075.

Chan, K., Chung, Y. P., and Johnson, H. (1993). Why option prices lag stock prices: A trading-based explanation. *Journal of Finance*, 48(5):1957–1967.

Chan, K., Chung, Y. P., and Johnson, H. (1995). The Intraday Behavior of Bid-Ask Spreads for NYSE Stocks and CBOE Options. *Journal of Financial and Quantitative Analysis*, 30(3):329–346.

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

Chatrath, A., Christie-David, R. A., Miao, H., and Ramchander, S. (2015). Short-term options: Clienteles, market segmentation, and event trading. *Journal of Banking and Finance*, 61:237–250.

Christoffersen, P., Goyenko, R., Jacobs, K., and Karoui, M. (2018). Illiquidity premia in the equity options market. *Review of Financial Studies*, 31(3):811–851.

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.

Cremers, K. and Weinbaum, D. (2010). Deviations from put-call parity and return predictability. *Journal of Financial and Quantitative Analysis*, 45(2):335–367.

Dalderop, J. (2020). Nonparametric filtering of conditional state-price densities. *Journal of Econometrics*, 214(2):295–325.

De Fontnouvelle, P., Fishe, R. P. H., and Harris, J. H. (2003). The behavior of bid-ask spreads and volume in options markets during the competition for listings in 1999. *Journal of Finance*, 58(6):2437–2463.

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

George, T. J. and Longstaff, F. A. (1993). Bid-Ask Spreads and Trading Activity in the S & P 100 Index Options Market. *Journal of Financial and Quantitative Analysis*, 28(3):381–397.

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

Harris, L. E. and Mayhew, S. (2005). Market fragmentation metrics. *Working Paper, University of Southern California.*

Holowczak, R., Hu, J., and Wu, L. (2014). Aggregating information in option transactions. *Journal of Derivatives*, 21(3):9–23.

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

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

Jacod, J., Li, Y., Mykland, P., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications*, 119:2249–2276.

Kalnina, I. and Xiu, D. (2017). Nonparametric estimation of the leverage effect: A trade-off between robustness and efficiency. *Journal of the American Statistical Association*, 112:384–396.

Kapetanios, G., Konstantinidi, E., Neumann, M., and Skiadopoulos, G. (2019). Jumps in option prices and their determinants: Real-time evidence from the E-mini S&P 500 options market. *Journal of Financial Markets*, 46:100506.

Lee, J. and Yi, C. H. (2001). Trade Size and Information-Motivated Trading in the Options and Stock Markets. *Journal of Financial and Quantitative Analysis*, 36(4):485–501.

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.

Lucca, D. O. and Moench, E. (2015). The Pre-FOMC Announcement Drift. *The Journal of Finance*, 70(1):329–371.

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

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

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.

Mishra, S. and Daigler, R. T. (2014). Intraday trading and bid-ask spread characteristics for SPX and SPY options. *Journal of Derivatives*, 21(3):70–84.

Muravyev, D. (2016). Order Flow and Expected Option Returns. *Journal of Finance*, 71(2):673–708.

Muravyev, D. and Pearson, N. D. (2020). Options Trading Costs Are Lower than You Think. *The Review of Financial Studies*, forthcoming.

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

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

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, 63(1):3–50.

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

Savor, P. and Wilson, M. (2013). How Much Do Investors Care About Macroeconomic Risk? Evidence from Scheduled Economic Announcements. *Journal of Financial and Quantitative Analysis*, 48(2):343–375.

Sheikh, A. M. and Ronn, E. I. (1994). A characterization of the daily and intraday behavior of returns on options. *Journal of Finance*, 49(2):557–579.

Simaan, Y. E. and Wu, L. (2007). Price discovery in the U.S. stock options market. *Journal of Derivatives*, 15(2):20–38.

Simon, D. P. (2013). The intraday and overnight behavior of SPY options and adjusted delta hedging. *Journal of Futures Markets*, 33(5):443–468.

Song, Z. and Xiu, D. (2016). A tale of two option markets: Pricing kernels and volatility risk. *Journal of Econometrics*, 190:176–196.

Stephan, J. A. and Whaley, R. E. (1990). Intraday price change and trading volume relations in the stock and stock option markets. *Journal of Finance*, 45(1):191–220.

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.

Todorov, V. (2019). Nonparametric spot volatility from options. *Annals of Applied Probability*, 29(6):3590–3636.

Vijh, A. M. (1990). Liquidity of the CBOE equity options. *Journal of Finance*, 45(4):1157–1179.

Zhang, Y. (2018). *Essays in Risk Management and Asset Pricing with High Frequency Option Panels*. PhD thesis, Durham University.

Web Appendix for "A Descriptive Study of High-Frequency Trade and Quote Option Data"

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

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

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

[c] *University of Manchester, Alliance Manchester Business School*

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

August 21, 2020

Abstract

This Web Appendix provides supplementary materials for the paper "A Descriptive Study of High-Frequency Trade and Quote Data".

Keywords: Options, 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 thank the guest editor Kris Jacobs and two anonymous referees for their invaluable comments which greatly improved the quality of this paper. 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.

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.
	MIAX Emerald	Princeton, NJ	2019	-	E	Based on a maker-taker pricing model and a pro-rata 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. 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.

Back to paper

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)
MIAX Emerald	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)
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).

Back to paper

	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.

Back to paper

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. Back to paper

Exchange	PG	WMT	K	XLP	XOM	CVX	MPC	XLE	JPM	BAC	BLK	XLF	AAPL	GOOG	EA	XLK	SPY
NYSE Amex	2.30	11.29	0.00	0.01	0.00	0.00	0.00	0.01	1,942.64	78.11	0.00	0.06	0.23	164.49	0.04	0.00	4,997.85
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.08*	*0.01*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.12*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.04*
BOX	0.44	2.95	0.00	0.01	0.00	0.00	0.00	0.01	685.81	31.46	0.00	0.01	0.00	18.80	0.02	0.00	1,116.04
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.09*	*0.01*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.10*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.03*
CBOE	2.65	15.55	0.00	0.04	0.00	0.00	0.00	0.34	1,889.39	203.65	0.01	0.08	0.61	180.42	0.01	0.01	10,319.90
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.07*	*0.01*	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.17*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.05*
ISE Gemini	2.37	4.79	0.00	0.01	0.00	0.00	0.00	0.00	688.67	1,212.16	0.16	1.81	0.06	25.22	0.00	0.00	4,546.27
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.06*	*0.07*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.13*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.02*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.05*
ISE	1.75	11.93	0.00	0.20	0.00	0.00	0.00	0.20	2,901.96	693.30	0.00	0.27	0.13	40.79	0.03	0.00	11,019.86
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.10*	*0.04*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.17*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.03*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.06*
MIAX	0.70	9.93	0.02	0.01	0.00	0.00	0.00	0.00	2,426.93	94.07	0.00	0.05	0.26	191.45	0.01	0.00	6,826.96
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.09*	*0.00*	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.14*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.04*
NYSE Arca	2.08	14.55	0.06	0.08	0.00	0.00	0.00	0.04	2,201.39	541.09	0.00	1.86	0.02	141.66	0.10	0.02	5,846.60
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.21*	*0.09*	*0.00*	*0.00*	*0.00*	*0.03*	*0.00*	*0.00*	*0.22*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.05*	*0.05*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.06*
Nasdaq OM	1.02	11.17	0.00	0.16	0.00	0.00	0.00	0.09	2,696.61	864.07	0.00	3.54	0.01	119.61	0.06	0.02	7,462.91
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.25*	*0.10*	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.25*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.05*	*0.04*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.06*
BX Options	0.29	2.37	0.00	0.01	0.00	0.00	0.00	0.00	908.62	141.02	0.00	0.00	0.02	69.31	0.00	0.00	2,466.56
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.09*	*0.01*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.11*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.03*
C2	0.35	0.93	0.00	0.01	0.00	0.00	0.00	0.00	342.37	8.07	0.00	0.30	0.03	11.29	0.00	0.00	3,430.56
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.04*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.10*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.03*
Nasdaq PHLX	0.96	14.16	0.00	0.10	0.00	0.00	0.00	0.02	1,645.32	190.94	0.01	0.21	0.90	207.92	0.01	0.00	8,008.16
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.07*	*0.01*	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.13*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.05*
BATS BZX	1.19	12.49	0.01	0.20	0.00	0.00	0.00	0.03	2,281.60	823.24	0.00	3.69	0.03	87.22	0.04	0.37	4,400.92
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.27*	*0.12*	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.21*
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.05*	*0.04*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.06*

Table A.5: Aggregate statistics for the option quotes which violate put-call parity in January 2015. All contracts expiring on February 20, 2015 are considered. The table contains the average number of option quotes (per day and strike price) which violate the put-call parity in January 2015 for each considered underlying asset and for each exchange market. The numbers in italics specify the fraction of records which violate the put-call parity (upper number) and the fraction of time when the put-call parity was violated (lower number), for a given underlying asset on a given exchange market.

Back to paper

Moneyness	$-3 \le m < -1$						$-1 \le m \le 1$						$1 < m \le 3$					
Weeks to expiry	7-6	6-5	5-4	4-3	3-2	< 2	7-6	6-5	5-4	4-3	3-2	< 2	7-6	6-5	5-4	4-3	3-2	< 2
Option classes																		
PG	0.077	0.105	0.107	0.082	0.084	0.093	0.078	0.073	0.109	0.100	0.091	0.131	1.137	1.009	1.248	1.098	1.162	1.562
WMT	0.117	0.072	0.108	0.095	0.096	0.102	0.080	0.069	0.071	0.065	0.075	0.081	1.118	0.956	1.038	1.051	1.087	1.038
K	0.138	0.152	0.122	0.149	0.142	0.173	0.247	0.254	0.234	0.238	0.172	0.561	1.643	1.574	1.525	1.492	1.498	1.853
XLP	0.135	0.141	0.144	0.141	0.145	0.150	0.110	0.135	0.156	0.120	0.134	0.266	1.833	1.852	1.741	1.715	1.772	1.871
XOM	0.167	0.179	0.203	0.194	0.270	0.183	0.086	0.094	0.119	0.112	0.152	0.141	1.405	1.355	1.592	1.261	1.659	1.606
CVX	0.136	0.146	0.160	0.175	0.181	0.176	0.074	0.068	0.097	0.100	0.118	0.141	1.481	1.345	1.210	1.452	1.514	1.592
MPC	0.161	0.177	0.192	0.178	0.190	0.257	0.273	0.318	0.285	0.315	0.326	0.481	0.964	1.323	1.390	1.629	1.485	1.918
XLE	0.137	0.127	0.105	0.127	0.104	0.106	0.092	0.083	0.120	0.101	0.097	0.119	1.393	1.456	1.407	1.331	1.382	1.456
JPM	0.062	0.074	0.081	0.091	0.098	0.084	0.054	0.051	0.055	0.052	0.064	0.095	1.368	1.312	1.121	1.295	1.438	1.510
BAC	0.016	0.015	0.013	0.014	0.022	0.027	0.053	0.057	0.078	0.073	0.109	0.237	1.377	1.425	1.365	1.416	1.639	1.894
BLK	0.063	0.067	0.074	0.085	0.090	0.105	0.189	0.245	0.241	0.281	0.393	0.629	1.626	1.710	1.707	1.795	1.925	1.971
XLF	0.030	0.032	0.035	0.035	0.100	0.102	0.127	0.111	0.109	0.126	0.189	0.383	1.947	1.785	1.823	1.839	1.922	1.992
AAPL	0.016	0.016	0.014	0.015	0.014	0.016	0.021	0.021	0.019	0.024	0.024	0.030	0.225	0.241	0.135	0.203	0.342	0.418
GOOG	0.033	0.037	0.042	0.047	0.060	0.072	0.050	0.046	0.047	0.061	0.066	0.085	1.101	1.142	1.024	0.780	0.665	0.653
EA	0.155	0.157	0.138	0.154	0.126	0.154	0.056	0.057	0.073	0.085	0.089	0.136	1.180	1.072	0.537	0.739	0.775	1.345
XLK	0.097	0.109	0.116	0.105	0.107	0.119	0.212	0.223	0.284	0.168	0.224	0.314	1.924	1.870	1.910	1.779	1.866	1.965
SPY	0.018	0.020	0.027	0.045	0.038	0.053	0.024	0.025	0.029	0.044	0.040	0.048	0.700	0.895	1.012	1.026	1.100	0.918
Exchange markets																		
Nasdaq PHLX	0.070	0.072	0.088	0.098	0.096	0.115	0.086	0.091	0.100	0.088	0.104	0.171	1.219	1.252	1.185	1.190	1.256	1.357
Nasdaq OM	0.085	0.087	0.099	0.114	0.111	0.120	0.083	0.088	0.097	0.085	0.099	0.162	1.212	1.239	1.180	1.171	1.244	1.331
BX Options	0.082	0.081	0.098	0.104	0.114	0.130	0.103	0.107	0.121	0.119	0.130	0.209	1.289	1.356	1.289	1.338	1.361	1.436
ISE	0.073	0.074	0.088	0.096	0.097	0.113	0.085	0.090	0.099	0.087	0.102	0.165	1.220	1.258	1.206	1.197	1.265	1.358
ISE Gemini	0.072	0.074	0.091	0.104	0.105	0.123	0.094	0.099	0.106	0.104	0.113	0.182	1.216	1.239	1.193	1.206	1.236	1.337
CBOE	0.081	0.080	0.104	0.111	0.117	0.124	0.085	0.090	0.102	0.089	0.103	0.168	1.236	1.280	1.229	1.241	1.284	1.374
C2	0.097	0.103	0.152	0.134	0.166	0.163	0.088	0.094	0.108	0.113	0.148	0.202	1.275	1.315	1.287	1.315	1.371	1.373
BATS BZX	0.073	0.075	0.094	0.103	0.103	0.116	0.086	0.090	0.100	0.092	0.103	0.162	1.230	1.253	1.185	1.176	1.233	1.327
NYSE Amex	0.071	0.067	0.090	0.095	0.100	0.111	0.086	0.089	0.102	0.091	0.105	0.169	1.256	1.222	1.187	1.229	1.308	1.356
NYSE Arca	0.074	0.072	0.093	0.096	0.100	0.111	0.083	0.085	0.098	0.086	0.099	0.158	1.223	1.155	1.127	1.152	1.240	1.288
MIAX	0.071	0.073	0.092	0.101	0.104	0.115	0.087	0.092	0.104	0.095	0.106	0.171	1.258	1.298	1.246	1.264	1.316	1.390
BOX	0.076	0.075	0.096	0.105	0.108	0.118	0.094	0.093	0.104	0.097	0.108	0.173	1.330	1.312	1.248	1.261	1.320	1.408

Table A.6: Average relative spread measures computed from the OPRA quote records for call option contracts traded between January 2nd and February 18th, 2015, which expire on February 20th, 2015. The moneyness is defined as $m = \log(K/F_t)\big/(\sigma_t\sqrt{\tau})$. Back to paper

Moneyness	$-3 \leq m < -1$						$-1 \leq m \leq 1$						$1 < m \leq 3$					
Weeks to expiry	7-6	6-5	5-4	4-3	3-2	< 2	7-6	6-5	5-4	4-3	3-2	< 2	7-6	6-5	5-4	4-3	3-2	< 2
Option classes																		
PG	0.317	0.380	0.398	0.384	0.625	0.874	0.057	0.060	0.108	0.083	0.081	0.109	0.100	0.114	0.100	0.116	0.109	0.116
WMT	0.341	0.269	0.286	0.337	0.522	0.729	0.065	0.060	0.065	0.060	0.067	0.081	0.123	0.100	0.127	0.128	0.093	0.128
K	1.538	1.527	1.628	1.741	1.817	1.968	0.219	0.216	0.191	0.245	0.211	0.570	0.105	0.116	0.126	0.137	0.119	0.135
XLP	0.460	0.516	0.379	0.539	0.687	1.198	0.094	0.131	0.142	0.116	0.124	0.212	0.145	0.167	0.205	0.180	0.172	0.206
XOM	0.252	0.226	0.287	0.335	0.478	0.480	0.077	0.070	0.101	0.089	0.106	0.108	0.179	0.179	0.261	0.221	0.272	0.279
CVX	0.267	0.257	0.246	0.304	0.399	0.546	0.062	0.059	0.089	0.085	0.095	0.109	0.138	0.133	0.197	0.195	0.203	0.208
MPC	0.705	0.789	0.883	1.377	1.510	1.566	0.258	0.251	0.354	0.370	0.470	0.578	0.167	0.137	0.166	0.216	0.239	0.318
XLE	0.651	0.469	0.425	0.515	0.558	0.583	0.100	0.094	0.103	0.082	0.083	0.108	0.137	0.107	0.128	0.137	0.142	0.170
JPM	0.285	0.214	0.255	0.349	0.396	0.672	0.035	0.037	0.042	0.042	0.056	0.083	0.171	0.144	0.147	0.118	0.141	0.145
BAC	0.513	0.511	0.725	0.932	1.147	1.620	0.040	0.047	0.060	0.069	0.094	0.195	0.015	0.015	0.015	0.016	0.044	0.048
BLK	0.845	0.835	0.769	1.008	1.304	1.598	0.142	0.191	0.209	0.239	0.328	0.443	0.056	0.062	0.073	0.082	0.086	0.106
XLF	0.297	0.341	0.385	0.437	0.804	1.739	0.062	0.059	0.074	0.076	0.143	0.226	0.028	0.033	0.037	0.038	0.139	0.146
AAPL	0.083	0.080	0.053	0.076	0.080	0.178	0.019	0.021	0.018	0.022	0.022	0.028	0.030	0.028	0.022	0.020	0.017	0.018
GOOG	0.338	0.325	0.178	0.237	0.273	0.420	0.037	0.037	0.043	0.056	0.061	0.085	0.031	0.036	0.043	0.046	0.066	0.093
EA	0.324	0.342	0.446	0.669	0.601	0.966	0.067	0.059	0.059	0.083	0.081	0.119	0.150	0.162	0.127	0.167	0.182	0.174
XLK	0.778	0.785	0.679	0.849	0.784	1.399	0.138	0.162	0.204	0.117	0.136	0.256	0.155	0.184	0.176	0.199	0.187	0.194
SPY	0.037	0.040	0.044	0.064	0.075	0.127	0.027	0.025	0.029	0.042	0.040	0.046	0.113	0.112	0.118	0.115	0.136	0.163
Exchange markets																		
Nasdaq PHLX	0.378	0.348	0.368	0.527	0.557	0.792	0.075	0.078	0.093	0.080	0.100	0.158	0.091	0.089	0.111	0.110	0.114	0.138
Nasdaq OM	0.366	0.337	0.354	0.514	0.544	0.767	0.077	0.079	0.091	0.079	0.096	0.152	0.105	0.102	0.123	0.131	0.131	0.140
BX Options	0.414	0.381	0.405	0.585	0.619	0.872	0.088	0.090	0.109	0.106	0.128	0.196	0.104	0.099	0.124	0.131	0.145	0.170
ISE	0.368	0.341	0.364	0.520	0.553	0.789	0.075	0.078	0.094	0.080	0.099	0.155	0.083	0.083	0.102	0.103	0.105	0.130
ISE Gemini	0.385	0.354	0.365	0.533	0.567	0.808	0.083	0.085	0.100	0.095	0.109	0.170	0.111	0.116	0.140	0.139	0.149	0.178
CBOE	0.376	0.346	0.363	0.530	0.572	0.792	0.074	0.077	0.097	0.083	0.097	0.154	0.085	0.084	0.107	0.114	0.119	0.143
C2	0.395	0.394	0.428	0.505	0.607	0.816	0.075	0.079	0.091	0.094	0.119	0.167	0.116	0.131	0.195	0.190	0.234	0.254
BATS BZX	0.380	0.357	0.359	0.525	0.566	0.780	0.078	0.080	0.098	0.087	0.099	0.152	0.099	0.098	0.120	0.125	0.133	0.151
NYSE Amex	0.380	0.326	0.347	0.514	0.574	0.791	0.075	0.078	0.098	0.084	0.099	0.156	0.085	0.078	0.107	0.108	0.115	0.136
NYSE Arca	0.363	0.312	0.333	0.497	0.553	0.760	0.074	0.076	0.095	0.079	0.094	0.147	0.082	0.076	0.102	0.102	0.108	0.127
MIAX	0.378	0.353	0.367	0.543	0.584	0.801	0.077	0.080	0.099	0.088	0.100	0.156	0.096	0.096	0.118	0.124	0.133	0.154
BOX	0.418	0.353	0.369	0.542	0.594	0.818	0.082	0.081	0.101	0.093	0.104	0.161	0.101	0.100	0.121	0.129	0.141	0.158

Table A.7: Average relative spread measures computed from the OPRA quote records for put option contracts traded between January 2nd and February 18th, 2015, which expire on February 20th, 2015. The moneyness is defined as $m = \log(K/F_t)\big/(\sigma_t\sqrt{\tau})$. Back to paper

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.[1]

Below we briefly discuss the information contained in the messages. A detailed specification can be found in the official documentation by OPRA.[2]

B.1 Data Content

Table B.8 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.

Back to paper

[1] 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.

[2] 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.8: Information contained in the OPRA dataset.

B.2 Data Examples

Table B.9 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.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.8.

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.10). 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. 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.10: 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.8. 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.

Back to paper

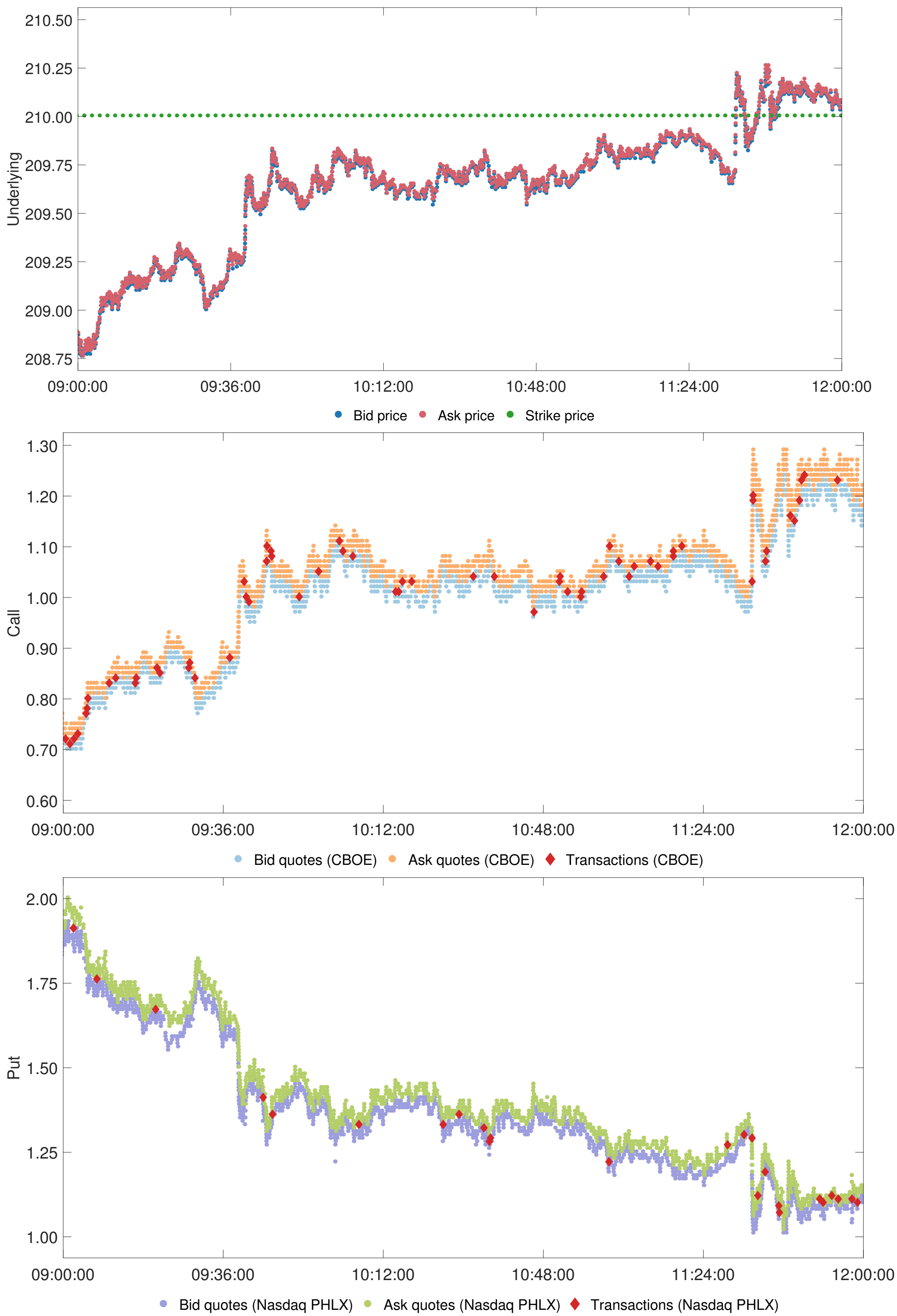


Figure B.1: An example of trade and quote data for at-the-money SPY options (strike price \$210, one week before expiration) traded on February 20, 2015 for the intra-daily interval between 9:00 and 12:00 CT. The underlying prices are depicted on the top plot. The middle plot contains 84,322 quotes and 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. Back to paper

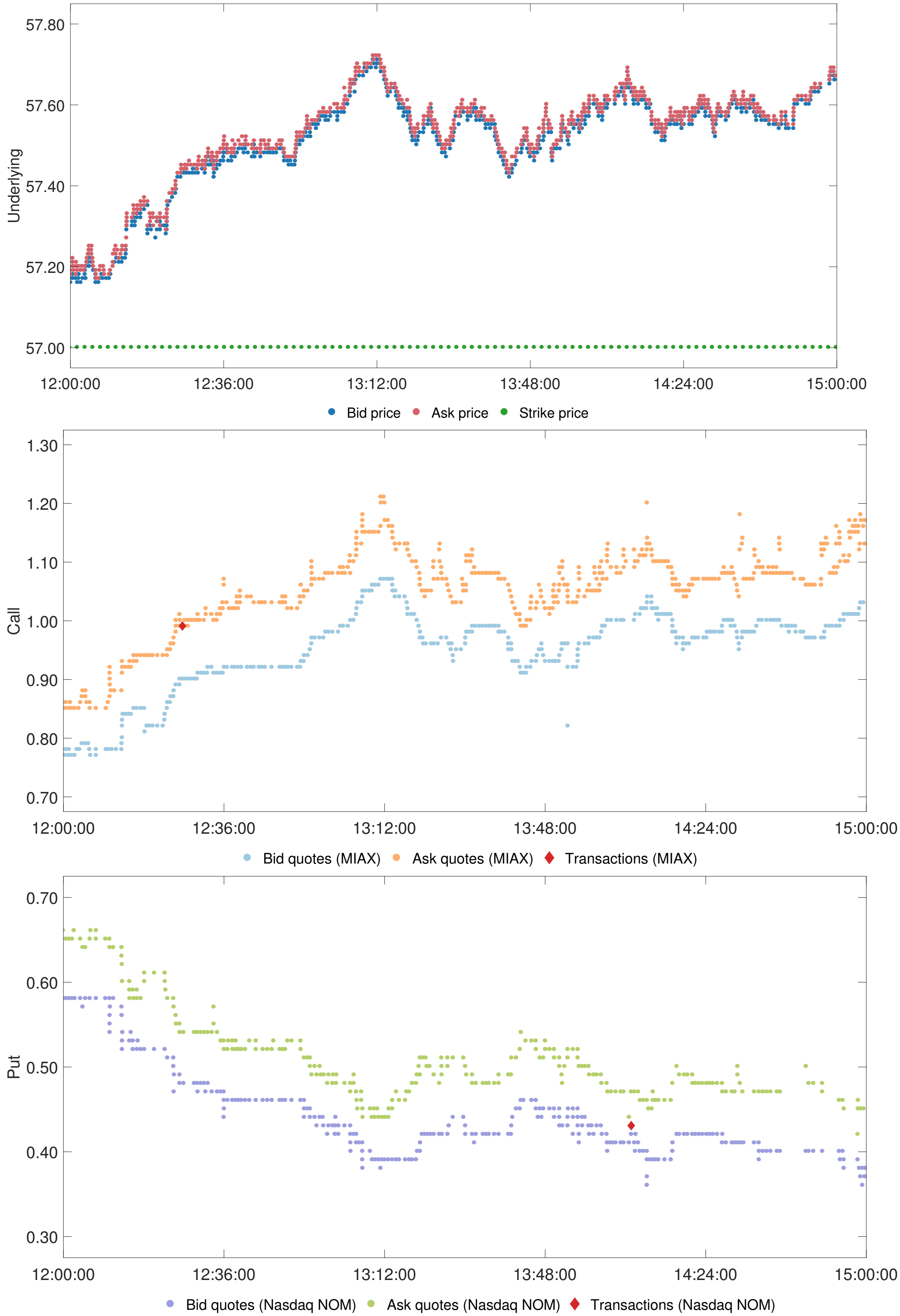


Figure B.2: An example of trade and quote data for at-the-money Electronic Arts options (strike price \$57, one week before expiration) traded on February 20, 2015 for the intra-daily interval between 9:00 and 12:00 CT. The underlying prices are depicted on the top plot. The middle plot contains 6,446 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. Back to paper

C Additional Evidence on Trading and Quoting Intensity

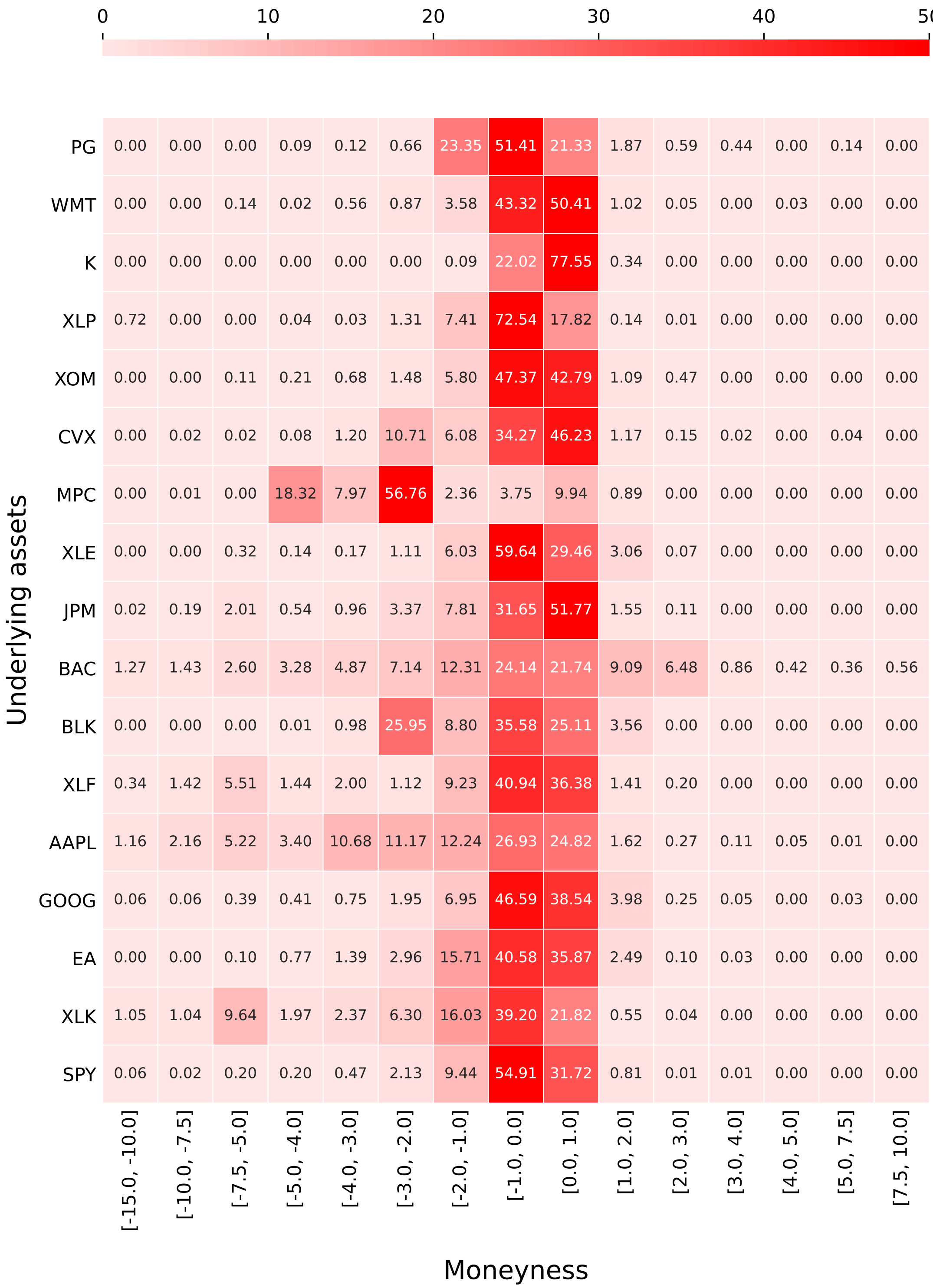
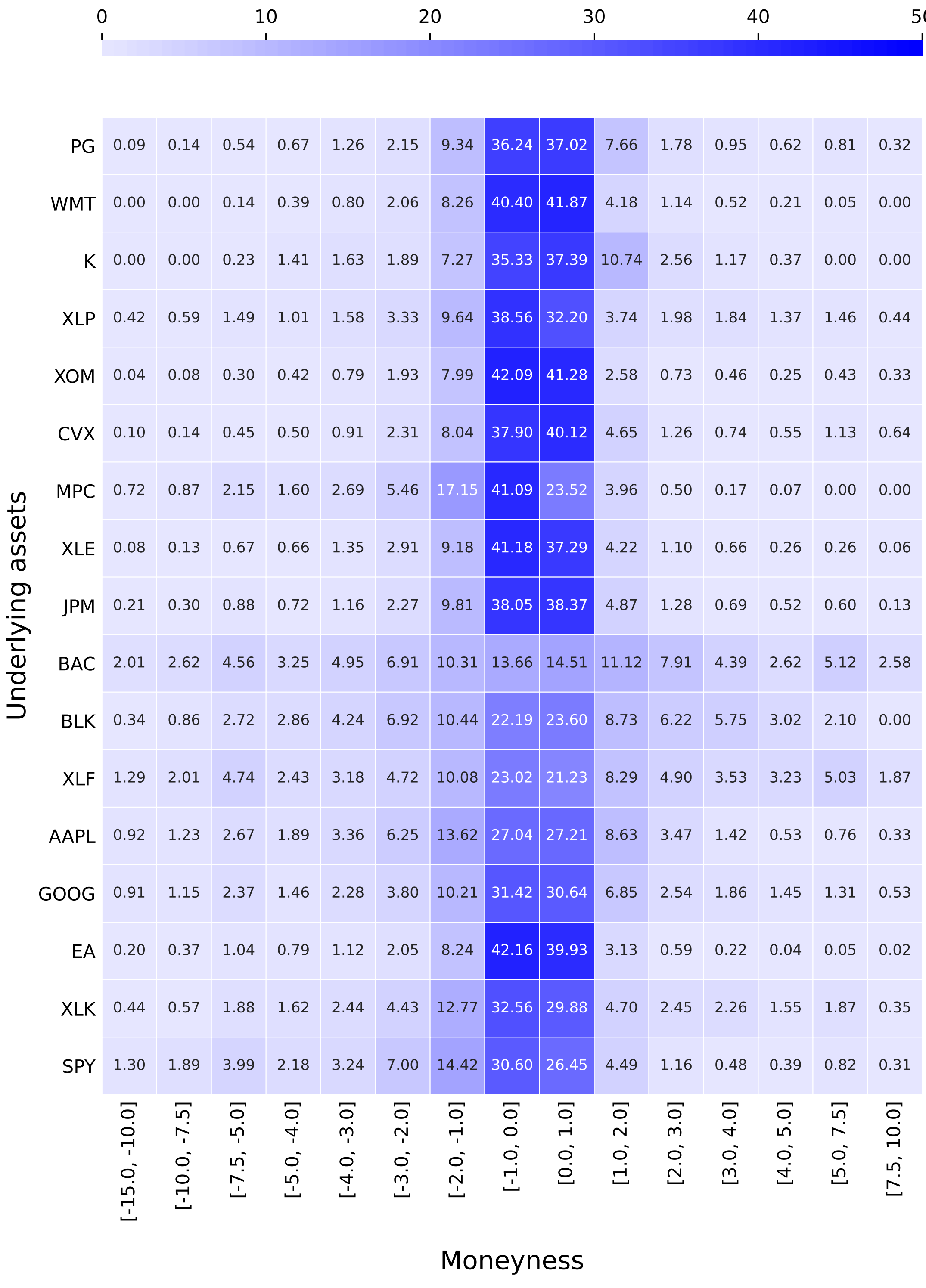


Figure C.3: Proportions of trading volumes (in %) by moneyness for the selected option classes. The results are obtained using the OPRA trade records for put and call option contracts traded between January 2nd and February 18th, 2015, which expire on February 20th, 2015. The moneyness is defined as $m = log(K/F_t)/(\sigma_t\sqrt{\tau})$. Back to paper



	[-15.0, -10.0]	[-10.0, -7.5]	[-7.5, -5.0]	[-5.0, -4.0]	[-4.0, -3.0]	[-3.0, -2.0]	[-2.0, -1.0]	[-1.0, 0.0]	[0.0, 1.0]	[1.0, 2.0]	[2.0, 3.0]	[3.0, 4.0]	[4.0, 5.0]	[5.0, 7.5]	[7.5, 10.0]
PG	0.09	0.14	0.54	0.67	1.26	2.15	9.34	36.24	37.02	7.66	1.78	0.95	0.62	0.81	0.32
WMT	0.00	0.00	0.14	0.39	0.80	2.06	8.26	40.40	41.87	4.18	1.14	0.52	0.21	0.05	0.00
K	0.00	0.00	0.23	1.41	1.63	1.89	7.27	35.33	37.39	10.74	2.56	1.17	0.37	0.00	0.00
XLP	0.42	0.59	1.49	1.01	1.58	3.33	9.64	38.56	32.20	3.74	1.98	1.84	1.37	1.46	0.44
XOM	0.04	0.08	0.30	0.42	0.79	1.93	7.99	42.09	41.28	2.58	0.73	0.46	0.25	0.43	0.33
CVX	0.10	0.14	0.45	0.50	0.91	2.31	8.04	37.90	40.12	4.65	1.26	0.74	0.55	1.13	0.64
MPC	0.72	0.87	2.15	1.60	2.69	5.46	17.15	41.09	23.52	3.96	0.50	0.17	0.07	0.00	0.00
XLE	0.08	0.13	0.67	0.66	1.35	2.91	9.18	41.18	37.29	4.22	1.10	0.66	0.26	0.26	0.06
JPM	0.21	0.30	0.88	0.72	1.16	2.27	9.81	38.05	38.37	4.87	1.28	0.69	0.52	0.60	0.13
BAC	2.01	2.62	4.56	3.25	4.95	6.91	10.31	13.66	14.51	11.12	7.91	4.39	2.62	5.12	2.58
BLK	0.34	0.86	2.72	2.86	4.24	6.92	10.44	22.19	23.60	8.73	6.22	5.75	3.02	2.10	0.00
XLF	1.29	2.01	4.74	2.43	3.18	4.72	10.08	23.02	21.23	8.29	4.90	3.53	3.23	5.03	1.87
AAPL	0.92	1.23	2.67	1.89	3.36	6.25	13.62	27.04	27.21	8.63	3.47	1.42	0.53	0.76	0.33
GOOG	0.91	1.15	2.37	1.46	2.28	3.80	10.21	31.42	30.64	6.85	2.54	1.86	1.45	1.31	0.53
EA	0.20	0.37	1.04	0.79	1.12	2.05	8.24	42.16	39.93	3.13	0.59	0.22	0.04	0.05	0.02
XLK	0.44	0.57	1.88	1.62	2.44	4.43	12.77	32.56	29.88	4.70	2.45	2.26	1.55	1.87	0.35
SPY	1.30	1.89	3.99	2.18	3.24	7.00	14.42	30.60	26.45	4.49	1.16	0.48	0.39	0.82	0.31

Figure C.4: Proportions of quotes (in %) by moneyness for the selected option classes. The results are obtained using the OPRA quote records for put and call option contracts traded between January 2nd and February 18th, 2015, which expire on February 20th, 2015. The moneyness is defined as $m = log(K/F_t)/(\sigma_t \sqrt{\tau})$.

Back to paper

D Additional Evidence on Spreads

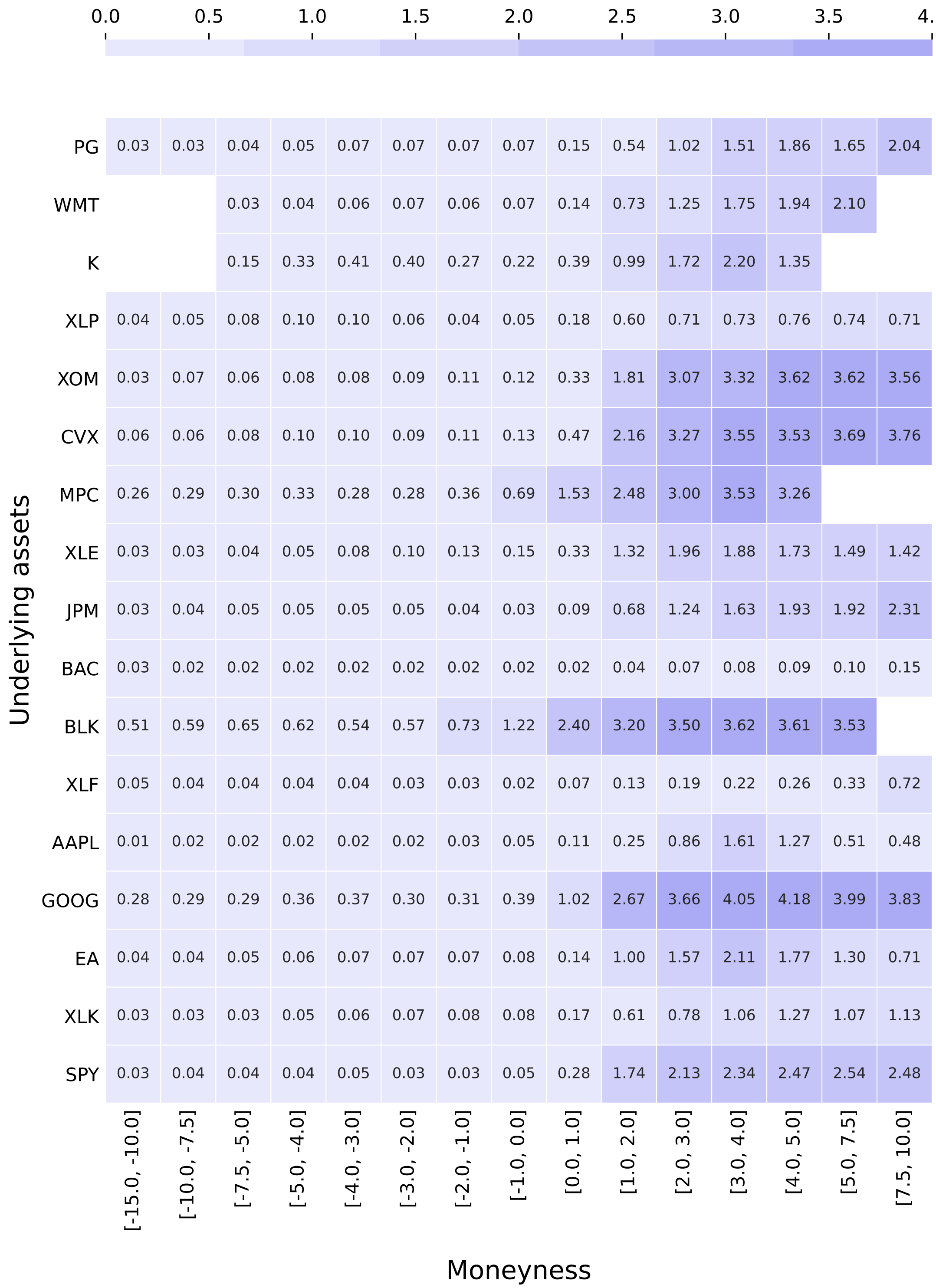


Figure D.5: Average spread levels by moneyness for the selected option classes. The results are obtained using the OPRA quote records for put option contracts traded between January 2nd and February 18th, 2015, which expire on February 20th, 2015. The moneyness is defined as $m = log(K/F_t)/(\sigma_t \sqrt{\tau})$.

Back to paper

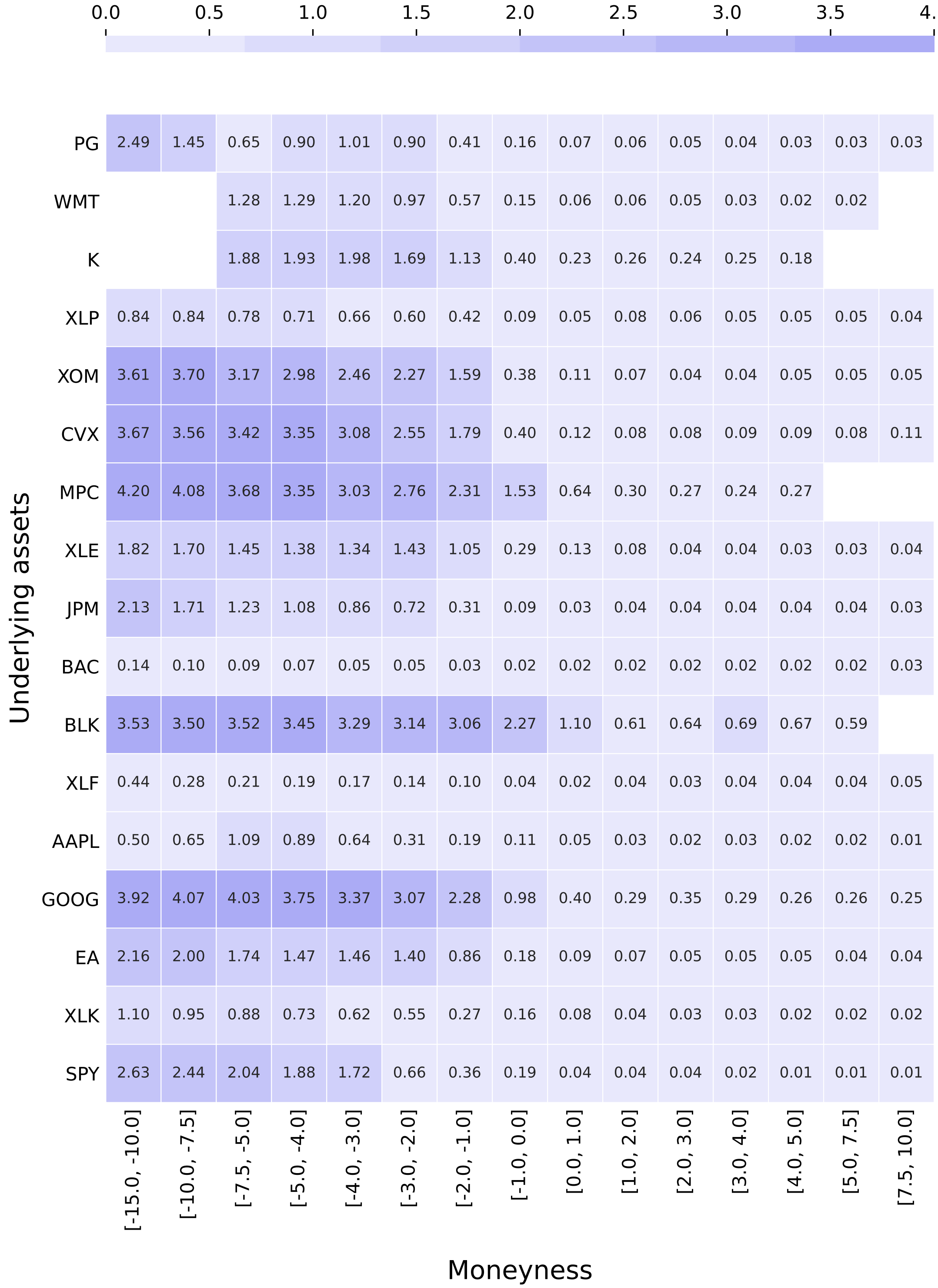
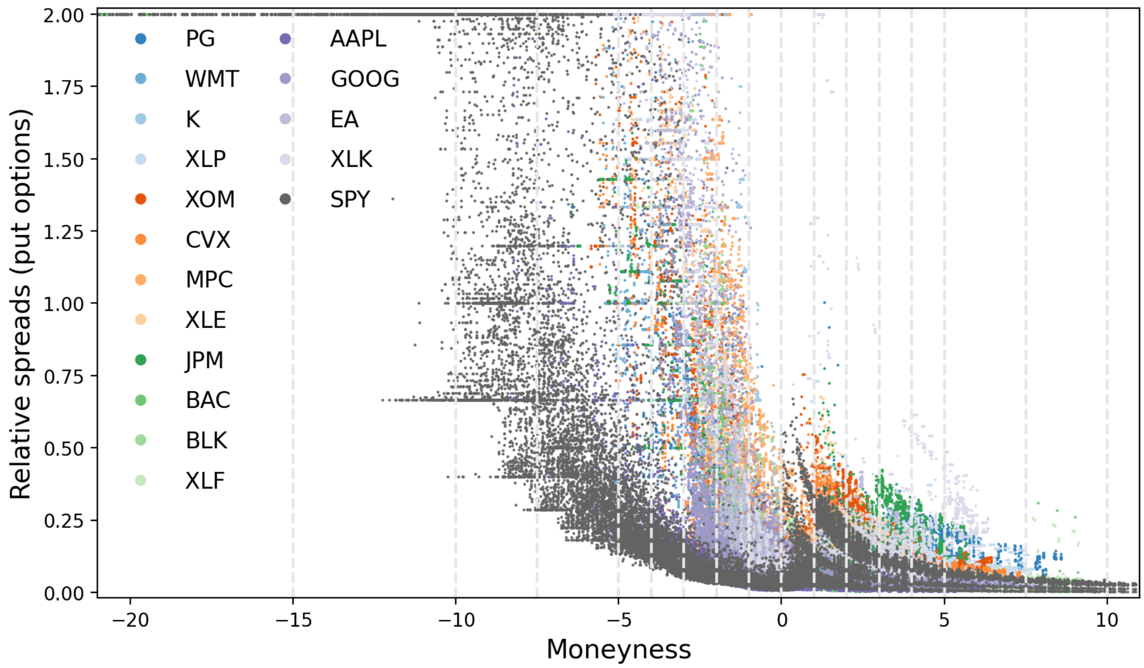


Figure D.6: Average spread levels by moneyness for the selected option classes. The results are obtained using the OPRA quote records for call option contracts traded between January 2nd and February 18th, 2015, which expire on February 20th, 2015. The moneyness is defined as $m = log(K/F_t)/(\sigma_t \sqrt{\tau})$.

Back to paper



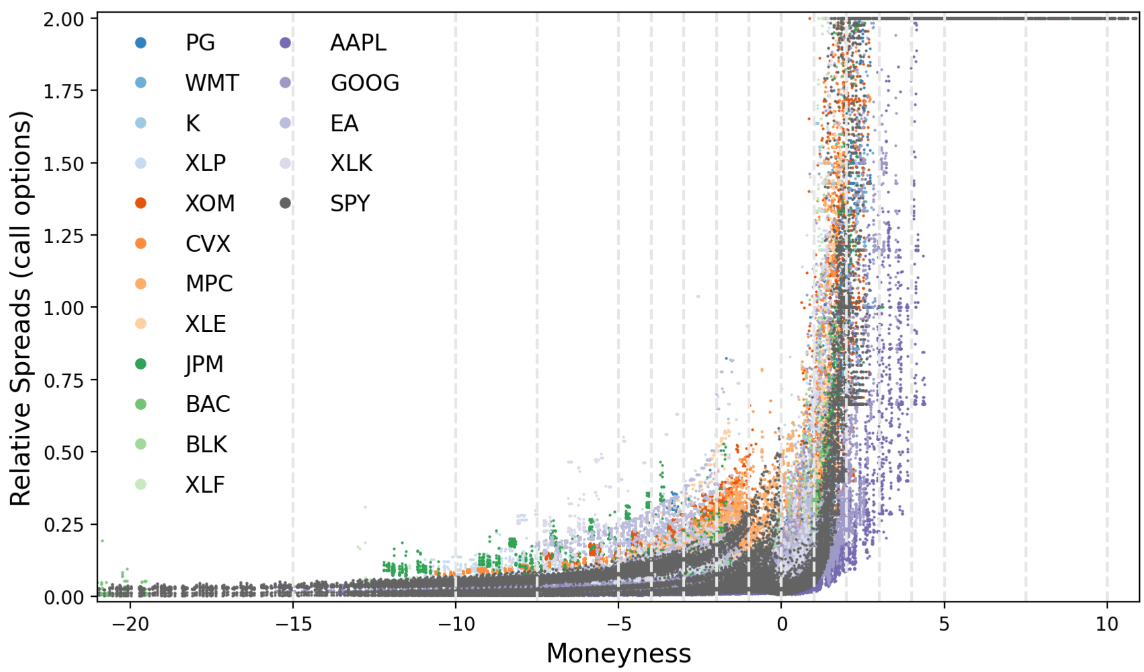


Figure D.7: Relative spread measures (average within 15 min intra-daily intervals observed between 9:00 and 10:00 CT) across underlying assets computed from the OPRA quote records for call (upper plot) and put (bottom plot) option contracts traded in the last week of January, 2015, which expire on February 20th, 2015 (3-4 weeks to maturity). The moneyness is defined as $m = log(K/F_t)/(\sigma_t \sqrt{\tau})$. Back to paper

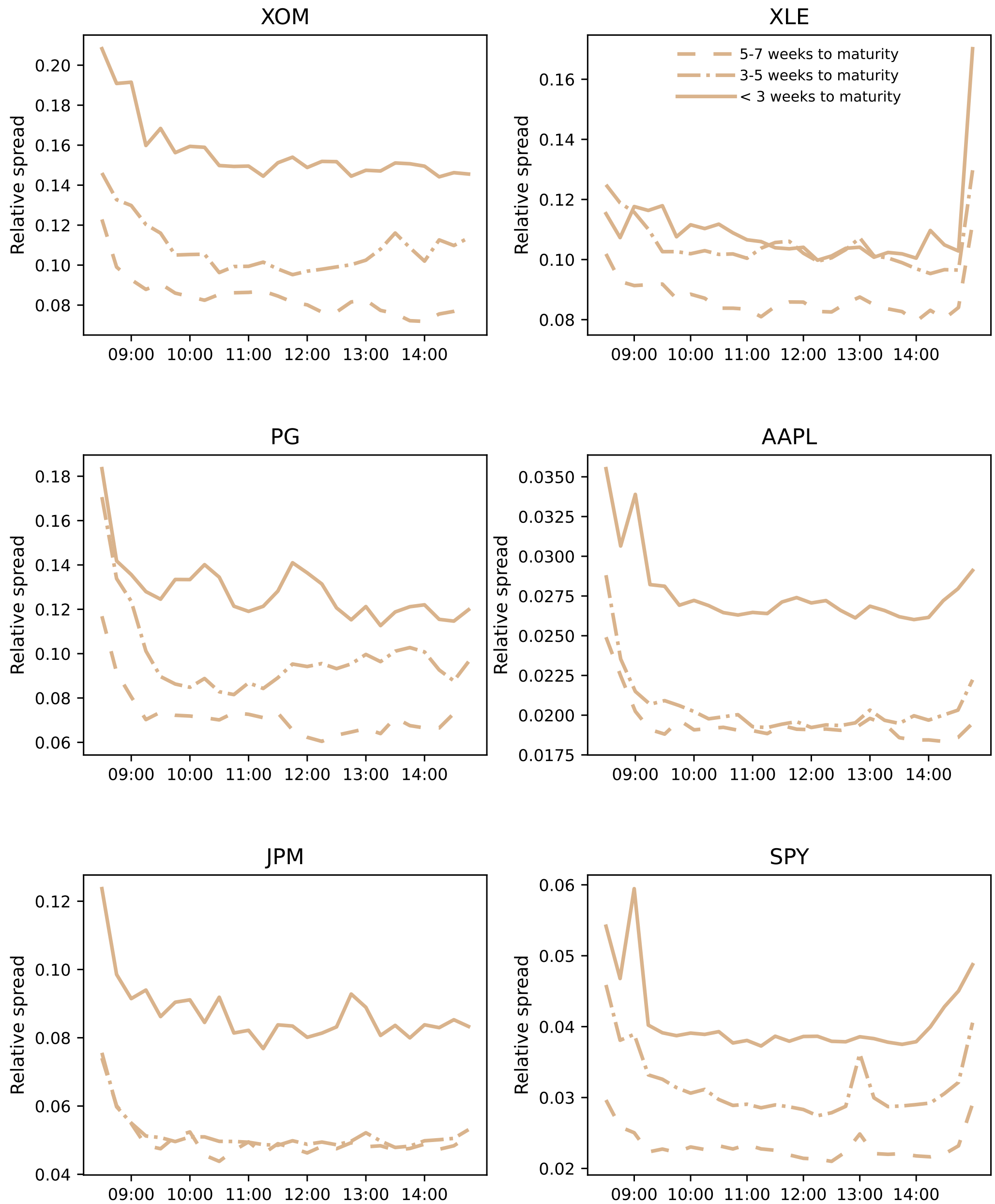


Figure D.8: Relative spread measures (average over 15 min intra-daily intervals) computed from the OPRA quote records for call option contracts traded in January and February, 2015, which expire on February 20th, 2015. The reported results are obtained using only those local 15-minute intervals, where moneyness is in the range $-1 \leq m \leq 1$. The moneyness is defined as $m = log(K/F_t)/(\sigma_t \sqrt{\tau})$. Back to paper

E MFIV Estimation

E.1 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 the 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 data 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 6.1, we use quote data for SPY and GOOG options traded on January 14, 2015 and expiring on January 17, 2015 (3 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 4.2, 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 platforms 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 the put-call parity for each calendar second, t. Based on F_t, select only out-of-the-money (OTM) options for each calendar second (puts with $K_i \leq F_t$ and calls with $K_i > F_t$). We denote mid-quote prices of such OTM 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 aim to extract. Since 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 3 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 $-2.5\sigma_t\sqrt{\tau} \leq \log(K_i/F_t) \leq 1.5\sigma_t\sqrt{\tau}$ 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 because OTM put options are presumably more liquid than OTM call options and have narrower spreads. Thus, the mid-quote prices of OTM puts are reliable for a wider moneyness range than the mid-quotes of OTM 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 $-15\sigma_t\sqrt{\tau} < \log(K_i/F_t) < -2.5\sigma_t\sqrt{\tau}$ 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 $1.5\sigma_t\sqrt{\tau} < \log(K_i/F_t) < 15\sigma_t\sqrt{\tau}$ 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 $|\log(K/F_t)| < 15\sigma_t\sqrt{\tau}$ with the strike step equal 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 option mid-quotes are extrapolated, by $\mathcal{I}_{l,t}$ and $\mathcal{I}_{r,t}$, respectively. Extrapolated mid-quotes 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 satisfying $-15\sigma_t\sqrt{\tau} < \log(K_i/F_t) < -2.5\sigma_t\sqrt{\tau}$ and $\mathcal{I}_{r,t}$ is a set of extrapolated strikes in the right tail satisfying $1.5\sigma_t\sqrt{\tau} < \log(K_i/F_t) < 15\sigma_t\sqrt{\tau}$.

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 day). Back to paper

E.2 Real-Time MFIV for Selected Stocks and ETFs

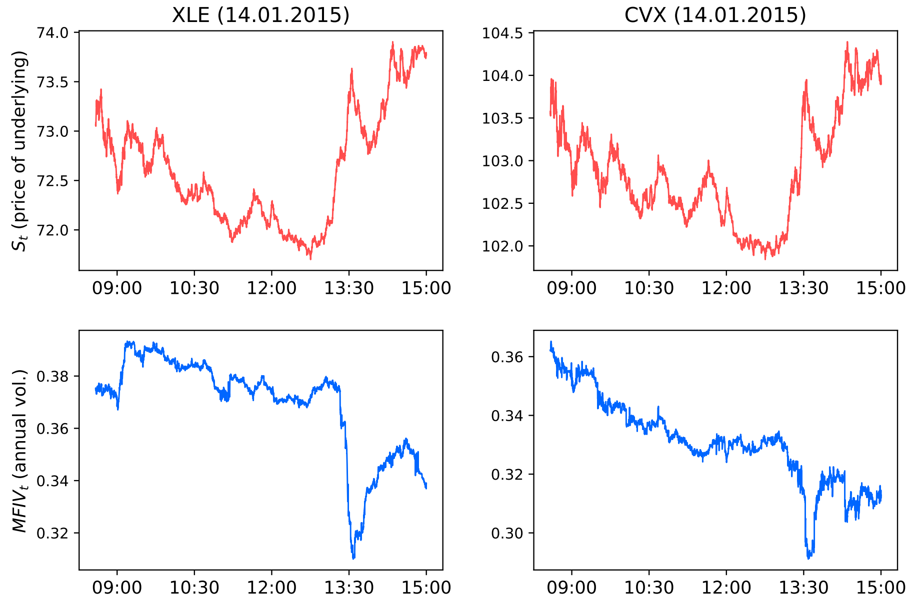


Figure E.9: On the top plots, intraday prices of XLE (left side) and CVX (right side) observed on January 14, 2015. On the bottom plots, intraday MFIV series constructed from XLE and CVX options on January 14, 2015 (expiring January 17, 2015). Real-Time MFIV is calculated on a second-by-second basis and normalized to annual volatility units.

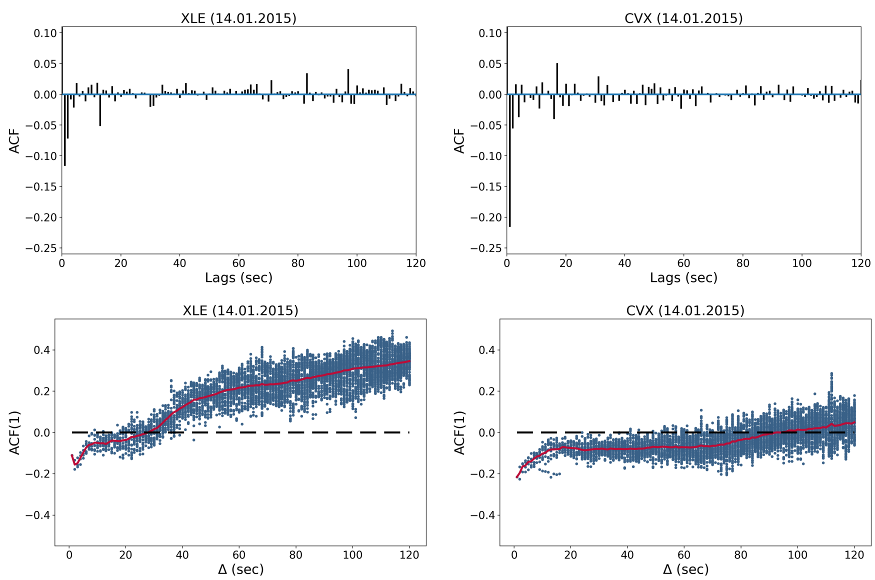


Figure E.10: Autocorrelation functions for MFIV series constructed with XLE options (left side) and with CVX options (right side) on January 14, 2015 (expired on January 17, 2015). Top plots show the autocorrelations as functions of lags constructed for the second-by-second increments of intraday MFIV (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.

Back to paper

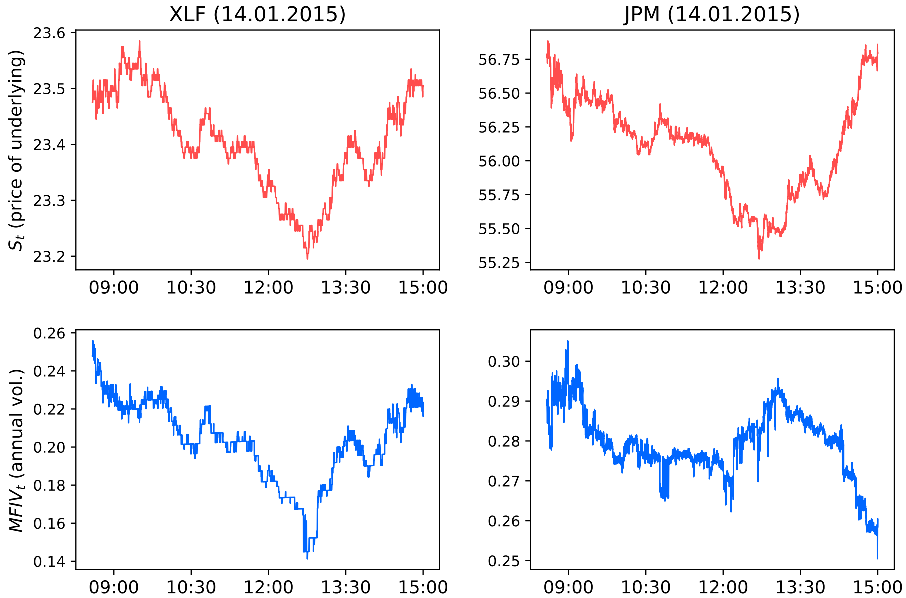


Figure E.11: On the top plots, intraday prices of XLF (left side) and JPM (right side) observed on January 14, 2015. On the bottom plots, intraday MFIV series constructed from XLF and JPM options on January 14, 2015 (expiring January 17, 2015). Real-Time MFIV is calculated on a second-by-second basis and normalized to annual volatility units.

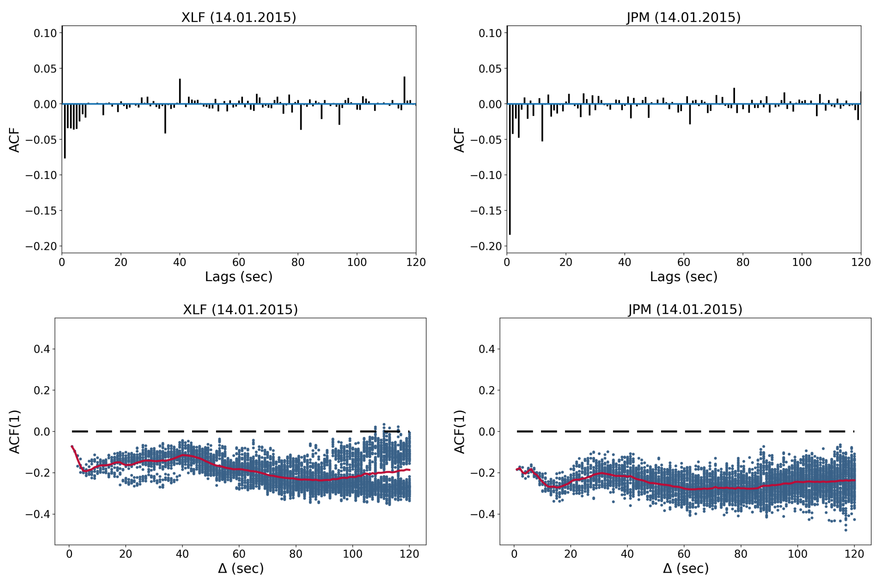


Figure E.12: Autocorrelation functions for MFIV series constructed with XLF options (left side) and with JPM options (right side) on January 14, 2015 (expired on January 17, 2015). Top plots show the autocorrelations as functions of lags constructed for the second-by-second increments of intraday MFIV (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. Back to paper

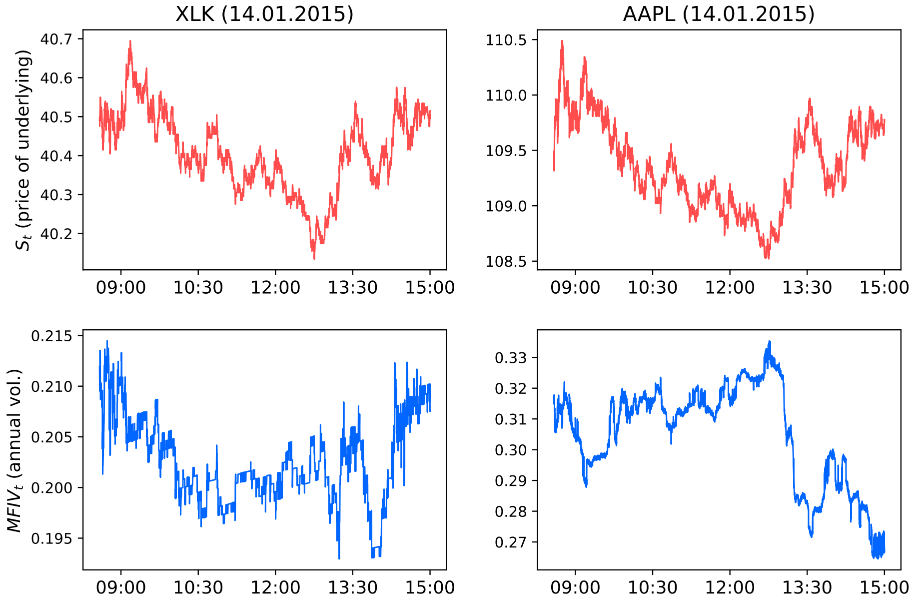


Figure E.13: On the top plots, intraday prices of XLK (left side) and AAPL (right side) observed on January 14, 2015. On the bottom plots, intraday MFIV series constructed from XLK and AAPL options on January 14, 2015 (expiring January 17, 2015). Real-Time MFIV is calculated on a second-by-second basis and normalized to annual volatility units.

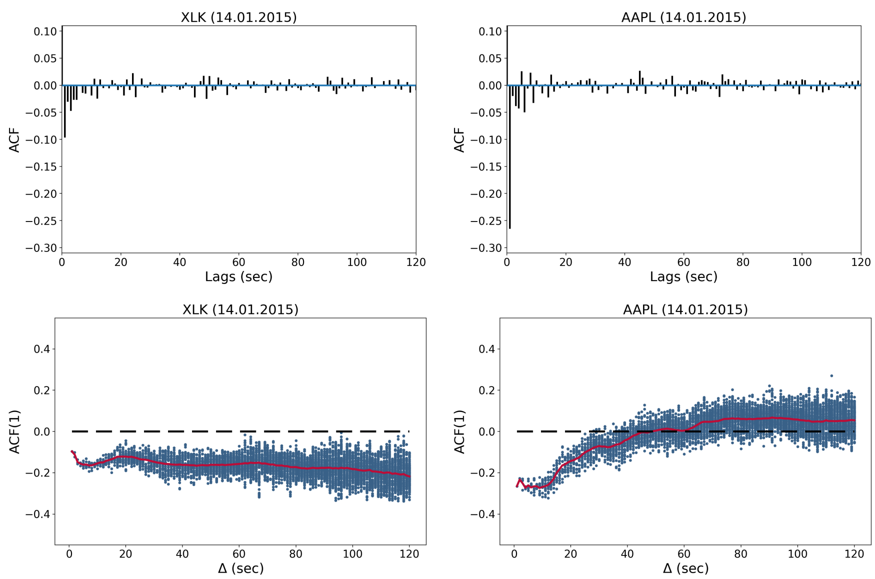


Figure E.14: Autocorrelation functions for MFIV series constructed with XLK options (left side) and with AAPL options (right side) on January 14, 2015 (expired on January 17, 2015). Top plots show the autocorrelations as functions of lags constructed for the second-by-second increments of intraday MFIV (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. Back to paper

F RND Estimation

F.1 Mixture of Lognormal Density Methodology

According to Ritchey (1990), Melick and Thomas (1997) and Liu et al. (2007), the RND $f_{\mathbb{Q}}(x)$ associated with a given time to maturity τ can be modeled by a mixture of M lognormal distributions as

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

with

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

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 $f_{\mathbb{Q}}(x|\boldsymbol{\theta})$ 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 $f_{\mathbb{Q}}(x|\boldsymbol{\theta})$ in Equation (F.1) has a flexible shape determined by $(3M-2)$ free parameters, as opposed to only one free parameter in the case where $f_{\mathbb{Q}}(x|\boldsymbol{\theta})$ is modeled with a single lognormal 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 $f_{\mathbb{Q}}(x|\boldsymbol{\theta})$ is simply the weighted average of the prices of call options with the same strike given by the Black (1976) option pricing formula, as implied by Equation (F.1). That is,

$$C(K|\boldsymbol{\theta}, r_f, \tau) = \sum_{i=1}^{M} w_i C_B(K, F_i, \sigma_i, r_f, \tau), \tag{F.3}$$

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 $f_{\mathbb{Q}}(x|\boldsymbol{\theta})$ in (F.1) 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,

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \left(C_{\text{market}}(K_i) - C(K_i|\boldsymbol{\theta})\right)^2, \tag{F.4}$$

where $C(\cdot|\boldsymbol{\theta})$ is the theoretical option price given in Equation (F.3).

The minimization in Equation (F.4) is a standard nonlinear least squares estimation problem which can be solved numerically by standard optimization algorithms. Back to paper

F.2 Data Preparation and MLN-based RND Estimation

Steps undertaken to prepare data for the RND estimation and information about how to choose M are described below:

1. For each cross-section of options at 30 minutes pre or post FOMC announcement on 18 March 2015, filter out potentially erroneous quote entries using the filtering rules F1, F2, F3, F5c and F6 detailed in Subsection 4.2. We use the median price to aggregate "simultaneous" bid or ask prices at the same millisecond for the filtering rule F6 and use the midquotes of the (aggregated) best bid and ask prices

of options as the observed market prices of the options. In order to reduce potential distortions in mid-quotes due to the bounded bid-ask corridor, we further discard quote records with zero bid prices.

2. Compute the "implied" forward price F_t of the underlying asset using the put-call parity relation. Specifically, $F_t = K^* + e^{r_f \tau} (C(K^*) - P(K^*))$, where r_f is the risk-free rate, τ is the time to maturity of option contracts, and $K^*$is the strike price at which the absolute put-call difference is smallest. We obtain the annualized risk-free rates available on 18 March 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.

3. Convert OTM puts into ITM calls using put-call parity and use call options to estimate the RNDs. This is because ITM options are typically less liquid than OTM options (e.g. Aït-Sahalia and Lo, 1998, Song and Xiu, 2016).

4. Remove the deepest OTM (ITM) call option when its strike is suspiciously far away from the nearest other strike, i.e., when the strike gap between the deepest and second deepest OTM (ITM) options is more than twice that between the second and third deepest OTM (ITM) options. Such an option which lies in the extreme tail and does not have any nearby strikes with quotes is likely an outlying option which should be removed to avoid distorting the estimated RND.

5. Scale both option prices and strike prices by the forward price F_t. This scaling method is in line with the assumption that option prices are homogeneous of order one in the forward price, which is often imposed in the prior literature (see, for example, Aït-Sahalia and Lo, 1998, Chen and Xu, 2014, Song and Xiu, 2016, Dalderop, 2020). With this scaling, the risk-neutrality requirement that $\sum_{i=1}^{M} w_i F_i = F_t$ simplifies to $\sum_{i=1}^{M} w_i F_i = 1$, where F_i is now interpreted as the ratio of the forward price of the i^{th} lognormal density to the forward price F_t.

6. Fit a constrained convex least squares regression proposed by Härdle and Hlávka (2009) to a given cross-section of scaled option prices and use the fitted option prices in the RND estimation. This helps preclude arbitrage opportunities and ensure that option prices are monotonic and convex with respect to strike prices (or moneyness).

7. Minimize the objective function in Equation (F.4) to estimate the RND implied by a cross-section of options. We consider various mixtures of M lognormal distributions, with M ranging from 1 to 4 and satisfying the identification condition that the number of free parameters $(3M - 2)$ in the mixture is no greater than the number of options (N) in the cross-section. To avoid the possibility of converging to a local minimum when minimizing the objective function in Equation (F.4) (which is highly nonlinear and not globally convex in the parameters $\boldsymbol{\theta}$ when $M > 2$ (Bondarenko, 2003)), we perform the minimization procedure at 20 different sets of starting values to ascertain whether a global minimum is reached. In addition, to avoid the issue that a lognormal mixture can overfit a small option cross-section and result in a spiky RND estimate (Bondarenko, 2003), for each cross-section we only report the estimated RND curve associated with the highest M (between 1 and 4) that is free from spurious spikes. Back to paper

F.3 Risk-neutral Density Plots for other Underlyings

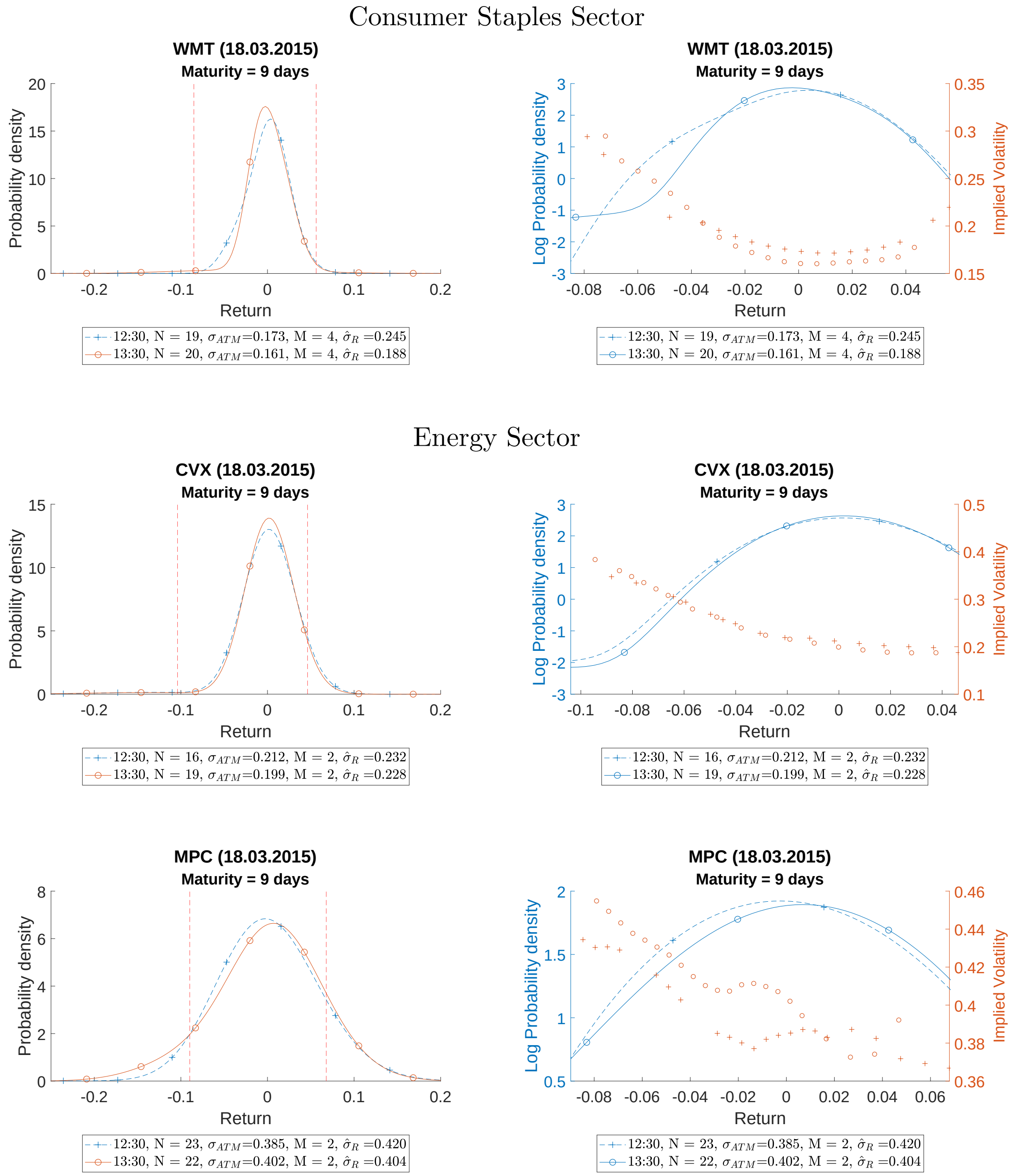


Figure F.15: The left plots depict the estimated RNDs, as a function of return $S_T/F_t - 1$, for the shortest time to maturity obtained from intraday OPRA data for options written on the underlyings in the **Consumer Staples** and **Energy** sectors at **30 minutes** before and after the FOMC announcement at 14:00 Eastern time (i.e. 13:00 CT) on 18 March 2015. Each RND curve is estimated from a mixture of M lognormal distributions. In each subplot, the number of strikes (N), the ATM Black-Scholes implied volatility (σ_{ATM}), the number of lognormal densities in each mixture (M), and the estimated annualized standard deviation ($\hat{\sigma}_R$) of each RND curve are reported. Vertical red dashed lines indicate the observed return ($K/F_t - 1$) range. Right plots show the logarithm of the RNDs (left axis) and the Black-Scholes implied volatility (right axis) over the observed return range. Time in each plot is CT. Back to paper

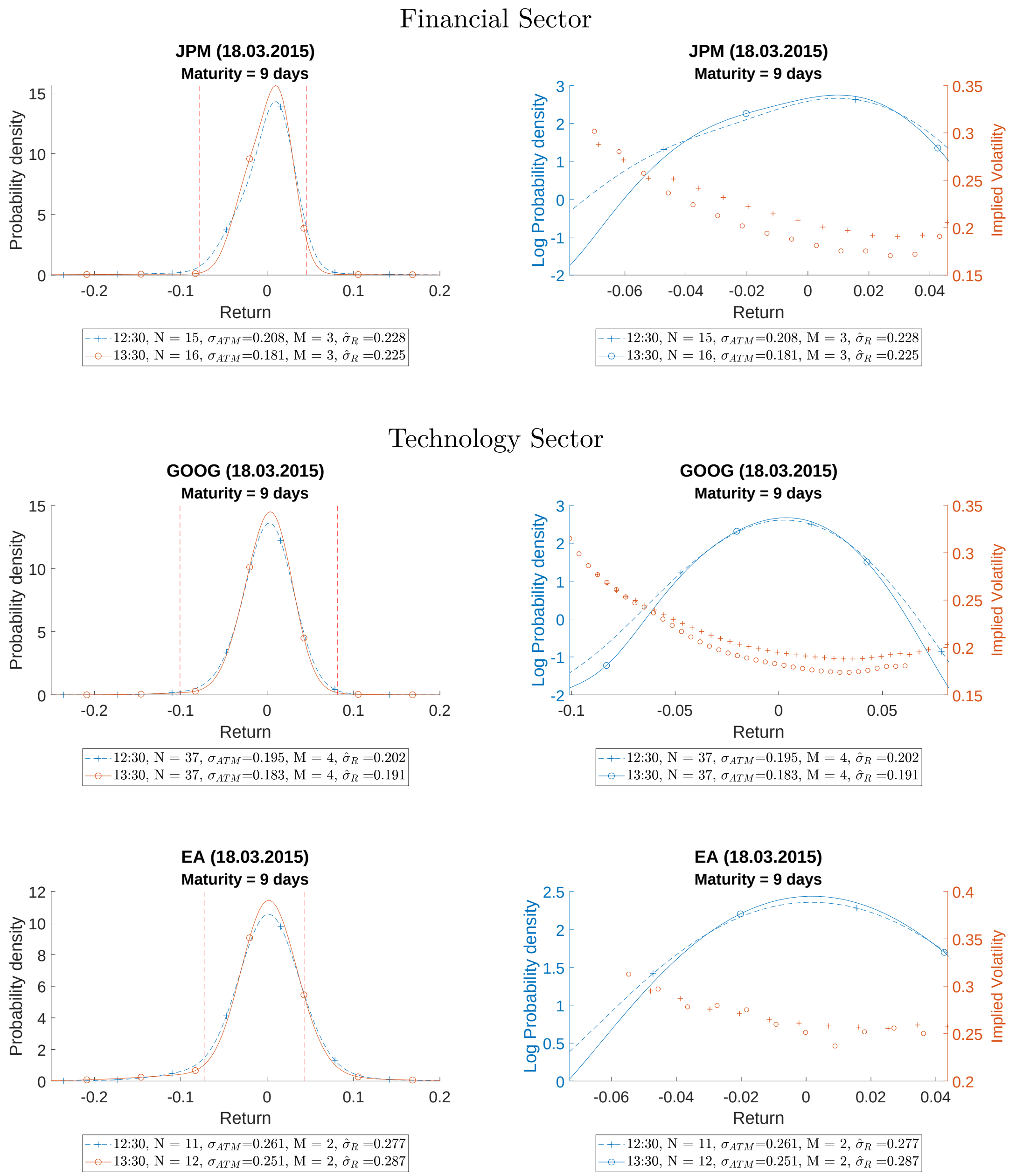


Figure F.16: The left plots depict the estimated RNDs, as a function of return $S_T/F_t - 1$, for the shortest time to maturity obtained from intraday OPRA data for options written on the underlyings in the **Financial** and **Technology** sectors at **30 minutes** before and after the FOMC announcement at 14:00 Eastern time (i.e. 13:00 CT) on 18 March 2015. Each RND curve is estimated from a mixture of M lognormal distributions. In each subplot, the number of strikes (N), the ATM Black-Scholes implied volatility (σ_{ATM}), the number of lognormal densities in each mixture (M), and the estimated annualized standard deviation ($\hat{\sigma}_R$) of each RND curve are reported. Vertical red dashed lines indicate the observed return ($K/F_t - 1$) range. Right plots show the logarithm of the RNDs (left axis) and the Black-Scholes implied volatility (right axis) over the observed return range. Time in each plot is CT. Back to paper

References

Aït-Sahalia, Y. and Lo, A. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, 3(2):499–547.

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

Bondarenko, O. (2003). Estimation of risk-neutral densities using positive convolution approximation. *Journal of Econometrics*, 116(1-2):85–112.

Chen, S. X. and Xu, Z. (2014). On implied volatility for options - Some reasons to smile and more to correct. *Journal of Econometrics*, 179(1):1–15.

Dalderop, J. (2020). Nonparametric filtering of conditional state-price densities. *Journal of Econometrics*, 214(2):295–325.

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

Härdle, W. and Hlávka, Z. (2009). Dynamics of state price densities. *Journal of Econometrics*, 150(1):1–15.

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.

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

Song, Z. and Xiu, D. (2016). A tale of two option markets: Pricing kernels and volatility risk. *Journal of Econometrics*, 190:176–196.