

Option Trade Classification, Demand Measurement, and Stock Return Predictability

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.)

von der KIT-Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Caroline Grauer, M. Sc.

Tag der mündlichen Prüfung: 06.08.2026

Referentin: Prof. Dr. Marliese Uhrig-Homburg

Korreferent: Prof. Dr. Christian Schlag

Karlsruhe 2026

Acknowledgments

This dissertation was written during my time as a research associate at the Institute for Finance at the Karlsruhe Institute of Technology (KIT), under the supervision of Prof. Dr. Marliese Uhrig-Homburg. I am especially grateful for her trust, support, and the opportunities she gave me throughout my doctoral studies, which allowed me to grow both as a researcher and as a person.

I would also like to thank Prof. Dr. Philipp Schuster, who supported me as a co-author and co-supervisor. I greatly appreciated our discussions, his thoughtful feedback, and the many insights that helped shape the research presented in this dissertation. I am also thankful to my colleagues in the research group for the many discussions, collaborations, and shared experiences over the years.

My thanks also go to Prof. Dr. Andreas Oberweis, Prof. Dr. Christian Schlag, and Prof. Dr. Julian Thimme for serving on my examination committee and for taking the time to engage with my work.

During my doctoral studies, I had the opportunity to spend four months as a visiting researcher at the Bauer College of Business at the University of Houston. I am very grateful to Dr. Steffen Hitzemann for welcoming me and for his support throughout my stay. The experience broadened my academic perspective and was equally enriching on a personal level.

Finally, I would like to thank my family for their unconditional love and support throughout the years. I am deeply grateful to know that I can always count on them.

Contents

List of Figures	v
List of Tables	vii
1 Introduction	1
1.1 Motivation	1
1.2 Structure of the Dissertation	4
2 Beyond the Limits of Trade Classification in Option Markets	7
2.1 Introduction	7
2.2 Option Market Structure and Limits of Stock-Based Trade Classification	10
2.2.1 Trade Initiation and Customer Trade Direction	11
2.2.2 Existing Trade Classification Methods	12
2.2.3 Option Market Features Limiting Trade Classification	15
2.2.4 Evidence from the Literature	17
2.3 Data and Benchmark Construction	19
2.3.1 Options Market Data	20
2.3.2 Benchmark for Customer Trade Direction	21
2.3.3 Representativeness of the Benchmark	22
2.3.4 Preregistered Two-Step Research Design	22
2.3.5 Sample Description	24
2.4 In-Sample Analysis of Classification Rules	28

Contents

2.4.1	Established Rules	28
2.4.2	Trade Size Rule for Limit Orders	31
2.4.3	Economic Mechanism	32
2.4.4	Trade Size Rule Refinements	34
2.4.5	Depth Rule for Midpoint Trades	36
2.5	Out-of-Sample Performance	37
2.5.1	Established Rules	37
2.5.2	Trade Size Rule	39
2.5.3	Subsample Evidence and Refinements	43
2.5.4	Depth Rule	46
2.6	Recommended Classification Algorithm	48
2.6.1	Combined Classification Procedure	48
2.6.2	Cross-Sectional Performance	49
2.6.3	Time-Series Variation in Classification Performance	52
2.7	From Classification to Customer Demand	54
2.7.1	Economic Interpretation of Identified Customer Side	54
2.7.2	Generalizability Beyond Sample Exchanges	56
2.8	Conclusion	57
3	Unmasking Option Demand:	
	Trade Classification and Stock Return Predictability	59
3.1	Introduction	59
3.2	Informed Option Trading and Stock Return Predictability	62
3.3	Signal Construction and Empirical Design	64
3.3.1	Data	64
3.3.2	Option-Based Signals	65
3.3.3	Empirical Design	66
3.3.4	Descriptive Statistics	68
3.4	The Impact of Trade Classification on Stock Return Predictability . .	70

3.4.1	Portfolio Sort Evidence	71
3.4.2	Time Variation in Predictability	72
3.4.3	Persistence and Incremental Information	74
3.5	Limit Orders and Classification Disagreement	77
3.5.1	Predictability and Classification Disagreement	77
3.5.2	Predictability and Customer Limit Order Activity	81
3.6	Robustness Across Option Demand Measures	86
3.7	Conclusion	89
4	When Measurement Matters: Coverage, Trader Identity, and Trade Classification in Option Markets	90
4.1	Introduction	90
4.2	Option Data Environments and Measurement Dimensions	94
4.2.1	Option Data Environments	95
4.2.2	Sources of Measurement Differences	95
4.2.3	Empirical Decomposition of Measurement Dimensions	100
4.3	Empirical Applications and Research Design	103
4.3.1	Option Order Imbalance and Daily Stock Returns	103
4.3.2	Put-Call Ratio and Weekly Stock Returns	104
4.3.3	Persistence and Incremental Information	105
4.3.4	Dealer Gamma Inventory and Stock Return Volatility	106
4.3.5	Option Expensiveness as a Price-Based Benchmark	107
4.3.6	Sample Construction and Summary Statistics	109
4.3.7	Correlations Across Alternative Measures	110
4.4	Measurement Dimensions and Predictive Inference	116
4.4.1	Exchange Coverage	116
4.4.2	Trader Identity	121
4.4.3	Trade Classification	125
4.5	Implications for Empirical Measurement in Option Markets	129

Contents

4.5.1	Relative Importance of Measurement Dimensions	129
4.5.2	Comparing Open/Close and LiveVol Measures	130
4.6	Comparing Quantity-Based Measures and Price-Based Proxies	133
4.7	Conclusion	134
5	Summary and Outlook	136

List of Figures

2.1	Hybrid methods	14
2.2	Time-series behavior of combined classification procedure	53
3.1	Cumulative returns by classification method	74
3.2	Predictive regressions by classification method	75
3.3	Cumulative returns by classification disagreement	79
3.4	Bivariate regressions in high disagreement subsample	80
3.5	Time-series and cross-sectional variation in customer limit order activity	81
3.6	Cumulative returns by customer limit order share	83
3.7	Bivariate regressions in high customer limit-order share subsample . .	85
3.8	Cumulative returns of option order imbalance strategies	87
3.9	Bivariate regressions using option order imbalances	88
4.1	Market fragmentation across U.S. option exchanges	96
4.2	Accuracy of alternative trade classification methods	100
4.3	Cumulative returns by exchange coverage	117
4.4	Predictive regressions by exchange coverage	119
4.5	Cumulative returns by trader identity	122
4.6	Predictive regressions by trader identity	123
4.7	Cumulative returns by classification method	126
4.8	Predictive regressions by classification method	127
4.9	Cumulative returns by data source	131

List of Figures

4.10 Bivariate regressions by data source	132
---	-----

List of Tables

2.1	Market design features of sample option exchanges	16
2.2	Empirical evidence on trade classification across markets	18
2.3	Trader-type pairings	23
2.4	Summary statistics for in-sample option trades	26
2.5	Summary statistics for out-of-sample option trades	27
2.6	In-sample performance of trade classification rules	30
2.7	In-sample performance with trade size rule by trade size	32
2.8	In-sample performance with trade size rule by trade condition	33
2.9	In-sample performance with trade size rule by quote location	34
2.10	In-sample performance of trade size rule refinements	35
2.11	Out-of-sample performance of trade classification rules	38
2.12	Out-of-sample performance of trade size rule	40
2.13	Out-of-sample performance with trade size rule	41
2.14	Out-of-sample performance with trade size rule by trade size	43
2.15	Out-of-sample performance with trade size rule by trade condition . .	44
2.16	Out-of-sample performance with trade size rule by quote location . .	45
2.17	Out-of-sample performance of trade size rule refinements	46
2.18	Out-of-sample performance of depth rule	47
2.19	Cross-sectional variation of combined classification procedure	50
2.20	Intraday returns to customer option trades	55
2.21	Market design features of non-sample U.S. options exchanges	56

List of Tables

3.1	Summary statistics for option-based signals and stock returns	69
3.2	Signal correlations across classification methods	70
3.3	Return predictability by classification method	72
3.4	Return predictability by classification disagreement	78
3.5	Return predictability by customer limit order share	82
3.6	Return predictability of option order imbalance	87
4.1	Trader identity assumptions and measured demand	98
4.2	Identification of measurement dimensions	102
4.3	Sample characteristics and measurement diagnostics	110
4.4	Correlations across measurement dimensions	111
4.5	Option order imbalance and next-day stock return predictability . . .	113
4.6	Put-call ratio and weekly stock return predictability	114
4.7	Dealer gamma inventory and next-day stock return volatility	115

Chapter 1

Introduction

1.1 Motivation

The rapid growth of U.S. options markets highlights the strong demand for option exposure. In 2025, total trading volume exceeded 15.2 billion contracts, marking the sixth consecutive annual volume record and nearly tripling the volume observed in 2019.¹ This growth reflects the broad economic role of options. They allow investors to express directional views, hedge risks, trade volatility, and obtain nonlinear exposure to underlying assets, making them attractive to a broad set of investors seeking to speculate, manage risk, or act on private information.

In frictionless settings such as the Black–Scholes model, options are redundant securities that can be replicated dynamically using the underlying asset and the risk-free asset (Black and Scholes, 1973; Merton, 1973). However, trading frictions, capital constraints, short-sale constraints, discrete hedging, and limits to arbitrage make option contracts distinct from synthetic positions in the underlying asset. As emphasized by Black (1975), options provide high embedded leverage and can therefore offer a capital-efficient way to obtain exposure to the underlying asset.

These features make options a natural venue for informed trading. Investors with positive private information can take positive-delta option positions, for example by buying calls or selling puts, to obtain leveraged exposure to future price increases. Conversely, investors with negative private information can take negative-delta option positions, for example by buying puts or selling calls, to benefit from expected

¹The Options Clearing Corporation, Historical Volume Statistics, <https://www.theocc.com/market-data/market-data-reports/volume-and-open-interest/historical-volume-statistics>, accessed on June 20, 2026.

price declines without directly short selling the stock. Consequently, option demand may reveal private information before it is fully incorporated into stock prices. A large empirical literature therefore examines whether option-market activity predicts subsequent stock returns and contributes to price discovery (e.g., Easley et al., 1998; Pan and Poteshman, 2006; Hu, 2014; Ge et al., 2016; Bondarenko and Muravyev, 2025). This literature provides evidence that signed option volume, option order imbalances, and put-call ratios contain information about future stock price movements.

Option trading also affects intermediaries, who often absorb customer demand and accumulate option inventory. These positions expose intermediaries to movements in the underlying asset. Intermediaries typically hedge the initial delta exposure by trading in the underlying stock, which can help transmit information contained in option demand to stock prices (Hu, 2014). As prices move, however, option deltas change, requiring intermediaries to rebalance their hedges dynamically. This rebalancing links option demand to subsequent stock-market dynamics. Theoretically, positive dealer gamma tends to dampen price movements, as intermediaries rebalance by trading against movements in the underlying, whereas negative dealer gamma can amplify price movements, as intermediaries hedge in the same direction as the underlying price change. Consistent with this mechanism, empirical evidence shows that dealer gamma exposure is related to future stock volatility (Ni et al., 2021; Dim et al., 2025). Option demand is therefore informative not only about investors' expectations, but also about intermediary risk and its implications for the underlying market.

Departures from the frictionless benchmark also imply that option demand can affect option prices. If options could be replicated costlessly and intermediaries were unconstrained, demand imbalances would not move prices independently of fundamentals. In practice, however, one-sided end-user demand can matter when intermediaries face inventory constraints, hedging costs, and limits to arbitrage (Shleifer and Vishny, 1997). Intermediaries who absorb this demand may require compensation for inventory risk and exposure to volatility and jump risk. Option prices may therefore reflect not only expectations about future volatility, but also the compensation required for absorbing end-user demand (Bollen and Whaley, 2004; Gârleanu et al., 2009). This demand-pressure component helps motivate the use of option expensiveness, such as the difference between implied and realized volatility, as a price-based proxy for option demand pressure.

Across these mechanisms, trade direction is central. The informational content

of option trading depends on whether investors buy or sell option exposure, because the sign of demand determines whether option order flow conveys positive or negative information about the underlying asset. Intermediary risk depends on the same distinction, because investor purchases and sales generate different dealer inventory positions whose hedge rebalancing may either reinforce or offset movements in the underlying asset and thereby affect future volatility. Option prices also depend on trade direction when demand imbalances require intermediaries to absorb net buying or selling pressure and adjust quotes to compensate for the resulting inventory risk and hedging costs. The empirical content of option-market activity therefore depends not only on trading volume, but also on who trades and on which side of the market.

This creates a central measurement challenge. Empirical research on option demand and dealer inventory typically relies on two data environments that capture different parts of trading activity. Open/Close data report buy and sell volumes by trader type at the exchange-day level, distinguishing between customers, firms, and broker-dealers, but they are available only for selected exchanges. Transaction-level data, by contrast, provide broad market coverage and detailed trade-level information, but they do not identify the customer side of a trade or the trader type behind the transaction. These data environments therefore capture different combinations of trade direction, trader identity, and exchange coverage, imposing different trade-offs when inferring option demand and dealer inventory.

In transaction-level data, researchers usually rely on trade classification algorithms to infer signed option demand from prices and quotes. However, conventional algorithms typically identify the aggressive side of the trade rather than the customer side (Ellis et al., 2000). This distinction is important because customer demand can be expressed through both marketable orders and resting limit orders, so the aggressive side of a trade need not coincide with the customer side. This is consistent with broader order-choice models in which informed traders may optimally use limit orders when information is longer-lived (Kaniel and Liu, 2006), and with recent evidence on limit-order usage in option markets (Barardehi et al., 2025). When customers submit passive limit orders, conventional classification methods may therefore assign economically relevant customer demand to the wrong side of the transaction. Such misclassification can directly affect measures of informed option trading, dealer inventory, and demand pressure.

The accuracy of trade classification methods may also depend on market design. U.S. options exchanges differ in fee structures, order-submission incentives, and

execution mechanisms. These differences can shape how investors submit orders and how brokers route those orders across exchanges (Battalio et al., 2021). In particular, maker-taker pricing can make passive limit-order submission more attractive by changing the relative execution costs of marketable and resting orders. Classification performance may therefore vary across exchanges with different pricing models.

For Open/Close data, a central limitation is exchange coverage. This limitation is especially important in fragmented markets, where trading activity is dispersed across many venues. U.S. options trading currently takes place across 18 operating exchanges, and fragmentation is expected to increase further with the planned launch of two additional venues in the second half of 2026.² Measures based on a limited set of trading venues may therefore capture only part of market-wide customer demand. In addition, even when trader types are observed, empirical studies must decide which categories represent the relevant demand. Firm proprietary traders may at times act as end-users of options and at other times engage in market-making or other intermediary activity. Similar ambiguity can arise for broker-dealer activity, which may reflect proprietary trading, hedging, or intermediary inventory management. Open/Close data therefore avoid transaction-level trade classification, but they still require choices about how limited exchange coverage and different trader-type categories enter empirical measures.

This dissertation studies when and why these measurement trade-offs matter for empirical inference in option markets.

1.2 Structure of the Dissertation

This dissertation is structured as follows:

Chapter 2 develops the methodological foundation underlying the working paper Grauer et al. (2026) by evaluating and improving trade classification methods for inferring customer trade direction from option transaction data. Building on the measurement challenge described above, the chapter focuses on transaction-level data, where the central challenge is to infer customer trade direction from prices and quotes. Yet existing classification algorithms are often adapted from equity markets and typically identify the trade initiator rather than the customer side. We construct a large benchmark sample by matching intraday option trades to exchange-reported

²See Securities Exchange Act Release Nos. 34-104152 and 34-103998, approving MX2 Options and IEX Options, respectively. For the planned September 2026 launch of MX2 Options and the anticipated October 2026 launch of IEX Options, see Release Nos. 34-105555 and 34-105546.

customer buy and sell volumes and use it to evaluate established classification rules. We show that existing methods systematically misidentify customer trade direction when customers express demand through resting limit orders and that this source of misclassification is closely linked to exchange pricing structures. Existing methods are also limited for midspread trades. We therefore propose two new classification rules based on quoted depth that address these limitations and improve classification accuracy. The rules are developed in-sample and preregistered before being validated out-of-sample across time, exchanges, and market settings. They provide an economically motivated classification procedure for measuring customer option demand more accurately.

Chapter 3 is based on the empirical application in the working paper Grauer et al. (2026) and examines how trade classification affects inference about the predictive power of option demand for future stock returns. Using the classification framework developed in Chapter 2, the chapter constructs directional option-demand measures from intraday option trades and compares them with measures based on established classification algorithms. The results show that more accurately classified customer option demand has stronger and more persistent predictive power for future stock returns, especially among less liquid stocks. In joint tests, demand measures based on established classification algorithms lose explanatory power once option demand is measured using the improved classification procedure. This superior performance suggests that the improved procedure recovers information embedded in customer limit orders that conventional algorithms miss. The chapter therefore links the methodological analysis of option trade classification to a central economic application and shows that better measurement of customer option demand can materially affect empirical conclusions about informed trading in option markets.

Chapter 4 broadens this perspective by comparing how empirical inference changes when measures of option demand and dealer inventory are constructed from transaction data versus exchange Open/Close data.³ While the previous chapter focuses on inferring customer option demand from transaction data, this chapter places that measurement problem in a broader framework. Exchange Open/Close data directly identify customer buying and selling activity but cover only a subset of trading venues, whereas transaction-level data provide broader market coverage but require trade classification algorithms and do not identify trader type. The chapter isolates three dimensions along which these proxies differ, exchange coverage, trader identity, and trade classification, and studies their consequences for directional-

³This chapter is based on joint work with Jan Harren, Philipp Schuster, and Marliese Uhrig-Homburg as part of Research Unit FOR 5230, *Financial Markets and Frictions*.

demand and dealer-inventory applications. We show that measurement constraints are economically meaningful but not uniform across applications. Exchange coverage and trade classification are particularly important for directional demand measures and stock return predictability, whereas trader identity is especially important for dealer-inventory measures and their relation to future stock volatility. The effects are most pronounced outside the S&P 500, where option markets are less liquid and trading activity is more heterogeneous. Overall, the chapter highlights that empirical inference in option-market research depends critically on how latent demand and inventory measures are constructed.

Chapter 5 recapitulates the main results of the dissertation and gives an outlook on potential questions for future research.

Chapter 2

Beyond the Limits of Trade Classification in Option Markets

2.1 Introduction

Option markets are driven by customer demand. Customers initiate positions to hedge risks or express views, while market makers primarily intermediate trades and hedge their exposures in the underlying asset. Understanding how customer demand affects option prices, liquidity, and market quality therefore requires identifying whether customers are buying or selling options at each point in time. In practice, however, this information is not directly observable. Transaction data record prices and quantities but do not indicate the customer side of a trade. Empirical studies therefore rely on classification algorithms, such as the quote rule, the tick test, the Lee and Ready (LR, 1991) algorithm, and the Ellis, Michaely, and O'Hara (EMO, 2000) rule, to infer trade direction. Yet these methods are designed to identify the trade initiator rather than the side taken by customers, raising the question of how accurately they recover customer demand in option markets.

We show that standard trade classification methods systematically misidentify the customer side of option trades. The source of this problem lies in a structural feature of option markets: customers frequently express demand through limit orders rather than market orders. When a market maker executes against a resting customer limit order, conventional algorithms classify the trade from the market maker's perspective, effectively reversing the sign of customer demand. As a result, customer purchases are often recorded as sells and vice versa. This misclassification is closely linked to exchange pricing models, as maker-taker fee structures can

strengthen incentives to submit or route customer limit orders.

Beyond this mechanism, additional features of option markets further limit the applicability of stock-based classification rules. Option markets are less liquid than equity markets, with many option series trading infrequently, which weakens the informativeness of tick tests. At the same time, trading is fragmented across multiple exchanges connected through a national market system, raising the question of whether quote rules should rely on exchange-specific best bid and offer (BBO) quotes or the National Best Bid and Offer (NBBO). Despite these challenges, systematic evidence on trade classification in option markets remains scarce.

To fill this gap, we construct a novel benchmark that identifies the customer side of trades at the transaction level by combining intraday transaction data with daily exchange-level reports of customer buy and sell volume. We focus on trades between customers and market makers and exploit exchange-day-series observations in which customer activity occurs on only one side of the market. Using this benchmark, we evaluate existing trade classification algorithms and develop new rules. To separate rule development from validation, we organize the analysis as a two-step design based on a preregistration on the Open Science Framework (OSF). After preregistering the in-sample evidence and proposed rules on OSF, we access additional Open/Close datasets for out-of-sample validation. This design allows us to evaluate the performance of classification algorithms for more than 274 million option trades from five U.S. exchanges over more than sixteen years, starting in May 2005.

Relative to this benchmark, we document that standard classification methods perform poorly in option markets. Quote rules outperform tick tests, but their overall accuracy remains limited and varies substantially across exchanges. In particular, quote rules achieve moderate accuracy on exchanges that rely on payment for order flow (PFOF) but perform extremely poorly on maker-taker exchanges, where classification accuracy falls below random assignment. Misclassification is concentrated in trades whose size matches quoted depth, consistent with executions of customer limit orders resting at the best bid or ask.

To address this limitation, we develop two new classification rules that exploit information contained in quoted depth. The trade size rule identifies trades associated with resting customer limit orders when transaction size matches quoted depth. The depth rule improves the classification of midpoint trades by comparing bid and ask depth. Both rules are transparent and grounded in the institutional structure of option markets. When applied within a combined classification procedure, these rules achieve an exchange-weighted average accuracy of approximately 72%, representing

an improvement of about 16 percentage points over standard methods. These gains are robust across settings and particularly large on maker-taker exchanges. In addition to increasing accuracy, the procedure reduces cross-sectional and time-series variation in classification performance, yielding more stable classification outcomes.

This chapter contributes to at least two strands of the literature. First, we contribute to the literature on trade classification in financial markets (e.g., Lee and Ready, 1991; Ellis et al., 2000; Odders-White, 2000; Lee and Radhakrishna, 2000; Chakrabarty et al., 2007) by providing the first comprehensive large-scale evaluation of classification methods in option markets. While prior studies document the performance of these methods primarily in equity markets, evidence for options is scarce and limited to early work based on small and dated samples (Savickas and Wilson, 2003). Moreover, this literature often does not distinguish between the classification of trade initiator and the identification of the customer side (Ellis et al., 2000; Savickas and Wilson, 2003). Using a novel benchmark that identifies the customer side of trades at the transaction level, we show that customer limit orders create a wedge between these concepts in option markets, leading to systematic misclassification by standard classification methods.

Second, our findings relate to the literature on limit order trading and order submission strategies (e.g., Demsetz, 1968; Biais et al., 1995; Foucault et al., 2005; Goettler et al., 2005). While this literature shows that limit orders are optimal when trading urgency is low or spreads are wide, recent evidence highlights their growing importance in option markets and their role in measuring execution costs (Muravyev and Pearson, 2020; Barardehi et al., 2025; Fu et al., 2025; Anand et al., 2026). We provide new evidence that customer demand in option markets is frequently expressed through limit orders and that their prevalence has first-order implications for the accuracy of classification algorithms, as standard classification methods fail precisely when limit orders are prevalent. We further show that limit order submission is closely linked to exchange pricing models, thereby connecting classification performance to trading incentives and market design. This result complements evidence that options-exchange fee structures affect quoting and routing decisions (Anand et al., 2016; Battalio et al., 2021).

Our findings also have broader implications for empirical research in derivative markets. A large literature relies on signed option trades to measure demand pressure, intraday order flow, and trading profits (e.g., Hu, 2014; Goyenko et al., 2015; Muravyev, 2016; Bryzgalova et al., 2023; Pederzoli et al., 2025). Systematic misclassification of the customer side can therefore bias these measures and distort

inference. By improving its identification, our approach provides a more reliable empirical foundation. Importantly, the economic mechanism underlying the trade size rule is not specific to the sample exchanges but reflects broader features of option markets. As a result, the improvements generated by our approach are likely to generalize across market settings, enabling consistent classification across exchanges, over extended sample periods, and at both daily and intraday frequencies using a single dataset.

The remainder of the chapter proceeds as follows. Section 2.2 clarifies the distinction between trade initiation and customer trade direction, reviews existing classification methods, and explains why option market structure limits their performance. Section 2.3 describes the data, introduces the benchmark, and outlines the preregistered research design. Section 2.4 evaluates standard classification rules in the in-sample data and develops two new rules that substantially improve classification accuracy. Section 2.5 assesses the out-of-sample robustness of these rules across time, exchanges, and trading environments. Section 2.6 presents the combined classification procedure and evaluates its cross-sectional and time-series performance. Section 2.7 discusses the economic interpretation of the identified customer side and the generalizability of the mechanism beyond the sample exchanges. Section 2.8 concludes.

2.2 Option Market Structure and Limits of Stock-Based Trade Classification

Standard trade classification algorithms were originally developed for equity markets, but option markets differ in several important respects that may limit their direct applicability to option trades. Despite their widespread use, there is little evidence on how accurately stock-based classification rules identify the customer side of option trades. In this section, we first clarify the concept of customer trade direction and distinguish it from trade initiation in option markets. We then review common stock-based classification methods, discuss key institutional and microstructural features of option markets that may limit their performance, and relate these considerations to the existing literature.

2.2.1 Trade Initiation and Customer Trade Direction

In financial markets, each trade involves a buyer and a seller, but identifying which party initiated the trade is often non-trivial. The trade initiator is commonly defined as the party that demands liquidity by submitting an aggressive order, such as a market order or a marketable limit order, while the counterparty supplies liquidity through a resting limit order (Lee and Radhakrishna (2000)). More generally, this notion corresponds to identifying the party whose order arrives last in chronological sequence (Odders-White (2000)). From an economic perspective, however, it is often more relevant to classify trades from the customer's perspective, that is, whether a transaction reflects customer buying or selling pressure, regardless of the chronological order (see Ellis et al. (2000), Savickas and Wilson (2003)). While trade initiation and customer trade direction coincide in some market settings, they often diverge in option markets, where sophisticated customers frequently submit limit orders rather than market orders.

This divergence arises from differences in the role of market makers across markets. In equity markets, market makers typically provide liquidity by continuously posting bid and ask quotes and may accumulate directional inventory in the underlying stock as they intermediate trades, effectively contributing to net demand or supply. In contrast, option market makers can dynamically hedge their positions in the underlying asset and therefore do not need to take directional views on the options themselves. As a result, trading pressure in option markets is primarily driven by customers, while market makers act as intermediaries who absorb customer demand and facilitate execution. This implies that trade initiation and customer trade direction diverge in option markets when customers use limit orders to express trading intent. If such orders are not executed immediately, they may rest in the order book and later execute against market makers. In these cases, the observed trade initiator need not coincide with the customer side that reflects demand for option positions.

This interpretation also suggests that customer limit orders in option markets need not represent liquidity provision in the same economic sense as market-maker intermediation. Although customers submitting limit orders may obtain better execution prices than customers using market orders, they do not necessarily earn positive short-horizon trading profits from supplying liquidity. We provide complementary evidence on customer trading outcomes in Section 2.7.1, supporting the view that customer limit orders primarily reflect demand for option positions rather than liquidity provision.

This divergence has important empirical implications. Trade classification methods that rely on identifying the initiator may systematically misclassify the customer side of option trades. Such misclassification can bias measures of order imbalances and distort inference about the price effects of customer option demand. Consequently, correctly identifying the customer side of transactions is central to empirical analyses of option demand (Bollen and Whaley (2004), Gârleanu et al. (2009)) and option price pressure (Goyenko and Zhang (2021)). The following sections review existing trade classification approaches and motivate the development of new methods tailored to the institutional and microstructural features of option markets.

2.2.2 Existing Trade Classification Methods

Because most option datasets do not report the customer side of a trade, empirical studies typically rely on heuristics to infer trade direction from transaction prices and quotes. These methods were originally developed for equity markets to identify the trade initiator but are widely applied in option markets to classify customer trade direction (see, e.g., Hu (2014); Goyenko et al. (2015); Muravyev (2016); Bryzgalova et al. (2023)). The dominant approaches fall into three broad categories: quote rules, tick tests, and hybrid methods.

Quote rules exploit the fact that market makers provide liquidity by continuously posting bid and ask quotes at which customers can transact. In this setting, customer purchases occur close to the ask and customer sales close to the bid. The customer side of a trade can therefore be inferred from the transaction price relative to the prevailing bid–ask midpoint (Hasbrouck, 1988; Harris, 1989). Specifically, trades above the midpoint are classified as customer buys, trades below the midpoint as customer sells, and trades at the midpoint remain unclassified:

$$D_{i,t}^{\text{Quote rule}} = \begin{cases} +1, & \text{if } p_{i,t} > m_t \quad (\text{customer buy}), \\ -1, & \text{if } p_{i,t} < m_t \quad (\text{customer sell}), \\ \text{undefined}, & \text{if } p_{i,t} = m_t. \end{cases} \quad (2.1)$$

Here $D_{i,t}$ denotes the direction of trade i at time t , $p_{i,t}$ is the transaction price of this trade, and m_t represents the midpoint of the bid–ask quote.

In practice, quote rules can be applied using either the NBBO or the BBO at the exchange where the trade occurs. Conceptually, it is not obvious which quote should be considered as a reference. To reduce the number of unclassified trades,

Muravyev and Ni (2020) propose a stacked approach that applies the quote rule first using the NBBO and then using exchange-level quotes, leaving trades unclassified only when both midpoints coincide. We also consider the reverse stacking order, applying the quote rule first to the BBO and then to the NBBO, an approach not previously considered in the literature.

Tick tests infer trade direction from changes in transaction prices, based on the intuition that buying pressure tends to push prices upward while selling pressure pushes them downward. The standard tick test compares the transaction price of a trade to that of the immediately preceding transaction (Holthausen et al., 1987; Hasbrouck, 1988). A trade is classified as a customer buy if its price exceeds the previous trade price and as a customer sell if it is lower. When consecutive trades occur at the same price, the classification follows the direction of the most recent price change, effectively assigning the same direction as the previous trade. If no prior trade with a different price exists, for example in the case of the first trade of an option, the tick test cannot be applied.¹ Formally, the tick test is defined as

$$D_{i,t}^{\text{Tick test}} = \begin{cases} +1, & \text{if } p_{i,t} > p_{i,t-1} \quad (\text{customer buy}), \\ -1, & \text{if } p_{i,t} < p_{i,t-1} \quad (\text{customer sell}), \\ D_{i,t-1}^{\text{Tick test}}, & \text{if } p_{i,t} = p_{i,t-1}, \\ \text{undefined}, & \text{if no prior price change exists.} \end{cases} \quad (2.2)$$

Here $p_{i,t-1}$ denotes the price of the immediately preceding trade in the same option.

An alternative variant is the reverse tick test, which compares a trade's transaction price to that of the subsequent transaction in the same option (Hasbrouck, 1988). A trade is classified as a customer buy if its price exceeds the next trade price and as a customer sell if it is lower. When both trades occur at the same price, the classification follows the direction of the next price change. If no subsequent price change exists, for example for the last trade of an option, the reverse tick test cannot be applied. Formally,

¹In equity markets, it is common to classify the first trade of a newly listed security as a buy, possibly reflecting a ceremonial purchase by the company president. We do not adopt this convention in the option market and instead leave the first trade unclassified.

$$D_{i,t}^{\text{Reverse tick test}} = \begin{cases} +1, & \text{if } p_{i,t} > p_{i,t+1} \quad (\text{customer buy}), \\ -1, & \text{if } p_{i,t} < p_{i,t+1} \quad (\text{customer sell}), \\ D_{i,t+1}^{\text{reverse tick test}}, & \text{if } p_{i,t} = p_{i,t+1}, \\ \text{undefined}, & \text{if no subsequent price change exists.} \end{cases} \quad (2.3)$$

Here $p_{i,t+1}$ denotes the price of the immediately subsequent trade in the same option.

The tick test and reverse tick test yield identical classifications at price reversals but differ when consecutive price changes occur in the same direction. In practice, both variants can be implemented using either consolidated transaction prices across all option exchanges or prices from a single trading venue. While consolidated prices provide more frequent updates and reflect most recent trade information, venue-specific prices may better capture the trading conditions and price dynamics on a given exchange.

Hybrid methods combine quote rules and tick tests to varying degrees. The most widely used algorithms are those proposed by Lee and Ready (1991) and Ellis et al. (2000). The LR algorithm applies the quote rule to all trades except those executed at the midpoint, which are classified using the tick test. In contrast, the EMO rule relies more heavily on the tick test, applying the quote rule only to trades executed exactly at the bid or ask and using the tick test for all other trades. Figure 2.1 provides a visual illustration of the two approaches.

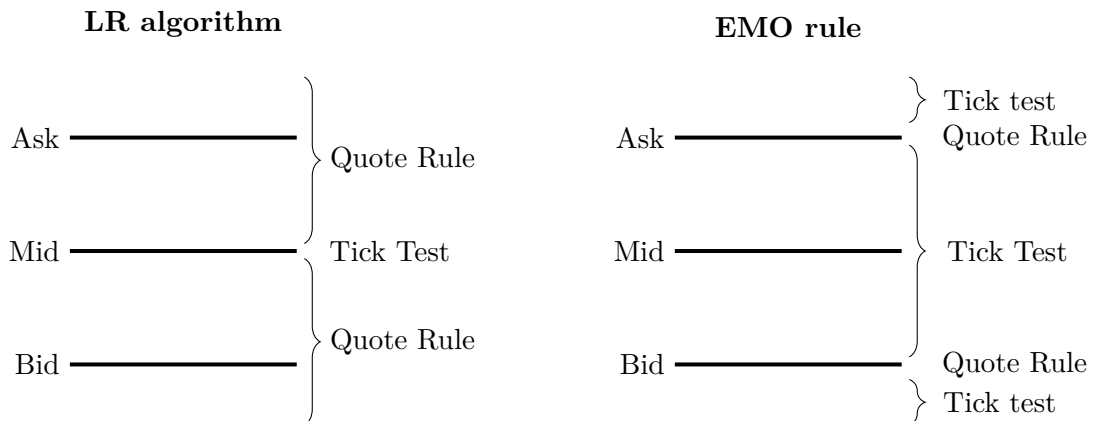


Figure 2.1: Hybrid methods

This figure illustrates how quote rules and tick tests are combined in two common hybrid methods, the LR algorithm and the EMO rule, depending on the position of the transaction price relative to the prevailing bid–ask quotes.

A natural extension is to combine the quote rule with the reverse tick test. We refer to these variants as ‘reverse LR’ and ‘reverse EMO’. Following the approach discussed above, we also consider stacked implementations that sequentially apply the quote rule using both the NBBO and BBO quotes before resorting to the standard or reverse tick test.

2.2.3 Option Market Features Limiting Trade Classification

Despite their widespread use, the applicability of stock-based trade classification rules to option markets is not obvious. Several institutional and microstructural features of option markets limit their direct application.

First, options are substantially less liquid than stocks, with many option series not trading for days or even weeks. In such thinly traded environments, tick tests that infer trade direction from price movements relative to preceding or succeeding trades may become unreliable. When trades occur infrequently, observed price changes may reflect stale information rather than contemporaneous trading pressure, weakening the link between price movements and the direction of customer demand.

Second, a large literature on optimal order submission (e.g., Demsetz (1968); Biais et al. (1995); Foucault et al. (2005); Goettler et al. (2005)) shows that when bid–ask spreads are wide or traders face low urgency, submitting a limit order can be optimal despite the risk of non-execution. These conditions are more prevalent in option markets than in highly liquid equity markets due to wider spreads and less frequent quote updates (see Muravyev and Pearson (2020)). Consistent with this prediction, Barardehi et al. (2025) document that limit orders account for about 17% of stock trading volume but more than 34% of options trading volume. As a result, customer demand in option markets is more often expressed through limit orders, complicating trade classification methods that aim to identify the aggressive side of the trade.

Third, the U.S. options market is fully exchange-based and comparatively transparent, but trading is fragmented across a large number of venues. Over our sample, the number of active options exchanges rose from six in May 2005 to sixteen in July 2021, while market shares became more evenly distributed across venues. All exchanges are linked through a national market system that disseminates both exchange-specific BBO quotes and the NBBO. In such a fragmented market, it is not obvious whether quote rules should reference exchange-level or national quotes.

Fourth, option markets are institutionally heterogeneous along several micro-structural dimensions that shape order submission and execution incentives. Table 2.1 summarizes key market design features of the five exchanges in our sample, namely Nasdaq International Securities Exchange (ISE), Chicago Board Options Exchange (CBOE), Nasdaq PHLX (PHLX), Nasdaq Options Market (NOM), and Nasdaq GEMX (GEMX). NOM operates a standard price–time priority limit order book, whereas the other exchanges rely on pro-rata allocation rules that favor larger orders. By increasing the execution probability of large liquidity providers, this mechanism strengthens incentives to quote aggressively in size. To protect retail participation, these exchanges also grant customer priority, giving customer limit orders execution precedence over broker–dealer orders at the same price. Market design further differs in trading mechanisms. Four exchanges offer price-improvement auctions that allow market makers to improve retail executions relative to the NBBO. In addition, some venues support complex order books that facilitate multi-leg option strategies, while others continue to operate open outcry trading floors. Exchanges also differ in market maker obligations, allocation privileges, and fee structures. NOM and GEMX employ maker–taker pricing models that reward liquidity provision through rebates, whereas CBOE and PHLX rely more heavily on PFOF arrangements designed to attract marketable retail order flow (Anand et al., 2016; Battalio et al., 2021). The ISE combines elements of both pricing models. These institutional differences shape order-submission and execution incentives and may therefore directly affect the performance of trade classification rules.

Table 2.1: Market design features of sample option exchanges

This table summarizes key market design features of the five U.S. options exchanges in our sample. The first column reports the year in which each exchange began operations. Market share refers to each exchange’s share of trading volume between January and July 2021. The pricing model distinguishes between payment for order flow (PFOF), maker–taker, and hybrid fee structures. Allocation priority indicates the primary execution rule, price–time or pro-rata, while customer priority denotes whether customer orders at the best price receive priority over other participants. The remaining columns indicate whether the exchange offers price improvement auctions, supports complex-order functionality for multi-leg strategies, operates an open outcry trading floor, or grants allocation advantages to designated lead market makers. See NYSE (2024) for a similar illustration.

	Year	Share (2021)	Pricing Model	Allocation Priority	Customer Priority	Auction	Multi Leg	Floor	Lead MM
ISE	2000	6%	Hybrid	Pro-rata	Yes	Yes	Yes	No	Yes
CBOE	1973	15%	PFOF	Pro-rata	Yes	Yes	Yes	Yes	Yes
PHLX	1974	12%	PFOF	Pro-rata	Yes	Yes	Yes	Yes	Yes
NOM	2008	7%	Maker–taker	Price–time	No	No	No	No	No
GEMX	2013	4%	Maker–taker	Pro-rata	Yes	Yes	No	No	Yes

Taken together, lower liquidity, widespread use of limit orders, market fragmentation, and institutional heterogeneity suggest that stock-based trade classification rules may perform fundamentally differently in option markets. Despite their widespread use, systematic evidence on the accuracy of these methods for identifying the customer side of option trades remains limited.

2.2.4 Evidence from the Literature

This section reviews existing evidence on the performance of trade classification rules across markets and over time, summarized in Table 2.2. Much of this evidence comes from equity markets, where a large literature evaluates the accuracy of competing classification methods. Early studies generally report high success rates, particularly for the LR algorithm, with accuracy often exceeding 80%. However, performance is highly sensitive to institutional settings, sample selection, and time period, raising concerns about the robustness and generalizability across markets.

Using data from the New York Stock Exchange (NYSE) from the early 1990s, Lee and Radhakrishna (2000) report an overall accuracy of approximately 93% for the LR algorithm, with even higher accuracy for trades executed at the bid or ask. In contrast, midpoint trades, classified using the tick test, are substantially less accurate, with success rates of only about 76%. Applying the LR algorithm to broader and less restrictive samples, Odders-White (2000) and Finucane (2000) document lower accuracy of roughly 85%, highlighting the sensitivity of reported performance to sample selection. Extending the analysis to Nasdaq markets, Ellis et al. (2000) find comparable overall accuracy of around 81%, but observe substantially weaker performance for trades inside the quotes, particularly midpoint trades, where success rates fall to about 60%. Peterson and Sirri (2003) further show that classification accuracy varies across market regimes surrounding a NYSE tick size change, emphasizing the importance of institutional settings.

More recent evidence questions whether these early results generalize to modern, high-speed, and fragmented markets. Easley et al. (2012b) argue that rapid order submission and cancellations reduce the reliability of traditional classification algorithms. Consistent with this view, Chakrabarty et al. (2007) report LR accuracy of at most 75% in 2005. In contrast, Carrion and Kolay (2020) show that quote information remains informative even in fast trading environments. Evidence from other markets also suggests that classification accuracy depends strongly on market structure. For example, Easley et al. (2012a) find that the tick rule performs

Table 2.2: Empirical evidence on trade classification across markets

This table summarizes existing evidence on the performance of trade classification rules across markets and over time. It reports success rates (in percent) for the quote rule, the tick test, the LR algorithm, and the EMO rule. For Peterson and Sirri (2003), (1) and (2) denote subperiods around the NYSE tick size change from 1/8 to 1/16. For Chakrabarty et al. (2012), (1) and (2) denote short and long sales, respectively.

Study	Market	Time	Quote	Tick	LR	EMO
U.S. equity market						
Lee and Radhakrishna (2000)	NYSE	1990	—	—	93.3	—
Odders-White (2000)	NYSE	1990	75.0	78.6	85.0	—
Finucane (2000)	NYSE	1990	—	83.0	84.4	—
Ellis et al. (2000)	NASDAQ	1996	76.4	77.7	81.1	81.9
Peterson and Sirri (2003)	NYSE	1997(1)	—	—	91.0	90.8
		1997(2)	—	—	87.4	88.2
Chakrabarty et al. (2007)	NASDAQ	2005	—	75.4	74.4	75.8
Chakrabarty et al. (2012)	NASDAQ	2005(1)	—	—	68.2	—
		2005(2)	—	—	69.3	—
Rosenthal (2012)	NASDAQ	2004	—	66.7	71.4	72.3
	NYSE	2004	—	60.7	75.6	79.6
Carrion and Kolay (2020)	NASDAQ	2008	—	78.6	86.9	—
International equity market						
Aitken and Frino (1996)	ASX	1992	—	74.4	—	—
Theissen (2001)	FSE	1996	75.4	72.2	72.8	—
Pöppe et al. (2016)	DB	2012	—	82.0	89.6	90.4
Emerging equity market						
Lu and Wei (2009)	TWSE	2006	92.8	74.2	96.5	95.0
Aktas and Kryzanowski (2014)	BIST	2008	95.0	90.4	96.4	86.9
Perlin et al. (2014)	BOVESPA	2009	—	72.0	—	—
Nowak (2020)	Warsaw	2019	84.4	19.9	84.3	80.4
	Budapest	2019	100	40.7	96.3	83.7
	Prague	2019	82.8	39.5	82.9	67.9
Other asset markets						
Easley et al. (2012a)	ES Futures	2010	—	86.4	—	—
	CL Futures	2010	—	79.0	—	—
	GC Futures	2010	—	67.2	—	—
Andersen and Bondarenko (2015)	ES Futures	2006	—	88.4	—	—
Omrane and Welch (2016)	EUR/USD	2005	77.0	67.0	—	—
Ma and Zhai (2021)	BTC	2017	—	76.9	—	—
Fedenia et al. (2024)	Corp.Bonds	2002	—	69.3	—	—
Option market						
Savickas and Wilson (2003)	CBOE	1995	79.0	60.8	79.0	74.4
Gayda (2025)	CBOE	2020	62.0	—	—	—
Gao et al. (2025)	PHLX	2024	56.5	—	—	—

reasonably well in E-mini S&P 500 futures but substantially worse in oil and gold futures.

More broadly, studies from international equity markets (Aitken and Frino, 1996; Theissen, 2001), emerging markets (Lu and Wei, 2009; Aktas and Kryzanowski, 2014; Perlin et al., 2014), and other asset classes (Easley et al., 2012a; Andersen and Bondarenko, 2015; Omrane and Welch, 2016; Ma and Zhai, 2021; Fedenia et al., 2024) document considerable heterogeneity in classification accuracy. These findings raise concerns about the generalizability of stock-based classification rules and underscore the importance of market-specific characteristics.

Against this background, evidence on trade classification in option markets remains scarce. Savickas and Wilson (2003) provide the only early assessment, showing that standard classification rules perform up to 24% worse for options than for stocks. However, their analysis relies on proprietary CBOE data from July to December 1995 and therefore predates the widespread adoption of electronic trading, auction mechanisms, and today’s fragmented option market structure. More recently, two studies examine trade classification in the context of option auctions. Gayda (2025) analyze intraday CBOE option auction data from 2020 to 2023 and show that incorporating minimum price improvement requirements increases classification accuracy from 62% to 76%, with particularly strong improvements for short-dated options. Gao et al. (2025) further demonstrate that machine learning approaches outperform traditional rules in auction settings, achieving an accuracy of 75.9% using proprietary PHLX data from January to April 2024.

While these studies provide valuable insights, they focus on narrow market segments, short sample periods, or auction-based trading environments. In contrast, we provide a comprehensive evaluation of trade classification accuracy in option markets. Using a large sample of matched option trades and quotes, we assess the performance of standard classification algorithms across a wide range of contracts, venues, and market conditions.

2.3 Data and Benchmark Construction

Evaluating trade classification algorithms requires a benchmark that reliably identifies the customer side of each trade. Because most option datasets do not contain this information at the transaction level, we construct a novel benchmark by combining intraday transaction data with exchange reports of customer buy and sell volume, allowing us to infer customer trade direction for a large set of trades with

high confidence. In this section, we first describe our data sources and the construction of the benchmark. We then outline our two-step research design before presenting summary statistics of the resulting matched samples.

2.3.1 Options Market Data

Our empirical analyses combine option data from several complementary sources. First, we use intraday transaction-level option data from LiveVol, provided by the CBOE. The dataset covers all option trades executed on U.S. exchanges and reports transaction prices, trading volumes, and contemporaneous quote information. Specifically, it includes the NBBO at the time of each trade, as well as exchange-level quotes and sizes for all exchanges on which the option is quoted, including the exchange executing the trade. We apply a minimal set of standard filters to ensure data quality, excluding trades with non-positive prices or volumes, volumes exceeding 10 million contracts, entries with multiple underlying symbols for the same option root, duplicate observations, and canceled trades.

Second, we use Open/Close Trade Profile data of five option exchanges: ISE, CBOE, PHLX, NOM, and GEMX. These data report daily buy and sell trading volumes by account type for each option series traded on the respective exchange. Trades are further categorized by whether they open or close positions and by account type, including customers, professional customers, firm proprietary traders, and firm broker-dealers. Market maker activity is reported indirectly under market-clearing conditions, and can be inferred as the residual side of customer and firm trades. We remove duplicate entries and separately aggregate buy and sell volumes for each account type, option series, and trading day, across opening and closing transactions and trade size categories. This procedure yields distinct daily buy and sell volumes for each account type and option series. For GEMX, we additionally use an intraday Open/Close dataset constructed at 10-minute intervals.

Third, to examine how classification accuracy varies with option and underlying characteristics, we obtain security identifiers and issue-type information from Ivy DB OptionMetrics and S&P 500 index membership from the Center for Research in Security Prices (CRSP). We match OptionMetrics to LiveVol at the contract-day level using standard contract identifiers, namely underlying symbol, strike price, expiration date, and option type, and retain the observation whose trading volume best matches the corresponding daily LiveVol volume.

2.3.2 Benchmark for Customer Trade Direction

Evaluating trade classification algorithms requires a benchmark that reliably identifies the true customer side of each trade. We construct such a benchmark by combining intraday transaction data from LiveVol with Open/Close data that report customer buy and sell volumes at the daily level for each exchange and option series. Our benchmark focuses on trades between customers and market makers. Because option market makers primarily intermediate trades and hedge their exposures rather than take directional views, this restriction provides a conservative benchmark that allows the customer side of each trade to be identified unambiguously.

To implement this benchmark, we apply the following matching procedure. For each exchange-day-series, we compare total trading volume in LiveVol with customer buy and sell volume reported in the Open/Close data. We retain observations for which total LiveVol volume exactly matches either total customer buy volume or total customer sell volume and for which no other trader categories are active. In these cases, all trades in the option series on that exchange-day must involve customers exclusively on one side of the market, with market makers taking the opposite side. The customer trade direction is therefore unambiguously identified. Specifically, if the Open/Close data report only customer buy (sell) activity for a given option series on a given exchange-day, we classify all corresponding LiveVol transactions as customer buys (sells). Trades are matched using a unique identifier based on trade date, expiration date, strike price, option type, and the root symbol of the underlying. This procedure yields five exchange-specific matched samples containing intraday transaction and quote data together with a benchmark indicator of customer trade direction. We further merge these samples with OptionMetrics to incorporate underlying characteristics.

To assess classification accuracy under different liquidity conditions, we additionally construct an intraday benchmark for GEMX using Open/Close data reported at 10-minute intervals. Applying the same matching logic at the intraday level, we identify LiveVol transactions within each interval for which total trading volume equals either customer buy or customer sell volume, and for which no other trader categories are active. To account for potential reporting delays, we allow for a 30-second time shift and retain only unique matches. This intraday benchmark substantially increases the number of matched trades while preserving unambiguous identification of customer trade direction.

2.3.3 Representativeness of the Benchmark

The benchmark construction focuses on trades between customers and market makers because this setting allows the customer side of each transaction to be identified transparently. A natural concern is whether this setting reflects a central trading interaction in option markets or instead captures a narrow subset of trading activity. To address this concern, we examine the frequency of different trader-type pairings in the Open/Close data.

We extend the matching procedure used for the matched sample to all trader categories reported in the Open/Close data. Specifically, we match intraday LiveVol transactions to Open/Close volumes at the exchange-day-series level and identify cases in which trading occurs exclusively between a single pair of trader categories. In these cases, all corresponding LiveVol transactions can be assigned to a unique trader-type pairing. Because market maker activity is not reported directly, we infer it when no other trader category is active on one side of the market. For the intraday GEMX sample, we apply the same logic at the 10-minute interval level.

Table 2.3 reports the resulting trader-type pairings. Customer–market maker trades are the most common interaction across all exchanges, accounting for 46% of trades at NOM and more than 85% at CBOE, with similarly high shares at ISE and GEMX. The intraday GEMX sample exhibits a distribution close to that of the daily GEMX sample, suggesting that these patterns are not driven by the daily benchmark construction but reflect the trading structure of option markets more generally.

These patterns support our focus on customer–market maker trades as the benchmark setting. In this setting, the customer side is most transparently identified, as customers represent the demand side while market makers intermediate trades. Other trader-type pairings occur much less frequently and do not permit an equally transparent identification of economic demand. The benchmark therefore captures the dominant and economically cleanest setting for evaluating whether trade classification rules recover the customer side of option trades.

2.3.4 Preregistered Two-Step Research Design

Our empirical design uses a preregistration on OSF to establish a clear separation between method development and validation.² We first use an in-sample dataset

²The preregistration was time-stamped on November 20, 2021 and documents the proposed classification rules and preliminary results from the in-sample dataset. It is available on OSF at

Table 2.3: Trader-type pairings

This table reports the distribution of trader-type pairings at the transaction level for exchange-days on which trading in a given option series can be uniquely attributed to a single pair of trader categories based on the Open/Close data, i.e., customer (Cust), market maker (MM), or ‘Other’, where ‘Other’ comprises professional customers, firm proprietary traders, and broker-dealers. On such exchange-days, all transactions in the LiveVol data can be mapped to one unique trader-type pairing. For the intraday GEMX sample, this matching requirement is imposed at the 10-minute interval level, i.e., within each interval trading must involve a single pair of trader categories. We report the share of transactions attributable to this subset relative to the full set of uniquely classified transactions (in percent).

Trader type [%]	ISE	CBOE	PHLX	NOM	GEMX	Intraday GEMX
Cust–MM	78.20	85.47	52.32	45.57	82.56	82.84
Cust–Other	1.39	0.92	29.70	11.90	1.56	1.81
Cust–Cust	1.44	0.65	1.35	1.73	1.94	3.35
MM–Other	6.41	4.62	8.04	16.02	5.33	4.86
MM–MM	12.32	8.26	8.36	23.48	8.42	6.97
Other–Other	0.24	0.08	0.23	1.31	0.19	0.18

to evaluate existing trade classification algorithms and develop new classification rules. After preregistering the in-sample results and proposed rules on OSF, we accessed additional Open/Close datasets for out-of-sample validation. This sequence strengthens the credibility of the validation exercise and provides a direct test of whether rules developed on one exchange generalize across time, exchanges, and market structures.

In the first step, we use ISE Open/Close data from May 2, 2005, to May 31, 2017. The start date reflects the earliest availability of ISE Open/Close data, while the end date corresponds to the data available to us at the time of method development and preregistration. The in-sample matched ISE dataset contains 49,203,747 option trades and serves as the basis for comparing existing classification algorithms and developing our two new rules.

In the second step, we validate the proposed algorithms using data accessed only after the OSF preregistration. These validation samples include ISE trades after May 2017 as well as four additional exchanges, CBOE, PHLX, NOM, and GEMX. The CBOE sample begins on January 1, 2011, following a structural change in its Open/Close reporting. The PHLX, NOM, and GEMX samples begin on January 1, 2009, November 11, 2011, and August 5, 2013, respectively, reflecting

https://osf.io/kj86r/?view_only=388a89b23254425a8271402e2b11fc4e.

data availability. All samples end on June 30, 2021. Together, the five matched validation datasets comprise more than 225 million option trades.

As an additional validation exercise, we use intraday GEMX Open/Close data to examine how classification accuracy varies with liquidity. The observation period matches that of the GEMX end-of-day data. The higher-frequency benchmark substantially expands the number of matched GEMX trades from approximately 40 million to more than 100 million.

2.3.5 Sample Description

This section describes the main characteristics of the matched samples. Table 2.4 reports summary statistics for the matched ISE sample used in the in-sample analysis in Panel A and compares it with the full universe of ISE trades during the same period in Panel B. The matched sample captures a meaningful share of market activity, accounting for approximately 15% of all ISE trades, 44% of option series per day, and 84% of underlyings per day. By construction of our benchmark, the matched sample focuses on option series with lower trading intensity, as the likelihood of observing exclusively customer buy or sell activity declines with trading intensity. Consistent with this selection, the matched sample exhibits lower liquidity, reflected in smaller trade sizes, longer times to maturity, fewer trades per option-day, and longer intervals between trades. While lower liquidity could potentially affect classification performance, we explicitly address this concern in our out-of-sample tests. Importantly, the fraction of trades with sizes equal to the quoted bid or ask size is comparable, and if anything slightly lower, in the matched sample relative to the full ISE sample. The matched sample therefore does not appear to overrepresent this feature of the trading environment.

Table 2.5 reports analogous summary statistics for the out-of-sample exchanges. While data availability leads to differences in time coverage across venues, the matched samples continue to represent a substantial share of market activity. Across exchanges, they account for 8% to 27% of trades, include 24% to 69% of option series, and cover 52% to 95% of underlyings. The last column reports statistics for the intraday GEMX matched sample. Relative to the end-of-day benchmark, this sample captures a broader cross-section of trades and contains trades from more frequently traded option series. Overall, it covers approximately 66% of all GEMX trades.

The full-sample summary statistics reveal noticeable cross-exchange heterogene-

ity in trading activity, contract characteristics, and execution patterns, consistent with differences in exchange design. CBOE exhibits the broadest cross-section of option series, consistent with its role as a large hybrid exchange with broad listings. PHLX is characterized by larger trade sizes, longer maturities, and longer intervals between trades. These patterns are consistent with more strategy-oriented or broker-mediated trading, which fits PHLX's hybrid floor-based structure, complex-order functionality, and pro-rata allocation model. By contrast, GEMX and NOM feature smaller trade sizes and shorter intervals between trades, suggesting more frequent trading in smaller quantities. The relatively high fraction of trades executed at the quoted bid or ask, with NOM reaching 52%, is consistent with a greater role for displayed liquidity provision on maker-taker venues, potentially including customer limit orders. Overall, these patterns provide descriptive evidence of heterogeneous trading environments across exchanges.

Table 2.4: Summary statistics for in-sample option trades

This table reports descriptive statistics for the matched ISE sample (Panel A) and compares them to all ISE trades in the LiveVol data (Panel B). All statistics are computed over the in-sample period from May 2, 2005 to May 31, 2017. We report the number of unique option series and underlyings per day, the number of trades per option-day, and trade-level characteristics including time between trades (hours), trade size (contracts), time to maturity (days), and moneyness (defined as underlying (strike) price relative to strike (underlying) price for calls (puts)). For each variable, we report the mean, standard deviation, and the 5%, 50% (median), and 95% quantiles. We also report the total number of observations (in millions) and the share of trades for which trade size equals quoted depth.

Panel A: Matched ISE trades					
	Mean	Std	5%	Median	95%
Option series per day	9,885.22	3,119.45	5,632	9,420	15,838
Underlyings per day	1,236.56	225.19	751	1,282	1,522
Trades per option day	1.64	2.49	1	1	4
Time between trades	36.12	160.51	0.00	0.59	149.57
Trade size	13.62	77.75	1	4	50
Days to maturity	107.29	150.08	2	46	459
Moneyness	0.99	2.45	0.69	0.97	1.29
Number of observations	49.20				
Trade size = quote size	22.28%				
Panel B: All ISE trades					
	Mean	Std	5%	Median	95%
Option series per day	22,388.14	4,773.00	14,192	22,579	29,792
Underlyings per day	1,468.24	293.97	778	1,535	1,795
Trades per option day	4.95	18.02	1	2	17
Time between trades	9.70	85.65	0.00	0.01	24.80
Trade size	23.47	289.92	1	5	72
Days to maturity	62.21	106.58	1	29	236
Moneyness	0.98	5.92	0.76	0.98	1.14
Number of observations	337.23				
Trade size = quote size	25.58%				

Table 2.5: Summary statistics for out-of-sample option trades

This table reports descriptive statistics for out-of-sample data across five option exchanges. The sample period for each exchange is reported at the top of the table. For each exchange, the first column presents statistics for the matched sample, while the second column reports the corresponding full sample of all option trades on that exchange in the LiveVol data. For GEMX, a third column reports the matched sample constructed using intraday Open/Close data. We report averages for the number of unique option series and underlyings per day, the number of trades per option-day, and trade-level characteristics including time between trades (hours), trade size (contracts), time to maturity (days), and moneyness (defined as underlying (strike) price relative to strike (underlying) price for calls (puts)). We also report the total number of observations (in millions) and the share of trades for which trade size equals quoted depth.

	ISE		CBOE		PHLX		NOM		GEMX		
Sample start	Jun 1, 2017		Jan 1, 2011		Jan 1, 2009		Nov 11, 2011		Aug 5, 2013		
Sample end	Jun 30, 2021		Jun 30, 2021		Jun 30, 2021		Jun 30, 2021		Jun 30, 2021		
	Matched	Full	Matched	Full	Matched	Full	Matched	Full	Matched	Full	Intraday
Option series per day	26,376	38,033	15,930	46,349	8,004	32,799	10,284	33,901	13,836	23,180	19,546
Underlyings per day	1,560	1,646	1,147	2,199	1,233	2,172	1,243	1,665	1,380	1,509	1,420
Trades per option day	1.73	4.33	1.75	5.82	1.39	3.99	1.34	5.39	1.47	3.88	2.58
Time between trades	24.11	7.72	33.82	7.83	41.35	11.23	36.88	6.36	24.19	6.76	10.35
Trade size	7.83	10.94	10.76	16.74	10.71	22.57	7.91	8.34	8.84	8.92	8.47
Days to maturity	72.80	42.52	90.89	42.46	99.66	51.29	104.24	45.89	88.42	42.64	55.90
Moneyness	0.98	0.97	0.98	0.97	0.99	0.98	1.07	1.02	0.99	0.97	0.98
Number of observations	45.05	169.44	73.64	712.43	34.85	411.38	33.42	442.88	40.41	179.14	100.03
Trade size = quote size	11.39%	13.62%	11.17%	13.94%	17.37%	16.15%	46.35%	40.63%	55.50%	51.68%	54.60%

2.4 In-Sample Analysis of Classification Rules

To assess how accurately existing methods identify the customer side of option trades, we evaluate their performance using our in-sample ISE dataset. We first examine the accuracy of established stock-based classification rules relative to our benchmark and compare alternative implementations that differ in their use of venue-specific versus cross-exchange information. Based on these results, we identify systematic weaknesses of existing methods and develop two new rules designed to address these limitations in option markets. Finally, we evaluate whether the proposed rules improve classification accuracy and examine their economic intuition through targeted subsample analyses, motivating potential refinements.

2.4.1 Established Rules

We begin by evaluating the ability of established trade classification rules to infer the customer side of option trades. Algorithmic classifications are compared with our Open/Close benchmark. To ensure comparability across methods that do not classify all trades, we assume that unclassified trades are correctly classified with a probability of 50%. Table 2.6 reports in-sample performance across implementations that rely on venue-level versus cross-exchange information. For quote rules and combined methods, we consider implementations using NBBO, BBO, or sequential combinations of both quotes. For tick tests, we compare specifications that use consolidated transaction prices across all exchanges with those that rely solely on prices from the trading venue. For brevity, hybrid methods are reported only for specifications that use consolidated prices across all exchanges.

Quote rules perform best among existing methods, consistently outperforming tick tests across all specifications and dominating corresponding hybrid methods. Within quote rule implementations, national quotes provide a modest advantage over exchange-specific quotes. Using NBBO yields a success rate of 63.72%, compared with 62.68% when relying on ISE quotes. The most accurate specification sequentially applies the quote rule first to NBBO and then to ISE quotes, yielding a success rate of 63.94%. The higher accuracy relative to using NBBO alone arises from a reduction in the fraction of trades that must be randomly assigned, from approximately 10.2% to 7.4%. Reversing the order and applying the quote rule first to ISE quotes and then to NBBO results in lower accuracy than using NBBO alone, suggesting that midpoint trades are difficult to classify. These results indicate that

national quotes provide a slightly more informative basis for inferring customer trade direction than exchange-specific quotes, although the effect is modest.

Tick tests perform substantially worse than quote rules. Using transaction prices from all exchanges improves performance relative to relying solely on prices from the trading venue, but success rates remain low. The tick test based only on ISE prices even performs worse than random assignment. Reverse tick tests perform somewhat better but still achieve a success rate of only 54.88%. These results suggest that price movements alone provide limited information about the customer side of option trades, likely reflecting lower liquidity and less frequent trading in option markets.

The performance ranking observed for quote and tick rule variants carries over to hybrid methods. The LR algorithm performs best when the quote rule is applied sequentially to NBBO and then to ISE quotes. Replacing the standard tick test with a reverse tick test further improves accuracy to 63.97%. By contrast, the conventional LR algorithm using NBBO quotes and the standard tick test achieves only 63.53%, underperforming the quote rule alone. This shortfall reflects the weak performance of tick tests in classifying midpoint trades, where accuracy falls below 50%. The EMO rule performs worse, as it relies more heavily on tick tests, applying them not only to midpoint trades but also to trades executed inside and outside the spread. The highest accuracy achieved by the EMO rule is 57.25%, obtained by combining the quote rule at the NBBO with the reverse tick test. Incorporating ISE quotes does not improve performance, suggesting that when NBBO and ISE quotes diverge, tick-based inference is more informative than exchange-specific quotes for determining customer trade direction.

Overall, the accuracy of established stock-based trade classification rules in identifying the customer side of option trades is remarkably low. This weak performance is driven largely by trades with sizes equal to the quoted bid or ask depth, for which quote rules achieve success rates of only about 29–32% in Table 2.6. Hybrid methods perform similarly poorly for this subsample, with success rates well below 50%, while tick tests perform no worse than in the overall matched ISE sample. Importantly, these trades account for 22.3% of trades in the sample (see the last row of Panel A in Table 2.4) and therefore contribute substantially to the overall low classification accuracy. Taken together, these results indicate that trade classification rules developed to identify the trade initiator in equity markets do not transfer well to identifying the customer side of trades in option markets, highlighting the need for new classification rules.

Table 2.6: In-sample performance of trade classification rules

This table reports the in-sample performance of standard trade classification rules and the trade size rule. The first four columns report the share of unclassified trades, overall classification success, and success rates for subsamples in which trade size equals the ask size or the bid size. The last two columns report the classification success and incremental improvement (Δ) when the trade size rule is applied prior to the standard classification rules. For quote rules and hybrid methods, we indicate the origin and ordering of quotes (NBBO or BBO). For (reverse) tick tests, we specify whether trade price information is taken across all exchanges or from the trading venue only. Success rates of hybrid methods are computed using trade prices across all exchanges. Trades that cannot be classified by a given rule are assigned a correct classification probability of 50%. All values are in percent.

	Standard classification rules				Trade size rule	
	Unclassified	Success	Ask size	Bid size	Success	Δ
Quote rule						
NBBO	10.22	63.72	31.82	30.93	74.53	10.81
BBO	10.86	62.68	29.67	28.63	73.98	11.30
NBBO, BBO	7.35	63.94	31.26	30.23	74.89	10.95
BBO, NBBO	7.35	63.50	30.06	28.96	74.73	11.22
Tick test						
Venue	8.42	49.90	47.48	44.28	57.46	7.56
All	2.15	53.29	47.79	46.64	60.56	7.28
Reverse tick test						
Venue	9.64	51.38	49.03	53.69	57.78	6.40
All	3.29	54.88	53.21	55.26	60.62	5.74
LR						
NBBO	0.24	63.53	32.09	30.95	74.31	10.77
BBO	0.24	62.54	29.99	28.67	73.80	11.26
NBBO, BBO	0.19	63.79	31.51	30.28	74.70	10.91
BBO, NBBO	0.19	63.35	30.31	29.01	74.54	11.19
Reverse LR						
NBBO	0.30	63.77	32.16	31.62	74.47	10.70
BBO	0.36	62.80	30.01	29.29	73.99	11.20
NBBO, BBO	0.23	63.97	31.52	30.77	74.83	10.86
BBO, NBBO	0.23	63.53	30.32	29.51	74.67	11.14
EMO						
NBBO	1.24	57.09	31.56	32.49	67.77	10.68
BBO	1.27	55.40	30.43	31.33	66.33	10.93
NBBO, BBO	1.22	57.03	31.06	31.98	67.82	10.79
BBO, NBBO	1.22	57.03	31.05	31.97	67.82	10.79
Reverse EMO						
NBBO	1.56	57.25	31.91	34.10	67.72	10.47
BBO	1.72	55.52	30.74	32.97	66.24	10.72
NBBO, BBO	1.53	57.17	31.38	33.50	67.77	10.59
BBO, NBBO	1.53	57.17	31.37	33.50	67.77	10.59

2.4.2 Trade Size Rule for Limit Orders

We argue that the poor performance of existing classification rules reflects frequent executions of customer limit orders resting on the limit order book. When such orders remain at the best bid or ask, they determine the displayed depth. In this case, the quoted depth reflects customers' willingness to buy the quoted bid size or sell the quoted ask size. When market makers subsequently execute these orders, they often fill them completely when profitable, resulting in trades whose size matches the quote depth. Because quote rules infer trade direction from the aggressive side of the transaction, such trades are systematically misclassified by existing methods. This mechanism implies that trade size relative to quoted depth contains information about the customer side of the transaction.

To exploit this information, we propose a simple trade size rule that infers the customer side of a transaction from the relationship between trade size and quoted depth. Trades whose size equals the quoted bid depth are classified as customer buys, while trades whose size equals the quoted ask depth are classified as customer sells. Conditional on trades for which the rule applies, the trade size rule achieves a classification accuracy of 79.92%, nearly 50 percentage points higher than the accuracy of conventional quote rules for the same subset of trades.

Motivated by this result, we implement the trade size rule as a first step and apply existing classification methods to the remaining trades, i.e., those whose size does not match either quoted depth or coincides with both. Incorporating the trade size rule substantially improves classification accuracy across all methods (see the last two columns of Table 2.6). Success rates increase by 10.5–11.8 percentage points for quote rules and hybrid methods, and by 5.7–7.6 percentage points for tick tests. Importantly, the relative ranking of methods remains unchanged. Quote rules continue to outperform tick tests and hybrid methods. The highest accuracy is achieved by the quote rule applied sequentially to NBBO and ISE quotes, which correctly classifies 74.89% of trades when applied after the trade size rule. Notably, incorporating the LR algorithm does not further improve the classification accuracy of the quote rule. Taken together, these results demonstrate that exploiting information contained in quoted depth provides a simple and effective way to identify the customer side of option trades. We next examine the economic mechanism underlying this rule through subsample analysis.

2.4.3 Economic Mechanism

We examine how the classification accuracy of the trade size rule varies across subsamples in which trades with sizes equal to quoted depth are more or less likely to originate from customer limit orders. Because our analysis focuses on broader patterns and variation within methods has only a minor impact on classification success, we report results for a subset of representative rules. Specifically, we consider quote rules applied sequentially to NBBO and BBO quotes, (reverse) tick tests based on consolidated transaction prices, and hybrid methods that combine both approaches.

First, we examine classification performance across trade size quintiles in Table 2.7. The trade size rule yields the largest gains for small to medium-sized trades, increasing success rates by 13–16 percentage points for the quote rule. In contrast, performance deteriorates slightly for trades in the highest size quintile, by about 2 percentage points. Notably, classification accuracy without the trade size rule generally increases with trade size. This pattern is consistent with the economic intuition that customer limit orders tend to be small. In contrast, large trades whose size equals the quoted depth more often reflect customer market orders that consume the full depth of the market maker’s quote.

Table 2.7: In-sample performance with trade size rule by trade size

This table reports the in-sample performance of trade classification rules across trade size quintiles (Q1–Q5) after applying the trade size rule first. We report the success rate (S) and the incremental improvement (Δ) relative to each standalone method without prior application of the trade size rule. Quote rules are applied sequentially to NBBO and BBO quotes, (reverse) tick tests rely on consolidated transaction prices, and hybrid methods combine both. All values are in percent.

	Q1		Q2		Q3		Q4		Q5	
	S	Δ	S	Δ	S	Δ	S	Δ	S	Δ
Quote rule	75.03	15.10	78.12	15.59	78.73	15.30	76.27	12.48	68.60	-1.95
Tick test	62.33	10.34	62.84	10.36	63.13	9.96	61.14	8.47	54.57	-1.79
Reverse tick test	62.93	7.86	63.19	7.91	63.37	7.86	61.62	6.43	53.06	-0.56
LR	74.86	15.05	77.94	15.56	78.53	15.26	76.07	12.46	68.43	-1.98
Reverse LR	75.00	14.97	78.09	15.50	78.69	15.20	76.23	12.41	68.45	-1.99
EMO	67.93	14.77	70.50	15.38	71.28	15.03	69.29	12.46	62.04	-1.98
Reverse EMO	68.13	14.44	70.74	15.14	71.34	14.83	69.44	12.29	61.14	-2.00

Second, we examine classification performance across LiveVol trade condition identifiers to capture differences in execution mechanisms in Table 2.8. We group trades into three categories: standard single-leg trades executed directly against

displayed quotes, multi-leg trades combining several option legs into a single transaction, and a residual category consisting primarily of intermarket sweep orders that access liquidity simultaneously across multiple exchanges. The trade size rule yields the largest gains for standard single-leg trades, increasing classification accuracy by 24 percentage points for the quote rule and hybrid methods. Improvements are much smaller for multi-leg trades, at about 3 percentage points, while performance deteriorates by roughly 10 percentage points for intermarket sweep orders. This pattern is consistent with the economic mechanism underlying the trade size rule. Because the rule is designed to identify executions of resting customer limit orders, it performs best when trades interact directly with displayed quotes and less well when executions occur through mechanisms that bypass the limit order book, such as bundled pricing across multiple legs or liquidity accessed across several exchanges.

Table 2.8: In-sample performance with trade size rule by trade condition

This table reports the in-sample performance of trade classification rules across trade condition IDs after applying the trade size rule first. Trade conditions are grouped into three categories: ‘single-leg’ includes regular (0) and electronically executed trades (18), ‘multi-leg’ contains spreads (35), straddles (36), and combos (38), and ‘other’ consists primarily of intermarket sweep orders (95) and remaining trades. We report the success rate (S) and the incremental improvement (Δ) relative to each standalone method without prior application of the trade size rule. Quote rules are applied sequentially to NBBO and BBO quotes, (reverse) tick tests rely on consolidated transaction prices, and hybrid methods combine both. All values are in percent.

	Single-leg		Multi-leg		Other	
	S	Δ	S	Δ	S	Δ
Quote rule	80.13	24.27	70.73	3.17	75.19	-9.67
Tick test	67.09	14.80	56.06	3.20	55.77	-6.74
Reverse tick test	68.20	11.34	56.03	2.66	50.15	-4.44
LR	79.93	24.28	70.52	3.10	75.27	-9.67
Reverse LR	80.22	24.27	70.59	3.00	74.85	-9.66
EMO	77.35	24.42	60.32	2.74	67.99	-9.59
Reverse EMO	77.93	24.42	60.17	2.36	64.83	-9.51

Third, we evaluate classification performance by the location of the transaction price relative to the prevailing bid and ask quotes at the ISE in Table 2.9. The trade size rule yields the largest gains for at-quote trades, increasing success rates by 19 percentage points for the quote rule and hybrid methods. Improvements are smaller for trades executed inside the spread or at the midpoint, while performance deteriorates slightly for trades executed outside the quotes, by about 3 percentage points. This pattern is consistent with the economic intuition that executions against resting customer limit orders most often occur at the quoted bid or ask. In contrast,

trades outside the spread typically reflect large market orders that consume liquidity across multiple price levels, so transaction prices are less tightly linked to the depth displayed at the top of the book. Notably, baseline classification accuracy under the quote rule without the trade size rule already exceeds 85% for outside-quote trades. By contrast, accuracy for midpoint trades remains low. Even after incorporating the trade size rule, success rates do not exceed 59%, indicating that such trades are inherently more difficult to classify.

Table 2.9: In-sample performance with trade size rule by quote location

This table reports the in-sample performance of trade classification rules across locations of the transaction price relative to the prevailing quotes at the trading venue, after applying the trade size rule first. We report the success rate (S) and the incremental improvement (Δ) relative to each standalone method without prior application of the trade size rule. Quote rules are applied sequentially to NBBO and BBO quotes, (reverse) tick tests rely on consolidated transaction prices, and hybrid methods combine both. All values are in percent.

	at quote		at mid		inside		outside	
	S	Δ	S	Δ	S	Δ	S	Δ
Quote rule	81.23	19.15	59.23	5.34	70.29	2.37	83.32	-2.57
Tick test	66.88	12.10	52.64	4.75	54.15	2.08	67.75	-1.52
Reverse tick test	67.11	9.41	54.12	4.08	53.61	1.73	67.93	-1.25
LR	81.23	19.15	57.21	4.98	70.29	2.37	83.32	-2.57
Reverse LR	81.23	19.15	58.38	4.41	70.29	2.37	83.32	-2.57
EMO	81.23	19.15	52.64	4.77	54.16	2.09	74.11	-2.10
Reverse EMO	81.23	19.15	54.12	4.10	53.61	1.74	74.44	-2.12

2.4.4 Trade Size Rule Refinements

The subsample evidence above indicates that the effectiveness of the trade size rule varies systematically with trade characteristics related to its underlying economic mechanism. In particular, classification improvements are concentrated among small to medium-sized trades, standard single-leg executions, and transactions occurring at the quoted bid or ask. By contrast, performance deteriorates for large trades, intermarket sweep orders, and transactions executed outside the prevailing quotes.

Motivated by this heterogeneity, we consider three potential refinements that condition the application of the trade size rule on observable trade characteristics. First, a ‘trade size refinement’ restricts the rule to the smallest 90% of trades within each underlying and month. Second, a ‘trade condition refinement’ applies the rule

only to standard single-leg trades. Third, a ‘quote location refinement’ excludes trades executed outside the prevailing bid or ask quotes from the trade size rule.

Under each refinement, the trade size rule is applied only when the respective condition is satisfied, thereby reducing the number of classifiable trades relative to the baseline trade size rule (Table 2.10). While the quote location refinement excludes only a small number of trades, the trade condition refinement is considerably more restrictive. Conditional on the trades to which the respective refinement applies, all three refinements slightly improve upon the baseline trade size rule, with the highest accuracy of 85% achieved under the trade size refinement.

We next implement these refinements as a first step and apply the quote rule, sequentially to NBBO and ISE quotes, to trades that do not meet the respective refinement criterion. Under this approach, the trade size and quote location refinements generate small in-sample gains, raising the classification accuracy by a few basis points relative to the baseline trade size rule. By contrast, the trade condition refinement slightly reduces performance by one percentage point. This result reflects the fact that the refinement excludes multi-leg trades, for which the trade size rule continues to outperform existing classification methods despite not being specifically designed to target such trades.

Table 2.10: In-sample performance of trade size rule refinements

This table reports the in-sample performance of three refinements of the trade size rule. The ‘baseline’ applies the rule to all trades. The ‘trade size refinement’ restricts it to the smallest 90% of trades within each underlying and month, the ‘trade condition refinement’ applies it only to standard single-leg trades, and the ‘quote location refinement’ excludes trades executed outside the prevailing BBO quotes. For each refinement, we report the share of classifiable trades and the corresponding classification success. We then report success rates of the quote rule applied sequentially to NBBO and BBO quotes after each refinement, along with the incremental improvement (Δ) relative to the success of the quote rule under the baseline trade size rule. All values are in percent.

	Baseline	Trade Size	Trade Condition	Quote Location
Classifiable trades	22.28	19.11	16.07	22.17
Success (classifiable)	79.92	84.82	82.23	80.07
Success (quote rule)	74.89	75.07	73.89	74.92
Δ vs. Baseline	–	0.18	-1.00	0.03

2.4.5 Depth Rule for Midpoint Trades

Despite the substantial improvement achieved by the trade size rule, classification accuracy for trades executed at the midpoint of the bid–ask spread remains low. By construction, quote rules cannot assign direction to trades executed exactly at the midpoint. In equity markets, this limitation is often addressed using tick tests. However, as shown in Section 2.4.1, tick tests perform poorly in option markets due to low trading frequency and long intervals between consecutive trades. We therefore introduce a depth rule that exploits information contained in quoted depth to classify midpoint trades.

The depth rule builds on the concept of the weighted midpoint, which provides an alternative proxy for the option’s fundamental value. Specifically, the weighted midpoint is defined as

$$\frac{P^B Q^A + P^A Q^B}{Q^A + Q^B}, \quad (2.4)$$

where P^A and Q^A denote the ask price and depth, and P^B and Q^B denote the bid price and depth (see Hagströmer, 2021). Intuitively, the weighted midpoint shifts toward the side with less quoted depth. This reflects the idea that market makers tend to post greater depth on the side of the market where they expect to earn higher compensation. Because minimum tick sizes constrain the granularity of price quotes, market makers cannot always quote symmetrically around the fundamental value (see Glosten, 1994; Anshuman and Kalay, 1998). Instead, they tend to place deeper quotes slightly farther from fair value and lighter quotes closer to it.

Based on this intuition, we classify midpoint trades as customer buys when ask depth exceeds bid depth and as customer sells when bid depth exceeds ask depth. The depth rule builds on the idea that larger quoted depth at the best bid or ask reflects greater liquidity provision and therefore a stronger willingness to trade on that side of the market, similar to the informational content of a tighter quote (see, e.g., Goldstein and Kavajecz, 2000). The depth rule applies to approximately 7% of ISE trades and classifies these trades with an accuracy of 68%. Although this standalone performance may appear moderate, it is economically meaningful because the rule targets midpoint trades for which alternative methods perform particularly poorly. Indeed, for this subset of trades, tick tests perform worse than random assignment.

Motivated by this result, we incorporate the depth rule into a sequential approach that first applies the trade size rule, followed by the quote rule sequentially to NBBO and ISE quotes, and finally the depth rule to classify midpoint trades. This approach

increases overall classification accuracy to 75.74%, representing an improvement of about 0.9 percentage points relative to the approach that combines the trade size and quote rules but excludes the depth rule. These results suggest that quoted depth contains more information than trade prices for classifying midpoint option trades.

2.5 Out-of-Sample Performance

To evaluate the robustness and generalizability of trade classification rules, we conduct out-of-sample tests along three dimensions: over time, across option exchanges, and in more liquid environments. We first test whether the relative accuracy of established classification rules documented in-sample persists. We then evaluate the performance of the newly proposed rules and examine whether the patterns identified in the subsample analysis extend to the broader out-of-sample settings. Finally, we assess the effectiveness of the proposed refinements.

2.5.1 Established Rules

We begin by examining the out-of-sample performance of established trade classification rules along three robustness dimensions. First, the time dimension asks whether the relative performance of the rules remains stable when they are evaluated in a later period. Second, the cross-exchange dimension tests whether the same rules perform similarly on other options exchanges with different market designs. Third, the intraday dimension examines whether the results change in a more liquid benchmark setting. Table 2.11 reports the results.

Using out-of-sample ISE data from May 2017 to June 2021, we first assess robustness over time and find that the relative performance patterns observed in the original ISE sample persist in this later period. Quote rules remain the most accurate. Variants based on NBBO outperform those based on venue-specific quotes, while sequentially combining the two yields modest additional gains. Tick tests and EMO rules continue to underperform, whereas LR algorithms maintain performance comparable to quote rules.

Next, we compare classification performance across four additional exchanges. Success rates on CBOE and PHLX are broadly comparable to those on the ISE, and the relative ranking of the methods remains qualitatively similar, with quote rules consistently outperforming tick tests and hybrid methods. Interestingly, quote rules

Table 2.11: Out-of-sample performance of trade classification rules

This table reports the out-of-sample performance of standard trade classification rules across the three robustness dimensions: over time, cross-exchange, and intraday. For comparison, the first column reports the corresponding in-sample ISE results. ‘Over time’ refers to the post-May 2017 ISE sample, ‘Cross-exchange’ to four additional exchanges (CBOE, PHLX, NOM, GEMX), and ‘Intraday’ to the GEMX intraday sample. For quote rules and hybrid methods, we indicate the origin and ordering of quotes (NBBO or BBO). For (reverse) tick tests, we specify whether trade price information is taken across all exchanges or from the trading venue only. Success rates of hybrid methods are computed using trade prices across all exchanges. Trades that cannot be classified by a given rule are assigned a correct classification probability of 50%. All values are in percent.

	In-sample	Over time	Cross-exchange				Intraday
	ISE	ISE	CBOE	PHLX	NOM	GEMX	GEMX
Quote rule							
NBBO	63.72	64.31	62.30	65.61	39.56	31.79	30.13
BBO	62.68	60.96	62.53	66.52	42.41	34.14	32.22
NBBO, BBO	63.94	64.64	63.22	66.36	40.74	32.78	31.08
BBO, NBBO	63.50	61.55	63.24	67.38	42.32	34.20	32.31
Tick test							
Venue	49.90	50.40	49.49	49.27	44.52	44.47	44.50
All	53.29	53.54	51.02	52.92	44.60	44.91	44.22
Reverse tick test							
Venue	51.38	50.16	50.48	51.54	52.38	52.32	53.53
All	54.88	53.69	52.30	56.45	56.47	57.00	58.68
LR							
NBBO	63.53	64.29	61.80	65.19	38.81	31.33	29.64
BBO	62.54	61.07	62.10	66.16	41.78	33.84	31.99
NBBO, BBO	63.79	64.67	62.92	66.09	40.31	32.65	30.96
BBO, NBBO	63.35	61.57	62.94	67.11	41.90	34.07	32.19
Reverse LR							
NBBO	63.77	64.05	62.20	65.61	40.10	32.11	30.56
BBO	62.80	60.96	62.46	66.75	42.78	34.32	32.45
NBBO, BBO	63.97	64.56	63.13	66.33	41.00	32.87	31.19
BBO, NBBO	63.53	61.47	63.15	67.35	42.59	34.29	32.42
EMO							
NBBO	57.09	56.50	53.17	58.94	34.93	28.76	27.38
BBO	55.40	54.56	50.62	57.73	34.67	28.14	26.83
NBBO, BBO	57.03	56.48	53.15	59.00	34.90	28.64	27.28
BBO, NBBO	57.03	56.48	53.15	59.00	34.90	28.64	27.28
Reverse EMO							
NBBO	57.25	56.58	53.66	60.89	40.47	31.71	30.34
BBO	55.52	54.51	50.98	59.65	40.22	31.18	29.95
NBBO, BBO	57.17	56.58	53.64	60.98	40.44	31.62	30.25
BBO, NBBO	57.17	56.58	53.64	60.98	40.44	31.62	30.25

perform better when applied to venue-specific quotes than to the NBBO on all four out-of-sample exchanges. Most notably, performance deteriorates sharply on NOM and GEMX. On these exchanges, quote rules and hybrid methods achieve success rates of only 28–43%, well below random classification, indicating that customer trade direction is systematically misclassified. These differences likely reflect differences in exchange pricing models. NOM and GEMX employ maker–taker fee structures that encourage the submission of customer limit orders. By contrast, CBOE and PHLX rely more heavily on PFOF arrangements that attract marketable retail order flow, while ISE operates a hybrid model combining elements of both structures.

Tick tests exhibit low classification accuracy across all exchanges. Standard tick rules often produce near-random classifications, while reverse tick tests perform only slightly better. This weak performance carries over to hybrid methods that use tick tests to classify midpoint trades. Although the reverse LR specification slightly outperforms the standard LR algorithm, neither approach exceeds the performance of quote rules on CBOE or PHLX. The apparent advantage of tick tests on NOM and GEMX reflects the particularly poor performance of quote rules on these exchanges rather than strong predictive power of tick tests.

Finally, we assess robustness in a more liquid environment using the intraday matched GEMX sample. Classification accuracy is similar to that observed in the daily GEMX sample, with quote rules achieving success rates of only 30–32%. If the poor performance were primarily driven by illiquidity, such as infrequent quote updates, accuracy should improve in this higher-frequency setting. Instead, performance deteriorates slightly. This evidence suggests that the limited accuracy of existing classification rules cannot be attributed to the lower liquidity of the daily matched sample.

2.5.2 Trade Size Rule

We next examine the out-of-sample performance of the trade size rule. We first evaluate the rule when applied on its own and then assess its incremental contribution when combined with existing trade classification algorithms.

Table 2.12 reports its standalone performance. In the post-2017 ISE sample, the rule applies to a smaller fraction of trades and achieves lower success rates than in the earlier in-sample period, reflecting structural changes in the option trading environment over time. Across exchanges, the rule classifies between 11–17% of trades on venues operating under PFOF pricing models and correctly infers customer trade

direction in 66–76% of the trades to which it applies. On maker–taker exchanges that incentivize limit order submission, the share of classifiable trades rises to 46–56%, with accuracy reaching 82–93%. Performance in the intraday GEMX sample is comparable to that in the daily GEMX sample, indicating that the effectiveness of the trade size rule is not driven by features specific to the daily matching procedure.

Table 2.12: Out-of-sample performance of trade size rule

This table reports the out-of-sample performance of the trade size rule across the three robustness dimensions: over time, cross-exchange, and intraday. For comparison, the first column reports the corresponding in-sample ISE results. For each sample, we report the share of classifiable trades and the corresponding classification success. All values are in percent.

Trade size rule	In-sample	Over time	Cross-exchange				Intraday
	ISE	ISE	CBOE	PHLX	NOM	GEMX	GEMX
Classifiable trades	22.28	11.39	11.17	17.37	46.35	55.50	54.60
Success (classifiable)	79.92	66.08	75.94	74.98	82.26	92.75	91.98

We next combine the trade size rule with existing classification algorithms by applying it first and then classifying the remaining trades using competing methods. Table 2.13 reports the resulting success rates (Panel A) as well as the incremental improvement relative to the methods applied without the trade size rule (Panel B). Incorporating the trade size rule improves classification accuracy across all robustness dimensions. In the post-2017 ISE sample, the gains are modest, increasing success rates by roughly 2 percentage points for quote rules and hybrid methods. In contrast, the gains are substantially larger across exchanges. For CBOE and PHLX, classification accuracy increases by approximately 5–7 percentage points, while gains are particularly pronounced on NOM and GEMX, where success rates rise by more than 28 and 46 percentage points, respectively. These large improvements reflect the poor performance of existing methods on these venues and indicate that the trade size rule corrects systematic misclassification. Similar patterns emerge in the intraday GEMX sample. Applying the trade size rule raises classification accuracy by around 45 percentage points, only slightly below the gains observed in the daily sample, indicating that the rule remains effective even in more liquid environments.

Taken together, these results show that the trade size rule substantially improves the performance of existing trade classification methods. While the magnitude of the gains varies across exchanges and time periods, overall success rates become considerably more stable once the rule is incorporated, suggesting that it is particularly valuable in settings where existing methods perform poorly.

Table 2.13: Out-of-sample performance with trade size rule

This table reports the out-of-sample performance of standard trade classification rules after applying the trade size rule first, across the three robustness dimensions: over time, cross-exchange, and intraday. Panel A reports the classification success, while Panel B shows the incremental improvement relative to each standalone method without prior application of the trade size rule. For comparison, the first column reports the corresponding in-sample ISE results. For quote rules and hybrid methods, we indicate the origin and ordering of quotes (NBBO or BBO). For (reverse) tick tests, we specify whether trade price information is taken across all exchanges or from the trading venue only. Success rates of hybrid methods are computed using trade prices across all exchanges. Trades that cannot be classified by a given rule are assigned a correct classification probability of 50%. All values are in percent.

Panel A: Classification success

	In-sample	Over time	Cross-exchange				Intraday
	ISE	ISE	CBOE	PHLX	NOM	GEMX	GEMX
Quote rule							
NBBO	74.53	66.01	67.16	72.63	67.53	78.16	75.07
BBO	73.98	63.08	67.44	74.04	69.98	80.31	76.97
NBBO, BBO	74.89	66.42	68.07	73.39	68.61	79.15	76.00
BBO, NBBO	74.73	63.63	68.13	74.75	69.92	80.38	77.05
Tick test							
Venue	57.46	52.23	52.85	54.40	62.35	71.68	71.02
All	60.56	55.01	54.37	57.81	62.81	72.65	71.76
Reverse tick test							
Venue	57.78	51.93	53.24	55.62	66.10	74.68	74.44
All	60.62	54.94	54.84	59.73	68.21	77.11	77.18
LR							
NBBO	74.31	65.96	66.67	72.21	66.85	77.71	74.61
BBO	73.80	63.16	67.02	73.63	69.41	80.03	76.75
NBBO, BBO	74.70	66.43	67.79	73.12	68.22	79.02	75.90
BBO, NBBO	74.54	63.64	67.84	74.48	69.53	80.25	76.95
Reverse LR							
NBBO	74.47	65.72	67.05	72.61	68.00	78.43	75.44
BBO	73.99	63.04	67.36	74.22	70.30	80.48	77.17
NBBO, BBO	74.83	66.31	67.98	73.34	68.84	79.23	76.10
BBO, NBBO	74.67	63.52	68.03	74.71	70.15	80.46	77.15
EMO							
NBBO	67.77	58.50	58.35	66.08	63.05	75.35	72.57
BBO	66.33	56.73	55.96	65.12	62.84	74.91	72.16
NBBO, BBO	67.82	58.52	58.38	66.16	63.06	75.36	72.57
BBO, NBBO	67.82	58.52	58.38	66.16	63.06	75.36	72.57
Reverse EMO							
NBBO	67.72	58.48	58.75	67.88	67.48	77.97	75.15
BBO	66.24	56.59	56.23	66.92	67.27	77.60	74.88
NBBO, BBO	67.77	58.51	58.78	67.97	67.49	78.00	75.16
BBO, NBBO	67.77	58.51	58.78	67.97	67.49	78.00	75.16

Table 2.13 continued

Panel B: Improvement							
	In-sample	Over time	Cross-exchange				Intraday
	ISE	ISE	CBOE	PHLX	NOM	GEMX	GEMX
Quote rule							
NBBO	10.81	1.70	4.86	7.02	27.97	46.37	44.94
BBO	11.30	2.12	4.91	7.52	27.57	46.17	44.75
NBBO, BBO	10.95	1.78	4.85	7.03	27.87	46.37	44.92
BBO, NBBO	11.22	2.08	4.89	7.37	27.60	46.18	44.74
Tick test							
Venue	7.56	1.83	3.36	5.13	17.83	27.21	26.52
All	7.28	1.47	3.35	4.89	18.21	27.74	27.54
Reverse tick test							
Venue	6.40	1.77	2.76	4.08	13.72	22.36	20.91
All	5.74	1.25	2.54	3.28	11.74	20.11	18.50
LR							
NBBO	10.77	1.67	4.87	7.02	28.04	46.38	44.97
BBO	11.26	2.09	4.92	7.47	27.63	46.19	44.76
NBBO, BBO	10.91	1.76	4.87	7.03	27.91	46.37	44.94
BBO, NBBO	11.19	2.07	4.90	7.37	27.63	46.18	44.76
Reverse LR							
NBBO	10.70	1.67	4.85	7.00	27.90	46.32	44.88
BBO	11.20	2.08	4.90	7.47	27.52	46.16	44.72
NBBO, BBO	10.86	1.75	4.85	7.01	27.84	46.36	44.91
BBO, NBBO	11.14	2.05	4.88	7.36	27.56	46.17	44.73
EMO							
NBBO	10.68	2.00	5.18	7.14	28.12	46.59	45.19
BBO	10.93	2.17	5.34	7.39	28.17	46.77	45.33
NBBO, BBO	10.79	2.04	5.23	7.16	28.16	46.72	45.29
BBO, NBBO	10.79	2.04	5.23	7.16	28.16	46.72	45.29
Reverse EMO							
NBBO	10.47	1.90	5.09	6.99	27.01	46.26	44.81
BBO	10.72	2.08	5.25	7.27	27.05	46.42	44.93
NBBO, BBO	10.59	1.93	5.14	6.99	27.05	46.38	44.91
BBO, NBBO	10.59	1.93	5.14	6.99	27.05	46.38	44.91

2.5.3 Subsample Evidence and Refinements

To examine whether the economic mechanism underlying the trade size rule persists out of sample, we analyze classification performance across subsamples defined by trade size, trade condition, and quote location. We then evaluate whether the refinements proposed in the in-sample analysis improve out-of-sample classification performance. For brevity, the subsample analyses focus on the quote rule, applied sequentially to NBBO and BBO quotes, and compare its success and incremental improvement when the trade size rule is applied first.

Table 2.14 reports out-of-sample performance across trade-size quintiles. The results broadly confirm the in-sample patterns. The largest improvements relative to the quote rule occur for small to medium-sized trades, while performance deteriorates in the highest size quintile on ISE and PHLX. For CBOE, NOM, and GEMX, improvements decline with trade size but remain positive, exceeding 25 percentage points even for the largest trades on NOM and GEMX. These large gains reflect the particularly poor standalone performance of the quote rule on maker-taker exchanges.

Table 2.14: Out-of-sample performance with trade size rule by trade size

This table reports the out-of-sample performance of the quote rule, sequentially applied to NBBO and BBO quotes, across trade size quintiles (Q1–Q5) after applying the trade size rule first. We report the success rate (S) and the incremental improvement (Δ) relative to the success of the quote rule without prior application of the trade size rule. All values are in percent.

Quote rule	Q1		Q2		Q3		Q4		Q5	
	S	Δ	S	Δ	S	Δ	S	Δ	S	Δ
ISE	65.50	2.64	65.98	2.19	67.04	1.69	66.77	0.78	69.57	-0.65
CBOE	64.04	4.64	67.86	4.93	69.35	5.27	70.15	5.84	72.62	3.59
PHLX	72.61	11.27	76.13	11.27	76.74	11.56	75.63	7.22	67.54	-7.85
NOM	69.31	30.84	68.42	29.40	68.17	27.58	69.84	28.97	65.37	17.43
GEMX	79.90	51.87	79.90	51.87	82.92	51.98	80.15	44.39	69.99	26.30
Intraday GEMX	75.41	49.92	75.41	49.92	78.59	49.72	78.52	42.95	69.73	24.76

Table 2.15 reports out-of-sample performance by trade condition. Consistent with the proposed mechanism, the trade size rule yields its largest improvements for standard single-leg trades, for which executions are most closely linked to displayed quotes. In contrast, improvements are close to zero or negative for trades that are less directly tied to displayed depth. This includes the SLIMC category, which was introduced only after the in-sample period and captures trades that are often

internalized or executed through price-improvement auctions and crossing mechanisms. For these trades, the link between trade size and quoted depth is therefore weaker, and the trade size rule slightly reduces classification accuracy. Intermarket sweep orders, grouped in the “other” category, exhibit the largest negative effects, while multi-leg trades remain largely unaffected. The relevance of these categories also varies with exchange-specific market structures. For example, NOM does not operate price-improvement auctions, and NOM and GEMX do not support complex-order functionality for multi-leg strategies.

Table 2.15: Out-of-sample performance with trade size rule by trade condition

This table reports the out-of-sample performance of the quote rule, sequentially applied to NBBO and BBO quotes, across trade condition IDs after applying the trade size rule first. Trade conditions are grouped into four categories: ‘single-leg’ includes regular (0) and electronically executed trades (18), ‘SLIMC’ refers to single-leg trades executed via price improvement (114) or other internal crossing mechanisms (115–118), ‘multi-leg’ contains spreads (35), straddles (36), combos (38) and other multi-leg flags (119–129, 133), and ‘other’ consists primarily of intermarket sweep orders (95) and remaining trades. Entries that are not applicable are indicated by –. We report the success rate (S) and the incremental improvement (Δ) relative to the success of the quote rule without prior application of the trade size rule. All values are in percent.

Quote rule	Single-leg		SLIMC		Multi-leg		Other	
	S	Δ	S	Δ	S	Δ	S	Δ
ISE	78.42	15.14	82.34	-6.18	64.76	0.84	73.50	-6.06
CBOE	75.15	19.73	65.23	-1.10	64.04	-0.31	70.23	-1.47
PHLX	77.70	13.49	74.90	-0.79	65.08	1.58	81.83	-14.72
NOM	68.43	29.45	–	–	–	–	74.20	-19.46
GEMX	79.15	49.42	81.85	-6.25	–	–	78.76	-14.33
Intraday GEMX	75.87	48.30	79.54	-6.82	–	–	78.18	-16.31

Table 2.16 reports out-of-sample performance by the location of the transaction price relative to the prevailing bid and ask quotes at the trading venue. Consistent with the in-sample evidence, the trade size rule yields its largest improvements for trades executed exactly at the quoted bid or ask. Trades executed inside the spread show smaller gains, with a slightly negative effect for CBOE. Classification remains challenging for midpoint trades, where success rates generally remain below 60% even after applying the trade size rule. Trades executed outside the quotes exhibit mixed outcomes, with small negative effects on most exchanges, a pronounced deterioration of 51.5 percentage points on NOM, and a modest improvement in the intraday GEMX sample.

Table 2.16: Out-of-sample performance with trade size rule by quote location

This table reports the out-of-sample performance of the quote rule, sequentially applied to NBBO and BBO quotes, across locations of the transaction price relative to the prevailing quotes at the trading venue, after applying the trade size rule first. We report the success rate (S) and the incremental improvement (Δ) relative to each standalone method without prior application of the trade size rule. All values are in percent.

Quote rule	at quote		at mid		inside		outside	
	S	Δ	S	Δ	S	Δ	S	Δ
ISE	82.35	5.98	53.94	1.87	64.71	0.61	79.44	-2.21
CBOE	77.32	16.27	57.45	0.19	65.45	-0.04	82.30	-2.44
PHLX	83.38	13.60	59.15	1.49	64.28	0.90	74.73	-1.55
NOM	73.16	44.04	65.08	2.17	62.15	8.15	43.36	-51.50
GEMX	82.41	56.28	74.05	0.82	61.13	2.63	81.37	-5.16
Intraday GEMX	78.96	56.18	54.26	2.98	65.76	1.25	83.81	4.98

Table 2.17 summarizes the out-of-sample performance of the three candidate refinements. The trade size refinement, which excludes the largest trades from the trade size rule, yields small improvements on ISE and PHLX but does not improve classification accuracy relative to the baseline trade size rule on CBOE, NOM, or GEMX. We therefore interpret this refinement as capturing exchange-specific patterns rather than a general structural effect and exclude it from the final recommendation.

The trade condition refinement, which restricts the trade size rule to standard single-leg trades, improves performance on most exchanges except ISE. These gains are consistent with the subsample evidence, which shows that standard single-leg trades are the setting in which transaction size is most closely linked to displayed depth. By contrast, auction, crossing, and multi-leg trades are less directly tied to the limit order book, weakening the economic mechanism underlying the rule.

The quote location refinement, which excludes trades executed outside the prevailing BBO quotes, improves accuracy across exchanges. However, the incremental gains relative to the baseline trade size rule remain small because such trades account for only a limited fraction of the sample. Consequently, excluding them has only a limited effect on overall classification accuracy, with somewhat larger gains on NOM, where the deterioration for outside-the-quote trades is most pronounced.

Overall, the out-of-sample evidence supports the economic mechanism underlying the trade size rule. At the same time, the candidate refinements provide only modest incremental gains, suggesting that the baseline rule captures most of the relevant classification improvement. We nevertheless retain the trade condition

and quote location refinements in the final recommendation because they are economically motivated and improve robustness by excluding trades for which the link between trade size and quoted depth is weaker.

Table 2.17: Out-of-sample performance of trade size rule refinements

This table reports the out-of-sample performance of three refinements of the trade size rule across the three robustness dimensions: over time, cross-exchange, and intraday. The ‘trade size refinement’ restricts it to the smallest 90% of trades within each underlying and month, the ‘trade condition refinement’ applies it only to standard single-leg trades, and the ‘quote location refinement’ excludes trades executed outside the prevailing BBO quotes. For each refinement, we report the share of classifiable trades and the corresponding classification success. We then report success rates of the quote rule applied sequentially to NBBO and BBO quotes after each refinement, along with the incremental improvement (Δ) relative to the success of the quote rule under the baseline trade size rule. For comparison, the first column reports the corresponding in-sample ISE results. All values are in percent.

	In-sample	Over time	Cross-exchange				Intraday
	ISE	ISE	CBOE	PHLX	NOM	GEMX	GEMX
Trade size refinement							
Classifiable trades	19.11	10.17	9.95	13.96	41.13	48.74	48.06
Success (classifiable)	84.82	67.36	76.92	82.85	83.36	94.24	93.29
Success (quote rule)	75.07	66.44	67.69	74.12	66.23	74.91	71.83
Δ vs. Baseline	0.18	0.02	-0.38	0.73	-2.38	-4.24	-4.17
Trade condition refinement							
Classifiable trades	16.07	3.12	7.80	14.42	45.53	54.35	53.36
Success (classifiable)	82.23	73.29	86.57	79.09	83.43	94.19	93.61
Success (quote rule)	73.89	65.95	68.54	73.76	69.24	79.80	76.79
Δ vs. Baseline	-1.00	-0.47	0.47	0.37	0.63	0.65	0.79
Quote location refinement							
Classifiable trades	22.17	11.25	11.12	17.18	45.83	55.44	54.58
Success (classifiable)	80.07	66.19	76.10	75.18	83.08	92.81	91.99
Success (quote rule)	74.92	66.46	68.09	73.43	69.06	79.16	76.00
Δ vs. Baseline	0.03	0.04	0.02	0.04	0.45	0.01	0.00

2.5.4 Depth Rule

We next evaluate the out-of-sample performance of the depth rule in Table 2.18. Specifically, we assess its standalone classification accuracy and its incremental contribution when incorporated into a sequential approach, which applies the trade size rule first, followed by the quote rule, sequentially to NBBO and BBO quotes, and finally the depth rule to classify midpoint trades. We compare its performance with

the approach that combines the trade size rule with the quote rule but excludes the depth rule.

Table 2.18: Out-of-sample performance of depth rule

This table reports the out-of-sample performance of the depth rule across the three robustness dimensions: over time, cross-exchange, and intraday. For comparison, the first column reports the corresponding in-sample ISE results. We report the share of classifiable trades (i.e., midpoint trades where bid and ask quote size do not coincide) and the corresponding classification success. We then report the success rate of the quote rule, applied sequentially to NBBO and BBO quotes after the trade size rule, when the depth rule is used to classify remaining midpoint trades, along with the incremental improvement (Δ) relative to the success of the quote rule under the trade size rule without the depth rule. All values are in percent.

	In-sample	Over time	Cross-exchange				Intraday
	ISE	ISE	CBOE	PHLX	NOM	GEMX	GEMX
Classifiable trades	7.03	6.99	7.01	6.51	4.75	1.34	1.31
Success (classifiable)	67.86	57.61	63.49	64.48	77.89	83.49	82.17
Success (quote rule)	75.74	66.78	68.98	74.23	69.77	79.55	76.37
Δ from depth rule	0.85	0.36	0.91	0.84	1.16	0.40	0.37

In the post-2017 ISE sample, the fraction of trades to which the depth rule applies remains stable at approximately 7%. Its standalone classification accuracy declines to about 58%, but remains economically meaningful given that midpoint trades are particularly difficult to classify and that tick tests perform poorly for this subset. When incorporated into a sequential approach, the depth rule therefore continues to improve overall classification accuracy, although the improvement is smaller than in the in-sample period. On the other PFOF exchanges, CBOE and PHLX, the share of classifiable trades is comparable and standalone classification accuracy remains close to that observed in the original ISE sample. On the maker-taker exchanges NOM and GEMX, the fraction of classifiable trades is smaller, but the standalone success rate is particularly high, exceeding 80% on GEMX. Nevertheless, the incremental gains within the sequential approach remain limited because the rule applies to only about 1.3% of trades on this exchange. The intraday GEMX sample exhibits success rates similar to those in the daily sample, indicating that the depth rule performs consistently even in more liquid trading environments.

Taken together, these results indicate that the depth rule improves the classification of midpoint trades out-of-sample, although its overall contribution to the classification accuracy is constrained by the relatively small share of trades to which the rule applies.

2.6 Recommended Classification Algorithm

This section presents our recommended trade classification algorithm for option markets. Building on the preceding analyses, we combine the trade size rule, quote rules, and the depth rule into a sequential classification procedure. We first describe the algorithm and then evaluate its performance across option and underlying characteristics and over time.

2.6.1 Combined Classification Procedure

The preceding analyses show that existing stock-based trade classification rules do not perform well in option markets. Quote rules remain more informative than tick tests but systematically misclassify an important subset of trades generated by resting customer limit orders. The trade size rule addresses this weakness directly, while the depth rule improves classification for midpoint trades that remain difficult to sign using tick tests. Taken together, these findings motivate a combined classification procedure that combines the most reliable elements of the preceding analyses into a single recommended algorithm.

The recommended combined classification procedure consists of four steps. In the first step, we classify trades using the trade size rule when trade size matches quoted depth. Trades whose size equals the quoted bid size are classified as customer buys, while trades whose size equals the quoted ask size are classified as customer sells. Based on the out-of-sample evidence, we restrict this rule to standard single-leg trades executed against displayed quotes and exclude trades executed outside the prevailing bid–ask spread. By contrast, we do not exclude large trades, as doing so provides only limited additional benefits and does not generalize consistently across exchanges.

Trades not classified in the first step are next classified using quote rules. We first apply the rule based on exchange-level quotes. If the trade remains unclassified, we apply the quote rule based on the NBBO. Although the in-sample results slightly favor the reverse ordering, broader out-of-sample evidence indicates that venue-specific quotes provide higher classification accuracy across exchanges.

Trades that remain unclassified after both quote-based steps are assigned direction using the depth rule. This rule targets midpoint trades, for which quote rules are mechanically uninformative and tick tests perform poorly in option markets. Midpoint trades are classified as customer buys when ask depth exceeds bid depth

and as customer sells when bid depth exceeds ask depth. Although the rule applies to only a small fraction of trades, it improves classification precisely for transactions that are otherwise difficult to sign.

This combined classification procedure offers three main advantages. First, it addresses the principal weakness of conventional quote rules by explicitly accounting for executions of resting customer limit orders. Second, it improves the classification of midpoint trades by exploiting information contained in quoted depth, thereby avoiding reliance on tick tests, which perform poorly in option markets due to infrequent trading. Third, it simplifies the widely used LR algorithm by relying exclusively on contemporaneous trade and quote information and eliminating the need for lagged price comparisons.

2.6.2 Cross-Sectional Performance

We next assess whether the gains from the combined classification procedure are concentrated in specific option-market segments or generalize across contract and trade characteristics. Table 2.19 reports the corresponding cross-sectional performance. Across all exchanges, the procedure achieves a classification accuracy of 72.3% on average, representing an improvement of 15.5 percentage points relative to the quote rule. Accuracy is broadly similar for call and put options, while the improvements relative to the quote rule are slightly larger for calls.

Across security types, classification accuracy is somewhat higher for equity options than for index options. Notably, however, both accuracy and improvements relative to the quote rule are particularly large for index options at the CBOE, where many major market indices trade exclusively. Differences by S&P 500 membership are modest, with slightly higher accuracy and improvements for options on non-S&P 500 stocks.

Performance also varies with option characteristics. Improvements increase almost monotonically with time to maturity and moneyness, where conventional quote rules tend to perform poorly. As a result, the combined procedure substantially reduces cross-sectional variation in classification accuracy.

Finally, improvements are larger for options with longer intervals between trades. Lower trading frequency increases the likelihood that customers submit limit orders. Overall, the results indicate that the combined classification procedure improves accuracy across a wide range of option characteristics.

Table 2.19: Cross-sectional variation of combined classification procedure

This table reports the cross-sectional variation in classification accuracy of the combined classification procedure, which combines the restricted trade size rule, the quote rule, sequentially applied to BBO and NBBO quotes, and the depth rule. Results are shown across subsamples defined by option and trade characteristics, including option and security type, S&P 500 membership, time to maturity, moneyness (i.e., underlying (strike) price relative to strike (underlying) price for call (put) options), and time from previous trade quintiles for each exchange. Results are reported separately for each exchange, as well as exchange-weighted aggregates. For ISE, the full sample period is used. We report the success rate of the combined classification procedure, along with the incremental improvement relative to the success of the quote rule without application of the new rules (in parentheses). All values are in percent.

	All	ISE	CBOE	PHLX	NOM	GEMX
Overall						
All	72.30 (15.54)	69.31 (6.74)	69.46 (6.23)	76.00 (8.62)	72.14 (29.82)	81.40 (47.20)
Option type						
Call options	72.66 (16.11)	69.74 (7.33)	69.72 (6.30)	76.93 (9.52)	72.06 (30.37)	81.65 (48.35)
Put options	71.83 (14.75)	68.76 (5.98)	69.14 (6.14)	74.78 (7.44)	72.25 (29.02)	81.00 (45.38)
Security type						
Stock options	71.69 (15.87)	68.44 (6.93)	68.56 (5.98)	75.68 (9.12)	72.02 (30.33)	81.25 (48.63)
Index options	66.07 (17.53)	65.40 (5.45)	71.94 (14.38)	65.51 (9.16)	59.85 (34.06)	62.56 (45.02)
S&P 500 membership						
S&P 500 stocks	71.19 (14.92)	67.82 (5.61)	68.06 (5.04)	74.84 (7.75)	72.34 (32.53)	80.63 (46.29)
Non-S&P 500 stocks	72.48 (17.31)	69.33 (8.83)	69.64 (8.03)	76.89 (11.07)	71.53 (26.90)	82.01 (51.47)

Table 2.19 continued

	All	ISE	CBOE	PHLX	NOM	GEMX
Time to maturity						
≤ 1 month	72.34 (15.04)	68.76 (5.68)	69.12 (5.49)	75.70 (7.89)	73.31 (30.28)	82.88 (47.83)
1-2 months	73.69 (15.67)	71.13 (6.99)	70.75 (6.54)	77.53 (8.25)	72.62 (30.45)	82.61 (46.73)
2-3 months	74.71 (16.66)	73.24 (8.59)	72.18 (7.61)	78.54 (8.98)	72.05 (30.75)	81.63 (46.90)
3-6 months	71.44 (14.91)	67.94 (5.94)	68.98 (5.86)	75.45 (8.19)	71.81 (29.23)	80.28 (46.30)
6-12 months	73.68 (17.55)	72.52 (10.20)	70.79 (8.03)	77.30 (10.50)	71.32 (29.68)	80.50 (48.09)
> 12 months	70.96 (18.72)	70.19 (11.57)	66.72 (8.41)	73.16 (12.92)	69.37 (29.51)	79.92 (50.28)
Moneyness						
DOTM ≤ 0.7	69.58 (14.20)	67.76 (7.26)	65.09 (6.76)	73.14 (7.84)	69.41 (19.02)	79.07 (45.47)
OTM (0.7-0.9]	72.50 (14.39)	70.64 (6.60)	68.73 (6.10)	77.03 (7.85)	71.96 (24.43)	80.24 (45.02)
ATM (0.9-1.1]	72.91 (15.51)	69.26 (6.21)	70.63 (5.97)	76.57 (8.28)	73.11 (32.06)	82.25 (47.14)
ITM (1.1-1.3]	70.52 (18.55)	68.40 (8.86)	66.68 (7.53)	73.02 (11.51)	70.06 (36.18)	80.71 (52.74)
DITM > 1.3	69.59 (19.02)	67.23 (9.15)	65.96 (8.35)	72.48 (13.36)	68.58 (33.65)	80.06 (54.26)
Time from previous trade (Q_1 : short, Q_5 : long)						
Q_1	69.09 (7.04)	68.20 (2.16)	71.11 (3.81)	70.72 (-0.51)	66.44 (12.66)	68.29 (26.13)
Q_2	70.89 (16.06)	66.96 (6.54)	67.07 (6.11)	74.09 (7.53)	73.68 (33.63)	82.00 (49.24)
Q_3	73.69 (16.72)	69.10 (6.71)	70.07 (5.55)	79.27 (8.31)	74.98 (35.02)	85.14 (52.58)
Q_4	74.27 (18.28)	71.29 (8.37)	70.02 (6.83)	78.49 (13.25)	73.83 (34.79)	85.71 (52.96)
Q_5	73.55 (19.38)	71.23 (10.06)	69.03 (8.71)	77.31 (13.79)	71.63 (32.23)	85.57 (54.73)

2.6.3 Time-Series Variation in Classification Performance

We now examine whether the gains from the combined classification procedure are stable over time. Figure 2.2 shows the success rate of the combined classification procedure (Panel A), its improvement relative to the quote rule applied sequentially to BBO and NBBO quotes (Panel B), and the share of trades eligible for the refined trade size rule (Panel C).

The classification accuracy of the combined classification procedure remains remarkably stable over time and across exchanges, whereas the improvements relative to the quote rule vary substantially and closely track the fraction of trades eligible for the restricted trade size rule. Periods with a larger share of eligible trades coincide with larger improvements in classification accuracy.

Two institutional events help illustrate the exchange-level mechanisms underlying this variation. At the CBOE, the share of eligible trades declines sharply in early 2015, coinciding with a change in the exchange’s rebate structure that reduced certain rebates and increased the trading volumes required to qualify for them.³ Similarly, at GEMX the fraction of eligible trades drops markedly in April 2017 following Nasdaq’s acquisition of the exchange and a subsequent platform migration that altered order allocation rules.⁴ Both events are consistent with the idea that exchange-level incentives affect the attractiveness of submitting customer limit orders. A more gradual pattern is visible for PHLX, where the share of eligible trades declines mildly after the exchange began moving selected option classes from maker–taker pricing to a PFOF fee structure in 2012.⁵ This pattern is less pronounced than the discrete shifts observed for CBOE and GEMX, but it is consistent with the view that changes in exchange-level fee incentives can affect the composition of customer order flow.

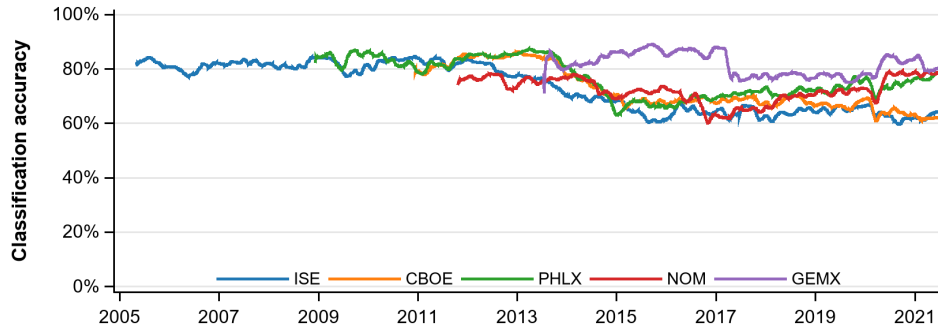
Beyond these institutional changes, the share of eligible trades increases noticeably at three exchanges after 2020. This period coincides with the well-documented rise in retail participation in option markets and is consistent with Barardehi et al. (2025), who document extensive limit order usage by retail traders after January 2020. The growing prevalence of customer limit orders further highlights the relevance of the trade size rule in modern option markets.

³See Securities Exchange Act Release No. 34-74134, File No. SR-CBOE-2015-005.

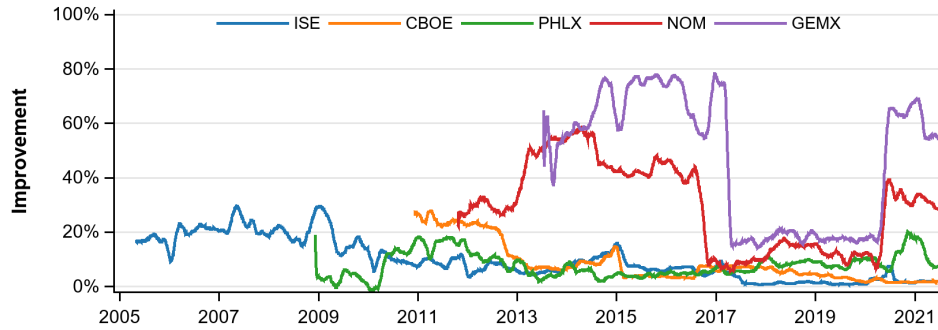
⁴See Nasdaq Options Trader Alert No. 2017-19, *Updated Symbol Migration Schedule for ISE Gemini transition to INET*, available at <http://www.nasdaqtrader.com/MicroNews.aspx?id=OTA2017-19>, and Securities Exchange Act Release No. 34-80239, File No. SR-ISEGemini-2017-14.

⁵See Securities Exchange Act Release Nos. 34-66252, 34-66488, 34-66756, 34-67439, 34-68674, and 34-69958. The corresponding operative dates range from February 1, 2012 to July 1, 2013.

Panel A: Classification accuracy of combined classification procedure



Panel B: Improvement in classification accuracy over the quote rule



Panel C: Share of classifiable trades by the restricted trade size rule

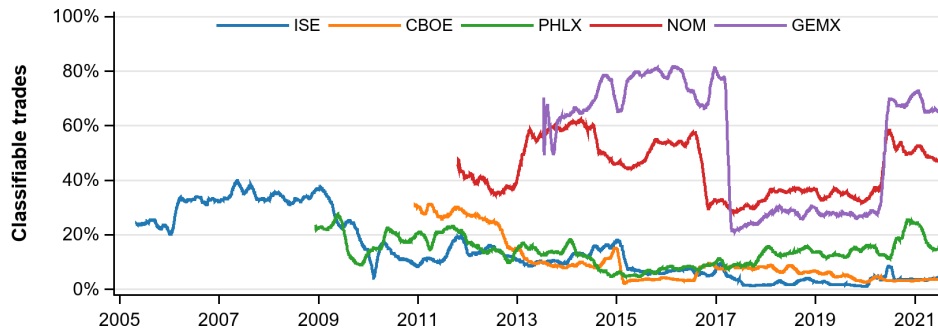


Figure 2.2: Time-series behavior of combined classification procedure

Panel A presents the classification accuracy of the combined classification procedure, which combines the restricted trade size rule, the quote rule, sequentially applied to BBO and NBBO, and the depth rule. Remaining trades are assigned a correct classification probability of 50%. Panel B shows the incremental improvements relative to the success of the quote rule without application of the new rules. Panel C depicts the share of classifiable trades by the restricted trade size rule, i.e., standard single-leg trades for which the trade size equals the quoted depth and the transaction price is not outside the prevailing BBO quotes. All series are reported as 30-day moving averages. The blue line shows data for the ISE, the orange line for the CBOE, the green line is for the PHLX, the red line for the NOM, and the purple line represents the GEMX.

Despite these variations, overall classification accuracy remains remarkably stable, indicating that the combined classification procedure not only improves average success but also stabilizes classification performance over time and across exchanges.

2.7 From Classification to Customer Demand

The preceding sections develop and validate a classification procedure that more accurately identifies the customer side of option trades. Before using this procedure to construct measures of customer option demand, two additional questions arise. First, does the identified customer side capture economically meaningful demand rather than liquidity provision? Second, is the mechanism underlying the trade size rule likely to extend beyond the sample exchanges used for validation? This section addresses both questions and thereby connects trade classification to the measurement of customer demand in option markets.

2.7.1 Economic Interpretation of Identified Customer Side

To assess whether the identified customer side reflects demand for option exposure rather than liquidity provision, we analyze short-horizon returns on customer option positions. This distinction is particularly relevant for trades captured by the trade size rule. These trades are likely to involve executions against resting customer limit orders and therefore mechanically resemble liquidity provision. If such orders represented liquidity provision comparable to market-maker intermediation, customers should earn positive short-horizon returns on average. If instead customers use limit orders primarily to obtain option exposure at more favorable prices, their returns may improve relative to other customer trades but need not become positive.

We compute returns from the customer perspective. Specifically, we assume that a customer enters a position at the transaction price and closes it at the midpoint prevailing at the next trade in the same option. If no subsequent trade occurs on the same day, we use the OptionMetrics end-of-day midpoint quote. Customer trade direction is identified using the Open/Close benchmark.

Panel A of Table 2.20 reports results for all customer trades in the matched samples. Average returns are negative across all exchanges, ranging from approximately -1.28% to -4.31%. These negative short-horizon returns indicate that customers lose money on average, consistent with market makers earning compensation for providing liquidity and intermediating customer demand.

Panel B focuses on trades whose size matches quoted depth and for which the trade size rule classification agrees with the Open/Close benchmark. These trades are the most likely candidates for executions against resting customer limit orders. Average returns remain negative across all exchanges, ranging from approximately -0.06% to -2.90%, although losses are substantially smaller than for customer trades overall. This pattern suggests that customers using limit orders obtain more favorable execution prices than the average customer trade but do not earn positive short-horizon trading profits. This evidence supports the interpretation that the customer side identified by the trade size rule reflects demand for option exposure rather than liquidity provision in the sense of market-maker intermediation.

Table 2.20: Intraday returns to customer option trades

This table provides summary statistics for returns to customer trades. Returns are constructed from a trading strategy that enters a position at the transaction price and unwinds it either at the midpoint of the next trade in the same option on the same day or, if no subsequent trade occurs, at the end-of-day midpoint quote. The return of buys (sells) is calculated as the positive (negative) difference between the midpoint at the next trade (or the end-of-day midpoint) and the buy (sell) price, divided by the buy (sell) price, where the customer buy or sell indicator is from Open/Close data. The return is winsorized at the 0.5 and 99.5 percentiles on a daily basis. Panel A provides statistics for all trades in the matched samples, while Panel B considers only trades with a trade size equal to the quote size for which our trade size rule matches the true trade side indicator. The return is given as a percentage.

	N	Mean	Std	5%	50%	95%
Panel A: All trades in matched samples						
ISE	94,235,548	-2.87	16.28	-27.27	-0.63	14.29
CBOE	73,627,477	-2.09	17.94	-27.27	-0.45	17.86
PHLX	34,810,338	-4.31	22.92	-35.71	-1.11	15.42
NOM	33,421,763	-2.04	19.05	-25.00	-0.37	16.67
GEMX	40,406,591	-1.28	17.54	-25.00	0.00	18.29
Panel B: Limit orders posted by customers						
ISE	12,144,484	-1.12	14.12	-19.25	0.00	15.06
CBOE	6,241,777	-0.06	17.11	-20.83	0.00	21.43
PHLX	4,530,523	-2.90	29.06	-28.57	-0.22	20.00
NOM	12,742,303	-1.21	15.83	-20.88	-0.05	16.67
GEMX	20,799,557	-2.35	25.95	-39.29	-0.63	30.65

2.7.2 Generalizability Beyond Sample Exchanges

The validation analysis relies on five exchanges for which Open/Close data allow us to identify the customer side of trades. To assess whether the proposed classification procedure can be applied more broadly, we examine whether these exchanges capture the main institutional features of the U.S. options market and whether the mechanism underlying the trade size rule is also present outside the sample exchanges.

Table 2.21 summarizes key design features of U.S. options exchanges not included in our matched samples. Together with Table 2.1, the evidence shows substantial institutional heterogeneity across trading venues. At the same time, the five sample exchanges span the main dimensions of this variation in market design, including differences in pricing models, allocation rules, and customer priority. This suggests that the validation sample covers a broad cross-section of relevant trading environments.

Table 2.21: Market design features of non-sample U.S. options exchanges

This table summarizes key market design features of U.S. options exchanges operating in 2021 that are not included in our sample. Variable definitions follow Table 2.1. The final column reports the share of trades for which trade size equals quoted depth, i.e., the set of transactions to which the trade size rule applies. Using LiveVol transaction data from January 2004 through June 2021, this share is computed as the number of such trades divided by the total number of trades on the respective exchange.

	Year	Share (2021)	Pricing Model	Execution Priority	Cust Prio	Auct	Multi Leg	Floor	Lead MM	Limit Order
AMEX	1974	8%	PFOF	Pro-rata	Yes	Yes	Yes	Yes	Yes	19%
ARCA	1976	11%	Maker-taker	Price-time	No	No	Yes	Yes	Yes	47%
BOX	2004	5%	Hybrid	Hybrid	Yes	Yes	Yes	Yes	No	32%
C2	2010	4%	Maker-taker	Pro-rata	No	No	Yes	No	No	20%
BZX	2010	8%	Maker-taker	Price-time	No	No	No	No	No	23%
MIAX	2012	5%	PFOF	Pro-rata	Yes	Yes	Yes	No	Yes	21%
NASDBX	2012	1%	Maker-taker	Hybrid	Yes	Yes	No	No	Yes	34%
EDGX	2015	4%	PFOF	Pro-rata	Yes	Yes	Yes	No	Yes	12%
MRX	2016	2%	Maker-taker	Pro-rata	Yes	Yes	Yes	No	Yes	17%
PEARL	2017	4%	Maker-taker	Price-time	No	No	No	No	No	56%
EMLD	2019	4%	Maker-taker	Pro-rata	Yes	Yes	Yes	No	Yes	55%

The final column reports the share of trades for which transaction size equals quoted bid or ask depth, capturing the potential scope of the trade size rule mechanism. This share is calculated from LiveVol transaction data from January 2004 through June 2021 as the number of such trades divided by the total number of trades on the exchange. These trades account for a substantial share of trading

activity across exchanges, with exchange-level shares ranging from 12% to 55%, and occur with frequencies comparable to those observed on the sample exchanges. Their prevalence is particularly high on maker-taker venues, in line with the incentives created by liquidity rebates. Overall, these patterns suggest that the trade size rule captures a trading feature that is not specific to the benchmark sample, but reflects a broader feature of trading activity in U.S. option markets.

2.8 Conclusion

This chapter develops a new procedure to classify customer trade direction in option markets. Because most option transaction datasets do not identify the customer side of trades, empirical studies typically rely on trade classification algorithms originally developed for equity markets. Using a novel benchmark constructed from matched intraday transactions and daily volume reports from five option exchanges, we show that these methods perform poorly in option markets and systematically misclassify trades arising from customer limit orders. Because conventional classification algorithms infer the aggressive side of a transaction, they frequently attribute passive customer trading activity to the counterparty.

To address this limitation, we introduce two new classification rules tailored to the institutional features of option markets: a trade size rule that identifies executions against resting customer limit orders and a depth rule that improves classification for midpoint trades. We combine these rules into a simple classification procedure that relies only on contemporaneous trade and quote information and substantially improves classification accuracy across exchanges, trading environments, and time. The gains are particularly large on maker-taker exchanges, where exchange pricing incentives encourage passive order submission. Relative to conventional quote-based methods, the procedure improves classification accuracy by roughly 16 percentage points while also delivering more stable performance across market environments.

These improvements have important implications for empirical research using signed option order flow. Many widely used measures, including option order imbalances and effective spreads, depend critically on correctly identifying the customer side of trades. Misclassification can therefore introduce substantial measurement error into empirical analyses of option trading and market quality. Consistent with this interpretation, Savickas and Wilson (2003) show that standard classification rules perform poorly when estimating effective spreads for options, while Barardehi

et al. (2025) demonstrate that conventional measures overstate effective spreads when customer limit orders are not properly accounted for. By substantially reducing these sources of error, the proposed procedure provides a more reliable framework for studying option trading and market quality.

More broadly, the chapter highlights that trade classification algorithms developed for equity markets do not necessarily transfer to option markets, where trading mechanisms, liquidity provision, and order submission incentives differ fundamentally. The findings further suggest that customer limit orders play a central role in modern option markets. Beyond improving trade classification, the mechanism underlying the trade size rule provides a simple empirical proxy for passive customer trading activity. This proxy provides a new framework for studying how passive customer trading activity shapes liquidity provision, execution costs, and information transmission in option markets.

Chapter 3

Unmasking Option Demand: Trade Classification and Stock Re- turn Predictability

3.1 Introduction

The informational content of option trading lies in its direction rather than in aggregate volume, that is, whether trades reflect buying or selling pressure (Easley et al. (1998)). Investors with positive information can buy call options or sell put options to obtain positive exposure to the underlying stock, whereas investors with negative information can buy put options or sell call options to obtain negative exposure. Accurately measuring trade direction is therefore essential for assessing whether option trades contain information about future stock prices.

A key challenge is that standard trade classification algorithms identify the trade initiator, not necessarily the customer side of the trade. This distinction matters in option markets because informed customers can express their demand through both market orders and limit orders. Theoretical models of order submission imply that traders choose between these order types by trading off immediacy against execution costs (Glosten, 1994; Foucault, 1999). When trading costs are large or information is sufficiently long-lived, informed traders may optimally submit limit orders rather than market orders (Chan et al., 2002). This mechanism is particularly relevant in option markets, where bid-ask spreads are relatively wide (Vijh, 1990). As a result, information may be revealed not only through aggressive trades but also through passive limit orders.

This order-submission choice has direct implications for empirical inference. Classification algorithms that identify the aggressive side of trades capture information only when it is revealed through market orders. If informed traders instead use limit orders, trades may be attributed to the counterparty, obscuring the true source of information and attenuating the relation between measured option order flow and future stock returns.

This chapter studies how trade classification affects the predictability of option-based signals for stock returns. We compare the conventional Lee and Ready (LR, 1991) algorithm with the combined classification procedure developed in Chapter 2, which identifies the customer side of trades more accurately by explicitly accounting for limit order activity. By holding the signal construction and empirical design fixed while varying only the classification method, we isolate how trade classification shapes empirical inference about stock return predictability.

We construct option-based signals using the put-call ratio of Pan and Poteshman (2006), which captures buying pressure in put options relative to call options and serves as our primary measure of directional option demand. To ensure that the results are not driven by this specific measure, we verify their robustness using the option order imbalance measure of Hu (2014), which captures the delta exposure embedded in signed option order flow. We evaluate predictive power using portfolio sorts following Bondarenko and Muravyev (2025) and complementary predictive regressions in the spirit of Pan and Poteshman (2006).

Our analysis yields three main findings. First, trade classification has a first-order effect on the predictability of option-based signals. Using the LR algorithm, the relation between option demand and future stock returns is weak and often insignificant, whereas the alternative classification reveals strong and statistically significant predictability. Second, in direct horse-race regressions, signals based on the alternative classification dominate conventional measures and retain predictive power for up to nine trading days, indicating that improved classification enhances both the strength and persistence of return predictability. Third, the differences are strongest in less liquid market segments, where limit order usage is more prevalent and disagreement between classification algorithms is greatest. In these settings, LR-based strategies perform poorly, whereas signals based on the alternative classification generate strong and persistent returns, providing direct evidence that customer limit orders convey economically meaningful information.

These findings contribute to the large empirical literature on whether option trading predicts future stock returns. Building on the model of Easley et al. (1998),

which implies that informed traders use option markets and that directional order flow should predict stock returns, empirical studies reach mixed conclusions. Chan et al. (2002) find no robust evidence of predictability, whereas Pan and Poteshman (2006) and Hu (2014) document significant effects using put-call ratios and option order imbalances. More recent evidence remains similarly heterogeneous. Bondarenko and Muravyev (2025) show that predictability disappears after 2009, while Bryzgalova et al. (2023) document return predictability associated with retail option trading after 2020. In other studies, return predictability is limited to specific settings, such as periods preceding takeover announcements (Cao et al., 2005) or episodes of high option illiquidity (Goyenko et al., 2015). We contribute to this literature by showing that conventional classification methods can understate the informational content of option demand and thereby materially affect conclusions about the strength and persistence of option-based stock return predictability.

This chapter also relates to the literature on order submission strategies and the informational role of limit orders (e.g., Demsetz, 1968; Biais et al., 1995; Foucault et al., 2005; Goettler et al., 2005). A central insight in this literature is that traders trade off immediacy against execution costs when choosing between market and limit orders. Kaniel and Liu (2006) show that the horizon of private information is central for order choice: longer-lived information increases the attractiveness of limit orders, whereas more urgent information raises the value of immediate execution. Our results are consistent with this broader order-choice perspective and show that the information in signed option demand depends on how that demand is expressed in the market and varies across liquidity segments.

The remainder of the chapter proceeds as follows. Section 3.2 reviews related work on informed option trading, stock return predictability, and the role of order-submission choices in determining the informational content of option trades. Section 3.3 describes the data, signal construction, and empirical design, and compares option-demand measures obtained from two alternative trade classification methods. Section 3.4 examines how these trade classification methods affect the predictability of option-based signals for future stock returns using portfolio sorts and predictive regressions. Section 3.5 investigates the role of customer limit orders and trade misclassification as the mechanism underlying the results. Section 3.6 presents robustness analyses using an alternative measure of option demand, and Section 3.7 concludes.

3.2 Informed Option Trading and Stock Return Predictability

Early theoretical work explains why options can serve as an important venue for informed trading. Black (1975) argues that options are attractive to investors with private information because they provide embedded leverage and make it easier to express negative views without directly short selling the underlying stock. Building on this idea, Easley et al. (2016) develop an asymmetric-information model in which informed investors allocate their trades across stock and option markets. Their model predicts that some informed traders choose to trade in options when option leverage and liquidity are sufficiently high. Relatedly, Mayhew et al. (1995) find that informed traders shift between stock and option markets when changes in option margin requirements alter the relative costs of trading in options.

If informed investors trade options before their information is fully incorporated into stock prices, option-market activity should contain information about subsequent stock returns. Because informed investors may trade on either positive or negative signals, the direction of option demand helps identify the information embedded in option trading. Investors with positive information can buy calls or sell puts to obtain positive exposure to the underlying stock, whereas investors with negative information can buy puts or sell calls to obtain negative exposure. Consistent with this interpretation, Pan and Poteshman (2006) show that put-call ratios constructed from option buy volume predict future stock returns, with stronger put buying relative to call buying predicting lower subsequent returns. Similarly, Ge et al. (2016) find that signed option volume contains predictive information for future stock returns. At the same time, the evidence suggests that option-based return predictability is strongest when option trading is more likely to reflect informed demand. Cao et al. (2005) find little evidence that option order imbalance predicts stock returns during normal times, but document significant predictive power before takeover announcements. Goyenko et al. (2015) show that option order imbalance predicts underlying stock returns primarily when the imbalance is more likely to reflect informed trading, as indicated by increasing option illiquidity. Bondarenko and Muravyev (2025) document significant return predictability from put-call ratios before 2009, but show that this predictability disappears abruptly afterward, coinciding with a large insider-trading enforcement campaign.

Market-maker hedging provides a mechanism through which information embedded in option order flow can be transmitted to stock prices. When market makers

take the opposite side of option trades, they may hedge the resulting exposure by trading the underlying stock, so option order flow can induce subsequent stock order flow. Hu (2014) emphasize this mechanism by showing that option-induced stock order imbalance predicts future stock returns without reversing at longer horizons, suggesting that hedging-related stock trading reflects information embedded in option demand. Dong (2025) provide further evidence that market-maker delta hedging transmits information across markets, but show that this transmission is selective. The hedging response is concentrated in single-leg and limit order book trades, whereas auction trades and multi-leg strategies generate little subsequent stock trading. This suggests that the stock-market impact of option order flow depends not only on the directional exposure created by the trade, but also on how the option order is executed.

The role of order submission is also central to Chan et al. (2002), who find no robust evidence that option net trade volume predict stock returns, but show that option-market information is incorporated into stock prices through quote revisions. They argue that informed investors may avoid demanding immediacy in the options market and instead submit passive limit orders, waiting for uninformed investors or liquidity traders to trade against them. More broadly, theoretical models of order submission imply that traders choose between market and limit orders by trading off immediacy against trading costs (Glosten, 1994; Handa and Schwartz, 1996; Foucault, 1999). This trade-off depends on both the value and the horizon of private information (Kaniel and Liu, 2006). When information is short-lived or highly valuable, informed traders have stronger incentives to demand immediacy and use marketable orders before their information is incorporated into prices (Barardehi et al., 2025). This trade-off is consistent with evidence that option trades contain information before earnings announcements (Amin and Lee, 1997; Hu, 2014), takeover announcements (Cao et al., 2005), and news events (Weinbaum et al., 2023), when private information is likely to be especially valuable and time-sensitive. In contrast, when information is longer-lived or trading costs are high, informed traders may prefer limit orders despite execution risk. These considerations are particularly relevant in option markets, where trading costs are relatively high (Vijh, 1990; Muravyev and Pearson, 2020). This motivates our empirical analysis of whether the return predictability of option demand depends on order-submission choices and on the ability to identify the customer side of limit-order trades.

3.3 Signal Construction and Empirical Design

This section describes the data, signal construction, and empirical design underlying the analysis. Our objective is to isolate the role of trade classification in measuring the information contained in option trading. To this end, we construct standard option-based signals under alternative classification procedures and evaluate their predictive performance within established empirical frameworks.

3.3.1 Data

We use LiveVol transaction data from January 2004 to June 2021 to construct measures of option demand for equity options written on common stocks. To ensure data quality, we apply standard filters and remove observations with non-positive prices or volumes, volumes exceeding 10 million contracts, duplicate records, canceled trades, and entries associated with multiple underlying symbols for the same option root.

Because customer trade direction is not directly observed in transaction-level option data, we construct signed option demand using two classification approaches. The conventional LR algorithm classifies trades executed above the prevailing NBBO midpoint as buyer-initiated, trades below the midpoint as seller-initiated, and midpoint trades using the tick test. This algorithm provides a standard benchmark for signing order flow, but it identifies the aggressive side of the trade rather than the customer side. We compare it to the combined classification procedure developed in Chapter 2, which combines the trade size rule, quote rules based on BBO and NBBO quotes, and the depth rule to more closely infer the customer side of option trades, including trades in which customer demand is expressed through resting limit orders. Comparing signals constructed under the two approaches allows us to assess whether return predictability depends on how trade direction and customer demand are measured.

We merge the classified LiveVol transactions with OptionMetrics Ivy DB to obtain option characteristics, including option deltas. For each option-day, we retain the OptionMetrics observation whose trading volume is closest to the corresponding daily LiveVol volume. Using the matched option characteristics, we aggregate classified trades to the stock-day level. We then merge the stock-day option measures with daily stock return data from the Center for Research in Security Prices (CRSP). Daily and weekly stock returns are computed using the midpoint of closing

bid and ask prices and adjusted for delistings. We also obtain shares outstanding to normalize option order imbalances and S&P 500 membership to facilitate subsample analyses. The Fama and French (1993) factors and the Carhart (1997) momentum factor are obtained from Kenneth French’s data library.

The final sample consists of approximately 10 million firm-day observations. The number of firms per day ranges from 1,531 to 3,428, with an average of 2,269, including on average 478 S&P 500 stocks and 1,792 non-S&P 500 stocks per day.

3.3.2 Option-Based Signals

We construct two stock-day measures of directional option demand. Our primary measure is the put-call ratio of Pan and Poteshman (2006), while the option order imbalance of Hu (2014) serves as a robustness check. To isolate the role of trade classification, we compute both measures separately under the LR algorithm and the combined classification procedure. We exclude observations for which trade direction cannot be inferred.

Put-call ratio. Our first signal follows Pan and Poteshman (2006) and measures the relative customer demand for put versus call options based on customer buy volume. For stock i on day t , we define the put-call ratio (PC) as

$$PC_{i,t} = \frac{PB_{i,t} + \varepsilon}{PB_{i,t} + CB_{i,t} + 2\varepsilon}, \quad (3.1)$$

where $PB_{i,t}$ and $CB_{i,t}$ denote the total number of put and call contracts classified as customer buys. The parameter $\varepsilon = 5$ contracts is added to both put and call volumes, following Bondarenko and Muravyev (2025). This adjustment avoids division-by-zero issues and shrinks the ratio toward 0.5 when option trading activity is low. By construction, the ratio is bounded between zero and one.

The economic interpretation follows from how investors express directional views in option markets. If traders with positive (negative) information buy calls (puts), PC decreases (increases). The ratio is therefore decreasing in positive information and increasing in negative information about the underlying stock. Consequently, if option trading reflects informed customer demand, higher values of PC should predict lower subsequent stock returns.

Option order imbalance. As a robustness check, we also consider the option order imbalance measure of Hu (2014), which captures directional exposure embedded in option order flow. For stock i on day t , we define option order imbalance (OOI) as

$$OOI_{i,t} = \frac{\sum_{j=1}^{n_{i,t}} 100 \text{Dir}_{i,j,t} \cdot \text{delta}_{i,j,t} \cdot \text{size}_{i,j,t}}{\text{shares_outstanding}_{i,t}}, \quad (3.2)$$

where $\text{Dir}_{i,j,t}$ is an indicator equal to one (minus one) if trade j is classified as a customer buy (sell), $\text{size}_{i,j,t}$ denotes the number of option contracts (each corresponding to 100 shares of the underlying), $n_{i,t}$ is the number of option trades, $\text{delta}_{i,j,t}$ measures the sensitivity of the option price to the underlying stock price, and $\text{shares_outstanding}_{i,t}$ denotes the number of shares outstanding. The measure is constructed using options with more than 10 days remaining to expiration.

The numerator represents the aggregate delta exposure in share equivalents, capturing the directional trading intent embedded in option order flow. Traders with positive (negative) information can take positions that generate positive (negative) delta exposure, e.g., by buying calls (puts) or selling puts (calls). As a result, OOI increases with positive information and decreases with negative information about the underlying stock. Scaling by shares outstanding mitigates mechanical imbalances in stocks with low option trading activity and facilitates comparability. If informed traders gain stock exposures in the options market, higher values of OOI should predict higher subsequent stock returns.

3.3.3 Empirical Design

We evaluate how the predictive content of option-based signals varies with the trade classification method. To assess the economic relevance of trade classification, we apply the empirical frameworks of Pan and Poteshman (2006) and Bondarenko and Muravyev (2025) and examine how return predictability changes when option demand is measured using the combined classification procedure rather than the standard LR algorithm. By holding the signal construction and empirical design fixed while varying only the classification method, we isolate the role of trade classification in shaping empirical inference and attribute differences in predictability directly to how trade direction is measured.

Portfolio sorts. We begin with portfolio sorts following Bondarenko and Muravyev (2025) to assess how trade classification affects the economic magnitude of return predictability. At the end of each trading day t , we sort stocks into decile

portfolios based on the option-based signal. We form equally weighted portfolios and compute returns over the subsequent week, corresponding to days $t + 2$ to $t + 6$. We skip one day between signal formation and return measurement to mitigate non-tradable microstructure effects. Portfolio returns are converted into alphas using the Carhart (1997) four-factor model. The portfolios are rebalanced daily by replacing one-fifth of the long-short portfolio and holding the newly added stocks for one week.

To assess predictability, we focus on the return spread between the highest and lowest signal deciles. For the put–call ratio, higher values correspond to relatively stronger demand for puts and are therefore expected to predict lower subsequent returns. For option order imbalance, higher values indicate positive directional demand and are expected to predict higher returns. Accordingly, we construct long–short portfolios that go long stocks with low PC (high OOI) and short stocks with high PC (low OOI).

We conduct the analysis separately for S&P 500 and non-S&P 500 stocks to account for differences in liquidity. To capture potential structural changes over time, we distinguish between the period before and after October 2009, associated with intensified insider-trading enforcement (Bondarenko and Muravyev, 2025), and the post-2020 period, which coincides with pronounced changes in market microstructure and increased retail participation (Bryzgalova et al., 2023).

Predictive regressions. We complement the portfolio analysis with predictive regressions following Pan and Poteshman (2006) to assess how trade classification affects the statistical significance and persistence of return predictability. Specifically, we estimate Fama and MacBeth (1973) regressions of the form

$$R_{i,t+h} = \alpha + \beta X_{i,t} + \varepsilon_{i,t+h}, \quad h = 1, \dots, 20, \quad (3.3)$$

where $R_{i,t+h}$ denotes the h -days-ahead risk-adjusted return of stock i , and $X_{i,t}$ is the option-based signal constructed from trading on day t .

Consistent with the literature, risk-adjusted returns are computed as residuals from the Carhart (1997) four-factor model, thereby isolating the idiosyncratic return component. The regression framework allows us to quantify how the statistical significance and persistence of return predictability depend on the trade classification method. Under the null hypothesis of no predictability, the coefficient β equals zero for all horizons. If option trading reflects informed customer demand, we expect β to be negative for PC and positive for OOI.

To directly compare the informational content of the alternative classification methods, we estimate bivariate Fama and MacBeth (1973) regressions of the form

$$R_{i,t+h} = \alpha + \beta X_{i,t}^{\text{LR}} + \gamma X_{i,t}^{\text{Combined}} + \varepsilon_{i,t+h}, \quad h = 1, \dots, 20, \quad (3.4)$$

where $X_{i,t}^{\text{LR}}$ and $X_{i,t}^{\text{Combined}}$ denote the option-based signals constructed using the LR algorithm and the combined classification procedure, respectively.

This specification allows us to assess whether improved trade classification enhances the informational content of option-based signals. If the combined classification more accurately captures customer demand, its associated coefficient should remain significant, while the coefficient on the LR-based signal is attenuated or becomes insignificant. Moreover, as the prediction horizon increases, we expect the predictive content of $X_{i,t}^{\text{Combined}}$ to persist over longer horizons than that of $X_{i,t}^{\text{LR}}$.

We estimate all regressions separately for S&P 500 and non-S&P 500 stocks to account for differences in liquidity across market segments, which may affect both trading behavior and the informativeness of option order flow.

3.3.4 Descriptive Statistics

Before turning to the predictive analyses, we summarize the properties of the option-based signals and examine how they vary across trade classification procedures.

Table 3.1 reports averages of daily cross-sectional summary statistics for the option-based signals and stock returns separately for S&P 500 firms (Panel A) and non-S&P 500 firms (Panel B). The put-call ratios exhibit very similar distributions across trade classification methods, with average values between 0.40 and 0.42. Consistent with Bondarenko and Muravyev (2025), the average ratios are slightly below 0.5, reflecting that call options on individual stocks are typically traded more actively than puts. In contrast, option order imbalance is centered around zero on average, consistent with Hu (2014), indicating that aggregate option trading is broadly balanced between buying and selling pressure. At the same time, option order imbalance exhibits substantially greater cross-sectional dispersion than the put-call ratio, particularly among non-S&P 500 firms. The summary statistics further highlight the greater heterogeneity of non-S&P 500 firms, which are substantially smaller and exhibit higher return volatility than S&P 500 firms.

Table 3.2 reports correlation patterns for the option-based signals constructed under the LR algorithm and the combined classification procedure. Correlations

Table 3.1: Summary statistics for option-based signals and stock returns

This table reports time-series averages of daily cross-sectional summary statistics for option-based signals and stock returns. We consider two versions of the put-call ratio (PC) and option order imbalance (OOI): one constructed using the standard LR algorithm and one constructed using the combined classification procedure. Market capitalization is reported in billions of U.S. dollars. Ret_{t+1} denotes next-day stock returns, and $\text{Ret}_{t+2:t+6}$ denotes cumulative stock returns from trading days $t + 2$ through $t + 6$. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks.

Panel A: Sample of S&P 500 stocks					
Variable	Mean	Std.	10%	50%	90%
PC (standard LR)	0.42	0.23	0.12	0.41	0.75
PC (combined procedure)	0.42	0.23	0.13	0.41	0.74
OOI (standard LR)	0.00	0.47	-0.14	0.00	0.14
OOI (combined procedure)	0.00	0.45	-0.12	0.00	0.12
Market Cap	33.26	59.98	5.38	15.23	72.48
Ret_{t+1}	-0.01	1.59	-1.58	-0.03	1.57
$\text{Ret}_{t+2:t+6}$	0.25	3.69	-3.63	0.21	4.17

Panel B: Sample of non-S&P 500 stocks					
Variable	Mean	Std.	10%	50%	90%
PC (standard LR)	0.41	0.25	0.08	0.41	0.77
PC (combined procedure)	0.40	0.24	0.08	0.40	0.75
OOI (standard LR)	-0.01	0.98	-0.15	0.00	0.13
OOI (combined procedure)	0.00	0.93	-0.12	0.00	0.11
Market Cap	2.61	5.66	0.20	1.18	5.37
Ret_{t+1}	-0.01	3.08	-2.67	-0.07	2.62
$\text{Ret}_{t+2:t+6}$	0.25	6.94	-6.18	0.07	6.67

between the put-call ratios are relatively high, with Pearson correlations of 0.79 for S&P 500 stocks and 0.75 for non-S&P 500 stocks, while the corresponding rank correlations are similarly large. The close relation between the two put-call ratio measures reflects the construction of the signal, which captures relative put versus call buying pressure, such that classification errors affecting put and call trades similarly tend to offset each other.

Correlations are weaker for option order imbalance, particularly in rank correlations, which decline to 0.37 for S&P 500 stocks and 0.27 for non-S&P 500 stocks, while Pearson correlations remain around 0.60. Because option order imbalance is based on signed net directional exposure, misclassified trades enter with the wrong sign and therefore affect the measure in the opposite direction. As expected, put-call ratios and option order imbalances are negatively correlated, reflecting that stronger

relative demand for puts is associated with more negative directional exposure.

Finally, the signals exhibit only modest persistence, with put-call ratios being more persistent than option order imbalances. Autocorrelations are nearly identical across classification methods, indicating that subsequent differences in return predictability are unlikely to be driven mechanically by differences in signal persistence.

Table 3.2: Signal correlations across classification methods

This table reports correlations and autocorrelations for option-based signals constructed using the standard LR algorithm and the combined classification procedure. The first panel reports time-series averages of daily cross-sectional Pearson and Spearman correlations between signals constructed under alternative classification methods. The second panel reports the corresponding correlations between put-call ratios and option order imbalances constructed under the same classification method. The final panel reports average first-order autocorrelations estimated separately for each stock and averaged across firms. Results are reported separately for S&P 500 and non-S&P 500 stocks.

	S&P 500 stocks		Non-S&P 500 stocks	
	Pearson	Spearman	Pearson	Spearman
<i>Correlation between signals under LR and combined classification procedure</i>				
Put-call ratio (PC)	0.79	0.78	0.75	0.74
Option order imbalance (OOI)	0.60	0.37	0.63	0.27
<i>Correlation between PC and OOI under the same classification method</i>				
Standard LR	-0.18	-0.38	-0.14	-0.41
Combined procedure	-0.13	-0.28	-0.10	-0.31
<i>Average first-order autocorrelations</i>				
PC (standard LR)	0.17	—	0.17	—
PC (combined procedure)	0.18	—	0.18	—
OOI (standard LR)	0.03	—	0.03	—
OOI (combined procedure)	0.03	—	0.03	—

3.4 The Impact of Trade Classification on Stock Return Predictability

This section examines whether the relation between option-based demand signals and future stock returns depends on how option trade direction is classified. We compare signals constructed using the conventional LR algorithm with signals based on the combined classification procedure, while keeping the empirical design fixed. We first present evidence from portfolio sorts, then examine how predictability evolves

over time, and finally use predictive regressions to analyze the persistence and incremental informational content of the signals.

3.4.1 Portfolio Sort Evidence

We begin by examining whether trade classification affects the ability of put-call ratios to predict future stock returns. Following Bondarenko and Muravyev (2025), we implement weekly portfolio sorts based on put-call ratios constructed under each classification method. Table 3.3 reports long-short returns, four-factor alphas, and Sharpe ratios for the resulting strategies.

The results reveal a clear contrast across classification methods. When trade direction is inferred using the LR algorithm, the put-call ratio strategy generates weak and often insignificant long-short alphas over the subsequent week. In contrast, the identical signal constructed using the combined classification delivers economically large and statistically significant alphas. Because the empirical design is held fixed, these differences can be attributed solely to how trade direction is measured.

This pattern holds across both liquid and less liquid stocks, although its magnitude varies systematically with market liquidity. For S&P 500 stocks (Panel A), the LR-based strategy exhibits little evidence of predictability, whereas the combined classification yields a statistically significant alpha of 5.51 basis points and increases the Sharpe ratio from 0.13 to 0.59. The contrast is substantially stronger among non-S&P 500 stocks (Panel B). While the LR-based strategy yields an alpha of 6.41 basis points, the combined classification increases the alpha to 15.84 basis points and more than doubles the Sharpe ratio. In both samples, the economic magnitude of the differences is substantial.

Overall, the evidence indicates that the predictive power of put-call ratios depends critically on how trade direction is inferred from option transactions. Conventional classification methods substantially weaken the measured relation between option demand and future stock returns, whereas a classification procedure that more accurately identifies the customer side of trades reveals large return predictability. The stronger effects among less liquid stocks further suggest that trade misclassification is particularly important in market segments with wider bid-ask spreads.

Table 3.3: Return predictability by classification method

This table reports the performance of weekly long-short strategies based on put-call ratios constructed using the standard LR algorithm and the combined classification procedure. Strategies buy stocks in the lowest decile and sell stocks in the highest decile of the put-call ratio distribution. Portfolios are formed one trading day after signal construction and held for one week. Performance is reported as average weekly excess returns in basis points, four-factor alphas from the Carhart (1997) model (FF4), corresponding Newey–West t -statistics (Newey and West, 1987), and annualized Sharpe ratios. The “difference” columns report the performance difference between the combined classification and LR-based strategies. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks. The sample is further divided into three subperiods associated with intensified insider-trading enforcement and changes in retail participation in option markets: Pre (Jan 1, 2004–October 15, 2009), Mid (October 16, 2009–December 31, 2019), and Post (January 1, 2020–June 30, 2021).

Panel A: Sample of S&P 500 stocks

	standard LR				combined procedure				difference			
	Full	Pre	Mid	Post	Full	Pre	Mid	Post	Full	Pre	Mid	Post
Return	1.25	14.27	-5.05	-5.89	5.34	17.50	-1.70	6.49	4.08	3.23	3.35	12.38
Alpha	1.95	15.07	-5.10	0.37	5.51	18.18	-2.39	8.51	3.56	3.11	2.71	8.14
t-stat	0.85	3.05	-2.94	0.04	2.48	3.56	-1.46	1.26	3.00	1.40	2.13	1.40
Sharpe	0.13	1.17	-0.86	-0.40	0.59	1.38	-0.31	0.59	0.77	0.54	0.83	1.43

Panel B: Sample of non-S&P 500 stocks

	standard LR				combined procedure				difference			
	Full	Pre	Mid	Post	Full	Pre	Mid	Post	Full	Pre	Mid	Post
Return	8.36	10.13	2.88	38.90	19.05	17.62	12.34	70.47	10.70	7.49	9.45	31.56
Alpha	6.41	10.46	0.63	25.36	15.84	17.58	9.05	51.15	9.44	7.13	8.42	25.79
t-stat	2.15	1.90	0.23	1.49	4.71	2.92	3.08	2.87	6.40	2.73	5.57	2.50
Sharpe	0.61	0.72	0.30	1.40	1.28	1.12	1.14	2.52	1.59	1.07	1.89	2.50

3.4.2 Time Variation in Predictability

We next examine whether the differences in predictability across classification methods vary systematically over time. To do so, we split the sample into three subperiods defined by the intensification of insider-trading enforcement after October 2009 (Bondarenko and Muravyev, 2025) and the changes in market structure and retail participation following January 2020 (Bryzgalova et al., 2023). This analysis allows us to connect our findings to the mixed evidence in the literature and assess whether improved trade classification helps reconcile the differing conclusions.

Across all subperiods reported in Table 3.3, strategies based on the combined classification procedure consistently outperform those based on the LR algorithm in

terms of long-short returns, four-factor alphas, and Sharpe ratios. The magnitude of predictability also varies systematically over time in ways that closely align with changes in market structure and trading behavior.

For S&P 500 stocks, both classification methods exhibit economically meaningful predictability prior to October 2009, consistent with earlier evidence in the literature, although strategies based on the combined classification deliver stronger performance throughout. After 2009, predictability weakens under both approaches, in line with the insider-trading enforcement documented by Bondarenko and Muravyev (2025). However, the deterioration is substantially more pronounced for the LR-based strategy, whose alphas reverse sign and become negative, whereas strategies based on the combined classification lose statistical significance. Following 2020, predictability under the combined classification strengthens again and turns positive, while the LR-based strategy continues to underperform.

The contrast is substantially stronger among non-S&P 500 stocks. In this less liquid segment, the LR-based strategy loses statistical significance after 2009, whereas strategies based on the combined classification continue to generate positive and statistically significant alphas throughout all subperiods. The divergence becomes particularly pronounced after 2020, when strategies based on the combined classification produce very large alphas and high Sharpe ratios, while LR-based strategies fail to capture these gains.

Figure 3.1 complements these results by plotting cumulative returns of the long-short portfolios. Strategies based on the combined classification consistently outperform those based on the LR algorithm, with the performance gap widening over time. For S&P 500 stocks (Panel A), both strategies perform similarly early in the sample but diverge after 2009, as LR-based strategies gradually erode their gains while strategies based on the combined classification remain comparatively stable and recover after 2020. For non-S&P 500 stocks (Panel B), the divergence is considerably stronger. Strategies based on the combined classification exhibit a persistent upward trend and continue to accumulate returns after 2009, whereas LR-based strategies largely stagnate.

Collectively, these findings suggest that the documented weakening of option-based return predictability after 2009 is partly driven by how option demand is measured. Under conventional methods, predictability appears to decline substantially or disappear. In contrast, when trades are more accurately attributed to the customer side, option demand continues to exhibit economically meaningful predictive content, particularly among less liquid stocks and during more recent periods.

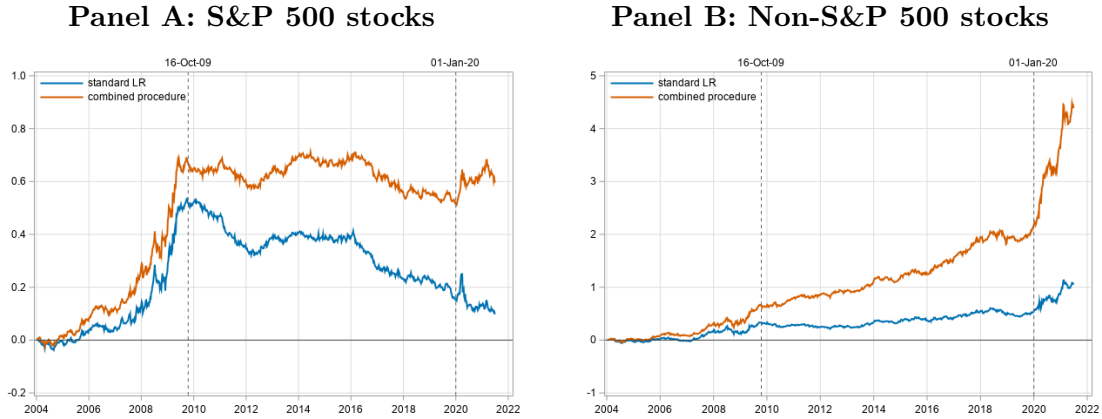


Figure 3.1: Cumulative returns by classification method

The figure plots cumulative returns of long-short strategies based on put-call ratios constructed using the LR algorithm (blue) and the combined classification procedure (orange). Strategies buy stocks in the lowest decile and sell stocks in the highest decile of the put-call ratio distribution. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks. The vertical lines mark October 16, 2009, associated with intensified insider-trading enforcement, and January 1, 2020, associated with changes in retail participation in option markets.

3.4.3 Persistence and Incremental Information

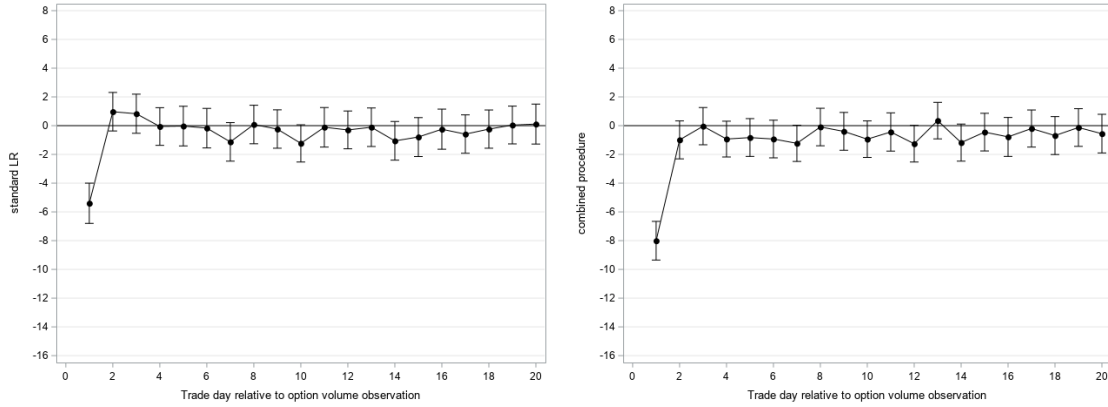
We next examine return predictability using predictive regressions. This allows us to assess the persistence and incremental informational content of option demand over horizons of up to 20 trading days.

The regression results in Figure 3.2 corroborate the portfolio-based evidence and show a clear contrast across classification methods. When put-call ratios are constructed using the LR algorithm, estimated coefficients are small, often statistically insignificant, and frequently reverse sign. In contrast, signals based on the combined classification yield coefficients with the expected sign that remain statistically significant over multiple trading days, indicating that more accurate trade classification substantially strengthens the measured persistence of return predictability.

For S&P 500 stocks (Panel A), univariate regressions show that both signals predict next-day returns, although coefficients based on the combined classification are consistently larger in magnitude. However, once both signals are included jointly, the LR-based measure loses statistical significance and frequently changes sign, whereas the signal based on the combined classification remains significant for horizons of up to six trading days. This pattern suggests that the predictive information captured by the LR-based signal is largely subsumed once option demand is measured more accurately.

Panel A: Sample of S&P 500 stocks

Univariate regressions using standard LR or combined classification procedure



Bivariate regression using standard LR and combined classification procedure

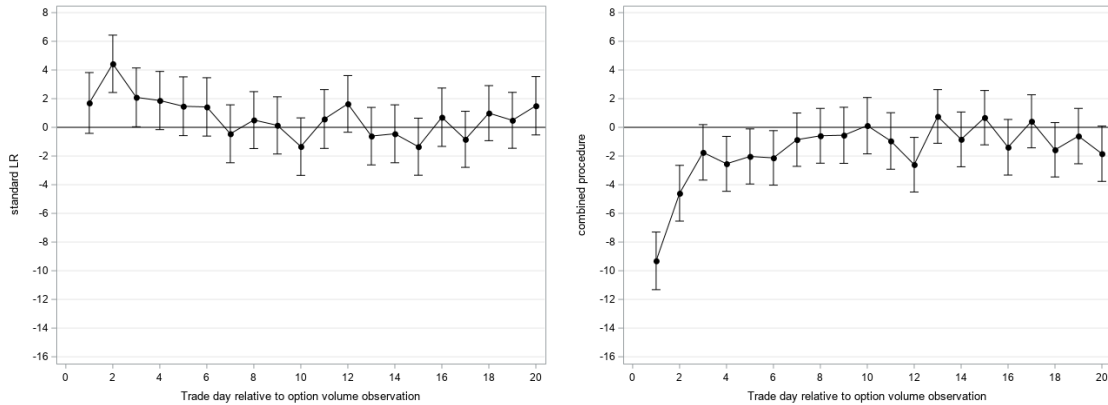
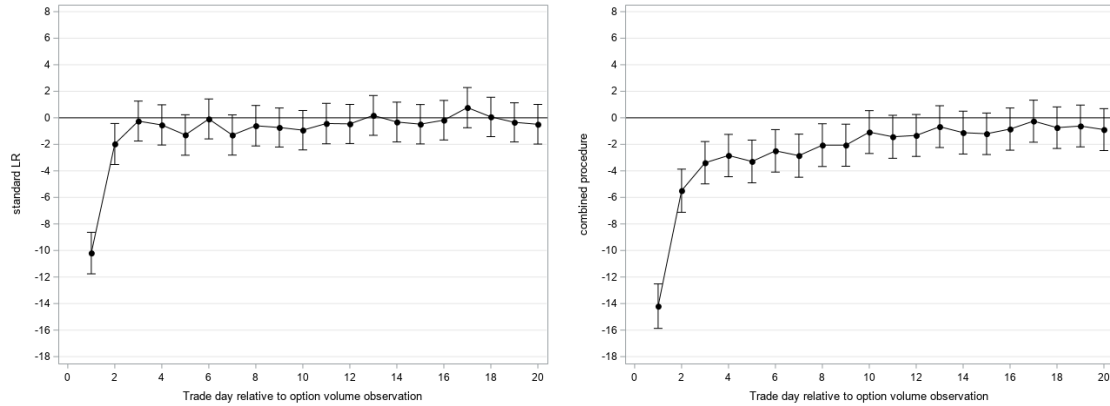
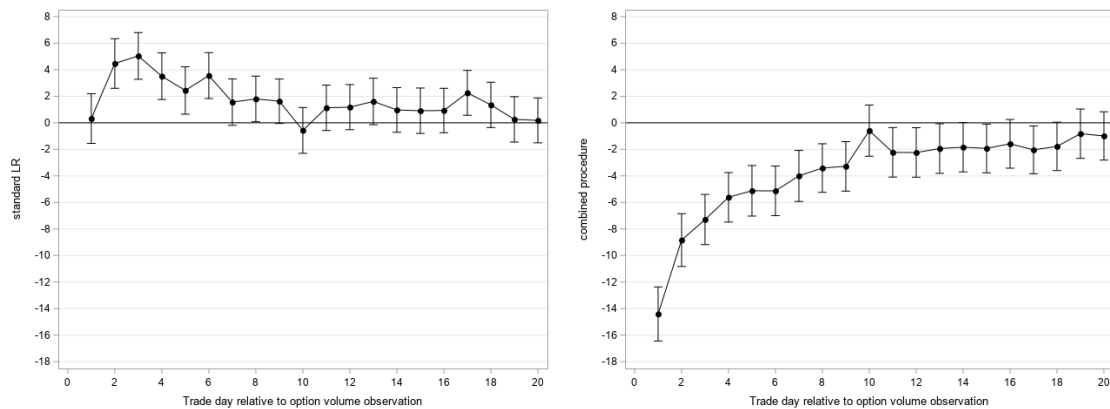


Figure 3.2: Predictive regressions by classification method

The figure reports predictive regressions of risk-adjusted stock returns on put-call ratios constructed using the standard LR algorithm (left) and the combined classification procedure (right) over return horizons of up to 20 trading days. The first row reports slope coefficients and 95% confidence intervals from univariate Fama–MacBeth regressions of future risk-adjusted stock returns on put-call ratios constructed using each classification method separately. The second row reports coefficients from bivariate Fama–MacBeth regressions that include both put-call ratio measures jointly. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks.

Figure 3.2 continued

Panel B: Sample of non-S&P 500 stocks*Univariate regressions using standard LR or combined classification procedure**Bivariate regression using standard LR and combined classification procedure*

The contrast is even stronger among non-S&P 500 stocks (Panel B). In univariate regressions, put-call ratios based on the combined classification predict returns over horizons of up to nine days, whereas the LR-based signal is informative only at short horizons. Once both signals are included jointly, the coefficient on the LR-based signal reverses sign, while the signal based on the combined procedure continues to predict returns with economically meaningful magnitude over extended horizons.

Overall, the regression evidence indicates that improved trade classification increases not only the magnitude but also the persistence of option-based return predictability. Signals constructed using the combined classification procedure contain information beyond that captured by conventional classification methods, whereas the predictive content of LR-based signals largely disappears once option demand is measured more accurately. The greater persistence of signals constructed under the combined classification further suggests that it more effectively captures the

information component of option trades associated with longer-lived information.

3.5 Limit Orders and Classification Disagreement

The previous section establishes that improved trade classification materially increases option-based return predictability. We now examine the mechanism underlying these results. Specifically, we test whether the predictive advantage of the combined classification is concentrated in settings characterized by greater disagreement between classification methods and a higher prevalence of customer limit order activity.

3.5.1 Predictability and Classification Disagreement

We first examine whether disagreement between classification methods reflects economically meaningful mismeasurement of option demand. To do so, we analyze how the predictive performance of put-call ratios varies with the degree of disagreement between signals constructed using the LR algorithm and the combined classification procedure. Specifically, we sort stocks each day into terciles based on the absolute difference between the two signals and compare return predictability across the lowest and highest disagreement terciles. If the combined classification captures informative customer demand that is mismeasured by the LR algorithm, its advantage over the LR-based signal should be largest in the high-disagreement tercile and smallest when the two signals are closely aligned.

The portfolio results in Table 3.4 strongly support this interpretation. When disagreement is low, differences between the two approaches are modest, and both signals exhibit similar qualitative patterns. In contrast, the performance differential becomes pronounced in the highest disagreement tercile. In this group, strategies based on the LR algorithm exhibit weak or negative performance, whereas strategies based on the combined classification generate economically large and statistically significant alphas. The difference between the high- and low-disagreement terciles is substantial for both S&P 500 and non-S&P 500 stocks, although the effects are considerably stronger among the less liquid non-S&P 500 firms. Figure 3.3 shows that the divergence between classification methods is concentrated in the high-disagreement tercile, where strategies based on the combined classification continue to accumulate returns while LR-based strategies largely stagnate or even decline.

Regression evidence in Table 3.4 further reinforces these findings. In bivariate regressions estimated within the high-disagreement tercile, signals based on the combined classification continue to predict returns with economically meaningful magnitude, particularly among non-S&P 500 stocks and over horizons of up to nine trading days. In contrast, the coefficient on the LR-based signal turns positive, consistent with trade misclassification distorting the measured relation between option demand and future stock returns.

Overall, the evidence indicates that disagreement between classification methods reflects economically meaningful differences in the measurement of option demand. The superior predictive performance of the combined classification becomes concentrated precisely in settings where disagreement between the signals is greatest, consistent with conventional classification methods misclassifying trades associated with customer limit order activity.

Table 3.4: Return predictability by classification disagreement

This table reports the performance of weekly long-short strategies based on put-call ratios constructed using the standard LR algorithm and the combined classification procedure (CC). Strategies buy stocks in the lowest decile and sell stocks in the highest decile of the put-call ratio distribution. Stocks are sorted each day into terciles based on the absolute difference between the two signals, with strategy performance reported separately for the lowest- and highest-disagreement terciles. The “Dif” columns report the performance difference between CC- and LR-based strategies within each tercile, while the “DD” column reports the difference-in-differences between the high- and low-disagreement terciles. The table reports average weekly excess returns in basis points, four-factor alphas from the Carhart (1997) model (FF4), corresponding Newey–West t -statistics (Newey and West, 1987), and annualized Sharpe ratios. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks.

Panel A: Sample of S&P 500 stocks							
	low disagreement			high disagreement			difference
	LR	CC	Dif	LR	CC	Dif	DD
Return	6.03	7.04	1.01	-7.33	3.67	11.00	10.00
Alpha	6.30	7.12	0.82	-5.78	3.47	9.25	8.42
t-stat	2.07	2.18	0.81	-2.31	1.58	3.01	2.83
Sharpe	0.48	0.53	0.24	-0.67	0.38	0.80	0.73

Panel B: Sample of non-S&P 500 stocks							
	low disagreement			high disagreement			difference
	LR	CC	Dif	LR	CC	Dif	DD
Return	21.36	22.94	1.59	-8.38	16.33	24.71	23.12
Alpha	18.35	19.66	1.31	-8.19	12.84	21.03	19.72
t-stat	4.45	4.67	1.50	-3.36	3.83	5.30	4.94
Sharpe	1.19	1.25	0.41	-0.67	1.11	1.33	1.26

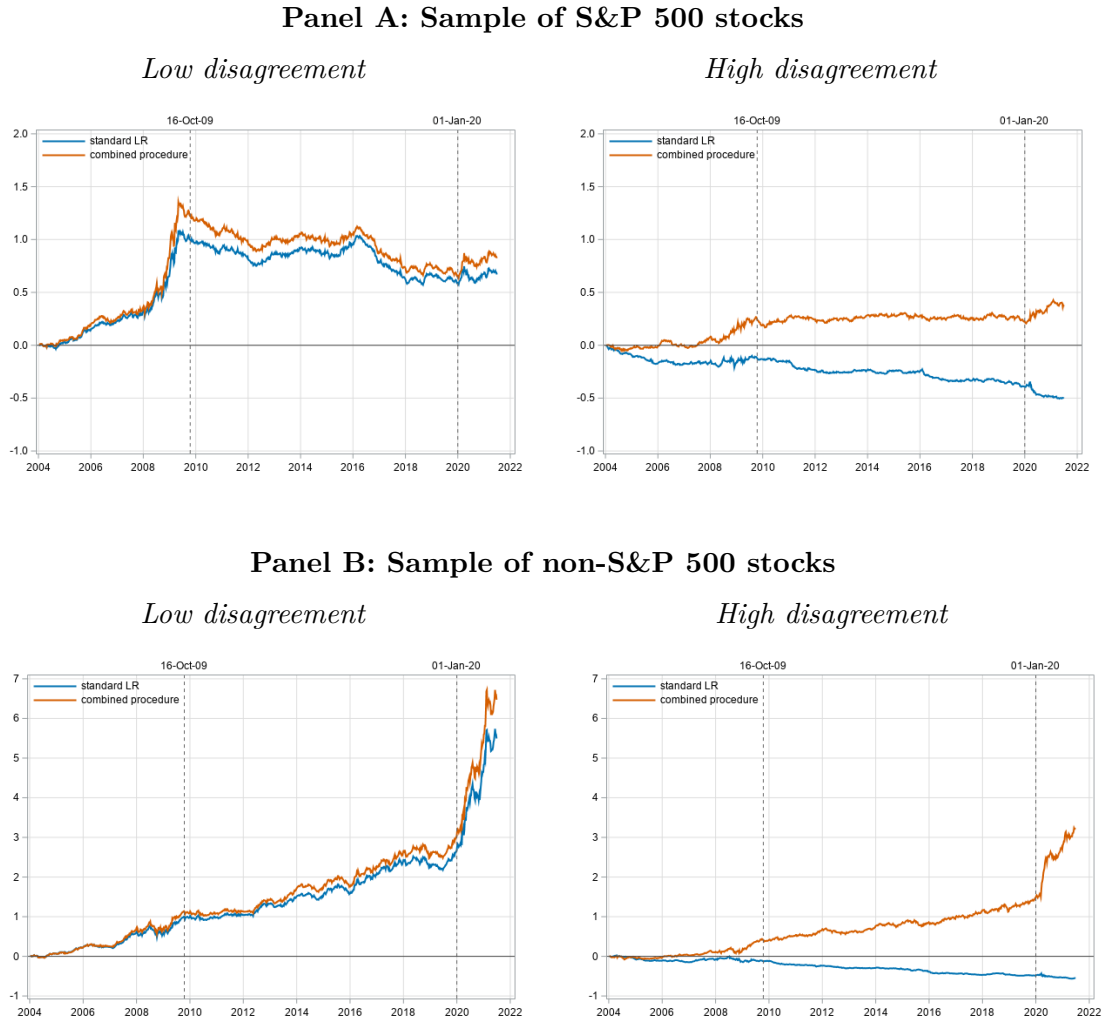
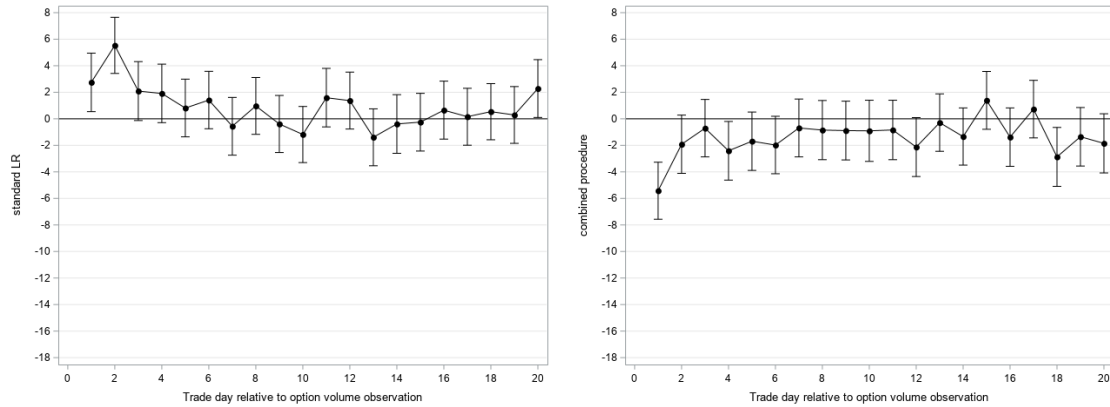


Figure 3.3: Cumulative returns by classification disagreement

The figure plots cumulative returns of long-short strategies based on put-call ratios constructed using the LR algorithm (blue) and the combined classification procedure (orange). Strategies buy stocks in the lowest decile and sell stocks in the highest decile of the put-call ratio distribution. Stocks are sorted each day into terciles based on the absolute difference between the two signals, with strategy performance reported separately for the lowest- and highest-disagreement terciles. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks. The vertical lines mark October 16, 2009, and January 1, 2020.

Panel A: Sample of S&P 500 stocks

Bivariate regression when LR and combined classification disagree



Panel B: Sample of non-S&P 500 stocks

Bivariate regression when LR and combined classification disagree

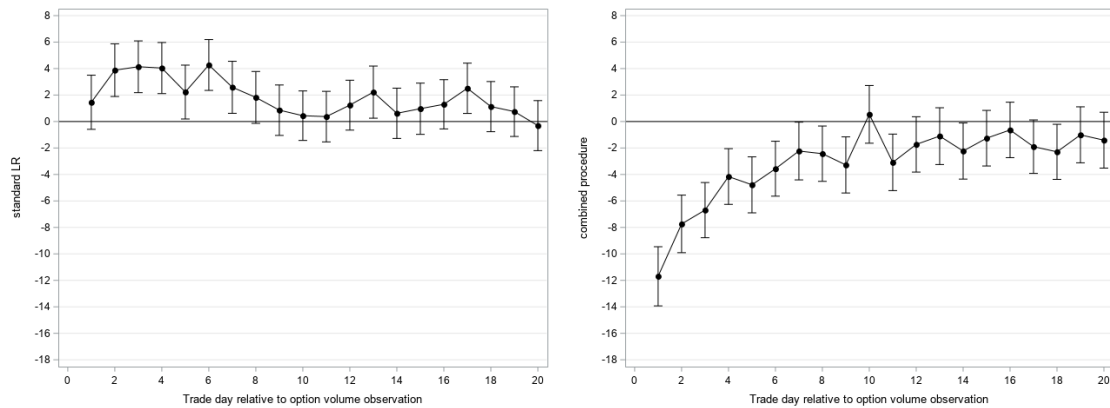


Figure 3.4: Bivariate regressions in high disagreement subsample

The figure reports slope coefficients and 95% confidence intervals from bivariate Fama–MacBeth regressions of risk-adjusted stock returns on put-call ratios constructed using the standard LR algorithm and the combined classification procedure over return horizons of up to 20 trading days. The regressions are estimated within the highest tercile of disagreement between the two put-call ratio measures, where stocks are sorted each day based on the absolute difference between the signals. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks.

3.5.2 Predictability and Customer Limit Order Activity

To examine whether customer limit orders drive the differences in return predictability across classification methods, we condition the analysis on their prevalence. Specifically, we proxy customer limit order activity by the daily volume share of trades for which trade size equals quoted depth, as identified under the trade size rule. We sort stocks each day into terciles based on this measure and compare predictive performance across the lowest and highest terciles. If the stronger predictive performance of the combined classification is driven by trades associated with customer limit-order activity, the difference between classification methods should be largest when such trades are most prevalent.

Figure 3.5 plots the time series of customer limit order activity for S&P 500 and non-S&P 500 stocks. The figure reports both the cross-sectional mean and the range between the lowest and highest terciles, thereby illustrating the heterogeneity that underlies the tercile analysis. Customer limit order activity remains economically important throughout the sample period, with average shares typically ranging between 20% and 40% of trading volume and exhibiting elevated levels both prior to 2009 and after 2020. Although average levels are broadly similar across the two samples, non-S&P 500 stocks exhibit substantially greater cross-sectional dispersion in customer limit order activity. This greater heterogeneity suggests that classification differences should be most pronounced within the non-S&P 500 tercile sorts.

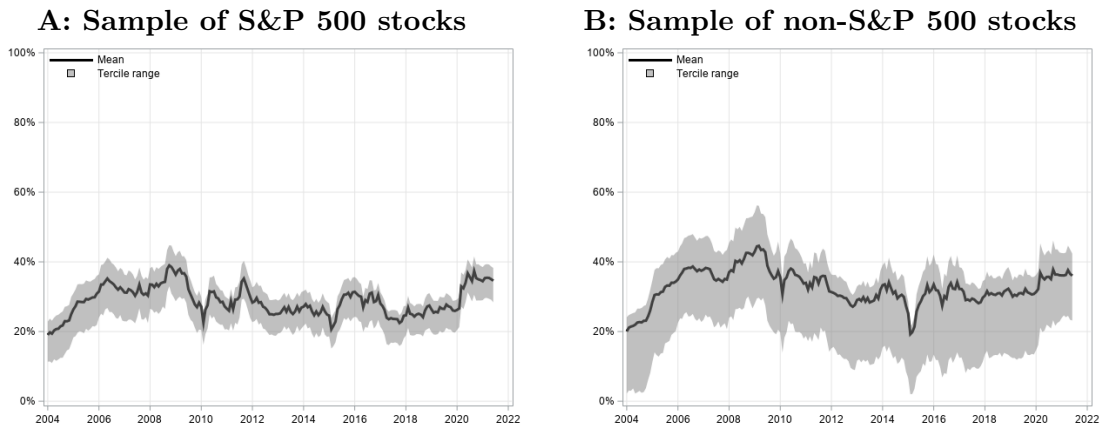


Figure 3.5: Time-series and cross-sectional variation in customer limit order activity

The figure plots customer limit order activity over time, measured as the daily volume share of trades identified under the trade size rule, where trade size equals quoted depth. The solid line reports the monthly average of the daily cross-sectional mean across stocks, while the shaded area captures the range between the lowest and highest terciles of the daily cross-sectional distribution. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks.

The portfolio results in Table 3.5 support both predictions. First, differences in return predictability across classification methods are substantially larger among stocks with a high share of limit-order-driven trades than among stocks with a low share. In the high-tercile portfolios, LR-based signals exhibit weak and unstable performance, whereas put-call ratios constructed under the combined classification procedure continue to generate economically meaningful returns. Second, consistent with the greater cross-sectional dispersion in customer limit-order activity, these differences are particularly pronounced among non-S&P 500 stocks. This pattern links the predictability results directly to the proposed mechanism and suggests that the informational content of option demand is most likely to be obscured in environments where customer limit orders play the largest role.

Table 3.5: Return predictability by customer limit order share

This table reports the performance of weekly long-short strategies based on put-call ratios constructed using the standard LR algorithm and the combined classification procedure (CC). Strategies buy stocks in the lowest decile and sell stocks in the highest decile of the put-call ratio distribution. Stocks are sorted each day into terciles based on the daily volume share of trades for which trade size equals quoted depth, with strategy performance reported separately for the lowest- and highest-disagreement terciles. The “Dif” columns report the performance difference between CC- and LR-based strategies within each tercile, while the “DD” column reports the difference-in-differences between the high- and low-disagreement terciles. The table reports average weekly excess returns in basis points, four-factor alphas from the Carhart (1997) model (FF4), corresponding Newey–West t -statistics (Newey and West, 1987), and annualized Sharpe ratios. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks.

Panel A: Sample of S&P 500 stocks

	low limit order share			high limit order share			difference
	LR	CC	Dif	LR	CC	Dif	DD
Return	1.54	5.13	3.59	-2.07	5.98	8.05	4.46
Alpha	1.70	5.39	3.68	-0.72	6.22	6.94	3.25
t-stat	0.60	1.86	2.29	-0.25	2.20	3.11	1.18
Sharpe	0.12	0.42	0.51	-0.18	0.52	0.83	0.39

Panel B: Sample of non-S&P 500 stocks

	low limit order share			high limit order share			difference
	LR	CC	Dif	LR	CC	Dif	DD
Return	16.49	20.55	4.06	-2.08	16.81	18.88	14.82
Alpha	14.73	18.08	3.35	-3.86	13.28	17.14	13.79
t-stat	4.31	4.96	2.44	-1.22	3.76	6.24	4.85
Sharpe	1.11	1.35	0.64	-0.14	1.04	1.63	1.29

Figure 3.6 complements the portfolio evidence by showing how these differences evolve over time. The divergence between classification methods is concentrated in stocks with a high share of limit-order-driven trades and widens persistently over the sample period. In these portfolios, LR-based strategies gradually deteriorate and often reverse sign, whereas strategies based on the combined classification continue to accumulate gains. This pattern indicates that the deterioration of LR-based predictability is closely tied to settings in which customer demand is more likely to

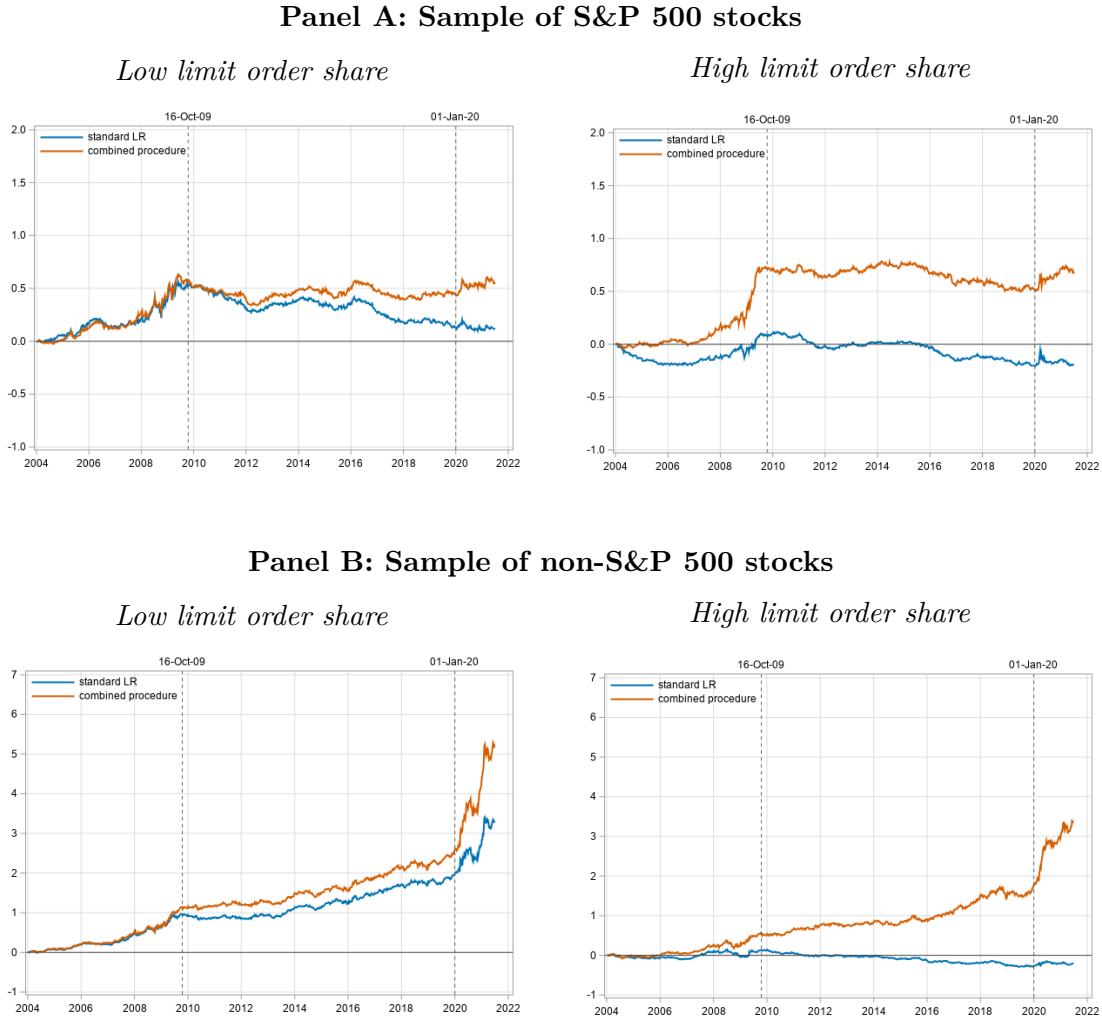


Figure 3.6: Cumulative returns by customer limit order share

The figure plots cumulative returns of long-short strategies based on put-call ratios constructed using the LR algorithm (blue) and the combined classification procedure (orange). Strategies buy stocks in the lowest decile and sell stocks in the highest decile of the put-call ratio distribution. Stocks are sorted each day into terciles based on the daily volume share of trades for which trade size equals quoted depth, with strategy performance reported separately for the lowest- and highest-disagreement terciles. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks. The vertical lines mark October 16, 2009, and January 1, 2020.

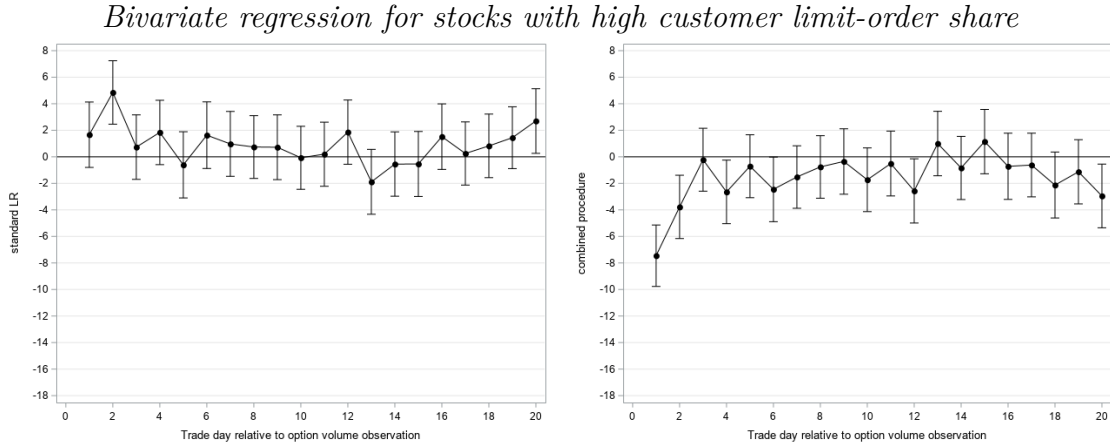
be expressed through passive order submission.

The figure also reveals important differences between S&P 500 and non-S&P 500 stocks. Under the combined classification, return predictability for non-S&P 500 stocks is strongest in low-limit-order environments and remains positive, though more muted, when the share of limit-order-driven trades is high. For S&P 500 stocks, by contrast, the strongest predictability arises in high-limit-order environments. This pattern is consistent with an order-choice interpretation. In the framework of Kaniel and Liu (2006), informed traders trade off execution certainty against price improvement. Marketable orders become more attractive when information is highly valuable or short-lived, because the opportunity cost of non-execution is larger. Limit orders become more attractive when information is longer-lived or less urgent, because the cost of waiting for execution is lower. The stronger overall return predictability from option demand outside the S&P 500 is consistent with information being more valuable, more slowly incorporated, or more difficult to arbitrage in this less liquid segment. In such settings, lower trading intensity increases execution risk for limit orders, making such orders less attractive when information is valuable or short-lived. In this case, the informed component of option demand may be more likely to be expressed through marketable orders. For liquid S&P 500 stocks, by contrast, execution risk is lower, so informed traders with longer-lived information can use limit orders to obtain price improvement without sacrificing much execution probability. This interpretation is also consistent with recent evidence in Barardehi et al. (2025), who document that limit-order usage is widespread in option markets and that option traders use limit orders to reduce trading costs despite non-execution risk. Thus, the evidence suggests that the informational content of limit-order demand depends on liquidity, execution risk, and the horizon or value of the underlying information.

Regression evidence in Table 3.7 provides a final test of whether customer limit-order activity is associated with stock-price information. In bivariate regressions estimated within the highest customer-limit-order tercile, the coefficient on the LR-based signal becomes positive, effectively reversing the expected relation between put-call ratios and future stock returns. In contrast, signals based on the combined classification continue to predict returns with the expected negative sign over extended horizons, particularly among non-S&P 500 stocks. This reversal indicates that conventional classification methods misattribute informative customer limit orders when passive order submission becomes more prevalent. More importantly, the persistence of the combined-classification signal suggests that customer limit-order activity contains information about future stock prices.

Overall, the evidence indicates that the superior predictive performance of the combined classification is concentrated precisely in settings where customer limit order activity is most prevalent. More broadly, the results suggest that conventional trade classification obscures information embedded in passive customer trading activity.

Panel A: Sample of S&P 500 stocks



Panel B: Sample of non-S&P 500 stocks

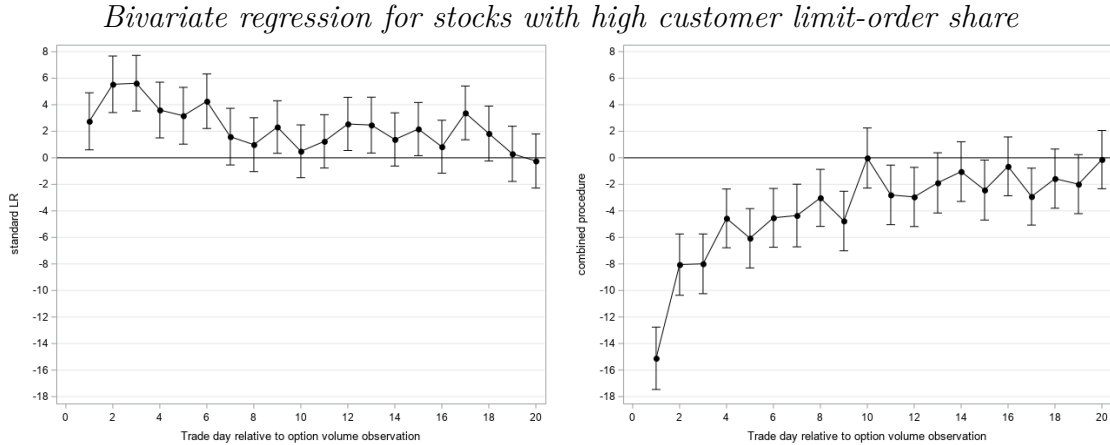


Figure 3.7: Bivariate regressions in high customer limit-order share subsample

The figure reports slope coefficients and 95% confidence intervals from bivariate Fama–MacBeth regressions of risk-adjusted stock returns on put-call ratios constructed using the standard LR algorithm and the combined classification procedure over return horizons of up to 20 trading days. The regressions are estimated within the highest tercile of customer limit order activity, where stocks are sorted each day based on the daily volume share of trades for which trade size equals quoted depth. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks.

3.6 Robustness Across Option Demand Measures

To assess whether our findings depend on the specific construction of option demand, we next examine whether trade classification affects the ability of option order imbalance to predict future stock returns. Unlike put-call ratios, which capture relative demand for put versus call options, option order imbalance measures the directional delta exposure embedded in option order flow and therefore captures both customer buying and selling activity.

Despite these differences in construction, the results closely mirror those obtained for put-call ratios. Table 3.6 shows that, across both S&P 500 and non-S&P 500 stocks, portfolio strategies based on option order imbalance constructed using the combined classification generally outperform those based on the conventional LR algorithm. Figure 3.8 further illustrates that strategies based on the combined classification accumulate persistent positive returns over time, with performance strengthening notably after 2020. In contrast, LR-based strategies largely stagnate after 2009 and gradually erode earlier gains during the post-2020 period. Similarly, regression evidence in Figure 3.9 shows that, in a direct horse race, the predictive content of the LR-based signal is largely confined to next-day returns and reverses sign at longer horizons. In contrast, the signal constructed using the combined classification retains positive and statistically significant predictive power for horizons of up to six trading days among non-S&P500 stocks. Among S&P 500 stocks, however, the combined-classification signal has no incremental predictive power once the LR-based signal is included.

Overall, these results demonstrate that the effect of trade classification on the measured relation between option demand and future stock returns is not specific to a particular proxy for option demand. Whether option demand is measured through relative put and call buying or through the directional exposure embedded in option positions, accurately identifying the customer side of trades is critical for recovering its predictive content. More broadly, the evidence reinforces our central conclusion that conventional trade classification obscures a substantial share of the information conveyed through customer limit order activity.

Table 3.6: Return predictability of option order imbalance

This table reports the performance of long-short investment strategies based on option order imbalances constructed using the standard LR algorithm and the combined classification procedure. Strategies buy stocks in the highest decile and sell stocks in the lowest decile of the option order imbalance distribution. Performance is reported as four-factor alphas from the Carhart (1997) model (FF4), expressed in basis points per week, together with corresponding Newey–West t -statistics (Newey and West, 1987). The “difference” columns report the performance difference between the combined classification and LR-based strategies. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks. The sample is further divided into three subperiods: Pre (Jan 1, 2004–October 15, 2009), Mid (October 16, 2009–December 31, 2019), and Post (January 1, 2020–June 30, 2021).

Panel A: Sample of S&P 500 stocks

	standard LR				combined procedure				difference			
	Full	Pre	Mid	Post	Full	Pre	Mid	Post	Full	Pre	Mid	Post
Alpha	1.37	7.03	-0.01	-9.27	6.48	13.74	-0.60	19.80	5.11	6.71	-0.59	29.07
t-stat	0.69	1.87	-0.01	-0.71	3.28	3.97	-0.33	1.54	1.74	1.44	-0.28	1.18

Panel B: Sample of non-S&P 500 stocks

	standard LR				combined procedure				difference			
	Full	Pre	Mid	Post	Full	Pre	Mid	Post	Full	Pre	Mid	Post
Alpha	4.94	13.84	2.42	-7.77	15.16	13.89	12.10	39.61	10.23	0.05	9.68	47.39
t-stat	2.37	4.37	1.12	-0.55	8.13	4.78	6.39	2.79	3.30	0.01	3.72	1.77

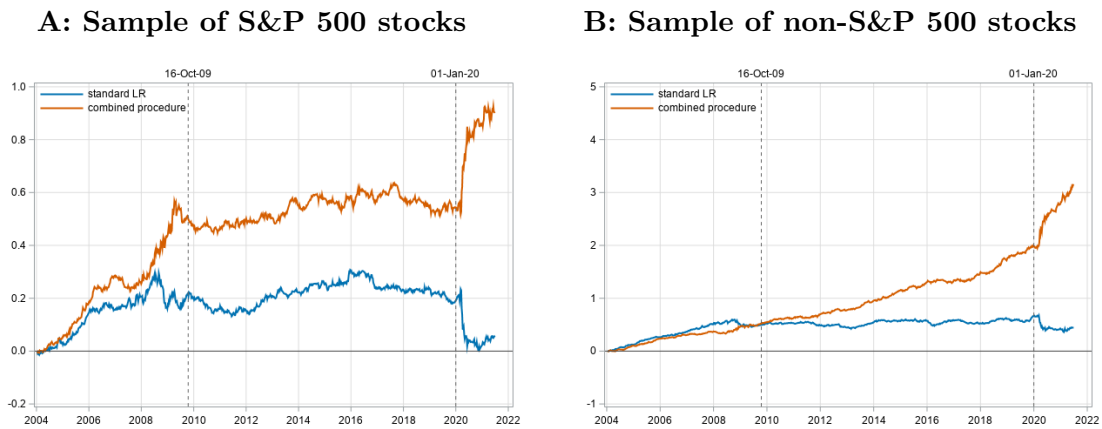
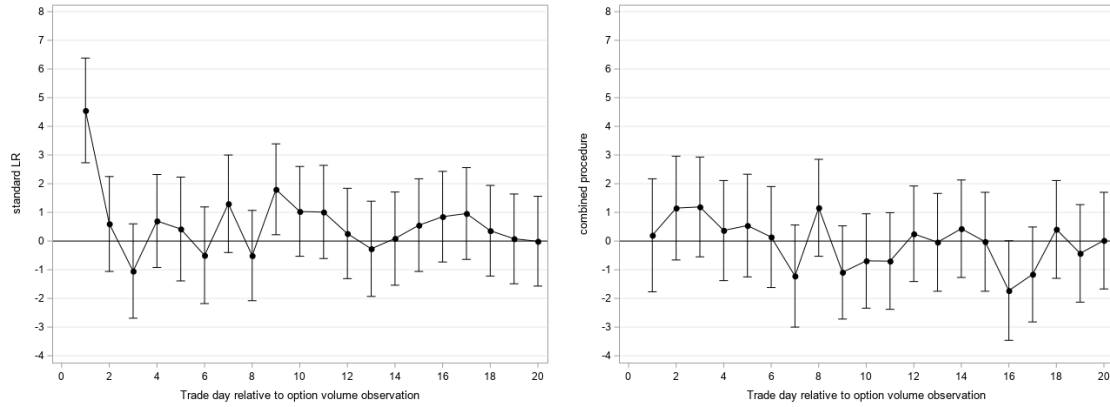


Figure 3.8: Cumulative returns of option order imbalance strategies

The figure plots cumulative returns of long-short strategies based on option order imbalances using the LR algorithm (blue) and the combined classification procedure (orange). Strategies buy stocks in the highest decile and sell stocks in the lowest decile of the option order imbalance distribution. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks. The vertical lines mark October 16, 2009, and January 1, 2020.

Panel A: Sample of S&P 500 stocks

Bivariate regression using standard LR and combined classification procedure



Panel B: Sample of non-S&P 500 stocks

Bivariate regression using standard LR and combined classification procedure

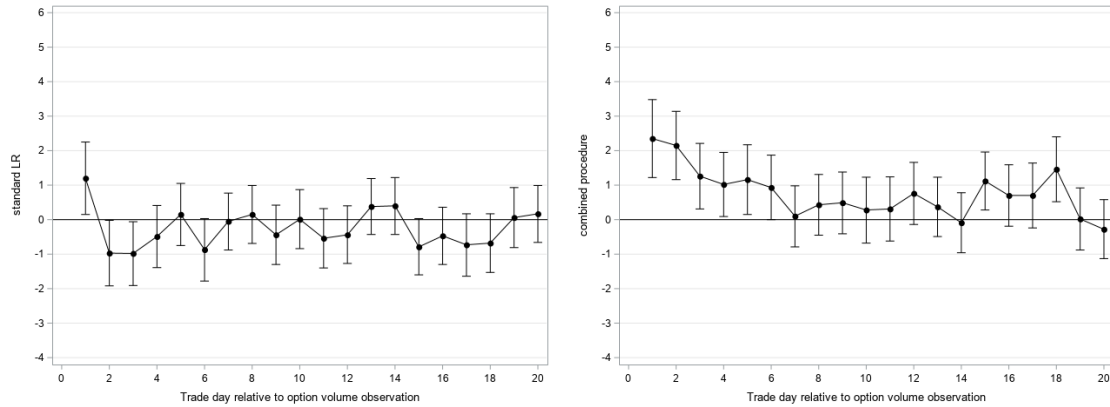


Figure 3.9: Bivariate regressions using option order imbalances

The figure reports slope coefficients and 95% confidence intervals from bivariate Fama–MacBeth regressions of risk-adjusted stock returns on option order imbalances constructed using the standard LR algorithm and the combined classification procedure over return horizons of up to 20 trading days. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks.

3.7 Conclusion

This chapter shows that the ability of option demand to predict future stock returns depends critically on how option trades are classified. Conventional trade classification methods substantially attenuate the predictive power of option-based signals. In contrast, signals constructed using the combined classification procedure outperform those based on conventional methods in predicting stock returns at both daily and weekly horizons, particularly among less liquid stocks and during the post-2020 period. In a direct horse race, the predictive content of conventional signals largely disappears once option demand is measured using the combined classification.

The superior performance of the combined classification appears to reflect its ability to identify information embedded in customer limit orders. When customer demand is expressed through passive order submission, conventional classification methods frequently attribute that demand to the counterparty. The combined classification substantially reduces this distortion by more accurately identifying the customer side of option trades. This evidence suggests that customer limit orders convey economically meaningful information about future stock returns. Our findings are consistent with theories of order submission in which traders with longer-lived information and lower execution urgency optimally rely on passive limit orders, particularly in environments where wider bid-ask spreads increase the value of price improvement (Chan et al., 2002; Kaniel and Liu, 2006; Barardehi et al., 2025).

The findings also help reconcile the mixed evidence in the literature on option-based return predictability. Under conventional classification methods, predictability appears to weaken substantially after 2009 (Bondarenko and Muravyev, 2025). Once option trades are more accurately attributed to the customer side, however, option demand continues to exhibit economically meaningful and persistent predictive power. The results therefore suggest that at least part of the documented decline in option-based predictability reflects measurement error arising from trade misclassification rather than a disappearance of informed trading in option markets.

The results have broader implications for empirical research using signed option order flow. Because trade classification fundamentally shapes the inferred direction of option demand, trade classification error can materially alter conclusions about return predictability and the role of option trading in price discovery. The findings therefore highlight that the informational content of financial markets depends not only on the presence of informed trading, but also on whether empirical measures correctly identify the economically relevant side of transactions.

Chapter 4

When Measurement Matters: Coverage, Trader Identity, and Trade Classification in Option Markets

4.1 Introduction

Option demand and dealer inventory play a central role in modern theories of information transmission, intermediary risk bearing, and demand-based asset pricing (Easley et al., 1998; Gârleanu et al., 2009). A large empirical literature uses option-market activity to study informed trading and stock return predictability (e.g., Pan and Poteshman, 2006; Hu, 2014; Bondarenko and Muravyev, 2025) and effects of intermediary risk on future stock volatility (e.g., Ni et al., 2021; Adams et al., 2025). Yet neither option demand nor dealer inventory is directly observable. Both are latent economic objects that must be inferred from imperfect option-market data.

This creates a first-order measurement problem. Researchers studying closely related questions often rely on different option-market datasets to infer the same latent economic objects but embed different sources of measurement error. Exchange-level Open/Close reports directly identify customer buying and selling activity but observe only a subset of trading venues, whereas transaction-level LiveVol data provide broader market coverage but require trade classification algorithms and do not identify whether trades originate from customers, firms, or broker-dealers. Different data sources therefore generate empirical proxies for option demand and dealer inventory that vary systematically in exchange coverage, trader identity, and trade classification.

These differences are not merely technical details. Exchange coverage determines whether observed trading activity reflects aggregate option demand, trader identity determines whether measured order flow captures end-user demand or also incorporates intermediary activity, and trade classification determines whether transactions are assigned the correct economic sign. As a result, measurement error can arise along multiple dimensions and directly affects how closely empirical proxies recover latent option demand and dealer inventory exposure. Consistent with this concern, prior evidence shows that option-order-flow measures constructed from different option-market data sources generate only weakly correlated signals (Muravyev and Ni, 2020).

These measurement concerns are especially relevant in modern option markets. Trading is increasingly dispersed across heterogeneous exchanges that differ in market structure, fee schedules, and execution mechanisms,¹ retail participation has changed the composition of market participants (Bryzgalova et al., 2023; Bogouslavsky and Muravyev, 2025), and price-improvement auctions and customer limit orders can increase classification error in conventional trade classification algorithms (Gayda, 2025; Grauer et al., 2026). These developments make measurement constraints central to empirical inference in option-market studies.

This raises a fundamental question: How do systematic measurement differences in option-market data affect the magnitude and interpretation of empirical relations? To answer this question, we decompose these differences into exchange coverage, trader identity, and trade classification, and examine how each dimension affects inference from option-market measures about subsequent stock-market outcomes. We study two channels. The first is an information channel, where directional option demand is used to study information transmission from option markets to future stock returns. The second is an intermediary channel, where dealer net gamma exposure is used to study how market-maker hedge rebalancing pressure relates to future stock volatility. Comparing these channels allows us to assess whether measurement differences affect inference from directional-demand and dealer-inventory measures in the same way.

Our empirical design combines exchange-level Open/Close reports from five option exchanges (ISE, CBOE, GEMX, NOM, and PHLX) with transaction-level LiveVol data for U.S. equity options from August 2013 to June 2021 and constructs measures that vary one measurement dimension at a time. We assess exchange coverage by comparing single-exchange and multi-exchange Open/Close measures,

¹By 2025, no single options exchange accounted for more than 20% of contract volume, and 18 venues each had more than 1% market share (Securities and Commission, 2026).

trader identity by comparing customer-only and broader trader-type measures, and trade classification by comparing alternative trade-signing procedures in LiveVol. Across comparisons, we hold the economic application fixed and evaluate how measurement differences affect cross-sectional predictability for future stock returns and stock volatility.

Our central finding is that measurement matters, but not uniformly across latent objects. Directional-demand measures are particularly sensitive to trade classification and exchange coverage, whereas dealer-inventory measures are more sensitive to trader identity and, among less liquid stocks, to trade classification accuracy. The consequences of measurement constraints therefore depend on whether the object of interest is investor demand or intermediary inventory exposure.

For directional-demand measures, broader exchange coverage and improved trade classification each add economically meaningful information beyond more restrictive measurement approaches. At the same time, customer order flow contains most of the informational content, while including firm and broker–dealer activity adds little incremental predictive power. Direct comparisons between Open/Close and LiveVol measures, which differ simultaneously in coverage, trader identity, and trade classification, show that both data environments lead to similar qualitative conclusions about the predictive content of directional option demand. However, neither data source fully subsumes the other. Instead, broader market coverage in LiveVol and direct identification of customer trading activity in Open/Close each contribute incremental predictive information.

Dealer-inventory measures are affected differently because they accumulate signed order flow over time. Measurement errors can therefore persist and compound as inferred inventory positions evolve. As a result, inventory-based measures are particularly sensitive to trader identity and, in less liquid stocks, trade classification. Across both channels, measurement effects are substantially larger among non-S&P 500 stocks, suggesting that accurate measurement becomes increasingly important in more heterogeneous and less liquid market segments.

Related Literature. These findings contribute to several strands of literature. First, we contribute to research on informed trading and return predictability in option markets. Existing evidence on directional option demand relies on two data environments that impose distinct measurement constraints. Studies based on transaction data rely on accurate trade classification (Easley et al., 1998; Chan et al., 2002; Cao et al., 2005; Hu, 2014; Bryzgalova et al., 2023), whereas Open/Close-based

studies rely on the representativeness of exchange-reported buy and sell volume (Pan and Poteshman, 2006; Ge et al., 2016; Weinbaum et al., 2023; Bondarenko and Muravyev, 2025). Although measurement constraints are often treated as implementation details, related evidence suggests that they can matter for empirical inference. Bondarenko and Muravyev (2025) show that option-based signals constructed from different exchanges contain independent information about future stock returns, while Pan and Poteshman (2006) and Bondarenko and Muravyev (2025) find little evidence of return predictability from firm trading activity. Related evidence from real estate investment trust option markets further shows that investor composition affects the informativeness of option trading for future returns (Cashman et al., 2025). We add to this literature by isolating the effects of exchange coverage, trader identity, and trade classification on return-predictability inference. We show that customer order flow contains most of the informational content in option trading activity, while improved trade classification and broader exchange coverage strengthen the predictive content of directional-demand measures for future stock returns.

Second, we contribute to the literature on option-market intermediation and its effects on the underlying stock market. Prior research shows that option-market activity can affect the underlying stock market through volatility information, dealer hedging, and intermediary risk bearing. Ni et al. (2008) show that non-market-maker demand for volatility predicts future realized stock volatility, while Ni et al. (2021) provide evidence that option market makers' hedge rebalancing affects underlying stock volatility. Related work links gamma hedging and dealer gamma imbalances to intraday momentum, reversals, and stock-market liquidity (Baltussen et al., 2021; Barbon and Buraschi, 2021; O'Donovan et al., 2023), with recent evidence on short-dated and zero-days-to-expiration SPX index options showing that positive market-maker gamma inventory can stabilize index volatility (Dim et al., 2025; Adams et al., 2025; Vasquez et al., 2025). However, the measurement constraints underlying estimates of dealer inventory remain less well understood. Ni et al. (2021) show that including firm proprietary positions can attenuate the estimated relation between dealer gamma inventory and stock volatility, while O'Donovan et al. (2023) show that limited coverage weakens the ability to identify the effect of option-market hedging on stock-market liquidity. We build on these insights by systematically isolating the roles of different sources of measurement error. We show that trader identity and exchange coverage materially affect the estimated strength of the dealer-inventory channel, with trade classification playing an additional role for less liquid stocks.

Third, this chapter contributes to demand-based option pricing and the broader literature that uses implied volatility and option expensiveness to study demand

pressure and intermediary risk-bearing constraints. Bollen and Whaley (2004) show that option order imbalance is related to changes in implied volatility, while Gârleanu et al. (2009) link market-maker positions to option expensiveness. Goyal and Saretto (2009) use the difference between historical realized volatility and implied volatility to predict option returns. Related studies use option expensiveness as a price-based proxy for demand pressure (Hitzemann et al., 2021; Kanne et al., 2023). We contribute by comparing empirical inference from price-based proxies with inference from quantity-based measures constructed from signed option trading activity. We show that price-based proxies are complements rather than substitutes, as they contain information not directly reflected in contemporaneous order flow and may reflect a mixture of demand pressure, expectations, risk premia, and compensation for intermediary risk bearing.

Finally, the chapter speaks to a broader measurement issue in empirical finance. Many central economic objects, such as option demand and dealer inventory, are latent and must be inferred from incomplete data. We show that measurement constraints have economically meaningful consequences that vary across applications. In option markets, increasing exchange fragmentation, heterogeneous trader types, and related trade classification challenges can materially affect the strength and interpretation of empirical relations.

The remainder of the chapter proceeds as follows. Section 4.2 describes the option-market data environments and introduces the three measurement dimensions. Section 4.3 describes the empirical applications and sample construction. Section 4.4 examines how exchange coverage, trader identity, and trade classification affect inference from directional-demand and dealer-inventory measures. Section 4.5 compares Open/Close and LiveVol measures and evaluates the combined effect of differences across data environments. Section 4.6 compares quantity-based measures with price-based proxies, and Section 4.7 concludes.

4.2 Option Data Environments and Measurement Dimensions

Closely related studies often use different option-market datasets to measure latent economic objects such as directional demand or dealer inventory. Because these data environments differ in the information available for measurement, empirical findings may reflect not only underlying economic mechanisms but also differences in how

empirical proxies are constructed. This section introduces the principal option-market data environments, identifies three sources of measurement differences, and develops a decomposition for isolating their effects on empirical inference.

4.2.1 Option Data Environments

Empirical studies of option demand and dealer inventory rely primarily on two data environments: exchange-level Open/Close volume reports and transaction-level OPRA data distributed through LiveVol. These datasets differ in the information available for measurement and therefore involve distinct trade-offs when constructing empirical measures of option demand and dealer inventory.

Open/Close data. Open/Close Volume Summary data report daily option trading volume separately for buys and sells by trader category and distinguish opening from closing transactions at the exchange-day-contract level. Reported trader categories include customers, professional customers, firms, and broker-dealers, while market-maker activity is observed indirectly through the market-clearing condition. A key advantage of Open/Close data is that both trade direction and trader identity are directly observed. Demand measures can therefore be constructed without trade classification algorithms and can be tailored to specific trader groups. The primary limitation is exchange coverage. Because each exchange reports only its own trading activity, studies using Open/Close data typically rely on one or a few venues and therefore capture only a subset of market-wide option demand.

LiveVol data. LiveVol provides transaction-level records of option trades across all U.S. option exchanges together with execution prices, trade sizes, timestamps, trade conditions, and contemporaneous quote information. Its primary advantage is broad market coverage and transaction-level granularity, which facilitate the construction of market-wide demand measures from individual trades. However, trader identity and trade direction are not directly observed. Demand measures therefore require trade classification algorithms to infer trade direction and assumptions about the trader categories represented in observed order flow.

4.2.2 Sources of Measurement Differences

The differences between Open/Close and LiveVol can be summarized along three dimensions that determine how latent option demand and dealer inventory are trans-

lated into empirical measures: exchange coverage, trader identity, and trade classification.

Exchange coverage. Because option trading is fragmented across exchanges, measures constructed from a subset of venues need not reflect aggregate option-market activity. Limited exchange coverage can omit economically relevant order flow and implicitly assumes that observed trading activity is representative of the broader market. This assumption is not innocuous, as U.S. option exchanges differ in market structure, liquidity provision, fee schedules, and order-routing incentives, which can generate systematic differences in trading activity across venues (Battalio et al., 2021)

Figure 4.1 illustrates the evolution of equity-option trading volume across exchanges. Empirical studies using Open/Close data often rely on ISE and/or CBOE. During the early 2000s, these exchanges together accounted for roughly 60% of trading volume in equity options. However, their combined market share declined steadily over time, falling to approximately 20% by June 2021. This decline coincided with the expansion of competing venues, including GEMX, NOM, and PHLX, and with substantial growth in market fragmentation. The number of operating option exchanges increased from six in 2004 to sixteen by the end of our sample period, and trading volume became increasingly dispersed across venues.

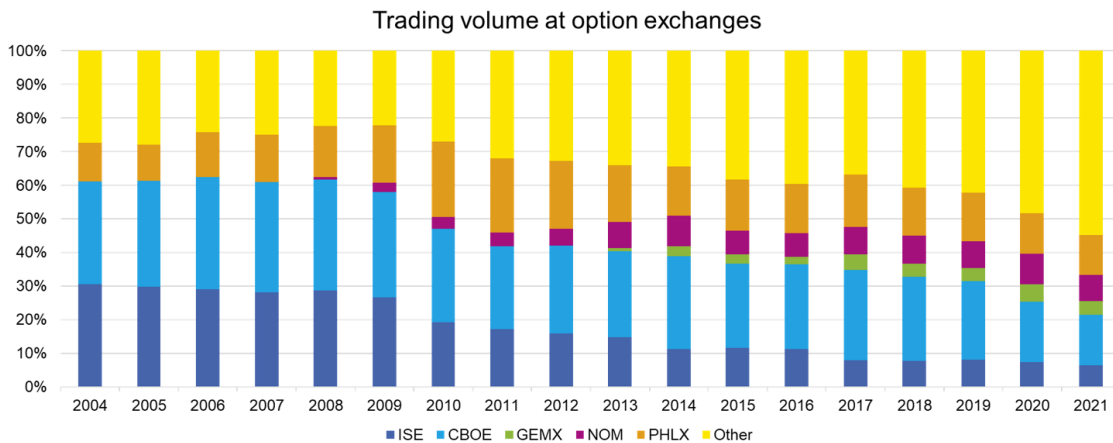


Figure 4.1: Market fragmentation across U.S. option exchanges

This figure reports the market shares of major U.S. option exchanges from January 2004 to June 2021, measured as the percentage of total option trading volume executed on each exchange. The stacked bars show the shares of ISE (dark blue), CBOE (light blue), GEMX (green), NOM (purple), PHLX (orange), and all remaining exchanges combined (yellow).

Motivated by this fragmentation, we compare two Open/Close-based measures that differ in exchange coverage. The narrow measure uses order flow from ISE, reflecting the single-exchange data environment commonly used in earlier studies. The broader measure aggregates order flow across the five exchanges for which Open/Close data are available in our sample: ISE, CBOE, PHLX, NOM, and GEMX. Together, these five exchanges account for more than 50% of trading volume during most of the sample period.

Measures based on a single exchange reflect only a subset of aggregate option-market activity. This creates measurement differences whenever economically relevant order flow occurs on omitted venues. If trading activity differs systematically across exchanges, because informed traders prefer particular venues or because market structure attracts different types of participants, limited exchange coverage can distort measured demand and inferred dealer inventory. The consequences are likely to be particularly pronounced in fragmented markets where omitted exchanges account for a substantial share of total trading volume.

Trader identity. Option markets involve heterogeneous market participants with different trading motives. Observing trader identity is therefore important for determining whose trading activity enters measured demand and whose counterparty positions are accumulated as inferred intermediary inventory. Open/Close data distinguish customer, professional customer, firm, and broker–dealer activity, allowing demand measures to be constructed for different trader groups. Customer trades include orders submitted by retail and institutional investors, such as hedge funds, while professional customers represent highly active customer accounts. Firm trades originate from proprietary trading desks of Options Clearing Corporation (OCC) members, whereas broker–dealer trades reflect activity by non-customer broker–dealers.

However, observing trader identity does not by itself determine whether a trader category represents end-user demand. This interpretation requires assumptions about the economic role of each trader type. Following Gârleanu et al. (2009), firm proprietary traders may at times act as end-users of options and at other times as liquidity suppliers. Similar ambiguity can arise for broker–dealer activity, which may reflect proprietary trading, hedging, or intermediary inventory management. We therefore construct option demand in two ways. First, we treat non-customer traders as liquidity suppliers and measure demand using customer order flow only. We refer to this definition as *customer-only* demand. Second, we treat professional customer, firm and broker–dealer activity as part of end-user demand and measure

demand using the combined order flow of customers, professional customers, firms, and broker-dealers. We refer to this broader definition as *customer+firm+broker-dealer* demand. The distinction is consequential because the same trading activity can enter measured demand under one interpretation but not under another, depending on the assumed role of the participating trader categories. For dealer-inventory measures, the same assumption determines which counterparty positions are interpreted as inventory risk borne by intermediaries.

Table 4.1 illustrates this distinction. Consider a transaction between a customer and a firm or broker-dealer. Under the customer-only definition, the transaction contributes to measured demand because the non-customer counterparty is treated as an intermediary. Under the broader customer+firm+broker-dealer definition, the same transaction does not contribute to net measured demand because both counterparties are treated as end-users expressing demand for option exposure, so their positions net out. Similarly, transactions between firms or broker-dealers and market makers contribute to measured demand only under the broader definition, which treats firm and broker-dealer trading as proprietary demand rather than intermediary activity. Thus, the same transaction can have different measurement implications for both option demand and inferred intermediary inventory depending on whether firms and broker-dealers are treated as liquidity suppliers or as end-users.

This distinction highlights a key limitation of LiveVol-based measures, as LiveVol does not observe trader identity. Demand must therefore be inferred from signed transactions alone. Trade classification algorithms implicitly assign demand to one

Table 4.1: Trader identity assumptions and measured demand

This table illustrates how alternative assumptions about trader identity affect whether transactions between different pairs of trader types contribute to measured demand. Under the *customer-only* definition, only customer trading contributes to measured demand, while under the *cust+firm+BD* definition, customers, professional customers, firms and broker-dealers are all treated as end-users. Because LiveVol does not observe trader identity, it assigns demand to one side of every signed transaction. *No net demand* indicates transactions that do not affect measured demand under the corresponding definition, either because both sides are included and net out or because neither side belongs to the included trader categories.

Trader-Type Pair	customer-only	cust+firm+BD	LiveVol
Customer – Market Maker	Net demand	Net demand	Net demand
Customer – Firm/BD	Net demand	No net demand	Net demand
Customer – Customer	No net demand	No net demand	Net demand
Firm/BD – Market Maker	No net demand	Net demand	Net demand
Firm – BD	No net demand	No net demand	Net demand

side of each trade and liquidity provision to the other, irrespective of the underlying trader identities. As a result, transactions between two customers contribute to option order imbalance measures in LiveVol but net out in Open/Close measures. More broadly, Open/Close measures can isolate demand from selected trader groups, but their interpretation still requires assumptions about the trading motives of those groups. By contrast, LiveVol-based measures capture signed order flow and inventory pressure transmitted to liquidity providers without observing the underlying trader identities (Muravyev and Ni, 2020).

Trader identity therefore affects both measured demand and inferred dealer inventory by determining which trader groups are treated as end-users and which counterparty positions are interpreted as intermediary exposure. These consequences may be especially important for inventory-based applications, where the interpretation of intermediary risk transfer depends on whether observed trades originate from end-users or from intermediaries themselves.

Trade classification. Directional-demand and net dealer-position measures both depend on trade direction. Trade direction determines whether a transaction is interpreted as buying or selling pressure and, in inventory-based applications, whether the transaction increases or decreases inferred intermediary exposure. Unlike Open/Close data, which directly report buy and sell volume by trader type, transaction-level LiveVol data do not observe trade direction. LiveVol-based measures therefore require trade classification algorithms to infer the relevant side of a trade.

We construct LiveVol-based measures using both the widely used Lee–Ready algorithm and the combined classification procedure. Figure 4.2 reports the accuracy of both procedures in the benchmark sample from Chapter 2, which evaluates classified LiveVol transactions against observed customer trade direction from Open/Close data. The figure shows that classification accuracy in option markets is limited and varies considerably over time and across exchanges, implying that exposure to classification error differs across samples and trading venues. The combined procedure improves classification accuracy across time and exchanges, but performance remains imperfect. Transaction-based measures therefore remain exposed to classification error, although less so under the combined procedure.

Classification errors directly affect both measured demand and inferred inventory because incorrectly signed transactions reverse the contribution of individual trades. For directional-demand measures, misclassification can obscure informational content by assigning true buying or selling pressure the wrong sign. For inventory-based

measures, the consequences may be even more severe because signed order flow is accumulated over time to infer intermediary positions. Misclassified transactions can therefore compound across periods and cause inferred inventory exposure to diverge from actual intermediary risk.

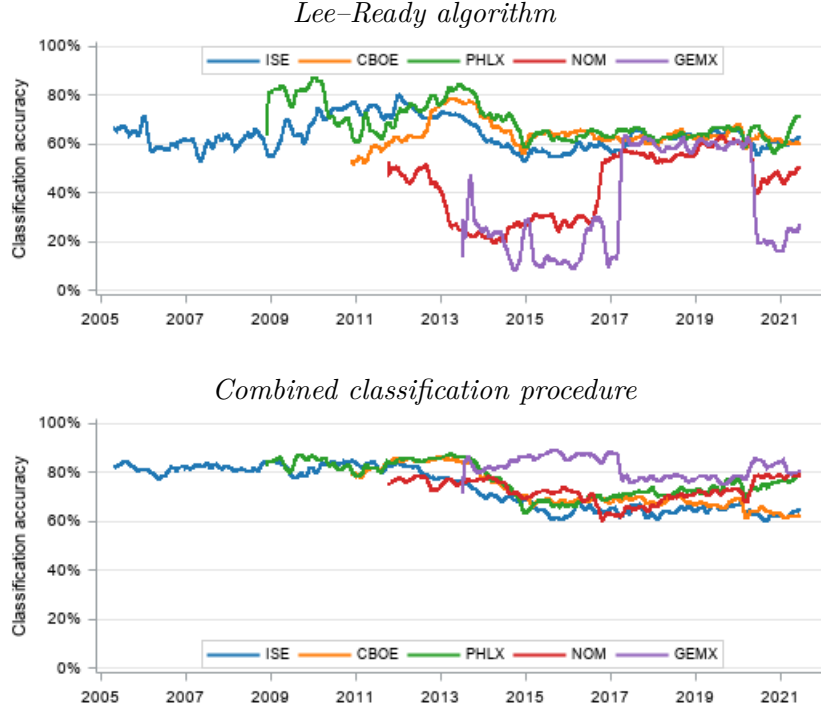


Figure 4.2: Accuracy of alternative trade classification methods

This figure illustrates the accuracy of the Lee–Ready algorithm and the combined classification procedure in the benchmark sample from Chapter 2. Classification accuracy is computed by comparing classified LiveVol transactions with observed customer trade direction from Open/Close data. All series are reported as 30-day moving averages. The blue line shows data for the ISE, the orange line for the CBOE, the green line is for the PHLX, the red line for the NOM, and the purple line represents the GEMX.

4.2.3 Empirical Decomposition of Measurement Dimensions

Empirical measures of option demand and dealer inventory are constructed from imperfect option-market data that differ in exchange coverage, trader identity, and trade classification. Consequently, measures constructed from different data environments need not recover the same empirical representation of option demand or dealer inventory, even when intended to capture the same underlying economic object. To organize this measurement problem, we summarize the relation between the latent object and its empirical measure with the following decomposition:

$$\widehat{D}_{i,t} = D_{i,t} + \varepsilon(\text{Coverage, Identity, Classification}), \quad (4.1)$$

where $\widehat{D}_{i,t}$ denotes the empirical stock-level measure, which may represent option demand or dealer inventory, $D_{i,t}$ is the corresponding latent economic object, and $\varepsilon(\cdot)$ captures deviations arising from incomplete exchange coverage, alternative assumptions about trader identity, and trade classification errors.

Equation (4.1) provides a conceptual decomposition of measurement constraints rather than a strict orthogonal partition. The three dimensions are linked through the institutional structure of option markets. Exchange design can affect both the composition of market participants and the information contained in observed transactions. Differences in fee schedules, customer-priority rules, market-making incentives, and execution mechanisms may attract different trader types across venues, while the accuracy of trade classification algorithms depends on trading protocols and execution behavior. The three dimensions should therefore be viewed as related sources of measurement differences whose effects may interact in practice.

Empirically, we isolate the contribution of each dimension by comparing alternative measures that differ along one dimension while holding the remaining dimensions fixed as closely as possible. This design does not impose full orthogonality, but it allows us to assess how exchange coverage, trader identity, and trade classification separately affect empirical measures of option demand and dealer inventory. Table 4.2 summarizes the alternative measurement specifications and the identification strategy used to isolate each measurement dimension.

First, we isolate exchange coverage using Open/Close data. We compare customer-only demand measured on ISE with otherwise identical customer-only demand aggregated across the five exchanges ISE, CBOE, PHLX, NOM, and GEMX. Because trader identity and trade direction are observed directly in both specifications, differences between the resulting signals primarily reflect the effect of exchange coverage.

Second, we isolate trader identity using the same five-exchange Open/Close data. We compare a customer-only measure with a broader measure that also includes professional customer, firm, and broker-dealer activity. This comparison holds exchange coverage fixed and examines how alternative assumptions about the relevant trader population affect measured demand.

Third, we isolate trade classification using LiveVol transaction data. We compare measures constructed with the Lee-Ready algorithm to measures constructed with the combined classification procedure. Because both measures are based on the

Table 4.2: Identification of measurement dimensions

This table summarizes the identification strategy used to isolate the three measurement dimensions. Panel A lists the alternative measurement specifications constructed from Open/Close and LiveVol data. Panel B reports the pairwise comparisons used to isolate exchange coverage, trader identity, and trade classification while holding the remaining dimensions fixed. The final comparison allows all three dimensions to differ simultaneously and captures the combined effect of using alternative option-data environments.

Panel A: Measurement specifications				
Measure	Data source	Coverage	Trader identity	Classification
(1) LR	LiveVol	All venues	Unobserved	Lee–Ready
(2) Combined	LiveVol	All venues	Unobserved	Combined
(3) ISE cust	Open/Close	ISE	Customer-only	Observed
(4) 5Ex cust	Open/Close	5 exchanges	Customer-only	Observed
(5) 5Ex cust+firm	Open/Close	5 exchanges	Cust+firm+BD	Observed

Panel B: Comparisons used to isolate measurement dimensions		
Dimension	Comparison	Interpretation
Exchange coverage	(4) vs. (3)	ISE vs. five Open/Close exchanges
Trader identity	(5) vs. (4)	Customer-only vs. customer+firm+bd
Trade classification	(2) vs. (1)	Lee–Ready vs. combined classification
Data environment	(4) vs. (2)	Open/Close vs. LiveVol

same transaction-level data, differences between them reflect how trade direction is inferred.

Finally, we compare five-exchange customer Open/Close measures with LiveVol measures based on the combined classification procedure. Unlike the preceding comparisons, this exercise allows exchange coverage, trader identity, and trade classification to differ simultaneously. It therefore captures the combined effect of using alternative option-data environments commonly employed in the literature.

The relative importance of the three measurement dimensions is ultimately an empirical question. Different dimensions may matter for different applications because the underlying measures rely on different economic interpretations of option trading. Directional-demand measures depend primarily on identifying buying and selling pressure, whereas dealer-inventory measures also require inferring intermediary inventory-risk exposure from signed trading activity. The empirical analysis therefore evaluates each measurement dimension separately and compares its role across directional-demand and inventory-based measures.

4.3 Empirical Applications and Research Design

This section introduces the empirical applications used to evaluate how the three measurement dimensions affect inference from option-market data. We consider two widely studied channels through which option-market activity is linked to subsequent stock-market outcomes: directional option demand, which can predict future stock returns, and dealer inventory exposure, which can affect future stock return volatility. Holding the underlying empirical design fixed, we vary exchange coverage, trader identity, and trade classification to isolate how alternative data environments affect measured signals and the resulting economic conclusions.

4.3.1 Option Order Imbalance and Daily Stock Returns

We begin with an application based on directional option demand and future stock returns. A large literature argues that option trading contains information about future stock prices because informed investors may preferentially trade in option markets. When market makers absorb option order flow, they subsequently hedge the resulting stock exposure in the underlying market. As a result, information contained in option trading can be transmitted to the stock market through dealer hedging activity, implying that measures of net directional option demand may predict future stock returns.

To measure directional option demand, Hu (2014) propose option order imbalance (OOI), a stock-level measure of net directional option demand expressed in delta-equivalent share units. For each stock i on day t , OOI is defined as

$$\text{OOI}_{i,t} = \frac{\sum_{j=1}^{N_{i,t}} 100 \cdot \text{Dir}_{i,j,t} \cdot \Delta_{i,j,t} \cdot \text{size}_{i,j,t}}{\text{SharesOutstanding}_{i,t}}, \quad (4.2)$$

where j indexes option trades on day t , $\text{Dir}_{i,j,t}$ equals $+1$ if trade j is buyer-initiated and -1 if it is seller-initiated, $\Delta_{i,j,t}$ is the option delta, and $\text{size}_{i,j,t}$ is trade size in option contracts, with one contract corresponding to 100 shares of the underlying. The numerator aggregates the signed, delta-equivalent share exposure of option trade initiators and therefore captures the net directional demand embedded in option trading. Scaling by shares outstanding facilitates cross-sectional comparisons and reduces noise when trading activity is sparse. Higher values of OOI therefore indicate stronger net positive directional demand. The measure is constructed using options with more than 10 days remaining to expiration.

To evaluate return predictability, stocks are sorted each trading day into equally weighted quintile portfolios based on OOI and held for one trading day. We then construct a long–short portfolio that buys stocks in the highest OOI quintile and sells stocks in the lowest OOI quintile. If directional option demand reflects information about future stock prices, stocks with high OOI should subsequently outperform stocks with low OOI, generating positive abnormal returns for the long–short portfolio. Performance is evaluated using four-factor alphas from Carhart (1997), Newey–West t -statistics (Newey and West, 1987), and annualized Sharpe ratios. Results are reported separately for S&P 500 and non-S&P 500 stocks in Table 4.5.

4.3.2 Put-Call Ratio and Weekly Stock Returns

We next consider a complementary measure of directional option demand based on relative buying pressure in put and call options. The underlying intuition is that investors who purchase call options express relatively bullish views on the underlying stock, whereas investors who purchase put options express relatively bearish views. Consequently, the relative demand for put and call options provides an alternative measure of investor expectations about future stock returns. If option traders possess information about future stock prices, stocks with relatively stronger call buying should subsequently outperform stocks with relatively stronger put buying.

To measure these expectations, we use the put-call ratio of Pan and Poteshman (2006), which compares investor demand for put and call options. For each stock i on day t , the put-call ratio (PC) is defined as

$$PC_{i,t} = \frac{PB_{i,t} + \varepsilon}{PB_{i,t} + CB_{i,t} + 2\varepsilon}, \quad (4.3)$$

where $PB_{i,t}$ and $CB_{i,t}$ denote put and call buy volume, respectively, measured in option contracts. Following Bondarenko and Muravyev (2025), we add $\varepsilon = 5$ contracts, which prevents division-by-zero issues and shrinks the ratio toward 0.5 when trading activity is low. The ratio is bounded between zero and one. Higher values indicate relatively stronger demand for puts than calls and therefore more bearish expectations about future stock returns. Under the informational interpretation of option trading, the put-call ratio should negatively predict subsequent stock returns.

While Pan and Poteshman (2006) and Bondarenko and Muravyev (2025) focus on opening buy volume, we construct put-call ratios using total buy volume, including both opening and closing transactions. This choice improves comparability

across Open/Close and LiveVol data environments by abstracting from differences arising from the direct observation of opening and closing activity in Open/Close data. Moreover, closing transactions may also convey information about investor beliefs and position adjustments, even if newly initiated positions are likely to be more informative on average (Ge et al., 2016). Using total buy volume therefore allows the analysis to focus on the measurement dimensions of exchange coverage, trader identity, and trade classification, rather than on differences stemming from the additional position-state information available in Open/Close data.

To evaluate weekly return predictability, stocks are sorted each trading day into equally weighted decile portfolios based on their put-call ratio. Portfolio positions are initiated one trading day after signal formation and held for one week. Following the portfolio-rotation approach of Bondarenko and Muravyev (2025), one-fifth of the long-short portfolio is rebalanced each day, and each newly added position is held for five trading days. The strategy buys stocks in the lowest put-call-ratio decile and sells stocks in the highest decile. If option-market demand contains information about future stock prices, the resulting long-short strategy should earn positive abnormal returns. Portfolio performance is evaluated using four-factor alphas from Carhart (1997), Newey–West t -statistics (Newey and West, 1987), and annualized Sharpe ratios. Results are reported separately for S&P 500 and non-S&P 500 stocks in Table 4.6. Figures 4.3, 4.5, 4.7, and 4.9 plot the corresponding cumulative returns.

4.3.3 Persistence and Incremental Information

We complement the portfolio-sort analysis with cross-sectional predictive regressions to examine the persistence of option-demand signals across forecasting horizons of up to 20 trading days and to assess whether alternative measurement specifications contain incremental predictive information. Following Pan and Poteshman (2006), we estimate both univariate and bivariate Fama and MacBeth (1973) regressions.

The univariate specification takes the form

$$R_{i,t+h} = \alpha + \beta \text{PC}_{i,t} + \varepsilon_{i,t+h}, \quad h = 1, \dots, 20, \quad (4.4)$$

where $R_{i,t+h}$ denotes the h -days-ahead risk-adjusted return of stock i , and $\text{PC}_{i,t}$ is the put-call ratio constructed under a given measurement specification. Consistent with the literature, risk-adjusted returns are computed as residuals from the Carhart (1997) four-factor model. If put-call ratios reflect informed option demand,

higher put demand relative to call demand should predict lower future stock returns, implying $\beta < 0$.

To compare alternative signal constructions directly, we estimate the following bivariate specification:

$$R_{i,t+h} = \alpha + \beta \text{PC}_{i,t}^{(1)} + \gamma \text{PC}_{i,t}^{(2)} + \varepsilon_{i,t+h}, \quad h = 1, \dots, 20, \quad (4.5)$$

where $\text{PC}_{i,t}^{(1)}$ and $\text{PC}_{i,t}^{(2)}$ denote put-call ratios constructed under alternative measurement specifications. These regressions provide a direct horse race between competing signal constructions and allow us to assess whether a given measurement restriction contributes incremental predictive information. If a signal captures information that is not fully reflected in competing specifications, its coefficient should remain significant after controlling for alternative measures. Conversely, if two specifications contain largely overlapping information, the predictive content of one signal should be subsumed by the other. All regressions are estimated separately for S&P 500 and non-S&P 500 stocks. Figures 4.4, 4.6, 4.8, and 4.10 report the corresponding coefficient estimates.

4.3.4 Dealer Gamma Inventory and Stock Return Volatility

The preceding applications focus on directional option demand and subsequent stock returns. We now consider a complementary channel linking dealer inventory exposure to future stock return volatility. Unlike directional-demand measures, inventory measures are designed to capture the transfer of risk to intermediaries through option trading. When dealers absorb option order flow, they accumulate inventory exposures that must be managed through dynamic hedging in the underlying stock. If dealers hold negative aggregate gamma positions, hedge rebalancing requires trading in the direction of underlying price movements, potentially amplifying return fluctuations. Consequently, stocks associated with larger dealer short-gamma positions may exhibit higher subsequent volatility.

Following Ni et al. (2021), dealer inventory positions are inferred from the evolution of aggregate non-market-maker open interest. For each option series j , open interest (OI) evolves according to

$$\text{OI}_{j,t} = \text{OI}_{j,t-1} + \text{Vol}_{j,t}^{\text{Buy}} - \text{Vol}_{j,t}^{\text{Sell}}, \quad (4.6)$$

where $\text{Vol}_{j,t}^{\text{Buy}}$ and $\text{Vol}_{j,t}^{\text{Sell}}$ denote buy and sell volumes in option series j on day t .

Dealer inventory is then inferred as the opposite side of aggregate non-market-maker positions,

$$\text{netOI}_{j,t}^{\text{MM}} = -\text{OI}_{j,t}, \quad (4.7)$$

reflecting the assumption that market makers absorb the residual order flow of other market participants.

Aggregating across all option series on stock i , dealer net gamma exposure is defined as

$$\text{net}_{i,t}\Gamma = 100 \left(\frac{S_{i,t}}{M_{i,t}} \right) \sum_{j=1}^{N_{i,t}} \text{netOI}_{j,t}^{\text{MM}} \Gamma_{i,j,t}(S_{i,t}), \quad (4.8)$$

where $S_{i,t}$ is the stock price, $M_{i,t}$ is shares outstanding, and $\Gamma_{i,j,t}$ denotes option gamma. More positive values of $\text{net}_{i,t}\Gamma$ indicate larger dealer long-gamma positions that may dampen volatility, whereas more negative values indicate larger dealer short-gamma positions and therefore greater potential for volatility-amplifying hedge rebalancing. The inventory analysis is restricted to underlying stocks with at least 500 trading days of available data.

To evaluate volatility predictability, stocks are sorted each trading day into equally weighted decile portfolios based on $\text{net}_{i,t}\Gamma$. We then compare next-day absolute stock returns between the highest and lowest decile portfolios. If dealer inventory affects future volatility through hedging activity, stocks associated with larger dealer long-gamma exposure should exhibit lower subsequent return volatility, whereas stocks associated with larger dealer short-gamma exposure should exhibit higher subsequent return volatility. Following Ni et al. (2021), we measure the volatility spread between the extreme decile portfolios and report the estimates in basis points together with Newey–West t -statistics (Newey and West, 1987). Results are reported separately for S&P 500 and non-S&P 500 stocks in Table 4.7.

4.3.5 Option Expensiveness as a Price-Based Benchmark

In addition to quantity-based measures constructed from option trading volume, we use option expensiveness as a price-based benchmark. While order-flow measures infer demand from signed trading activity, option expensiveness captures whether options are priced high or low relative to a reference volatility level. It therefore provides a complementary measure of option-market pressure that does not require trade classification or observed trader identity, and is not tied to a particular exchange.

Motivated by the demand-based option-pricing intuition that options with stronger

net end-user demand tend to become more expensive (Gârleanu et al., 2009), we first define option-level expensiveness as

$$\text{expensiveness}_{j,t} = \log(\text{iv}_{j,t}) - \log(\text{hv}_{i(j),t}), \quad (4.9)$$

where $\text{iv}_{j,t}$ is the implied volatility of option j and $\text{hv}_{i(j),t}$ is the historical volatility of the underlying stock i , computed as the standard deviation of the underlying's daily log returns over the preceding 365 calendar days, following Goyal and Saretto (2009). Positive values of $\text{expensiveness}_{j,t}$ indicate that the option is expensive relative to the underlying stock's historical volatility, whereas negative values indicate that it is cheap relative to this benchmark.

To obtain a stock-level price-based proxy of directional demand, $\Delta\text{-exp}_{i,t}$, we average delta-weighted expensiveness across all options on the respective stock:

$$\Delta\text{-exp}_{i,t} = \frac{1}{N_{i,t}} \sum_{j=1}^{N_{i,t}} \Delta_{i,j,t} \cdot \text{expensiveness}_{j,t}, \quad (4.10)$$

where $\Delta_{i,j,t}$ is the option delta. This aggregation maps option expensiveness into directional stock exposure. Expensive call options contribute positively to $\Delta\text{-exp}_{i,t}$, whereas expensive put options contribute negatively. Higher values therefore indicate stronger price-based evidence of bullish option-market pressure.

To construct a corresponding price-based proxy of dealer gamma inventory, $\Gamma\text{-exp}_{i,t}$, we average gamma-weighted expensiveness across all options written on the respective stock:

$$\Gamma\text{-exp}_{i,t} = -\frac{1}{N_{i,t}} \sum_{j=1}^{N_{i,t}} \Gamma_{i,j,t} \cdot \text{expensiveness}_{j,t}, \quad (4.11)$$

where $\Gamma_{i,j,t}$ denotes the option's gamma. The gamma weighting assigns greater weight to expensive call and put options with larger convexity exposure. Under the interpretation that option expensiveness reflects end-user demand pressure, this term captures price-based evidence of demand for long gamma exposure. The negative sign maps this end-user demand into the corresponding market-maker position, reflecting the assumption that market makers take the opposite side of customer option demand. More negative values of $\Gamma\text{-exp}_{i,t}$ therefore correspond to stronger price-based evidence of dealer short-gamma exposure and greater potential hedge-rebalancing pressure.

4.3.6 Sample Construction and Summary Statistics

The empirical analysis combines transaction-level option data from LiveVol and exchange-level Open/Close volume data from five option exchanges (ISE, GEMX, NOM, PHLX, and CBOE). These data are merged with option characteristics from OptionMetrics, stock returns from CRSP, and risk factors from Kenneth French's data library. Option-demand and dealer-inventory measures are constructed separately under alternative measurement specifications and aggregated to the stock-day level. Standard data cleaning filters are applied to both datasets, including the removal of duplicate observations, canceled trades, records with non-positive prices or volumes, and trades with reported volume exceeding ten million contracts.

The sample consists of equity options written on common stocks and spans August 5, 2013 through June 30, 2021, the period for which Open/Close data are available for all five exchanges following the launch of GEMX. To facilitate comparisons across data environments, we restrict the sample to underlying stocks that trade at least once on one of the five Open/Close exchanges during the sample period, while retaining stock-days for which option trading occurs exclusively on other exchanges in the LiveVol data. Throughout the analysis, results are reported separately for S&P 500 and non-S&P 500 stocks.

Table 4.3 reports summary statistics together with diagnostics related to the three measurement dimensions. The final sample contains approximately 1.0 million stock-days for S&P 500 firms and 6.1 million stock-days for non-S&P 500 firms, corresponding to 503 and 3,060 underlying stocks, respectively.

Measurement diagnostics for the three measurement dimensions highlight the potential importance of data constraints. First, option trading is highly fragmented across exchanges, particularly for S&P 500 stocks, which trade on an average of 11.5 exchanges per day compared with 5.8 exchanges for non-S&P 500 stocks. Consistent with this fragmentation, the ISE accounts for only about 16% of total option volume in both samples, while the five exchanges covered by our Open/Close data jointly account for approximately 53% and 57% of volume for S&P 500 and non-S&P 500 stocks, respectively. Second, professional customer, firm and broker-dealer activity represents a meaningful share of non-market-maker trading, accounting for 18% of buy volume among S&P 500 stocks and 14% among non-S&P 500 stocks. Finally, the Lee-Ready and combined classification procedures disagree on approximately 36% to 38% of classified volume, indicating substantial scope for trade classification differences to affect measured option demand and dealer inventory.

Table 4.3: Sample characteristics and measurement diagnostics

This table reports sample characteristics and measurement diagnostics separately for S&P 500 and non-S&P 500 stocks. Market capitalization and daily stock returns are reported as time-series averages of daily cross-sectional means. The average number of exchanges is the mean number of option exchanges with trading activity in a stock-day. Volume shares are reported relative to total option trading volume. Exchange volume shares are measured relative to total option trading volume in the corresponding stock group. The non-customer buy-volume share is the combined buy volume of professional customers, firms, and broker-dealers relative to total buy volume. The LR/combined disagreement share is the fraction of LiveVol transactions for which the Lee-Ready algorithm and the combined classification procedure assign different trade directions.

	S&P 500	Non-S&P 500
Sample characteristics		
Observations (stock-days)	1,016,329	6,107,783
Number of stocks	503	3,060
Market capitalization (\$bn)	44.67	2.43
Daily stock return (bps)	5.81	6.89
Measurement Diagnostics		
Average number of exchanges	11.50	5.82
ISE volume share (%)	16.05	16.37
Five-exchange volume share (%)	53.23	56.69
Non-customer buy-volume share (%)	17.94	14.40
LR/Combined disagreement share (%)	35.83	38.33

4.3.7 Correlations Across Alternative Measures

Before examining the economic implications of alternative measurement dimensions, we assess the similarity of the constrained measures. Although they are designed to capture the same underlying economic quantities, differences in exchange coverage, trader identity, and trade classification may generate systematic variation in the resulting signals. Pairwise correlations provide a descriptive assessment of the extent to which each measurement dimension affects measures of option demand and dealer inventory.

Table 4.4 reports time-series averages of daily cross-sectional Spearman correlations between alternative implementations of option order imbalance, put-call ratios, and dealer net gamma exposure. Correlations are generally positive and economically large, indicating that alternative measures capture substantial common variation in underlying option-market activity. However, the degree of similarity varies considerably across measurement dimensions.

Table 4.4: Correlations across measurement dimensions

This table reports time-series averages of daily cross-sectional Spearman correlations between alternative implementations of option order imbalance (OOI), the put-call ratio (PC) and dealer net gamma exposure (net Γ). Each row isolates one source of measurement variation by comparing two measures that differ along the corresponding dimension, as defined in Table 4.2. Trade classification compares measures based on the Lee–Ready and combined classification procedures. Exchange coverage compares customer-only measures constructed from ISE and five-exchange Open/Close data. Trader identity compares customer-only and customer+firm measures across the five Open/Close exchanges. Data source compares the combined LiveVol measure with the five-exchange customer Open/Close measure. Price vs. quantity compares option expensiveness with the corresponding five-exchange customer Open/Close measure.

	S&P 500 stocks			Non-S&P 500 stocks		
	OOI	PC	net Γ	OOI	PC	net Γ
Trade Classification	0.648	0.798	0.641	0.640	0.742	0.546
Exchange Coverage	0.462	0.568	0.913	0.551	0.511	0.908
Trader Identity	0.706	0.799	0.649	0.766	0.828	0.738
Data Source	0.319	0.581	0.294	0.325	0.548	0.336
Price vs. Quantity	-0.022	-0.006	0.147	-0.008	-0.054	0.065

Differences arising from trade classification are most pronounced for measures that rely on signed net demand. Correlations between measures constructed using the Lee–Ready algorithm and the combined classification procedure range from 0.55 to 0.80. Correlations are highest for put-call ratios and lower for option order imbalance and dealer inventory measures, consistent with classification errors having a larger effect on measures that rely on signed net demand, where misclassified transactions enter with the wrong sign.

Differences arising from exchange coverage are substantially larger for directional-demand measures than for dealer inventory measures. While correlations between inventory measures constructed from ISE and five-exchange Open/Close data exceed 0.90 in both samples, the corresponding correlations for option order imbalance and put-call ratios range only from 0.46 to 0.57. These results indicate that expanding exchange coverage changes directional-demand signals much more strongly than inferred dealer inventory positions.

By contrast, differences arising from trader identity are larger for dealer inventory measures. Correlations between customer-only and customer-plus-firm measures range from 0.71 to 0.83 for option order imbalance and put-call ratios, indicating that the inclusion of firm and broker–dealer activity has only a limited effect on directional-demand measures. Correlations are lower for dealer inventory

measures, ranging from 0.65 to 0.74, suggesting that inferred intermediary positions are more sensitive to assumptions about which trader categories are interpreted as representing underlying demand.

The largest differences emerge when comparing measures constructed from different data environments. Correlations between the combined LiveVol measure and the corresponding five-exchange customer Open/Close measure are only about 0.30 for option order imbalance and dealer inventory measures and only moderately higher for put-call ratios. These low correlations suggest that differences in exchange coverage, trader identity, and trade classification compound when signals are constructed from alternative data environments. As a result, transaction-level and exchange-level data can produce substantially different signals despite being designed to capture the same underlying economic phenomenon.

Finally, price-based and quantity-based measures exhibit little common variation. Correlations between option expensiveness and the corresponding five-exchange customer Open/Close measures are close to zero for option order imbalance and put-call ratios and only weakly positive for dealer inventory measures. These results suggest that price-based and quantity-based measures capture distinct aspects of option-market activity.

Overall, the correlations indicate that the consequences of measurement choices depend on the economic object being measured. Trade classification differences are most pronounced for measures that rely on signed net demand, exchange coverage primarily affects directional-demand measures, and dealer inventory measures exhibit greater sensitivity to trader-identity assumptions. At the same time, the low correlations across data environments suggest that the combined effects of multiple measurement dimensions can substantially alter the resulting signals. The subsequent analyses examine whether these differences in signal construction translate into economically meaningful differences in return and volatility predictability.

Table 4.5: Option order imbalance and next-day stock return predictability

This table reports next-day stock return predictability from option order imbalance measures constructed using alternative data sources and measurement specifications. Stocks are sorted daily into equally weighted quintile portfolios and held for one trading day. Long–short portfolios buy stocks in the highest quintile and sell stocks in the lowest quintile. The table reports long–short portfolio alphas from the Carhart (1997) model in basis points, Newey–West t -statistics (Newey and West, 1987), and annualized Sharpe ratios. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks. *LiveVol* refers to signals constructed from transaction-level data. Columns (1) and (2) use the Lee–Ready and combined classification procedures, respectively. *Open/Close* refers to signals constructed from exchange Open/Close data. Column (3) uses ISE customer trades, column (4) aggregates customer trades across five exchanges, and column (5) additionally includes professional customer, firm, and broker–dealer trading volume. *Measurement Effects* report differences between pairs of specifications that isolate the corresponding measurement dimension. Column (6) reports results for delta-weighted option expensiveness as a price-based benchmark.

Panel A: Sample of S&P 500 stocks

	LiveVol		Open/Close			Measurement Effects				Benchmark
	LR	Combined	ISE Cust	5Ex Cust	5Ex Cust+Firm	Classification	Coverage	Trader identity	Data source	Δ -exp
	(1)	(2)	(3)	(4)	(5)	(2)–(1)	(4)–(3)	(4)–(5)	(4)–(2)	(6)
Alpha	2.08	4.81	2.91	5.34	4.96	2.74	2.52	0.38	0.52	0.88
t -stats	2.45	5.86	4.04	5.78	5.83	2.19	3.38	0.84	0.75	0.57
Sharpe	0.87	2.01	1.39	1.95	1.98	–	–	–	–	0.18

Panel B: Sample of non-S&P 500 stocks

	LiveVol		Open/Close			Measurement Effects				Benchmark
	LR	Combined	ISE Cust	5Ex Cust	5Ex Cust+Firm	Classification	Coverage	Trader identity	Data source	Δ -exp
	(1)	(2)	(3)	(4)	(5)	(2)–(1)	(4)–(3)	(4)–(5)	(4)–(2)	(6)
Alpha	1.48	7.77	6.59	8.79	7.91	6.30	2.22	0.88	1.02	13.09
t -stats	1.98	13.77	10.79	11.33	11.66	6.29	3.05	2.89	1.62	8.71
Sharpe	0.78	5.26	3.68	4.34	4.50	–	–	–	–	3.32

Table 4.6: Put-call ratio and weekly stock return predictability

This table reports weekly return predictability from put-call ratios constructed using alternative data sources and measurement specifications. Stocks are sorted daily into equally weighted decile portfolios based on the put-call ratio. Long-short portfolios buy stocks in the lowest decile and sell stocks in the highest decile. Portfolios are formed one trading day after signal construction and held for one week, with daily rebalancing through the replacement of one-fifth of portfolio holdings. The table reports long-short portfolio alphas from the Carhart (1997) model in basis points, Newey–West t -statistics (Newey and West, 1987), and annualized Sharpe ratios. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks. *LiveVol* refers to signals constructed from transaction-level data. Columns (1) and (2) use the Lee–Ready and combined classification procedures, respectively. *Open/Close* refers to signals constructed from exchange Open/Close data. Column (3) uses ISE customer trades, column (4) aggregates customer trades across five exchanges, and column (5) additionally includes professional customer, firm, and broker–dealer trading volume. *Measurement Effects* report differences between pairs of specifications that isolate the corresponding measurement dimension. Column (6) reports results for delta-weighted option expensiveness as a price-based benchmark.

Panel A: Sample of S&P 500 stocks

	LiveVol		Open/Close			Measurement Effects				Benchmark
	LR	Combined	ISE Cust	5Ex Cust	5Ex Cust+Firm	Classification	Coverage	Trader identity	Data source	- Δ -exp
	(1)	(2)	(3)	(4)	(5)	(2)–(1)	(4)–(3)	(4)–(5)	(4)–(2)	(6)
Alpha	-3.40	0.28	3.33	-1.78	-2.16	3.68	-5.09	0.39	-2.05	-1.33
t -stats	-1.39	0.12	1.32	-0.66	-0.99	2.17	-2.34	0.24	-1.09	-0.28
Sharpe	-0.67	-0.13	0.28	-0.28	-0.43	–	–	–	–	-0.25

Panel B: Sample of non-S&P 500 stocks

	LiveVol		Open/Close			Measurement Effects				Benchmark
	LR	Combined	ISE Cust	5Ex Cust	5Ex Cust+Firm	Classification	Coverage	Trader identity	Data source	- Δ -exp
	(1)	(2)	(3)	(4)	(5)	(2)–(1)	(4)–(3)	(4)–(5)	(4)–(2)	(6)
Alpha	11.43	21.86	11.22	24.92	20.33	10.43	13.71	4.59	3.06	14.07
t -stats	2.55	4.19	3.05	4.21	3.81	4.33	3.48	2.73	1.19	1.62
Sharpe	0.86	1.60	1.12	1.65	1.45	–	–	–	–	0.26

Table 4.7: Dealer gamma inventory and next-day stock return volatility

This table reports next-day stock return volatility across net gamma-weighted inventory measures constructed using alternative data sources and measurement specifications. Stocks are sorted daily into decile portfolios based on normalized net gamma-weighted inventory. The table reports the difference in average absolute stock returns on day $t + 1$ between the lowest and highest deciles, expressed in basis points, together with Newey–West t -statistics (Newey and West, 1987). Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks. *LiveVol* refers to signals constructed from transaction-level data. Columns (1) and (2) use the Lee–Ready and combined classification procedures, respectively. *Open/Close* refers to signals constructed from exchange Open/Close data. Column (3) uses ISE customer trades, column (4) aggregates customer trades across five exchanges, and column (5) additionally includes professional customer, firm, and broker–dealer trading volume. *Measurement Effects* report differences between pairs of specifications that isolate the corresponding measurement dimension. Column (6) reports results for gamma-weighted option expensiveness as a price-based benchmark.

Panel A: Sample of S&P 500 stocks

	LiveVol		Open/Close			Measurement Effects				Benchmark
	LR (1)	Combined (2)	ISE Cust (3)	5Ex Cust (4)	5Ex Cust+Firm (5)	Classification (2)–(1)	Coverage (4)–(3)	Trader identity (4)–(5)	Data source (4)–(2)	Γ -exp (6)
Return	15.53	16.66	28.90	31.22	25.44	1.13	2.32	5.78	14.55	16.30
t -stats	15.75	13.83	20.18	19.97	20.40	1.18	4.78	6.11	12.19	7.91

Panel B: Sample of non-S&P 500 stocks

	LiveVol		Open/Close			Measurement Effects				Benchmark
	LR (1)	Combined (2)	ISE Cust (3)	5Ex Cust (4)	5Ex Cust+Firm (5)	Classification (2)–(1)	Coverage (4)–(3)	Trader identity (4)–(5)	Data source (4)–(2)	Γ -exp (6)
Return	10.77	28.28	35.15	42.81	35.56	17.51	7.65	7.24	14.53	97.79
t -stats	12.39	26.00	34.52	35.76	32.08	20.86	14.63	11.71	16.54	39.01

4.4 Measurement Dimensions and Predictive Inference for Stock Returns and Volatility

This section evaluates how the measurement dimensions discussed above influence the construction and predictive content of option-market signals. Holding the underlying empirical application fixed allows us to isolate the effects of exchange coverage, trader identity, and trade classification on empirical inference. For each dimension, we examine implications for both directional-demand measures and dealer inventory measures.

4.4.1 Exchange Coverage

We first examine how expanding coverage from a single exchange to multiple exchanges affects empirical inference from option-market signals.

Directional Demand Measures. Exchange coverage affects the measurement of directional option demand because single-exchange measures observe only a subset of aggregate option-market activity. The economic consequences of this limitation depend on whether trading activity is distributed uniformly across venues. If trading activity contains exchange-specific information, single-exchange measures will not simply be noisier versions of aggregate demand but may systematically misrepresent the underlying signal.

Consistent with this interpretation, broader exchange coverage strengthens the predictive content of directional-demand measures. Comparing single-exchange and five-exchange measures reveals economically meaningful differences in return predictability, although the effects vary across directional-demand measures and stock universes. For option order imbalance measures in Table 4.5, aggregating customer trading activity across five exchanges strengthens return predictability in both samples, with long–short alphas increasing from 2.91 to 5.34 basis points among S&P 500 stocks and from 6.59 to 8.79 basis points among non-S&P 500 stocks. The evidence for put–call ratios in Table 4.6 is more nuanced. Among non-S&P 500 stocks, broader coverage substantially strengthens predictability, with the long–short alpha more than doubling from 11.22 to 24.92 basis points. By contrast, return predictability remains weak among S&P 500 stocks, and the single-exchange measure performs slightly better than its five-exchange counterpart in weekly trading strategies.

The results for option order-imbalance measures indicate that buy and sell activity contains information that is not uniformly distributed across exchanges, causing single-exchange measures to omit economically relevant order flow. The substantially larger effects for put-call ratios suggest that put and call trading is even more heterogeneous across venues, particularly among non-S&P 500 stocks. The contrast between S&P 500 and non-S&P 500 stocks further suggests that the benefits of broader exchange coverage are not driven solely by the quantity of additional volume observed. Although expanding coverage increases observed volume by a similar amount in both samples, the predictive gains are concentrated for put-call ratios among non-S&P 500 stocks. This pattern suggests that option trading activity is more heterogeneous across trading venues outside the S&P 500 universe, causing single-exchange measures to be less representative of market-wide option demand. Broader exchange coverage therefore improves not only the completeness of measured option demand but also its ability to capture the information embedded in aggregate option order flow.

The cumulative returns in Figure 4.3 provide additional evidence that exchange heterogeneity has become increasingly important over time. The performance gap between single- and multi-exchange measures widens noticeably after 2016 and con-

Exchange coverage: ISE only versus five-exchange coverage

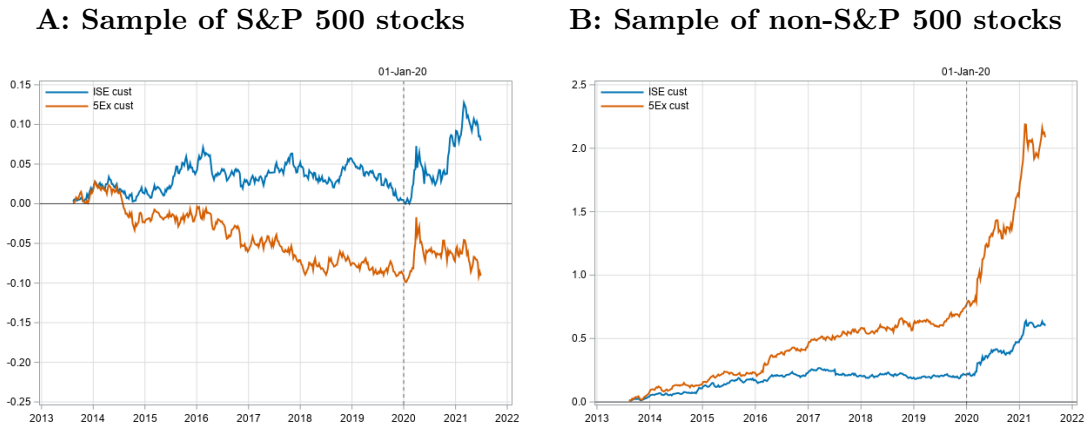


Figure 4.3: Cumulative returns by exchange coverage

The figure plots cumulative returns of long-short strategies based on put-call ratios constructed using ISE customer trades (orange) and customer trades aggregated across five exchanges (blue). Stocks are sorted daily into decile portfolios based on the put-call ratio. Long-short portfolios buy stocks in the lowest decile and sell stocks in the highest decile. Portfolios are formed one trading day after signal construction and held for one week. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks. The vertical line marks January 1, 2020, associated with increased retail participation in option markets.

tinues to grow throughout the remainder of the sample. This pattern is consistent with the increasing fragmentation of the U.S. options market following the introduction of several new exchanges, including EDGX in 2015, MRX in 2016, PEARL in 2017, and EMLD in 2019. The growing performance advantage of multi-exchange measures suggests that economically relevant order flow became increasingly dispersed across trading venues over time.

In contrast, cumulative returns remain comparatively similar in the S&P 500 sample, with ISE-based strategies performing modestly better over parts of the sample. Viewed in isolation, this evidence might suggest that broader exchange coverage adds little value for highly liquid stocks. The direct horse races in Figure 4.4, however, show that broader coverage captures information beyond that contained in single-exchange signals. In bivariate regressions, five-exchange measures consistently retain predictive power after controlling for their ISE counterparts, whereas the predictive content of ISE-based measures largely disappears. This pattern holds across both samples and is particularly pronounced among non-S&P 500 stocks, where five-exchange measures remain informative over horizons of up to six trading days. The results therefore indicate that broader coverage does not merely reduce measurement noise. Instead, order flow observed on additional exchanges contains distinct information about future stock returns that is not captured by single-exchange measures. Exchange coverage therefore affects not only the strength of estimated effects but also the information captured by directional-demand measures.

Dealer Inventory Measures. Dealer inventory measures depend on observing the customer order flow that generates intermediary risk exposure. Because dealers accumulate and manage inventory positions across multiple trading venues, measures constructed from a limited set of exchanges may provide an incomplete representation of aggregate dealer inventories.

The empirical evidence indicates that broader exchange coverage strengthens the relation between dealer inventory imbalances and future stock return volatility. Comparing the single-exchange and five-exchange specifications in Table 4.7 shows that expanding coverage from one to five exchanges increases the volatility spread from 28.90 to 31.22 basis points among S&P 500 stocks and from 35.15 to 42.81 basis points among non-S&P 500 stocks. By capturing a larger share of the customer order flow that generates dealer inventory positions, broader coverage provides a more complete measure of the net gamma exposures underlying future hedge rebalancing. These differences are economically meaningful despite the very high correlations between single- and multi-exchange inventory measures documented in the diagnostic

Panel A: Sample of S&P 500 stocks

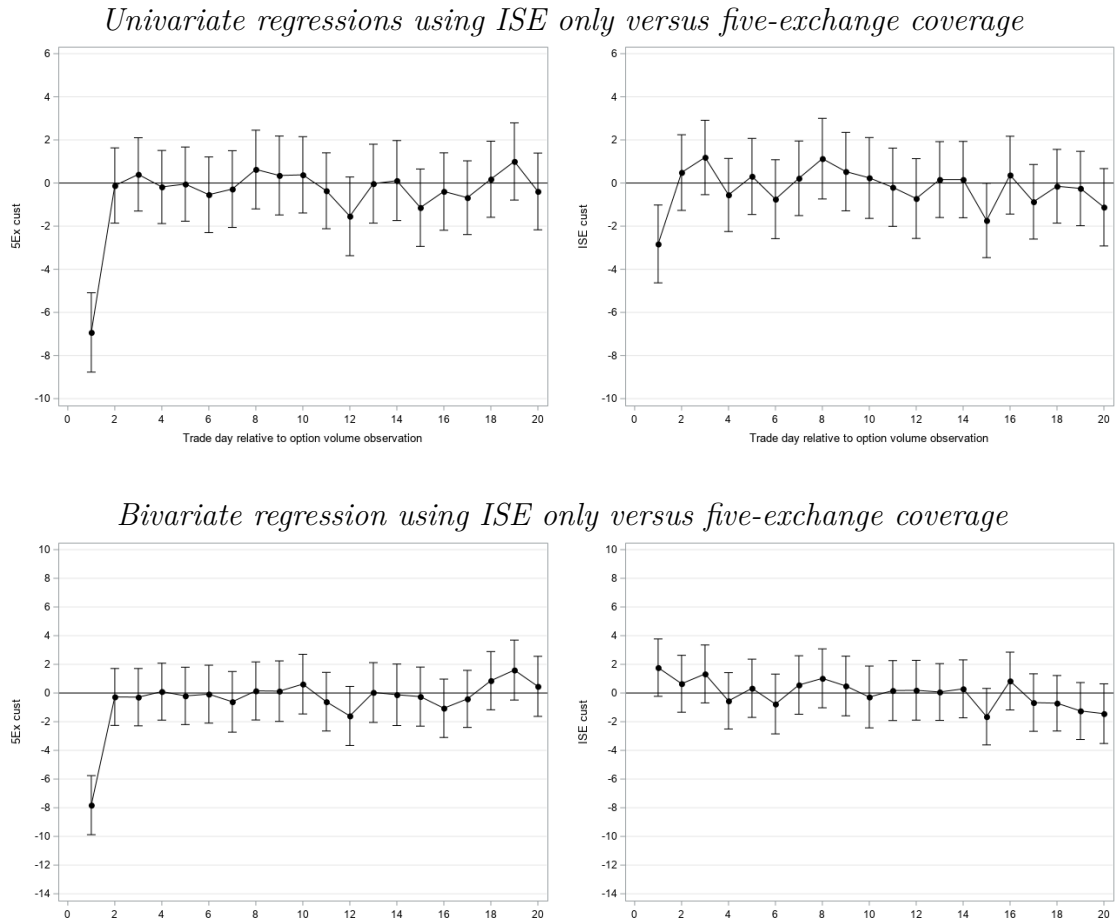


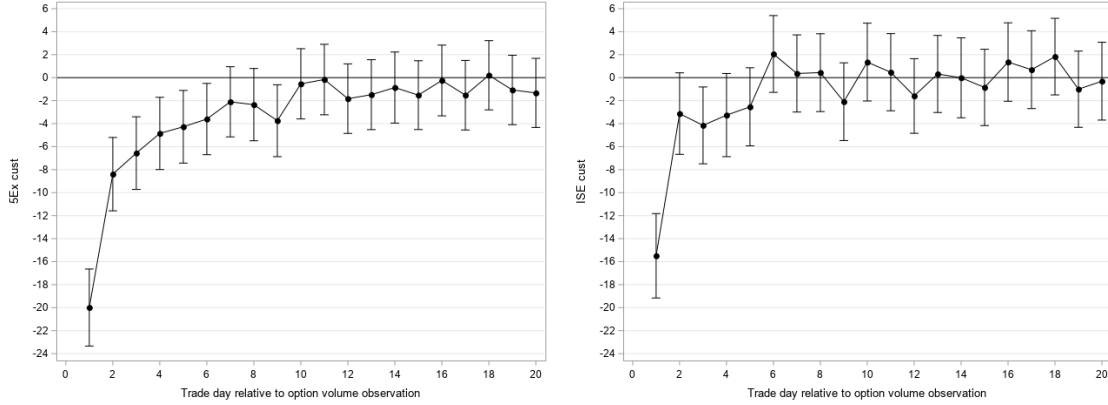
Figure 4.4: Predictive regressions by exchange coverage

The figure reports predictive regressions of risk-adjusted stock returns on put-call ratios constructed using ISE customer trades (left) and customer trades aggregated across five exchanges (right) over return horizons of up to 20 trading days. The first row reports slope coefficients and 95% confidence intervals from univariate Fama–MacBeth regressions of future risk-adjusted stock returns on put-call ratios constructed using each classification method separately. The second row reports coefficients from bivariate Fama–MacBeth regressions that include both put-call ratio measures jointly. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks.

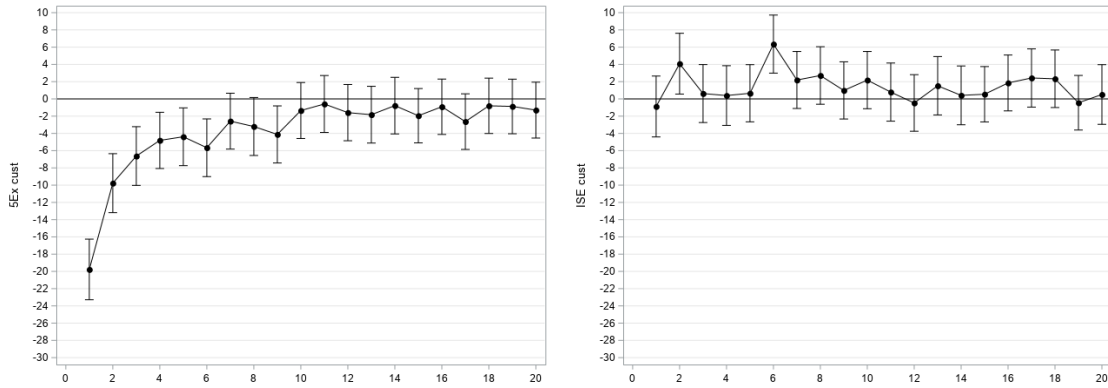
Figure 4.4 continued

Panel B: Sample of non-S&P 500 stocks

Univariate regressions using ISE only versus five-exchange coverage



Bivariate regression using ISE only versus five-exchange coverage



analysis. The results therefore imply that inventory-based applications are sensitive to measurement differences that appear small in the cross section but nevertheless affect economically relevant variation in dealer risk exposure.

The larger effects among non-S&P 500 stocks further suggest that the benefits of broader exchange coverage cannot be explained solely by the quantity of additional trading activity observed. Because the increase in covered volume is similar across the two samples, the stronger improvement in volatility predictability outside the S&P 500 universe points to differences in how inventory positions evolve over time. In particular, dealer inventories are likely to persist for longer periods in less liquid stocks. When inventories unwind more slowly, omitted trades have a more persistent effect on measured dealer exposure, causing relatively small differences in observed order flow to translate into larger differences in inventory-based volatility predictions.

4.4.2 Trader Identity

We next examine how alternative assumptions about trader identity affect the measurement of option demand and dealer inventory exposure.

Directional Demand Measures. Trader identity determines whose trading activity enters measured demand. If different trader categories reflect distinct economic motives, including or excluding them can alter the informational content of option-market signals. Relatively high correlations between measures based solely on customer trading activity and measures that additionally incorporate firm and broker–dealer trades suggest that alternative trader-type specifications share a substantial common component, while still capturing some distinct variation in directional option demand.

Consistent with these correlations, Tables 4.5 and 4.6 show that trader identity has only a modest effect on the predictive content of directional-demand measures. Restricting attention to customer trading activity generally strengthens return predictability, but the economic magnitudes are small. Among S&P 500 stocks, differences between trader type specifications are economically and statistically insignificant for both option-order-imbalance and put-call-ratio strategies. The effects are somewhat larger among non-S&P 500 stocks, where excluding firm and broker–dealer activity has little effect on option-order-imbalance strategies but increases put-call-ratio portfolio returns by approximately 4.6 basis points. The relatively small share of non-market-maker buy volume and the high correlations between alternative trader-type specifications suggest that these measures capture largely similar variation in option demand, helping to explain the modest differences in return predictability. This similarity is also evident in Figure 4.6, where cumulative returns track each other closely across trader-type specifications.

The predictive regressions in Figure 4.6 provide a more direct assessment of the informational contribution of different trader categories. The univariate regressions reveal broadly similar predictive patterns across specifications, consistent with the modest differences observed in portfolio sorts. The bivariate regressions, however, show that customer-only signals retain predictive power after controlling for measures that additionally incorporate firm and broker–dealer activity. By contrast, other trader types contribute little incremental information and frequently reverse sign at short horizons. These results indicate that the predictive content of directional-demand measures originates primarily from customer trading activity. Firm and broker–dealer trades, in contrast, appear to reflect inventory management,

hedging activity, and liquidity provision rather than directional information about future stock returns.

Trader identity: customer only versus customer+firm+broker-dealer

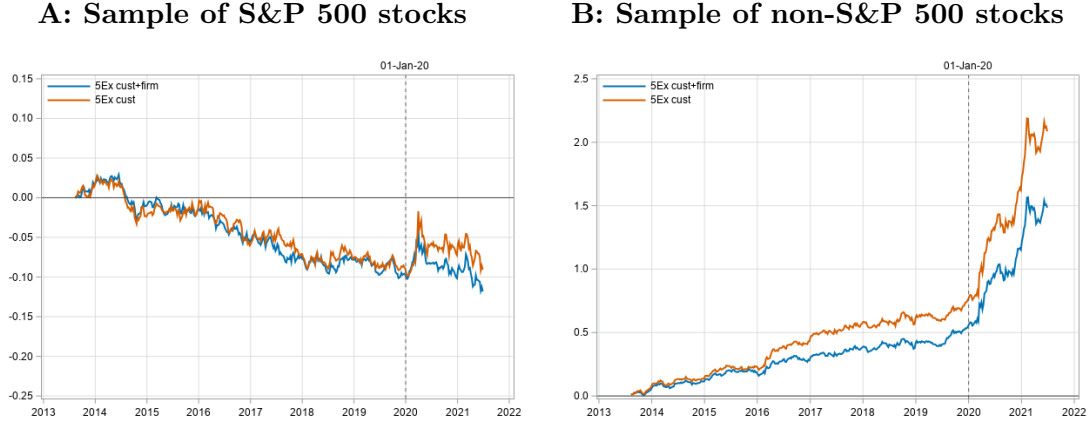


Figure 4.5: Cumulative returns by trader identity

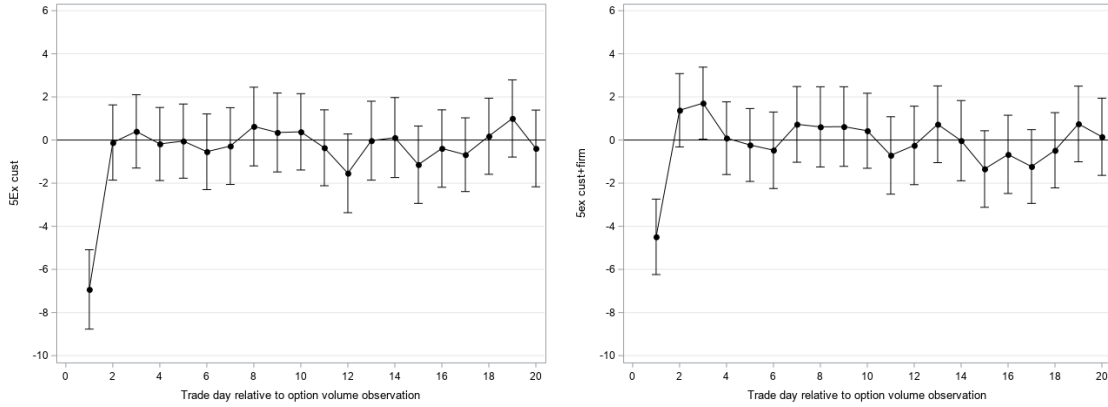
The figure plots cumulative returns of long–short strategies based on put–call ratios constructed using customer trades aggregated across five exchanges (orange) and from five-exchange customer, professional customer, firm, and broker–dealer trades (blue). Stocks are sorted daily into decile portfolios based on the put–call ratio. Long–short portfolios buy stocks in the lowest decile and sell stocks in the highest decile. Portfolios are formed one trading day after signal construction and held for one week. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks. The vertical line marks January 1, 2020, associated with increased retail participation in option markets.

Dealer Inventory Measures. Dealer inventory measures depend directly on assumptions about which trader categories represent underlying demand. Because dealers take the opposite side of customer trades and hedge the resulting exposures in the underlying stock, trader identity determines which transactions contribute to inferred dealer inventory positions. If firm and broker–dealer trades primarily reflect intermediary activities such as hedging, inventory management, and liquidity provision rather than underlying demand, including these trades can materially distort measured net gamma exposure and its relation to future stock return volatility.

Consistent with this mechanism, inventory measures inferred from customer trading activity exhibit substantially stronger relations with future stock return volatility than measures that additionally incorporate professional customer, firm and broker–dealer trades. Table 4.7 shows that the volatility spread between the lowest and highest deciles sorted on net gamma-weighted dealer inventory increases from 18.63 to 31.22 basis points among S&P 500 stocks and from 20.56 to 42.81 basis points among non-S&P 500 stocks. The differences are economically large and highly sta-

Panel A: Sample of S&P 500 stocks

Univariate regressions using customer only versus customer+firm+broker-dealer



Bivariate regression using customer only versus customer+firm+broker-dealer

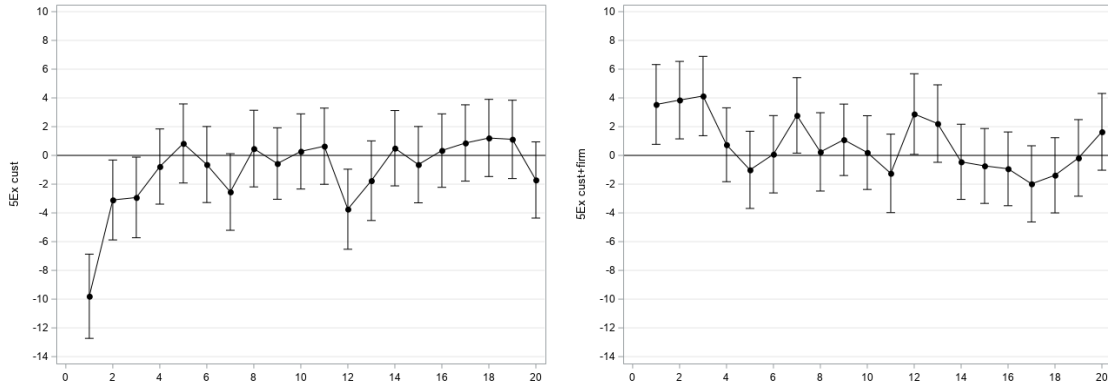


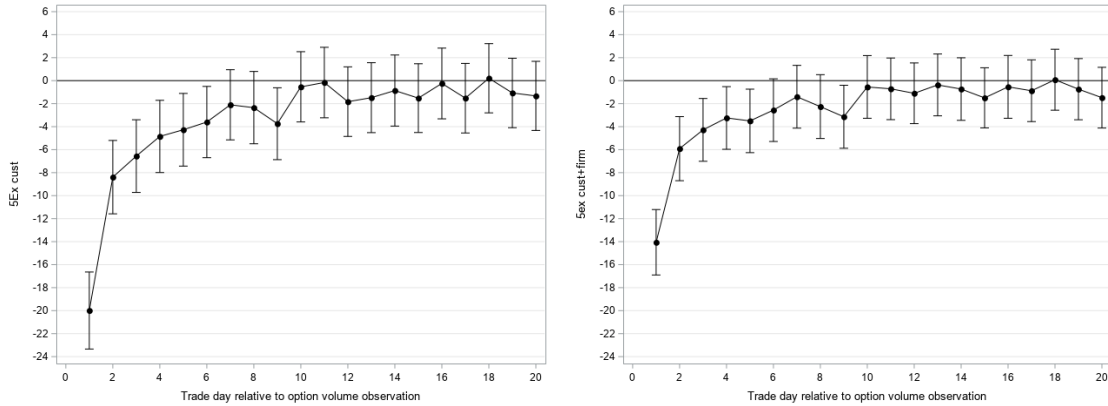
Figure 4.6: Predictive regressions by trader identity

The figure reports predictive regressions of risk-adjusted stock returns on put-call ratios constructed using customer trades aggregated across five exchanges (left) and from five-exchange customer, professional customer, firm, and broker-dealer trades (right) over return horizons of up to 20 trading days. The first row reports slope coefficients and 95% confidence intervals from univariate Fama–MacBeth regressions of future risk-adjusted stock returns on put-call ratios constructed using each classification method separately. The second row reports coefficients from bivariate Fama–MacBeth regressions that include both put-call ratio measures jointly. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks.

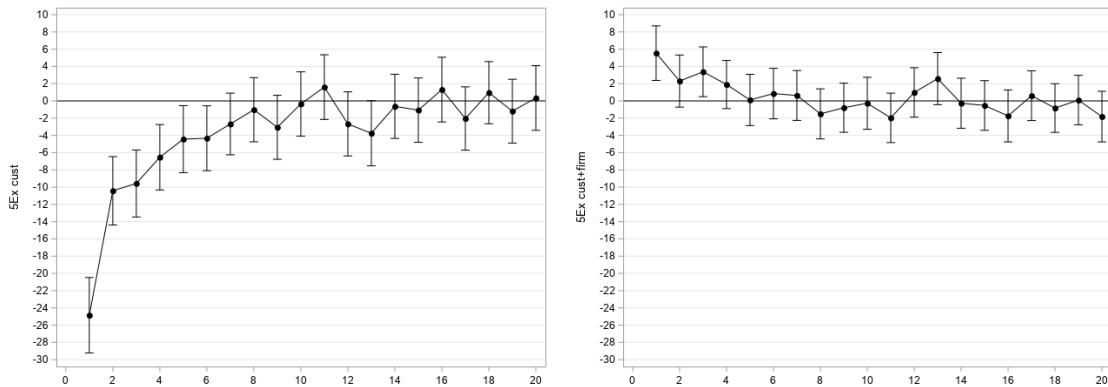
Figure 4.6 continued

Panel B: Sample of non-S&P 500 stocks

Univariate regressions using customer only versus customer+firm+broker-dealer



Bivariate regression using customer only versus customer+firm+broker-dealer



tistically significant in both samples. These results indicate that measured dealer inventory exposure is highly sensitive to assumptions about which trader categories represent underlying demand. The substantially stronger performance of customer-based inventory measures suggests that firm and broker-dealer trades largely reflect intermediary activities rather than the customer demand that generates dealer inventory risk.

The larger effects among non-S&P 500 stocks suggest that inventory measures become particularly sensitive to trader identity in less liquid market segments, where intermediary positions are likely to unwind more slowly. In such settings, positions generated by firm and broker-dealer trading activity remain embedded in inferred dealer inventories for longer periods, making it more difficult to isolate the customer demand that drives future hedging pressures and stock return volatility.

The magnitude of these effects stands in sharp contrast to the comparatively

modest role of trader identity in directional-demand applications. While directional-demand measures remain broadly similar across trader-type specifications, inventory measures are substantially more sensitive to how trader categories are interpreted. More broadly, the results highlight that the consequences of measurement choices depend fundamentally on the economic object being measured.

4.4.3 Trade Classification

We finally examine how alternative trade classification algorithms affect empirical inference from transaction-level option data.

Directional Demand Measures. Trade classification determines how observed transactions are mapped into signed demand. If customer trading activity is systematically misclassified, measured order flow may provide a distorted picture of directional option demand. Building on Chapter 3, which shows that improved trade classification strengthens option-based return predictability, the analysis here revisits this dimension in a different setting to facilitate comparison with the other measurement dimensions. Specifically, we restrict the cross-section to underlying stocks that trade at least once on one of the five Open/Close exchanges during the sample period, which begins in 2013. This analysis therefore also serves as a robustness test and confirms that trade classification remains a first-order determinant of the predictive content of directional-demand measures.

Comparing the conventional LR algorithm with the combined classification procedure shows that improving trade classification strengthens return predictability across directional-demand measures. For option-order-imbalance strategies in Table 4.5, long-short alphas increase by 2.7 basis points among S&P 500 stocks and by 6.3 basis points among non-S&P 500 stocks. For put-call-ratio strategies in Table 4.6, the corresponding improvements are 3.7 and 10.4 basis points. These economically large gains suggest that classification error materially attenuates the information contained in measured option demand. The larger effects among non-S&P 500 stocks are consistent with conventional classification methods performing less reliably in less liquid markets. In the case of option-order-imbalance measures, improved classification recovers economically and statistically significant next-day return predictability that is largely absent from conventional LR-based measures.

The cumulative returns in Figure 4.7 suggest that the economic benefits of improved trade classification have increased over time. The performance gap between

the combined classification and the conventional LR algorithm widens noticeably after 2020 and continues to grow thereafter. This pattern is consistent with increased retail participation and a growing prevalence of customer limit-order activity, a setting in which conventional trade classification methods perform poorly.

Trade classification: LR algorithm versus combined procedure

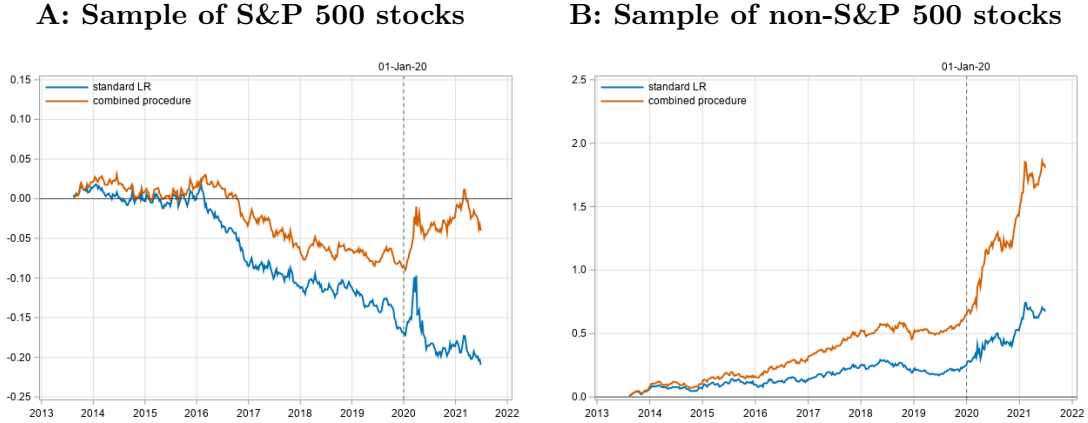


Figure 4.7: Cumulative returns by classification method

The figure plots cumulative returns of long–short strategies based on put–call ratios constructed from trades classified using the combined classification procedure (orange) and the LR-algorithm (blue). Stocks are sorted daily into decile portfolios based on the put–call ratio. Long–short portfolios buy stocks in the lowest decile and sell stocks in the highest decile. Portfolios are formed one trading day after signal construction and held for one week. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks. The vertical line marks January 1, 2020, associated with increased retail participation in option markets.

The predictive regressions in Figure 4.8 indicate that improved trade classification recovers information that is largely missed by conventional classification methods. In direct horse races, signals constructed using the combined classification retain predictive power after controlling for LR-based measures, whereas LR-classified signals contribute little incremental information and frequently reverse sign at short horizons. The results therefore indicate that classification error does not merely weaken directional-demand measures but causes them to miss economically relevant information embedded in option order flow.

Dealer Inventory Measures. Dealer inventory measures require accurately signing the trading activity used to infer intermediary risk exposure. Misclassified trades enter inventory measures with the wrong sign, causing errors in inferred dealer positions that can persist as inventories accumulate over time.

Comparing dealer inventory measures constructed using the conventional LR

Panel A: Sample of S&P 500 stocks

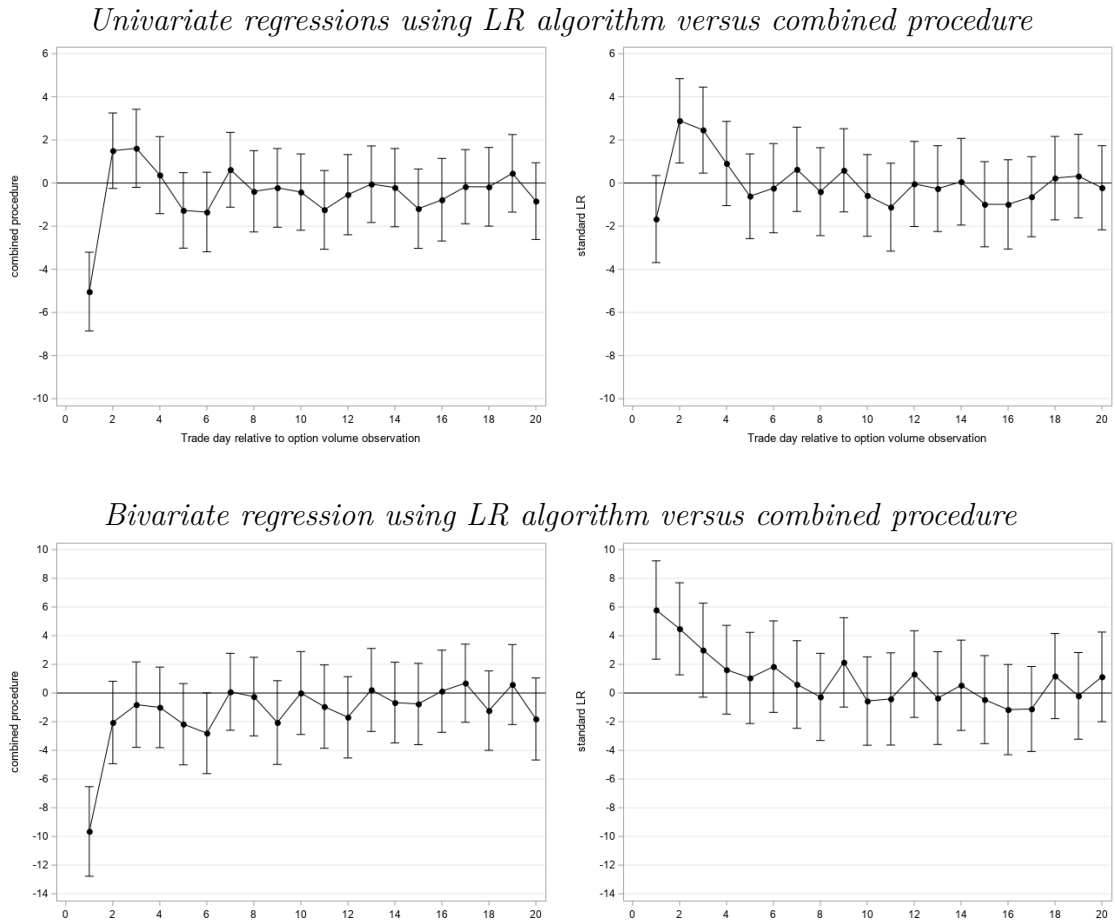
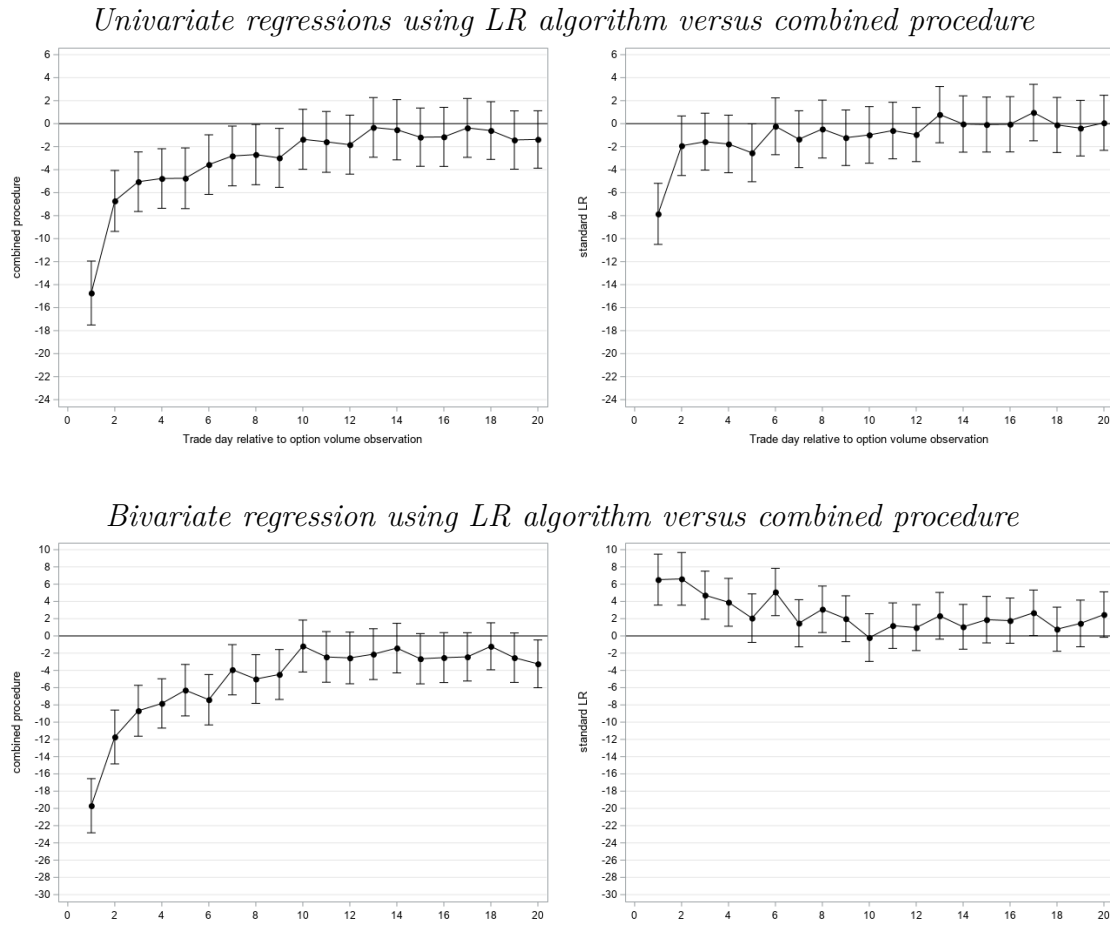


Figure 4.8: Predictive regressions using put-call ratios by classification method

The figure reports predictive regressions of risk-adjusted stock returns on put-call ratios constructed using the combined classification procedure (left) and the LR-algorithm (right) over return horizons of up to 20 trading days. The first row reports slope coefficients and 95% confidence intervals from univariate Fama–MacBeth regressions of future risk-adjusted stock returns on put-call ratios constructed using each classification method separately. The second row reports coefficients from bivariate Fama–MacBeth regressions that include both put-call ratio measures jointly. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks.

Figure 4.8 continued

Panel B: Sample of non-S&P 500 stocks

algorithm with those based on the combined classification procedure in Table 4.7 reveals that trade classification can materially affect the measured relation between dealer inventory imbalances and future stock return volatility. The consequences, however, differ markedly across liquidity segments. Among S&P 500 stocks, improved trade classification has an economically small and statistically insignificant impact, increasing the volatility spread by only 1.1 basis points. By contrast, among non-S&P 500 stocks, the volatility spread increases by almost 18 basis points, a difference that is both economically large and highly statistically significant.

The larger effects among non-S&P 500 stocks are consistent with conventional classification methods performing less reliably in less liquid markets. Because dealer inventory measures accumulate signed order flow over time, classification errors can persist and compound as inventory positions evolve. As a result, even moderate misclassification can translate into economically meaningful differences in inferred

dealer exposures and their relation to future stock return volatility. The evidence therefore suggests that conventional LR classification may materially understate intermediary inventory effects when classification errors are persistent.

4.5 Implications for Empirical Measurement in Option Markets

This section synthesizes the preceding evidence to assess the implications of measurement error for empirical research in option markets. We first compare the relative importance of exchange coverage, trader identity, and trade classification across directional-demand and dealer-inventory applications. We then examine whether alternative option-market data environments recover similar economic information despite their distinct measurement trade-offs.

4.5.1 Relative Importance of Measurement Dimensions

The evidence indicates that exchange coverage, trader identity, and trade classification all influence empirical inference in option markets. Their relative importance, however, differs markedly across applications and depends fundamentally on the economic object being measured.

For directional-demand applications, trade classification and exchange coverage emerge as the most important measurement dimensions. Improved trade classification substantially strengthens return predictability by reducing errors in signed order flow, while broader exchange coverage incorporates information dispersed across trading venues. By contrast, trader identity has only modest effects. Although customer-only measures generally outperform specifications that additionally incorporate professional customer, firm and broker–dealer activity, the resulting differences are small relative to those associated with trade classification and exchange coverage.

The ranking differs for dealer-inventory applications. Here, trader identity becomes an important measurement dimension. Because dealer-inventory measures depend on identifying the customer demand that generates intermediary risk exposure, including firm and broker–dealer activity materially weakens the relation between inferred dealer inventories and future stock return volatility. Exchange coverage also remains important because dealer inventories are accumulated and

managed across fragmented trading venues. Trade classification matters among non-S&P 500 stocks, where classification errors are more pronounced and inventory positions persist for longer periods.

A second pattern emerges across all applications: measurement effects are consistently larger among non-S&P 500 stocks. Improvements in exchange coverage, trader identification, and trade classification generally have stronger consequences outside the S&P 500 universe, suggesting that measurement error becomes more consequential in less liquid and more heterogeneous market segments.

Overall, the results indicate that directional option demand is most sensitive to trade classification and exchange coverage, whereas dealer inventory exposure is also sensitive to trader identity. More broadly, the findings suggest that the consequences of measurement error depend fundamentally on the economic object being measured.

4.5.2 Comparing Open/Close and LiveVol Measures

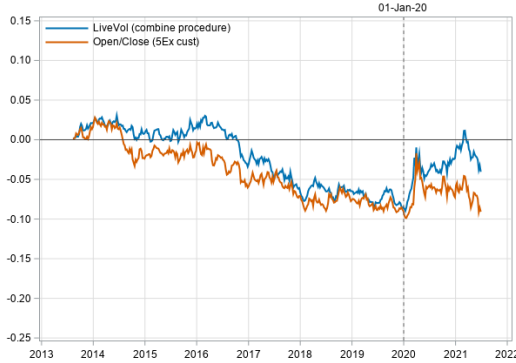
Because Open/Close and LiveVol data differ simultaneously along the measurement dimensions examined above, contrasting the two environments provides a useful assessment of their overall economic implications. We therefore compare measures constructed from customer trading volume aggregated across five Open/Close exchanges with corresponding measures constructed from LiveVol data using the combined classification procedure. The resulting evidence reveals whether alternative option-market data environments recover similar economic information despite their distinct measurement trade-offs.

Directional Demand Measures. Despite embodying fundamentally different measurement trade-offs, Open/Close and LiveVol measures lead to remarkably similar conclusions about the predictive content of directional option demand. Tables 4.5 and 4.6 show that differences in long-short portfolio performance are generally modest and statistically insignificant relative to the much larger effects associated with exchange coverage and trade classification examined in the previous sections. This similarity is also evident in the cumulative-return results in Figure 4.9, where Open/Close and LiveVol strategies track each other closely over time.

The bivariate regressions in Figure 4.10 provide a more nuanced picture. Among S&P 500 stocks, the predictive content of the two measures largely overlaps, although Open/Close measures retain some incremental information at the shortest horizon. Among non-S&P 500 stocks, both measures remain informative after controlling for

Data source: Open/Close customer trades versus LiveVol combined classification

A: Sample of S&P 500 stocks



B: Sample of non-S&P 500 stocks

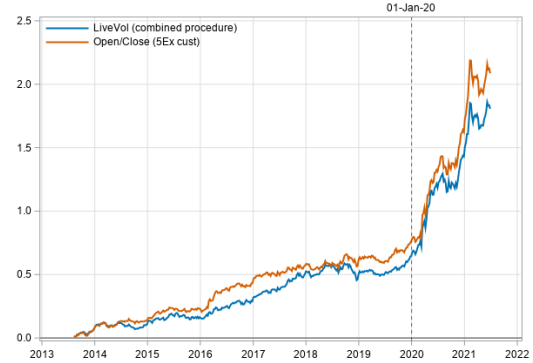


Figure 4.9: Cumulative returns by data source

The figure plots cumulative returns of long–short strategies based on put–call ratios constructed from Open/Close customer trades across five exchanges (orange) and from LiveVol trades classified using the combined classification procedure (blue). Stocks are sorted daily into decile portfolios based on the put–call ratio. Long–short portfolios buy stocks in the lowest decile and sell stocks in the highest decile. Portfolios are formed one trading day after signal construction and held for one week. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks. The vertical line marks January 1, 2020, associated with increased retail participation in option markets.

the alternative signal, with Open/Close measures predicting returns for up to three trading days and LiveVol measures for up to seven days. These results suggest that neither data source fully subsumes the other. Instead, broader market coverage in LiveVol and direct identification of customer trading activity in Open/Close each contribute incremental information, particularly in less liquid market segments. At the same time, the substantial overlap in predictive content suggests that both data environments recover largely the same underlying directional-demand signal despite their distinct measurement trade-offs.

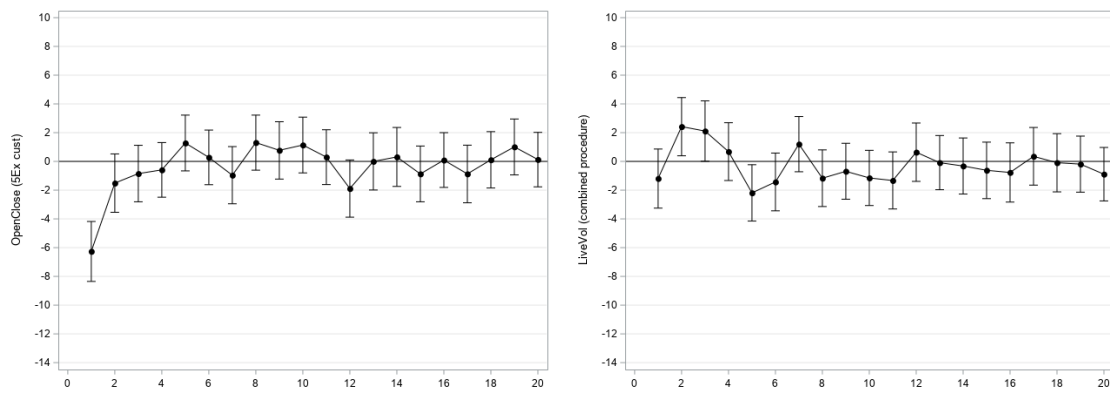
Dealer Inventory Measures. The comparison reveals larger differences across data environments for inventory-based applications. Although both data sources generate economically large and highly significant relations between dealer inventory imbalances and future stock return volatility, the estimated magnitudes differ substantially. Table 4.7 shows that volatility spreads constructed from five-exchange customer Open/Close data exceed those based on LiveVol measures by approximately 14.5 basis points among both S&P 500 and non-S&P 500 stocks.

The larger differences are consistent with the evidence presented in the preceding sections that dealer inventory measures are particularly sensitive to trader identity

and trade classification. Unlike directional-demand measures, inventory measures accumulate signed order flow over time, causing classification errors to persist as inventory positions evolve. Moreover, Open/Close data directly identify the customer trading activity that generates dealer inventory exposure, whereas LiveVol measures must infer customer demand without observing trader identity. As a result, measurement differences that have only modest effects on directional-demand applications can translate into economically meaningful differences in inferred dealer inventories.

Panel A: Sample of S&P 500 stocks

Bivariate regression using Open/Close versus LiveVol



Panel B: Sample of non-S&P 500 stocks

Bivariate regression using Open/Close versus LiveVol

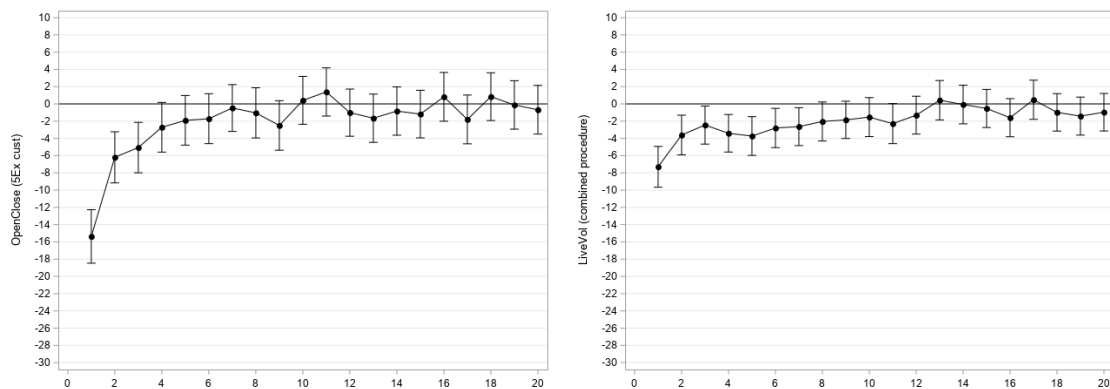


Figure 4.10: Bivariate regressions by data source

The figure reports slope coefficients and 95% confidence intervals from bivariate Fama–MacBeth regressions of risk-adjusted stock returns on put-call ratios constructed from Open/Close customer trades across five exchanges (left) and from LiveVol trades classified using the combined classification procedure (right) over return horizons of up to 20 trading days. Panel A reports results for S&P 500 stocks, and Panel B reports results for non-S&P 500 stocks.

4.6 Comparing Quantity-Based Measures and Price-Based Proxies

Whereas the preceding analyses compare alternative quantity-based measures of option demand and dealer inventory, this section examines option expensiveness as a price-based benchmark. Unlike order-flow measures, expensiveness does not rely on observed trading activity and is therefore not directly exposed to trade classification error, unobserved trader identity, or limited exchange coverage. The comparison is motivated by the demand-based option-pricing literature, in which end-user demand and intermediary risk bearing are reflected in equilibrium option prices (Gârleanu et al., 2009). This allows us to assess whether price-based proxies generate empirical relations with subsequent stock-market outcomes comparable to those obtained from quantity-based measures.

In applications that test the relation between directional-demand measures and future stock returns, quantity-based measures constructed from signed order flow generally provide the stronger signal. Option expensiveness generates little evidence of return predictability among S&P 500 stocks and produces economically meaningful next-day predictability only among non-S&P 500 stocks (Table 4.5). Even in the latter sample, its risk-adjusted performance remains below that of quantity-based measures, and its predictive content largely disappears at weekly horizons (Table 4.6). These findings suggest that signed order flow captures directional information more directly than option prices, which aggregate multiple valuation components beyond directional demand.

The comparison differs when we examine the relation between dealer gamma inventory exposure and future stock volatility. Among S&P 500 stocks, gamma-weighted option expensiveness generates smaller and less precisely estimated volatility spreads than quantity-based inventory measures. Among non-S&P 500 stocks, however, expensiveness produces a volatility spread of 97.8 basis points, substantially exceeding the spreads obtained from quantity-based measures (Table 4.7).

One interpretation is that price-based measures are less sensitive to the measurement constraints involved in inferring dealer inventory from transaction data. Quantity-based inventory measures accumulate signed trades over time, so that measurement errors can persist in the inferred position. In addition, early exercise and other unobserved position adjustments are not captured by transaction-based measures. These frictions can cause inferred dealer gamma positions to diverge from latent dealer inventory exposure. Price-based measures avoid reconstructing

inventory positions trade by trade and instead infer inventory-related pressure from equilibrium option prices.

At the same time, the exceptionally large effects among non-S&P 500 stocks suggest that option expensiveness captures more than dealer inventory exposure. Option prices may also reflect expectations about future volatility, uncertainty, jump risk, and compensation for intermediary risk bearing, all of which may be especially important in less liquid stocks. The evidence therefore indicates that price-based measures are complements rather than substitutes for quantity-based measures. While transaction-based measures provide a more direct estimate of dealer inventory exposure, price-based measures capture a broader equilibrium component associated with demand pressure and intermediary risk bearing.

4.7 Conclusion

Option demand and dealer inventory play a central role in a growing literature on information transmission, intermediary asset pricing, and demand-based option valuation. Yet neither is directly observable. Instead, both must be inferred from data that differ in exchange coverage, trader identity, and trade classification. As a result, empirical measures reflect not only the latent economic objects of interest but also the data constraints that shape their construction.

This chapter provides a systematic evaluation of how alternative measurement approaches affect empirical inference in option markets by isolating three distinct sources of measurement error. The results show that these measurement dimensions matter economically and that their importance depends fundamentally on the object being measured. Directional-demand measures are particularly sensitive to trade classification and exchange coverage, whereas dealer-inventory measures are more sensitive to trader identity and, in less liquid stocks, classification accuracy. Across applications, these effects are substantially larger among non-S&P 500 stocks, suggesting that measurement constraints become increasingly important in more heterogeneous market segments.

The results suggest that the choice between transaction-level and Open/Close data should be guided by the economic question under consideration rather than by a presumption that one data environment uniformly dominates the other. Alternative data environments involve different trade-offs. Transaction-level data provide market-wide coverage but require demand to be inferred through trade classification procedures, whereas Open/Close data directly identify customer trading activity but

observe only a subset of aggregate market activity. For directional-demand applications, improved trade classification procedures and broader exchange coverage each contribute economically meaningful information beyond more restrictive measures. As a result, multi-exchange Open/Close measures and transaction-level measures constructed using improved trade classification recover remarkably similar qualitative relations despite their different construction. The conclusions differ somewhat different for dealer-inventory applications, where trader identification remains particularly valuable because measurement errors can accumulate through time and distinguishing customer demand from intermediary activity is central to the economic interpretation of the signal.

More generally, the results suggest that option demand and dealer inventory should be viewed as latent economic objects rather than directly observed variables. This perspective helps reconcile differences across studies that employ alternative option-market datasets and measurement procedures, as differences in reported effect sizes may reflect measurement variations rather than differences in the underlying economic mechanisms. Understanding how alternative data environments map into these underlying objects is therefore essential for interpreting existing evidence, comparing results across studies, and designing empirical measures that more closely align observed option-market activity with the underlying economic objects of interest.

Chapter 5

Summary and Outlook

This dissertation examines how option trade classification shapes the inference of customer option demand from transaction-level data and how inferred customer demand predicts movements in underlying stock prices.

Chapter 2 evaluates existing trade classification algorithms for inferring customer trade direction in option markets. Using a novel benchmark based on matched intraday transactions and Open/Close volume reports from five option exchanges, we show that established methods have low classification accuracy in option markets. Quote rules perform best among conventional approaches, whereas tick tests are less reliable due to infrequent trading and stale prices. However, quote rules systematically misidentify the customer side when customer demand is expressed through limit orders because they infer the aggressive rather than the customer side of a trade. To address this limitation, we introduce two new rules that substantially improve classification accuracy across exchanges, trading environments, and time, with particularly large gains on maker-taker exchanges.

Chapter 3 analyzes how trade classification affects conclusions about the predictive relation between option demand and future stock returns. Comparing demand measures based on the Lee and Ready (1991) algorithm with measures based on the improved procedure developed in Chapter 2, we show that conventional classification understates the information content of option trading. When customer demand is measured more accurately, option order flow exhibits stronger and more persistent return predictability, while conventional measures lose explanatory power in joint tests. The gains are strongest in less liquid market segments, where limit-order usage is more prevalent and disagreement between classification algorithms is greatest. This pattern suggests that customer limit orders convey economically meaningful information that is obscured by conventional trade classification.

Chapter 4 investigates how empirical inference changes when option demand and dealer inventory are measured from transaction-level data versus exchange Open/Close reports. The two data sources differ in the information available for measuring latent option demand and therefore involve distinct trade-offs. Open/Close reports directly identify customer buying and selling activity but have limited exchange coverage, whereas transaction-level data provide broader market coverage but require trade classification to infer the customer side of each trade. We show that both data sources contain economically meaningful information, but that differences in coverage, trader identity, and trade classification affect measures of option demand, dealer inventory, and their relation to underlying stock returns and volatility. Demand inferred from transaction data contains substantial predictive information for future stock returns and performs similarly to customer demand measured from Open/Close reports. Measures of dealer inventory constructed from both data sources are related to future stock volatility, although the strength of this relation differs across measures. Overall, the results show that neither data source fully dominates the other and that measurement choices materially affect empirical inference in option markets.

Our results prompt a variety of interesting follow-up questions for future research. A first direction is to further improve option trade classification using more flexible prediction methods and richer transaction-level information. The goal of this dissertation was not to maximize classification accuracy with the most complex model, but to develop simple and economically motivated rules that explain why conventional algorithms fail in option markets. As market structure becomes more fragmented and trading protocols more heterogeneous, however, machine-learning methods may become useful for combining a broader set of trade, quote, and market-design characteristics. Future work could, for example, use trade condition identifiers, as in Bryzgalova et al. (2023), together with routing information, exchange fee models, order-flow characteristics, and liquidity measures to better distinguish between trader types, order-submission strategies, and trading environments. Such approaches could improve classification accuracy while also providing new evidence on how market design shapes order-submission behavior.

A second promising direction is to study customer limit-order activity more directly. The findings of this dissertation suggest that passive customer trading contains economically meaningful information, but the determinants of this order-submission choice remain less well understood. Theoretical models of order submission emphasize the strategic choice between execution certainty and execution costs (Foucault, 1999), while empirical work on market design shows that fees, rebates,

and priority rules can affect routing incentives and thereby influence where and how orders are executed (Battalio et al., 2021; Hendershott et al., 2026). Future research could connect these literatures by using identified customer limit-order executions as a proxy for passive customer demand in option markets. This would allow researchers to study when investors choose to express information through limit orders, how this choice varies with liquidity, volatility, information events, and exchange fee structures, and whether passive customer demand contributes differently to price discovery than aggressive customer trading.

A third direction for future research is to develop better measures of latent customer option demand in fragmented markets. Although this dissertation improves trade classification in option markets, raising average accuracy across the sample exchanges from 56.8% to 72.3%, customer trading activity remains imperfectly observed in standard transaction-level data. The results also show that option-demand proxies are sensitive to trade classification, exchange coverage, and trader identity, and that no single data source fully resolves these measurement challenges. Ideally, researchers would have access to market-wide Open/Close reports or transaction-level data with customer trade direction. In the absence of such complete data, measuring market-wide customer option demand remains an open challenge. Future work could address this challenge by combining multiple imperfect signals to estimate latent customer option demand. A joint measurement framework could help isolate the common demand component from variation driven by classification error, exchange coverage, trader identity, liquidity, and option expensiveness. Institutional changes, such as exchange fee changes, new exchange launches, or modifications to routing rules, could provide useful variation for examining how trade classification accuracy, customer limit-order activity, and the relation between transaction-based and Open/Close demand measures vary with market design, as in Battalio et al. (2021). Simulation environments could provide a complementary way to assess how classification errors interact with trader-type composition, order-submission behavior, exchange design, and routing incentives. Together, these approaches could improve the measurement of customer option demand and strengthen empirical tests of its role in stock return predictability.

A further direction for future research is to study how the measurement challenges documented in this dissertation arise in other option markets. This dissertation focuses on U.S. options because the available data make it possible to benchmark trade classification across exchanges and to compare alternative measures of option demand. In many other option markets, including European markets, comparable transaction-level data and trader-type information remain difficult to obtain. At the

same time, the role of measurement is likely to depend on market structure. European option markets, for example, differ from U.S. markets in terms of exchange organization, liquidity, and the degree of market fragmentation. In particular, lower liquidity may make measurement challenges especially important, because wider spreads and less frequent trading can make customer order submission and trade direction more difficult to infer from transaction-level data. This concern is reinforced by the evidence in this dissertation, which shows that the consequences for empirical inference are most pronounced in less liquid and more heterogeneous market segments. If comparable data became available, studying option demand in these markets would therefore provide useful evidence on how market design shapes measurement and whether empirical conclusions drawn from U.S. data generalize to different trading environments.

Bibliography

- Adams, Greg, Jean-Sebastien Fontaine, and Chayawat Ornthanalai, 2025, The market for 0DTE: The role of liquidity providers in volatility attenuation, Working Paper.
- Aitken, Michael, and Alex Frino, 1996, The accuracy of the tick test: Evidence from the Australian stock exchange, *Journal of Banking & Finance* 20, 1715–1729.
- Aktas, Osman Ulas, and Lawrence Kryzanowski, 2014, Trade classification accuracy for the bist, *Journal of International Financial Markets, Institutions and Money* 33, 259–282.
- Amin, Kaushik I, and Charles MC Lee, 1997, Option trading, price discovery, and earnings news dissemination, *Contemporary Accounting Research* 14, 153–192.
- Anand, Amber, Jian Hua, and Tim McCormick, 2016, Make-take structure and market quality: Evidence from the US options markets, *Management Science* 62, 3271–3290.
- Anand, Amber, Mehrdad Samadi, Jonathan Sokobin, and Kumar Venkataraman, 2026, Retail limit orders, *Review of Finance* 30, 459–488.
- Andersen, Torben G, and Oleg Bondarenko, 2015, Assessing measures of order flow toxicity and early warning signals for market turbulence, *Review of Finance* 19, 1–54.
- Anshuman, V Ravi, and Avner Kalay, 1998, Market making with discrete prices, *Review of Financial Studies* 11, 81–109.
- Baltussen, Guido, Zhi Da, Sten Lammers, and Martin Martens, 2021, Hedging demand and market intraday momentum, *Journal of Financial Economics* 142, 377–403.
- Barardehi, Yashar H, Vincent Bogousslavsky, and Dmitriy Muravyev, 2025, Unpacking retail trading costs: the role of options trading and limit order usage, Working Paper.
- Barbon, Andrea, and Andrea Buraschi, 2021, Gamma fragility, Working Paper.

- Battalio, Robert, Todd Griffith, and Robert Van Ness, 2021, Do (should) brokers route limit orders to options exchanges that purchase order flow?, *Journal of Financial and Quantitative Analysis* 56, 183–211.
- Biais, Bruno, Pierre Hillion, and Chester Spatt, 1995, An empirical analysis of the limit order book and the order flow in the Paris Bourse, *Journal of Finance* 50, 1655–1689.
- Black, Fischer, 1975, Fact and fantasy in the use of options, *Financial Analysts Journal* 31, 36–41.
- Black, Fischer, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of political economy* 81, 637–654.
- Bogousslavsky, Vincent, and Dmitriy Muravyev, 2025, An anatomy of retail option trading, Working Paper.
- Bollen, Nicolas PB, and Robert E Whaley, 2004, Does net buying pressure affect the shape of implied volatility functions?, *Journal of Finance* 59, 711–753.
- Bondarenko, Oleg, and Dmitriy Muravyev, 2025, What information do informed traders use?, Working Paper.
- Bryzgalova, Svetlana, Anna Pavlova, and Taisiya Sikorskaya, 2023, Retail trading in options and the rise of the big three wholesalers, *Journal of Finance* 78, 3465–3514.
- Cao, Charles, Zhiwu Chen, and John M. Griffin, 2005, Informational content of option volume prior to takeovers, *Journal of Business* 78, 1073–1109.
- Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.
- Carrion, Allen, and Madhuparna Kolay, 2020, Trade signing in fast markets, *Financial Review* 55, 385–404.
- Cashman, George D, David M Harrison, Benjamin Scheick, and Hainan Sheng, 2025, Reit option volume, retail investors, and return predictability, *Real Estate Economics* 53, 808–840.
- Chakrabarty, Bidisha, Bingguang Li, Vanthuan Nguyen, and Robert A. Van Ness, 2007, Trade classification algorithms for electronic communications network trades, *Journal of Banking & Finance* 31, 3806–3821.
- Chakrabarty, Bidisha, Pamela C Moulton, and Andriy Shkilko, 2012, Short sales, long sales, and the lee-ready trade classification algorithm revisited, *Journal of Financial Markets* 15, 467–491.
- Chan, Kalok, Y. Peter Chung, and Wai-Ming Fong, 2002, The informational role of stock and option volume, *Review of Financial Studies* 15, 1049–1075.

- Demsetz, Harold, 1968, The cost of transacting, *Quarterly Journal of Economics* 82, 33–53.
- Dim, Chukwuma, Bjorn Eraker, and Grigory Vilkov, 2025, 0DTEs: Trading, gamma risk and volatility propagation, Working Paper.
- Dong, Yunjiang, 2025, Cross market price discovery and selective delta hedging in the option market, Working Paper.
- Easley, David, Marcos Lopez de Prado, and Maureen O’Hara, 2016, Discerning information from trade data, *Journal of Financial Economics* 120, 269–285.
- Easley, David, Marcos Lopez de Prado, and Maureen O’Hara, 2012a, Bulk classification of trading activity, *Johnson School Research Paper Series* 8, 14.
- Easley, David, Marcos M López de Prado, and Maureen O’Hara, 2012b, Flow toxicity and liquidity in a high-frequency world, *Review of Financial Studies* 25, 1457–1493.
- Easley, David, Maureen O’Hara, and P. S. Srinivas, 1998, Option volume and stock prices: Evidence on where informed traders trade, *Journal of Finance* 53, 431–465.
- Ellis, Katrina, Roni Michaely, and Maureen O’Hara, 2000, The accuracy of trade classification rules: Evidence from Nasdaq, *Journal of Financial and Quantitative Analysis* 35, 529–551.
- Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of political economy* 81, 607–636.
- Fedenia, Mark, Tavy Ronen, and Seunghan Nam, 2024, Machine learning and trade direction classification: insights from the corporate bond market, *Review of Quantitative Finance and Accounting* 63, 1–36.
- Finucane, Thomas J., 2000, A direct test of methods for inferring trade direction from intra-day data, *Journal of Financial and Quantitative Analysis* 35, 553–576.
- Foucault, Thierry, 1999, Order flow composition and trading costs in a dynamic limit order market, *Journal of Financial Markets* 2, 99–134.
- Foucault, Thierry, Ohad Kadan, and Eugene Kandel, 2005, Limit order book as a market for liquidity, *Review of Financial Studies* 18, 1171–1217.
- Fu, Lei, Li Su, David K Musto, and Neil D Pearson, 2025, Hope at a reasonable price: Customer use of limit orders in the 0dte market, Working Paper.
- Gao, Si, Xiaoxiao Tang, Guofu Zhou, and Zhaoque Zhou, 2025, Inferring trade directions in options via machine learning, Working Paper.

- Gârleanu, Nicolae, Lasse Heje Pedersen, and Allen M. Potesman, 2009, Demand-based option pricing, *Review of Financial Studies* 22, 4259–4299.
- Gayda, Leander, 2025, Inferring the trade direction in option auctions, Working Paper.
- Ge, Li, Tse-Chun Lin, and Neil D Pearson, 2016, Why does the option to stock volume ratio predict stock returns?, *Journal of Financial Economics* 120, 601–622.
- Glosten, Lawrence R, 1994, Is the electronic open limit order book inevitable?, *Journal of Finance* 49, 1127–1161.
- Goettler, Ronald L., Christine A. Parlour, and Uday Rajan, 2005, Equilibrium in a dynamic limit order market, *Journal of Finance* 60, 2149–2192.
- Goldstein, Michael A, and Kenneth A Kavajecz, 2000, Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the NYSE, *Journal of Financial Economics* 56, 125–149.
- Goyal, Amit, and Alessio Saretto, 2009, Cross-section of option returns and volatility, *Journal of Financial Economics* 94, 310–326.
- Goyenko, Ruslan, Chayawat Ornthanalai, and Shengzhe Tang, 2015, Options illiquidity: determinants and implications for stock returns, Working Paper.
- Goyenko, Ruslan, and Chengyu Zhang, 2021, Demand pressures and option returns, Working Paper.
- Grauer, Caroline, Philipp Schuster, and Marliese Uhrig-Homburg, 2026, Unmasking option demand: New classification methods and stock return predictability, Working Paper.
- Hagströmer, Björn, 2021, Bias in the effective bid-ask spread, *Journal of Financial Economics* 142, 314–337.
- Handa, Puneet, and Robert A Schwartz, 1996, Limit order trading, *Journal of Finance* 51, 1835–1861.
- Harris, Lawrence, 1989, A day-end transaction price anomaly, *Journal of Financial and Quantitative analysis* 24, 29–45.
- Hasbrouck, Joel, 1988, Trades, quotes, inventories, and information, *Journal of Financial Economics* 22, 229–252.
- Hendershott, Terrence, Saad Khan, and Ryan Riordan, 2026, Option auctions, *Review of Financial Studies*, forthcoming.
- Hitzemann, Steffen, Michael Hofmann, Marliese Uhrig-Homburg, and Christian Wagner, 2021, Margin requirements and equity option returns, Working Paper.

- Holthausen, Robert W, Richard W Leftwich, and David Mayers, 1987, The effect of large block transactions on security prices: A cross-sectional analysis, *Journal of Financial Economics* 19, 237–267.
- Hu, Jianfeng, 2014, Does option trading convey stock price information?, *Journal of Financial Economics* 111, 625–645.
- Kaniel, Ron, and Hong Liu, 2006, So what orders do informed traders use?, *Journal of Business* 79, 1867–1913.
- Kanne, Stefan, Olaf Korn, and Marliese Uhrig-Homburg, 2023, Stock illiquidity and option returns, *Journal of Financial Markets* 63, 100765.
- Lee, Charles M. C., and Balkrishna Radhakrishna, 2000, Inferring investor behavior: Evidence from TORQ data, *Journal of Financial Markets* 3, 83–111.
- Lee, Charles M. C., and Mark J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733–746.
- Lu, Y-C, and Y-C Wei, 2009, Classification of trade direction for an equity market with price limit and order match: Evidence from the taiwan stock market, *Investment Management and Financial Innovations* 135–147.
- Ma, Donglian, and Pengxiang Zhai, 2021, The accuracy of the tick rule in the bitcoin market, *SAGE Open* 11, 1–13.
- Mayhew, Stewart, Atulya Sarin, and Kuldeep Shastri, 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, 1635–1653.
- Merton, Robert C., 1973, Theory of rational option pricing, *Bell Journal of Economics and Management Science* 4, 141–183.
- Muravyev, Dmitriy, 2016, Order flow and expected option returns, *Journal of Finance* 71, 673–708.
- Muravyev, Dmitriy, and Xuechuan Charles Ni, 2020, Why do option returns change sign from day to night?, *Journal of Financial Economics* 136, 219–238.
- Muravyev, Dmitriy, and Neil D. Pearson, 2020, Options trading costs are lower than you think, *Review of Financial Studies* 33, 4973–5014.
- Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.
- Ni, Sophie X, Jun Pan, and Allen M Poteshman, 2008, Volatility information trading in the option market, *Journal of Finance* 63, 1059–1091.
- Ni, Sophie X., Neil D. Pearson, Allen M. Poteshman, and Joshua White, 2021, Does option trading have a pervasive impact on underlying stock prices?, *Review of Financial Studies* 34, 1952–1986.

Bibliography

- Nowak, Sabina, 2020, The accuracy of trade classification rules for the selected cee stock exchanges, *Springer Proceedings in Business and Economics* 65–75.
- NYSE, 2024, U.S. equity options market models (by Steven W. Poser), *Data insights – New York Stock Exchange*, <https://www.nyse.com/data-insights/us-equity-options-market-models>, accessed July 20, 2026.
- Odders-White, Elizabeth R, 2000, On the occurrence and consequences of inaccurate trade classification, *Journal of Financial Markets* 3, 259–286.
- O'Donovan, James, Gloria Yang Yu, and Jinyuan Zhang, 2023, Option market maker hedging and stock market liquidity, Working Paper.
- Omrane, Walid Ben, and Robert Welch, 2016, Tick test accuracy in foreign exchange ecn markets, *Research in International Business and Finance* 37, 135–152.
- Pan, Jun, and Allen M. Poteshman, 2006, The information in option volume for future stock prices, *Review of Financial Studies* 19, 871–908.
- Pederzoli, Paola, Hitesh Doshi, and Saim Ayberk Sert, 2025, Risky intraday order flow and option liquidity, Working Paper.
- Perlin, Marcelo, Chris Brooks, and Alfonso Dufour, 2014, On the performance of the tick test, *Quarterly Review of Economics and Finance* 54, 42–50.
- Peterson, Mark, and Erik Sirri, 2003, Evaluation of the biases in execution cost estimation using trade and quote data, *Journal of Financial Markets* 6, 259–280.
- Pöppe, Thomas, Sebastian Moos, and Dirk Schiereck, 2016, The sensitivity of VPIN to the choice of trade classification algorithm, *Journal of Banking & Finance* 73, 165–181.
- Rosenthal, Dale WR, 2012, Modeling trade direction, *Journal of Financial Econometrics* 10, 390–415.
- Savickas, Robert, and Arthur J. Wilson, 2003, On inferring the direction of option trades, *Journal of Financial and Quantitative Analysis* 38, 881–902.
- Securities, U.S., and Exchange Commission, 2026, Roundtable on options market structure: Supporting data, *Staff report, Office of Analytics and Research*, <https://www.sec.gov/files/roundtable-options-market-structure.pdf>, accessed July 20, 2026.
- Shleifer, Andrei, and Robert W Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35–55.
- Theissen, Erik, 2001, A test of the accuracy of the lee/ready trade classification algorithm, *Journal of International Financial Markets, Institutions and Money* 11, 147–165.

Bibliography

- Vasquez, Aurelio, Diego Amaya, Neil D. Pearson, and Pedro Angel Garcia-Ares, 2025, 0DTE index options and market volatility: How large is their impact?, Working Paper.
- Vijh, Anand M, 1990, Liquidity of the cboe equity options, *Journal of Finance* 45, 1157–1179.
- Weinbaum, David, Andrew Fodor, Dmitriy Muravyev, and Martijn Cremers, 2023, Option trading activity, news releases, and stock return predictability, *Management Science* 69, 4810–4827.