

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**HIGH-FREQUENCY TRADING DYNAMICS IN TURKISH INDEX FUTURES:
A COMPREHENSIVE ANALYSIS**



Ph.D. THESIS

Onur OLGUN

Department of Management

Management Programme

MARCH 2025

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**HIGH-FREQUENCY TRADING DYNAMICS IN TURKISH INDEX FUTURES:
A COMPREHENSIVE ANALYSIS**

Ph.D. THESIS

**Onur OLGUN
(403182010)**

Department of Management

Management Programme

Thesis Advisor: Assoc. Prof. Dr. Cumhuri EKİNCİ

MARCH 2025

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**TÜRK ENDEKS VADELİ İŞLEM SÖZLEŞMELERİNDE YÜKSEK
FREKANSLI İŞLEM DİNAMİKLERİ ÜZERİNE KAPSAMLI BİR ANALİZ**

DOKTORA TEZİ

**Onur OLGUN
(403182010)**

İşletme Anabilim Dalı

İşletme Programı

Tez Danışmanı: Doç. Dr. Cumhuri EKİNCİ

MART 2025

Onur OLGUN, a Ph.D. student of ITU Graduate School student ID 403182010, successfully defended the thesis/dissertation entitled “HIGH-FREQUENCY TRADING DYNAMICS IN TURKISH INDEX FUTURES: A COMPREHENSIVE ANALYSIS”, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Assoc. Prof. Dr. Cumhur EKİNCİ**
İstanbul Technical University

Jury Members : **Prof. Dr. Burç ÜLENGİN**
İstanbul Technical University

Prof. Dr. Oktay TAŞ
İstanbul Technical University

Prof. Dr. Recep BİLDİK
İstanbul Ticaret University

Assoc. Prof. Dr. Oğuz ERSAN
Kadir Has University

Date of Submission : 07 January 2025
Date of Defense : 11 March 2025





To my eternal beloveds, Tamay and Ada,



FOREWORD

This thesis is the product of extensive research conducted under the expert guidance of Assoc. Prof. Dr. Cumhuri EKİNCİ. His profound knowledge of financial market microstructure and steadfast support allowed me to pursue novel concepts in high-frequency trading, enriching the scholarly field. I feel privileged to have had his mentorship throughout the every stage of the process.

I extend my gratitude to the jury members, Prof. Dr. Burç ÜLENGİN and Assoc. Prof. Dr. Oğuz ERSAN, for their insightful feedback and expert critique. Moreover, I sincerely thank my colleague, Ramazan ARIKAN, whose technical contribution greatly aided the development of the Java programme utilized in the study.

Finally, I would like to dedicate this work to my family—my wife, Tamay, and my daughter, Ada—whose unwavering encouragement and love have been the foundation of my persistence in this endeavor. Thank you for being my greatest inspiration.

March 2025

Onur OLGUN



TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	ix
TABLE OF CONTENTS	xi
ABBREVIATIONS	xiii
LIST OF TABLES	xv
LIST OF FIGURES	xvii
SUMMARY	xix
ÖZET	xxi
1. INTRODUCTION	1
2. LITERATURE REVIEW	5
2.1 Brief Theoretical Background for Major HFT Concepts	5
2.2 HFT Identification.....	11
2.3 HFT – Liquidity Relationship and Tick Size Changes	15
2.4 HFT – Volatility Dynamics.....	18
3. STRUCTURAL OVERVIEW OF MARKET PROPERTIES AND TRADING ECOSYSTEM	23
4. DATA WORKFLOW: ACQUISITION, PROCESSING AND TRANSFORMATION	29
5. THE PERFORMANCE OF SELECTED HIGH-FREQUENCY TRADING PROXIES: AN APPLICATION ON TURKISH INDEX FUTURES MARKET	33
5.1 Background	33
5.2 Data and Variables	35
5.3 Methodology	36
5.4 Empirical Results	37
5.4.1 Descriptive statistics	37
5.4.2 Machine learning metrics.....	39
5.4.3 Regression analysis	40
5.5 Conclusion.....	42
6. IMPACT OF REVERSE DECIMALIZATION ON HFT-DRIVEN LIQUIDITY IN TURKISH INDEX FUTURES MARKET: A COMPARATIVE TICK SIZE ANALYSIS	45
6.1 Scope	45
6.2 Data and Models.....	47
6.2.1 Model 1	48
6.2.2 Model 2	49
6.2.3 Model 3	49
6.2.4 Robustness check	50
6.3 Results and Discussion.....	50
6.3.1 Descriptive statistics	50
6.3.2 Model findings	54
6.3.2.1 Model 1	54

6.3.2.2 Model 2	55
6.3.2.3 Model 3	55
6.3.3 Robustness check	56
6.4 Final Remarks.....	57
7. DO HIGH-FREQUENCY TRADERS FUEL MARKET HAYWIRE WHEN VOLATILITY PEAKS? EVIDENCE FROM THE TURKISH INDEX FUTURES MARKET	59
7.1 Context	59
7.2 Empirical Framework.....	63
7.2.1 Sample.....	63
7.2.2 Variables.....	63
7.2.3 Model formation.....	64
7.3 Findings and Interpretations	66
7.3.1 Summary statistics.....	66
7.3.2 Models results	68
7.3.3 Robustness assessment.....	71
7.4 Concluding Implications	73
8. CONCLUSIVE REMARKS.....	75
REFERENCES.....	79
APPENDICES	89
APPENDIX A	90
APPENDIX B.....	101
APPENDIX C.....	113
CURRICULUM VITAE	115

ABBREVIATIONS

2SLS	: Two-Stage Least Squares
AIC	: Akaike Information Criterion
ARE	: All Runs In Extended Algorithm
ARP	: All Runs In Process Algorithm
AT	: Algorithmic Trading
BBO	: Best Bid and Offer
BIST	: Borsa Istanbul
CCP	: Central Counterparty
EPM	: Extreme Price Movement
EQ	: Equation
ES	: Effective Spread
ETF	: Exchange Traded Fund
FAK	: Fill and Kill Order Validity
FIX	: Financial Information Exchange Protocol
FOK	: Fill or Kill Order Validity
FX	: Foreign Exchange
GL	: Ghost Liquidity
GMM	: Generalized Method of Moments
GTC	: Good-Till-Cancel Order Validity
GTD	: Good-Till-Date Order Validity
HFT	: High-Frequency, or Low Latency, Trading
HFTs	: High-Frequency, or Low Latency, Traders (Investors)
ISE	: Istanbul Stock Exchange
ITCH	: Integrated Trading Clearinghouse Protocol
IV	: Instrumental Variable
LOB	: Limit Order Book
MAE	: Mean Absolute Errors
ML	: Machine Learning
MLH	: Maximum Likelihood
OBS	: Observation

OLS	: Ordinary Least Squares
OTR	: Order To Trade Ratio
PTRM	: Pre-Trade Risk Module
RMSE	: Root Mean Square Errors
ROC	: Receiver Operating Characteristics
SD	: Standard Deviation
TRY	: Turkish Lira
TSD	: Two Step Detection Algorithm
QS	: Quoted Spread
US	: United States
USD	: United States Dollar
VIF	: Variance Inflation Factor
VIOP	: Borsa Istanbul Derivatives Market
VPIN	: Volume Synchronized Probability of Informed Trading
VWAP	: Volume Weighted Average Price
WFE	: World Federation Exchanges

LIST OF TABLES

	<u>Page</u>
Table 3.1 : Global Market Statistics for Top Index Futures.	25
Table 5.1 : Explanatory Notes on the Construction of HFT Variables.	36
Table 5.2 : Summary Statistics for Avg. Daily Nb of Order Messages.	38
Table 5.3 : Summary Statistics for Avg. Daily Trade Volume.	38
Table 5.4 : Daily HFT Participation Estimates of Real HFT and Proxy Measures. .	39
Table 5.5 : Pairwise Correl's. of Real HFT vs. Proxies and Tests of Equality.	39
Table 5.6 : Machine Learning (ML) Parameters.	40
Table 5.7 : OLS Regression Results.	42
Table 6.1 : Trading Patterns in BIST-30 Index Futures.	51
Table 6.2 : Overview of Summary Statistics for Model Variables.	52
Table 6.3 : Correlation Analysis.	53
Table 6.4 : Returns and HFTs Active Volume.	54
Table 6.5 : Spread Dynamics and HFT.	56
Table 6.6 : Relative Tick Size and HFT Activity.	57
Table 7.1 : Selected Trading Dynamics in Major Global Stock Exchanges.	62
Table 7.2 : Descriptive Statistics for Variables.	67
Table 7.3 : Correlation Matrix.	68
Table 7.4 : Volatility and HFT Dynamics.	69
Table 7.5 : Volatility - HFT Relationship based on Market Price Behavior.	71
Table 7.6 : Robustness Results of Alternative Models.	72



LIST OF FIGURES

	<u>Page</u>
Figure 3.1 : Traded Value Progression in VIOP.....	24
Figure 4.1 : HFT ProxyChecker Application Diagram.....	30
Figure 5.1 : ROC Plots of Proxy Measures.....	41
Figure A.1 : Raw Orders (XLSX Screenshot).	91
Figure A.2 : Raw Trades (XLSX Screenshot).	92
Figure A.3 : Data Cleaning Process (XLSX Screenshot).	93
Figure A.4 : HFT Accounts and Users (XLSX Screenshot).....	94
Figure A.5 : Cached Orders (XLSX Screenshot).....	95
Figure A.6 : Cached Trades (XLSX Screenshot).....	96
Figure A.7 : Proxy Rules (XLSX Screenshot).....	97
Figure A.8 : Calculated RHFT and Proxies (XLSX Screenshot).	98
Figure A.9 : Machine Learning Parameters (XLSX Screenshot).	99
Figure A.10 : Printed Aggregated Output (XLSX Screenshot).	100



HIGH-FREQUENCY TRADING DYNAMICS IN TURKISH INDEX FUTURES: A COMPREHENSIVE ANALYSIS

SUMMARY

The emergence of scientific interest in high-frequency trading (HFT) dates to the 1980s as global exchanges adopted fully electronic order book systems. However, academia began scrutinizing HFT's impact more intensely after the 2010 Flash Crash in US, where it was partly blamed as a systematic risk factor in the market, associated with volatility spikes and liquidity drain. Since then, technological advancements have underscored the importance of speed in trading, establishing HFT as a pivotal research focus.

This thesis aims to reveal the mechanics of HFT in Turkish index futures by examining its economic implications for market quality across multiple dimensions. In this manner, structured as a self-contained, theme-based framework, the study includes three individual chapters that align with the primary research objectives and have some interconnected aspects. Although each part retains its own distinct motives and novel insights, the thesis makes a broader contribution to the literature by expanding the product scope vertically beyond the traditionally dominant equity asset class.

The dataset, granted under special consent by Borsa Istanbul (BIST), contains tick-by-tick order and trade series for derivative securities, with a unique feature: designated HFT-flags derived from the encrypted information about self-declared low-latency accounts at each electronic message instance. As such, it enables direct filtering of these investors (HFTs) alongside relevant variables, which mitigates reliability concerns and eases the burden often attributed to indirect HFT measurement in the existing literature. Most related studies struggle to access such comprehensive data.

Chapter 1 (Section 5), titled “The Performance of Selected High-Frequency Trading Proxies: An Application on Turkish Index Futures Market”, intends to provide evidence for how well HFT proxies capture low-latency activity in the market. The suggested identification algorithms are applied to order and trade message data in order to produce HFT estimates of the models. These estimates are then contrasted with exchange-provided classification tags, which are considered as accurate representations of real HFT messages. Through this comparison, the soundness and consistency of the proxies are critically evaluated in line with various reference metrics, adopting a rigorous statistical procedure. The findings indicate that certain proxies more effectively determine low-latency trading behavior than others, thereby reinforcing their credibility as robust indicators of HFT.

Chapter 2 (Section 6), titled “Impact of Reverse Decimalization on HFT-Driven Liquidity in Turkish Index Futures Market: A Comparative Tick Size Analysis”, investigates the effects of a minimum tick size increase, prompted by the reverse decimalization of the BIST-30 index, on HFT liquidity provision and trading activity in the Turkish index futures market. Despite no change in the monetary value of the smallest tick, due to adjustments in underlying value and contract size, the minimum

price tick for the instrument expanded tenfold. This amendment offers a rare opportunity to inquire HFT behavior, with an emphasis on both minimum and relative tick size factors. The results demonstrate that HFT activity negatively affects market liquidity as shown by a depletion in tradeable volumes and a concurrent widening of effective spreads. The interaction strengthens post the minimum tick size rule change. Moreover, it is highlighted that there is a meaningful relationship between HFT order tendency and relative tick size variables, which is confirmed across both treatment and control models.

Chapter 3 (Section 7), titled “Do High-Frequency Traders Fuel Market Haywire When Volatility Peaks? Evidence from the Turkish Index Futures Market”, explores the influence of HFT on the price volatility of BIST-30 index futures. The study initially analyzes the interaction between low latency trading and volatility across the entire sample. It then assesses whether institutional HFTs exacerbate market stress during turbulent periods, potentially disadvantaging unsophisticated non-HFT retail investors. The reported estimates display that HFT amplifies volatility on trading days with negative returns in stable market conditions. Yet, no similar impact is observed in times of heightened market stress and extreme volatility.

To conclude, the thesis enhances the understanding of recently surged HFT dynamics in the BIST index futures market by utilizing a multi-layered empirical approach. The resulting propositions not only strengthen the theoretical foundation of this field but also carry practical implications from both business and regulatory viewpoints.

TÜRK ENDEKS VADELİ İŞLEM SÖZLEŞMELERİNDE YÜKSEK FREKANSLI İŞLEM DİNAMİKLERİ ÜZERİNE KAPSAMLI BİR ANALİZ

ÖZET

Yüksek frekanslı işlemler tanım gereği gelişmiş algoritmalar üzerinden saniyenin binde birine varan ultra hızlar eşliğinde, piyasalardaki en küçük fiyat dalgalanmalarından kar elde etmek amacıyla yapılan çok sayıdaki alım satım işlemi olarak ifade edilmektedir. Bu işlem türü milisaniye ölçeğinde düşük gecikme süresi sağlayan kapsamlı bir teknolojik altyapı ile üst düzey finansal bilgi ve deneyim gerektirdiğinden çoğunlukla kurumsal tabanlı yatırım firmaları tarafından uygulanmaktadır. Başlıca yüksek frekanslı işlem stratejileri piyasa yapıcılık, arbitraj ve momentum sacayakları, üzerinde şekillenirken hız avantajıyla gönderilen çok sayıdaki alım satım emrinin olumlu taraftan piyasadaki likiditeye katkı sağlayabileceği gibi olumsuz taraftan ani fiyat dalgalanmalarına yol açarak oynaklığı da arttırabileceği değerlendirilmektedir.

Yüksek frekanslı işlemlere yönelik bilimsel ilginin ortaya çıkışı küresel borsaların tamamen elektronik emir iletimini benimsediği 1980’li yıllara dayanmaktadır. Fakat akademi, bu işlemlerin etkilerini daha yoğun bir şekilde araştırmaya, ABD borsalarında 2010’da yaşanan algoritmik sistemlerin bozulması neticesinde piyasalarda gözlenmiş sıçrayıcı nitelikteki volatilité artışı ve likidite daralmasıyla ilişkilendirilen “Flash Crash” olayında, yüksek frekanslı işlemlerin kısmi bir sistematik risk faktörü olarak algılanmasından sonra başlamıştır.

Yüksek frekanslı işlemlerin finansal piyasalar üzerindeki etkinliğinin ivmelenerek süregelen teknolojik gelişmeler ve değişim gösteren yatırımcı ihtiyaçları doğrultusunda kararlı bir şekilde arttığı gözlemlenmektedir. Bu bağlamda, yüksek frekanslı işlemlerin piyasa oluşumu için marjinal faydaya sahip olup olmadığı genel önermesi şiarında çok sayıda araştırmacı olgunun özelliklerini ortaya koymayı ve likidite, oynaklık ve işlem maliyetleri gibi piyasa kalitesi göstergeleri ile etkileşimini incelemeyi amaçlamıştır. Ancak yüksek frekanslı işlemlerin piyasa katılımcıları üzerindeki net etkisi hala muğlak ve tartışmalı olup sosyal değeri arttırıp arttırmadığına dair bir görüş birliği henüz oluşmamıştır.

Bu tez, yüksek frekanslı işlemlerin Türk endeks vadeli işlem piyasasındaki mekaniklerini ortaya koymayı amaçlayarak, piyasa kalitesi üzerindeki etkilerini derinlemesine ve çok boyutlu olarak analiz etmektedir. Tematik bir çerçeve kapsamında yapılandırılmış çalışma, ana araştırma hedefleriyle uyumlu ve dolaylı olarak birbirleriyle bağlantılı üç ayrı bölüm (bildiri) ihtiva ederek sonuçlandırılmıştır. Her bölüm bilimsel motivasyon unsurları yanında özgün bulgu ve çıkarımlarını bireysel olarak kendi içinde bulundursa da tez, ürün kapsamını, vadeli işlem sözleşmeleri ile, dikey bir biçimde genişletip çoğunlukla konu edinilen hisse senedi varlık sınıfının ötesine taşımakta ve üst perdede literatüre geniş bir katkı sunmaktadır.

Çalışmada kullanılan veri seti Borsa İstanbul (BIST) özel izni ile edinilmiş olup incelemeye konu vadeli işlem sözleşmelerine ait 02 Ocak 2019 – 31 Aralık 2021

tarikhleri arasındaki 741 işğünündeki milisaniye ölçekli tüm alım satım faaliyetini yansıtan emir ve işlem elektronik mesajlarını içermektedir. Bu minvalde, yaklaşık 342 milyon emir ve 95 milyon ham işlem mesajı uygun süreçlerden geçirilip işlenmiş ve hesaplamalara hazır hale getirilmiştir. Temizlenen emir ve işlem zaman serileri bünyelerinde ID, tarih, aralık, sözleşme, açılış, kapanış, uzlaşma, en iyi alış, en iyi satış, toplam mesaj sayısı, işlem sayısı, işlem hacmi, işlem parasal değeri gibi tanımlayıcı nitelikteki kolonlardan ilgili olanları bulundurmaktadır. Bu noktada veri setini özel kılan temel unsurdan bahsetmek önem taşımaktadır. Öyle ki, Borsa'ya yapılan gönüllü beyana dayalı olarak belirlenmiş yüksek frekanslı işlem hesap sahibi belirteci her bir emir ve işlem mesajı üzerine kimlik şifrelemesi ile önceden atanmıştır. Bu sayede, herhangi bir emir veya işlem mesajının yüksek frekanslı işlem gerçekleştiren hesaptan mı yapıldığı bilgisinin direk olarak kaynaktan ayrıştırılması sağlanmış ve bu işlem türüne ait yeni tanımlayıcı istatistiksel metrikler de öncül nihai veriye eklenmiştir.

Kapsam dahilinde çalışan akademisyenlerin karşılaştığı en önemli zorluğun, bu tezde sahip olunan, yüksek frekanslı işlemlerin doğrudan filtrelenmesine imkan sunacak detaylı veriye erişimde yaşandığı vurgulanmaktadır. Araştırmacılar bu engeli aşabilmek için seçilen (halka açık) emir ve işlem parametrelerini girdi olarak kullanan çeşitli vekil göstergeler geliştirmiş ve bunlarla yapılacak hesaplamalar üzerinden piyasa yüksek frekanslı işlem aktivitesine yakınsadıklarını varsaymıştır. Diğer taraftan ilişkili literatürde, endirekt ölçüm yöntemlerin doğruluğu ve tutarlılığına ilişkin kaygıları ortaya koyan azımsanmayacak sayıda bilimsel yayın mevcuttur. İşbu çalışma, yukarıda belirtilen veri seti özelliği sayesinde bu sorunların üstesinden gelmekte ve yüksek frekanslı işlem etkilerine haiz sonuçların güvenilirliğini sağlamlaştırmaktadır. Ancak, tezdeki veri setinin içerdiği Borsa tarafından tanımlanmış yüksek frekanslı işlem aittliğine dair sınıflandırmanın gönüllü beyan temelinde yürütülmesi sebebiyle, bu değişkenin istatistiksel açıdan 1. ve 2. tür hatadan muaf olmayacağı hususu çalışma kısıtı olarak saklı tutulmaktadır. Teknik manada detaylı açıklanırsa, bir takım yüksek frekanslı işlem hesaplarının gizli kalmayı tercih edip beyanda bulunmayabileceği ya da beyanda bulunmuş bir takım hesapların yüksek frekans işlem faaliyetlerini geçici ya da kalıcı olarak sürdürmeyebileceği olasılıkları çıkarımda bulunurken göz ardı edilmemelidir.

Çalışmanın, “Seçilmiş Yüksek Frekanslı İşlem Vekil Göstergelerinin Performansı: Türk Endeks Vadeli İşlemler Piyasası Üzerine Bir Uygulama” başlıklı, birinci kısmı (Bölüm 5) akademide yaygın olarak kullanılan en bilindik yüksek frekanslı işlem vekil göstergelerinin bu işlem türüne ait piyasa aktivitesini hangi etkinlikte belirleyebildiğini ortaya koymayı amaçlamaktadır. Buna göre, önerilen ve sıklıkla başvuru alan beş yüksek frekanslı işlem vekil göstergesi seçilmiş ve bunlara ait tanımlama algoritmaları emir ve işlem mesaj verileri üzerinde çalıştırılarak her biri için yüksek frekanslı işlem tahminlerine ulaşılmıştır. Ardından, bu tahminler çeşitli istatistiksel teknikler ve kıstaslar ışığında, gerçek yüksek frekanslı işlem mesajlarının doğru bir temsili olarak kabul edilen Borsa tarafından ayrıştırılmış referans aittik verisi ile karşılaştırılarak değerlendirilmiştir. Ulaşılan sonuçlar, bazı vekil göstergelerin yüksek frekanslı işlem davranışlarını saptamada diğerlerine oranla daha başarılı olduklarını ve böylelikle güvenilirliklerini pekiştirdiğini göstermektedir.

İkinci kısım (Bölüm 6), “Türk Endeks Vadeli İşlem Piyasasında Ters Desimalizasyonun Yüksek Frekanslı İşlem Odaklı Likidite Üzerindeki Etkisi: Karşılaştırmalı Bir Fiyat Adımı Analizi” başlıklı çalışmaya ayrılmıştır. 26 Temmuz 2020 tarihinde BIST-30 fiyat endeksinde yapılan sıfır atılması (ya da ondalık

küçültme) operasyonu dahilinde bu dayanak varlığı kullanan endeks vadeli işlem sözleşmeleri için minimum fiyat adımı değiştirilmiştir. Çalışma, piyasa kurallarında yapılan bu değişikliğin yüksek frekanslı işlemler tarafından sağlanan likidite faaliyetleri özelindeki etkilerini incelemektedir. Bu düzenleme ile dayanak varlık ve sözleşme büyüklüğünde bir takım ayarlamalar yapılmış ve minimum fiyat adımı on katına çıkarılmıştır. Açık pozisyonlara ve geçmiş işlem hacmi istatistiklerine ilişkin olumsuzluk veya sapma yaşanmaması adına, sözleşmelerin parasal değerinin aynı bırakılması sağlanmıştır. Bahse konu değişiklik, yüksek frekanslı işlem likidite davranışlarını hem minimum hem de göreceli fiyat adımı faktörleri açısından değerlendirebilmek için nadir bir fırsat sunmaktadır. Raporlanan bulgular, yüksek frekanslı işlemlerin piyasa likiditesini iki açıdan olumsuz etkilediğini ortaya koymaktadır. Arz edilen hacim miktarı düşmüş ve efektif alım satım marjları açılmıştır. Etkileşimin minimum fiyat adımı kural değişikliğinden sonra daha da güçlendiği tespit edilmiştir. Ayrıca, yüksek frekanslı işlem emir eğilimleri ile göreceli fiyat adımı değişkenleri arasında istatistiksel olarak anlamlı bir ilişkinin olduğu, sırasıyla endeks ve altın sözleşmeleriyle kurulmuş uygulama ve kontrol modelleri üzerinde doğrulanmıştır.

“Volatilitenin Zirve Yaptığı Dönemlerde Yüksek Frekanslı İşlemler Piyasayı Dengesizleştiriyor mu? Türk Endeks Vadeli İşlem Piyasasından Kanıtlar” başlıklı üçüncü kısımda (Bölüm 7), BIST-30 endeks vadeli işlem sözleşmelerinde yüksek frekanslı işlemler ve piyasa volatilitesi arasındaki korelasyon sorgulanmaktadır. Değişkenler arasındaki etkileşim, öncelikle tüm örneklem üzerinden analiz edilmektedir. İkinci kısımda ise yüksek frekanslı işlemlerin çalkantılı dönemlerde piyasa stresini artırıp arttırmadığı ve bunun, sofistike olmayan bireysel yatırımcılar aleyhine bir durum oluşturup oluşturmadığı araştırılmıştır. İstatistiksel modellerin tahminleme sonuçları neticesinde, yüksek frekanslı işlemlerin negatif getiri görülen istikrarlı piyasa koşulları için volatilitiyi arttırdığı fakat yüksek piyasa stresine haiz aşırı volatilité dönemlerinde benzer bir etkiyi tetiklemediği bulunmuştur.

Sonuç olarak, tez, BIST endeks vadeli işlem piyasasında hızla artan yüksek frekanslı işlem dinamiklerini çok katmanlı bir ampirik yaklaşımla ele alarak bu konuya yönelik anlayışı derinleştirmektedir. Elde edilen bulgular, yalnızca bu alanın teorik temellerini güçlendirmekle kalmayıp, sektör ve düzenleyici perspektifler açısından da pratik çıkarımlara sahiptir.



1. INTRODUCTION

High-frequency trading (HFT) refers to executing vast numbers of orders at remarkable speeds using advanced technologies and intricate algorithms. Operating within milliseconds, HFT firms employ sophisticated computational systems to exploit minimal price discrepancies. As technological progress reshapes the global investment landscape, HFT has emerged as a dominant force in modern trading environments. However, a tenacious challenge lies in accurately identifying and analyzing the orders and trades associated with HFT participants, underscoring the inherent complexity and critical importance of this research area.

HFT is often characterized by its opaque nature within financial ecosystems. Consequently, numerous studies have sought to examine its multifaceted dimensions thus far. The long-standing debate about whether HFT enhances or undermines market structure continues to engage scholars and practitioners alike. In particular, efforts to assess the impact of HFT on market quality indicators such as liquidity, volatility, and transaction costs remain pivotal to academic and industry research. Furthermore, the ongoing discourse has steadily intensified as the race for reduced latency and greater technological innovation accelerates. Yet, the ambivalence surrounding HFT's role—whether as a contributor to or detractor from market stability and social value—persists, leaving unanswered questions about its broader social and economic implications.

This thesis seeks to uncover the dynamics of HFT in the Turkish index futures market. It offers an in-depth and multifaceted analysis of the phenomenon's effects on various aspects of market quality. Aligned with the purpose, the study is framed by three interconnected chapters, as outlined below, which are presented in the form of separate researches. Each chapter contains its own research objectives, unique approaches, and novel implications, while adhering to the overarching theme of the thesis.

The first chapter (Section 5) intends to evaluate the performance of certain widely applied HFT proxies. Initially, the algorithms of selected proxies are implemented to generate binary identification estimates for indirect methods. These are subsequently

compared against direct HFT information from the Exchange (i.e., real HFT) employing various accuracy metrics in an empirical framework. Moreover, both direct and indirect methods are jointly scrutinized to elucidate their viability and predictive power, with a focus on their compatibility with stylized facts.

In the next chapter (Section 6), the effects of tick-size increase (or reverse decimalization) on HFT-driven liquidity are analyzed. Using the rule change date as reference point, the full sample is divided into ex-ante and ex-post subsamples. The bidirectional interactions between HFT activity and liquidity in each period are then investigated and contrasted through regression models. Additionally, the relative tick size factor is incorporated to explore whether it can be adopted as a supplementary explanatory variable in predicting HFT order tendency.

The final chapter (Section 7) aims to reveal the causality between HFT and volatility during stable and extreme market conditions. The study first examines the relationship between low-latency trading and volatility across the full sample. Next, it considers whether institutional HFTs exacerbate market stress during turbulent periods, thereby disadvantaging non-HFT retail investors.

The contributions, novelties, and key premises of the thesis should be emphasized from multiple perspectives. Starting with a broader viewpoint, it enriches the existing literature both vertically and horizontally by addressing the under-researched asset class of futures and the rarely studied emerging market of Türkiye. Given the recent advancements, the BIST platform has become one of the leading trading venues in the region among its peers and thus warrants greater attention. Nevertheless, the thesis has a distinct strength in working with a unique dataset, provided with the Exchange's special consent, which enables the precise sorting of transactions, bypassing the reliability issues and efforts associated with indirect HFT identification.

On an individual basis and discretely, to the best of our knowledge, there is no prior inquiry offering a quantitative performance assessment of HFT proxies similar to the approach in the first chapter (Section 5). It is essential to validate the propositions and practicality of analogous research in the field that relies solely on proxies to measure HFT in the market. The second chapter (Section 6), on the other hand, sheds light on the consequences of a highly exceptional case—a minimum price tick widening instead of a narrowing— in financial markets, while maintaining attention on the

interplay between HFT and liquidity. Finally, the observations and justifications of Chapter 3 (Section 7) entail critical risk management insights for business practitioners and policymakers, including recommendations for establishing appropriate regulative boundaries for this type of investor.

The remainder of the thesis is structured as follows. Section 2 conducts an in-depth analysis of the existing literature. Section 3 presents a brief overview of the market properties and the trading ecosystem. Section 4 describes the workflow involved in preparing the data series up to the final stage. Chapters 1, 2, and 3, mentioned earlier, are covered in Sections 5, 6, and 7, respectively. Section 8 wraps up the concluding remarks of the thesis.





2. LITERATURE REVIEW

2.1 Brief Theoretical Background for Major HFT Concepts

HFT has profoundly revolutionized the landscape of financial markets since the 2000s by harnessing innovative technologies and computerized systems. Characterized by sophisticated algorithms and ultra-fast execution speeds, the phenomenon has evolved into a game-changing force within the global trading universe. Today, HFT's reach and market penetration are unprecedentedly high. Due to its ever-growing dominance, HFT has attracted considerable academic interest across a wide spectrum of microstructure research areas.

Before proceeding further, it is essential to articulate a precise description of HFT, which has been a focal point of early academic inquiries. As a subset of algorithmic trading (AT), HFTs differ from other traders because of their use of automated technology for processing information and trading quickly besides submission of high numbers of orders and/or quotes (Haldane, 2011; Foucault et al., 2013; O'Hara, 2015 and Pagnotta and Philippon, 2018). Research by Hendershott et al. (2011) and Boehmer et al. (2012) indicate that arbitrage strategies within AT significantly enhance price formation. Through a structural model, Foucault et al. (2013) report that HFTs receive and act on information marginally ahead of the broader market. In this vein, Carrion (2013) and Brogaard et al. (2014) demonstrate that such traders can predict short-term price movements over horizons less than 3 to 4 seconds. Moreover, HFT activity correlates with two key sources of public information: macroeconomic announcements (Jovanovic and Menkveld, 2016) and order book imbalances (Cao et al., 2009). Serrano (2020) contends that the term "high-speed" merits particular emphasis, as millisecond-level intervals should reflect very rapid timeframes beyond human comprehension.

At the outset, numerous scientific endeavors have concentrated on situating HFT within theoretical constructs and appraising its social welfare repercussions. Hoffmann (2014), Biais et al. (2015), Foucault et al. (2016), Jain et al. (2016), Menkveld and

Zoican (2017), Pagnotta and Philippon (2018) and Baldauf and Mollner (2020) represent novel works that have shaped the foundational modeling of HFT.

The primary topics associated with the HFT motive in financial literature have so far been categorized, though not exhaustively, into price discovery (efficiency); market quality (liquidity, transaction costs, and volatility); systemic risks and market (flash) crashes; as well as regulatory frameworks. Using an inductive approach, the following part summarizes the oft-cited works of scholars belonging to each of these groups.

One of the earliest and most widely investigated aspects of HFT is its influence on price discovery. Supporting foundational studies by Aitken et al. (2012), Brogaard et al. (2014) and Biais et al. (2015) underscore that HFT facilitates price discovery by integrating information into market prices more swiftly than conventional trading mechanisms. Their work asserts that increased low-latency activity mitigates price inefficiencies and fosters market transparency. According to Brogaard (2010), HFT trades furnish investors with superior access to private information relative to non-HFT trades. Blocher et al. (2016) reveal that HFTs prioritize information processing through cancellations rather than executions, which the authors argue is more effective for price discovery as it avoids the need for monetary exchanges. Despite the predominant view in prior research favouring HFT's positive impact on price efficiency, controversies persist, with Linton and O'Hara (2012) and Hasbrouck and Saar (2013) questioning its welfare implications.

Market quality effects, another prominent theme in HFT research, cover key aspects such as liquidity, transaction costs, and volatility. Yet, the lack of a consensus in the literature fuels ongoing debates regarding whether HFT's stabilizes or destabilizes these features. Brogaard (2010), Friederich and Payne (2011), Menkveld (2013), Hagstromer and Norden (2013) and Chaboud et al. (2014) advocate the view that HFTs are advantageous liquidity providers who, on balance, tend to supply more liquidity than they take from the market. Barker and Pomeranets (2011) claim that the liquidity-enhancing impact of HFT is closely tied to the rise of trading volumes across global trading venues. On the other hand, opponents like Hasbrouck and Saar (2009), Kirilenko and Lo (2013), Arnoldi (2016) and Jovanovic and Menkveld (2016) acknowledge that although HFTs' multi-platform quotes may initially appear to add depth, they are often withdrawn promptly once the first quote is executed. As a result, the market is misinformed with price levels that do not actually exist, which is called

as “Ghost Liquidity” (GL). Such withdrawals may cause temporary liquidity drops in the short run, but if they persist, they can lead to greater disruptions in market dynamics, as other participants' willingness to trade diminishes (Ersan et al., 2021). Degryse et al. (2021) explore GL traits of HFTs in distinct fragmented markets. They show that HFTs have the highest GL rate among all type of traders, with the rate being particularly higher for large market-cap stocks but lower in more volatile markets. Apart from GL discussion, critics such as Cvitanic and Kirilenko (2010), Easley et al. (2011), and Baron et al. (2012) emphasize HFT's adverse influence on liquidity. They suggest that HFT is more inclined to take liquidity rather than supplying it contrary to commonly shared perception.

Foucault et al. (2013) document that algorithmic trading reduces monitoring burden due to a heightened trading rate. Similar findings on how HFT activity decreases both direct and indirect transaction costs for institutional and retail investors are also stated by numerous researchers, including Harris (2013), Menkveld (2013), Conrad et al. (2015), and Baron et al. (2019). In pioneering research, Menkveld (2016) affirms the general opinion by presenting new evidence and comparing both explicit and implicit transaction costs as identified by various researchers. Stocks listed on NASDAQ and NYSE in 2001 (presumably with low or no HFT activity) and 2011 (certainly featuring significant HFT) exhibit at least a 50% decline in all four examined measures: effective spread, commissions to retail investors, commissions to institutional investors and implementation shortfall.

Considering the bid-ask spread as a benchmark of market quality, Brogaard (2010), Friederich and Payne (2011) and Menkveld (2013) express that HFTs often quote the best bid and ask, which narrows the spread. As regards, Menkveld and Zoican (2017) introduce a model that simultaneously incorporates liquidity-provider and liquidity-taker HFTs. By regressing the adverse selection parameter of the spread aggregated across stocks, they highlight a reasonable positive improvement in the spread, approximately 7%, after a reduction in market latency. Hasbrouck and Saar (2013) further confirm that intensified low-latency activity corresponds to reduced posted and effective spreads in NASDAQ during the period from Q4 2007 to Q2 2008. Though, Brogaard (2010) maintains that HFTs prefer excessive trading in tighter spreads, which eventually causing them to widen. Hendershott and Mouton (2011) offer another contradictory perspective. They establish that the decline of execution time (from 10

second to 1 on average) in US markets expands the bid-ask spreads as a consequence of the rise in adverse selection.

Ding et al. (2014) posit that less than 2 milliseconds lasting order book data feeds penalize active slow traders consisting of institutional investors. Additionally, as noted by Hoffmann (2014) and Hirschey (2018), the negative front-running implications of HFTs may harm other market participants since their submitted orders face a lower execution probability, raising concerns about trading costs. Biais and Foucault (2014) similarly point out that HFTs inflict higher transaction costs on slower traders for their own gain. The authors justify the conclusion by underscoring HFTs' privileged access to valuable information ahead of non-HFTs, which enlarges their exposure to adverse selection. Van Kervel (2015) corroborates that the speed advantage of HFTs enables them to outperform human traders, who frequently trade against outdated quotes in a market where prices change very rapidly. Nevertheless, Brogaard (2010) explains that HFT's impact on transaction costs remains negligible around macroeconomic news announcements. Their findings assert that as the number of HFTs grows exponentially in the market, the likelihood of matching with each other naturally extends, resulting in a zero-sum scenario for these traders. To put differently, the profits of one low-latency trader translate into losses for another.

The last market quality feature, volatility, has underpinned a vast amount of academic work in the existing HFT literature from diverse viewpoints. Gsell (2008) postulates that HFT engagement curtails volatility within a hypothesized market structure combining informed, momentum and noise traders. Zervoudakis et al. (2012) state that there is no meaningful causality between HFT participation and the rise in volatility. Aldridge (2014) arrives at a more skeptical interpretation, suggesting that high volatility does not necessarily provoke a crash. And the difference between a crash and strong noise is often subtle. Moreover, Linton and Mahmoodzadeh (2018) intend to analyze whether HFT induces volatility or if volatility attracts HFT. They fail to yield a definite outcome in the empirical context. Hasbrouck and Saar (2013) find that increased HFT activity is linked to lower short-term volatility. The pattern is evident in both stable and stressful market conditions. Hagstromer and Norden (2013) contend that HFT, regardless of its approach—market-making or opportunistic—lessens intra-day price volatility. Myers and Gerig (2014) draw parallel conclusions, validating HFT's volatility-reducing effect through an agent-based simulation approach.

However, Zhang (2010) observes a divergent result, identifying a positive relationship between HFT and volatility, particularly in large-cap stocks. In addition, the studies by Brogaard (2010) and Chaboud et al. (2014) demonstrate that HFTs apply identical algorithmic strategies, which could lead to abnormal price fluctuations driven by herding behavior. Extending the product type to futures, Bollen and Whaley (2015) endorse the shared view that HFT activity negatively affects volatility.

Another central domain of HFT scholarly attention involves its conceivable connection with systemic risks and market (flash) crashes. The first systemic risk factor points to adverse selection in fragmented markets assuming HFTs as fast informed traders (Carrion, 2013; Brogaard et al., 2014; Biais et al., 2015; and Foucault et al., 2016). Consequently, an asymmetric information structure is emerged between HFT and non-HFT participants. The interaction between HFT and non-HFT market-makers accentuates the adverse selection risks. Slower non-HFT market makers are at risk of trading against more informed and faster HFT counterparts. This situation is called as “order toxicity” (Easley et al., 2012) or “toxic arbitrage” (Foucault et al., 2017). Chakabarty et al. (2020) substantiate the relationship among HFT and adverse selection in a contemporary analysis. Menkveld (2013) delves into the role of HFT firms as pseudo-market makers. He underscores their ability to provide liquidity under normal conditions but also their tendency to worsen liquidity shocks during crises. Easley et al. (2011), Huh (2014) and Malceniace et al. (2019) reveal that in stressful market conditions, where information asymmetry widens, HFTs’ liquidity provision reasonably declines. Such a reduction could deepen the extent of turbulence risks.

HFT’s second risk factor emanates from closely aligned strategies that could foster herd behavior. The evidence of high correlation and common statistical patterns in HFT order flows, transactions and strategies are documented by numerous authors in the literature, including Chaboud et al. (2014), Lafarguette (2016), Boehmer et al. (2018) and Malceniace et al. (2019). Intuitively, HFT’s increasing penetration in global markets gradually erodes the human role in trading. This shift has been promoting more homogenous, algorithm-centric, investment decisions over time. As outlined by Serrano (2020), the discussed vulnerability impacts systemic exposure in two ways: (i) a shock affecting the solvency of multiple HFT firms simultaneously could heighten counterparty risk and contagion, and (ii) convergence among HFT

strategies might provoke overreaction to particularly unexpected new information by producing abnormally high volatility in prices with a more transient and severe nature.

Biais and Foucault (2014) and Jain et al. (2016) convey that overreaction to new information and herding could be associated with greater risk of tail events or market (flash) crashes. The disruption of cross-market arbitrage activity intensifies the fragility of markets and raises the likelihood of such crashes (Menkveld and Yueshen, 2018). Johnson et al. (2013) undertake a thorough investigation of all recorded crashes and spikes in the US equity markets from 2006 to 2011. Their findings highlight a clear interaction between the onset of algorithmic trading and the occurrence of these events. Kirilenko et al. (2017) explore HFT's role in the notorious 6 May 2010 Flash Crash. They present concrete evidence of how low-latency trading facilitated extreme price movements during the collapse through rapid order cancellations. Scholars indicate that the inventory changes of HFTs are positively correlated with contemporaneous prices, whereas the co-movement direction is negative for non-HFT market makers. Thus, if some traders (i.e., HFTs) can react more quickly than others (i.e., non-HFTs) to a signal, they may exploit "latency arbitrage" to the disadvantage of slower traders. With a further analysis on 6 May 2010 Flash Crash event, Easley et al. (2011) underline the liquidity-suppressing impact of amplified "order toxicity" phenomena (Easley et al., 2012) throughout the duration of the event. In this context, a significant portion of trades originates from informed low-latency traders, while non-HFT market makers incur losses. As the adjustment process unfolds, non-HFT market makers are more likely to seek liquidity than to supply it. This swing exacerbates volatility and unsteadiness inevitably particularly during market turmoil events like the 6 May 2010 Flash Crash.

Finally, the regulatory frameworks, policy considerations and supervisory mechanisms governing HFT have also been subject to detailed scrutiny in academic discourse. Legislative authorities worldwide have enacted a range of measures such as, minimum resting times, circuit breakers and order-to-trade ratio thresholds and penalties to manage HFT-related risks. The interplay between continually evolving HFT strategies and regulatory constraints remains a fluid area of inquiry, with no clear consensus. Alongside ongoing discussions, researchers offer mixed evaluations regarding the effectiveness of implemented regulations and designed structures to strike a balance between market efficiency and systemic stability. Haldane (2011)

recommends establishing “resting rules,” which impose restrictions on trading speed, as a safeguard against scenarios akin to the 6 May 2010 Flash Crash. This initiative would ease the burden of adverse selection on non-HFT market makers. In alignment, Ye et al. (2013) confirm that speed-limiting policies could yield advantages across all market participants, HFTs included. Hu (2018) and Chakabarty et al. (2018) advocate the use of speed bump applications in favor of observed improvements in price discovery and declines in trading expenses.

Circuit breakers are frequently proposed as an alternative stabilizing tool to cap speed of trading during periods of financial stress (Haldane, 2011). Still, critics like Hautsch and Horvath (2019) allege that there is a regulatory trade-off exists between the protective role of circuit breakers and their negative impact on market quality. These tools often fail to break volatility spirals and resolve price uncertainties, with their positive effects being offset by their detrimental influence on market quality and price stability. Cumulatively, Harris (2013) and Budish et al. (2015) emphasize intricate complexities involved in regulatory decision-making. They remark that such measures could bring about challenges as reduced liquidity and increased trading costs for policymakers. Pagnotta and Philippon (2018) further attest that unfitting interventions would not optimise the market welfare in the equilibrium.

In summary, academic research on HFT spans a broad spectrum of topics, such as price discovery, market quality, systemic risk, and regulatory responses. Foundational studies have laid the groundwork for understanding HFT’s complex role in modern markets, driving a nuanced dialogue. As HFT develops, sustained academic scrutiny is vital for shaping policies and practices that maximize its benefits while properly addressing its risks and drawbacks.

2.2 HFT Identification

In HFT research, the major challenge scholars faced is unequivocally the accurate classification of low-latency activity. It represents the most fundamental prerequisite step for rigorous empirical analysis on the subject. Yet, associated HFT data with distinguishing labels (flags) at each order and trade message instance is either non-existent, inaccessible or prohibitively costly for theoreticians and practitioners’ thorough scrutiny. As noted by Brogaard et al. (2014, p.2270), “Markets and regulators

are the only sources of these and high-frequency traders and other traders often oppose releasing identifying data.”

HFT studies frequently utilize order (quote) and trade data derived from proprietary limit order book (LOB) information. Notable instances of such work include Hendershott et al. (2011), Friederich and Payne (2015), Frino et al. (2017), Brogaard and Garriott (2019), Ben Ammar et al. (2020), and Boehmer et al. (2021), among others. At this point, concisely recapping the main characteristics of the LOB, as systematically elaborated by Hossain (2022), is warranted before advancing. That is, each new information arrival triggers updates in the LOB. Such occurrences may stem from new orders, modifications, cancellations, trade executions, or similar events. The limit order and market order are the two core order types that investors place in the system based on their liquidity needs. Investors have the flexibility to adjust orders based on their needs when submitting them to the LOB. The successful execution of a limit order relies on finding a counterparty with opposing needs, which is why it may sometimes remain unexecuted. In contrast, a market order is executed instantly when it matches an outstanding limit order on the opposite side of the LOB. On both sides of the LOB, limit orders are sorted according to their price and time priority. The best bid and offer prices, BBO, refers to the highest bid and lowest offer prices available on both sides of the limit order book at a specific point in time.

Due to the factors like voluminous nature, computational burden, time constraints and resource commitment, the scientific handling of LOB data is essentially complex and demanding (Hossain, 2022). To ease such obstacles to an extent, researchers typically optimize the required data fields by eliminating certain dimensions through simplification approaches. Focusing solely on the BBO depth level of the LOB is the most widely applied method in the area (e.g., Hendershott, 2011; Boehmer et al., 2021). This approach risks misidentifying true HFT activity since the quotes beyond the top of the LOB are disregarded. Nonetheless, the source-dependent characteristic of proprietary LOB datasets may also lead additional externalities regarding the fields and depth levels accessible to users.

The scarce and expensive accessibility of HFT data narrows the range of alternatives for researchers. Consequently, pertinent scientific endeavors tend to adopt, at least partially, imperfect HFT identification methodologies. Revisiting Hossain’s (2022) seminal study, scholars pursuing empirical HFT analysis face constrained pathways

and must either (i) use an exchange-supplied HFT-flagged dataset (direct method), providing a straightforward but potentially limited solution, or (ii) apply a proxy to trace HFT activity (indirect method), requiring more complex assumptions and analytical rigor.

An HFT-flagged dataset comprises exchange system-generated binary tags for each order and trade electronic message, which enable in-house classification of low-latency transactions at source. The NASDAQ HFT dataset is one of the most recognized in this group. Papers by Carrion (2013), Brogaard et al. (2014), and Brogaard and Garriott (2019) are important examples utilizing the dataset to analyze various HFT implications. The dataset covers tick-by-order transactions of 26 manually classified HFT firms across 120 U.S. equities listed on either NASDAQ or New York Stock Exchange (NYSE) between 2008 and 2009. Each data point is assigned one of four order execution types based on two dimensions: liquidity-taking (aggressive) versus liquidity-providing (passive) and HFTs versus non-HFTs. Brogaard et al. (2014) highlight two shortcomings of the NASDAQ HFT dataset as the exclusion of HFT firms acting as brokers while implementing proprietary trading strategies, and the omission of routed HFT orders transmitted through larger non-HFT intermediaries.

Comparable studies that identify HFT through direct method or exchange-flagged datasets can be grouped by the origin countries of investigated equity markets: Hendershott and Riordan (2013) for Germany, Hagstromer and Norden (2013) for Sweden, Brogaard et al. (2014) for the United Kingdom, and Brogaard and Garriott (2019) for Canada. However, Conrad et al. (2015) present critical interpretations regarding the limitations of such pre-identified datasets. They assert that the direct method is not free from bias, and true HFT activity might be obscured because of likely sampling issues arising from exchanges' subjective selection criteria and participants' conflict of interest.

The indirect method, a secondary approach for HFT identification, relies on proxies to discern low-latency activity. Since the majority of accessible data naturally lacks HFT-flags, this method is more prevalent in the literature, as Hossain (2022) reveals. The definition of an HFT proxy should target one or more HFT-identifying features and formulate the measure accordingly. Zaharudin et al. (2021) emphasizes the

replicability of proxy tools as a significant advantage. With limited dependence on internal expertise, it can be adapted to diverse markets and longer durations.

Hendershott et al.'s (2011) AT proxy, calculated as number of electronic messages per \$100 trading volume, stands as one of the pioneering alternatives. The 'speed' factor is central to their HFT identification, which employs BBO depth level information. Another influential contribution by Hasbrouck and Saar (2013) unveils *RunsInProcess* proxy, which reflects HFT behavior through dynamic order placement strategies. This measure is based on the number of strategic runs which consist of sequences of rapid order submissions, cancellations, and executions, assumed to be carried out by HFT algorithms. Orders are considered linked if (1) the cancelled order and the new order are in the same direction, (2) the sizes of the orders are identical, and (3) the cancellation of an order is swiftly followed by another within 100 millisecond. Runs consisting of fewer than 10 messages are discarded to reduce the effects of agency algorithms. Ersan and Ekinici (2016) extend Hasbrouck and Saar's (2013) *RunsInProcess* measure by incorporating modification messages to the existing procedure. They designate the new proxy as *RunsInProcessExtended* and assess it with Turkish equities data. Similarly, Ekinici and Ersan (2018) build a two-step detection algorithm (TSD) to gauge HFT activity. In the first step, the HFT-label is assigned to messages if any potential combinations (submission followed by modification; submission followed by cancellation; modification prior to cancellation; and two successive modifications) in the sequence within the same second are observed. In the second step, additional message(s) are included to the prior HFT subset if another unlabeled order message of the same size as the HFT-labeled ones is found within 1 second, assuming it is also submitted by the same high-frequency trader.

The order-to-trade (OTR) ratio serves as another widely accepted HFT proxy in empirical research (Hagstromer and Norden, 2013; Friederich and Payne, 2015; Frino et al., 2017). Intuitively, large OTRs in financial instruments correspond to increased HFT activity. On the other hand, Yao and Ye (2018) and O'Hara et al. (2019) report that a greater percentage of liquidity provided by HFTs actually correlates with a lower OTR. Against conventional wisdom, the authors observe that HFTs tend to supply more (or less) liquidity and cancel orders less (or more) frequently in securities with larger (or smaller) tick sizes. Thus, OTR may inadequately encapsulate HFT patterns.

Less sophisticated and computationally less demanding HFT proxies, derived from various identifying measures, have also been introduced in the literature. For instance, Van Ness et al. (2015) recommend the cancellation rate, which is the proportion of withdrawn limit orders relative to the total limit orders, as a measure to track HFT activity. In tandem, a range of identifying factors, such as quote updates (Conrad et al., 2015), daily net positions and end-of-day inventories (Krilenko et al., 2017; Brogaard et al., 2019), order imbalance (Comerton-Forde et al., 2018), and order durations (Bouveret et al., 2014), have been utilized by academics.

In essence, supporters of the proxy identification approach (indirect method) point to the functional content of HFT information embedded in trading activity, which tends to be free from prejudgment bias when compared to the alternative direct method. However, this technique faces criticism primarily due to the high dispersion among suggested proxies and the sensitivity to cut-off levels, both contributing to inconsistencies in reported findings (Elizarov et al., 2017).

Overall, both direct (flag-based) and indirect (proxy-driven) HFT identification methods show unique merits and drawbacks, as documented in the cited literature. Hence, at some stage in HFT analysis, scholars unavoidably construct their assumptions and logical reasoning to fortify the associated theoretical groundwork. Nevertheless, the absence of a unified standard for identifying the phenomenon has led to various concerns about the precision of studies evaluating HFT impacts on markets (Zaharudin et al., 2021).

2.3 HFT – Liquidity Relationship and Tick Size Changes

The implications of HFT involvement in price formation have been extensively studied across a wide range of markets. Hendershott et al. (2011), in their seminal work, reveal that the rise of algorithmic trading in the NYSE narrowed bid-ask spreads, thereby improving liquidity. Brogaard et al. (2014) demonstrate that HFTs enhance price efficiency by aligning trades with permanent price changes and countering transitory errors via liquidity-demanding orders. Similarly, Boehmer et al. (2021) analyze 42 equity markets globally, examining the influence of algorithmic trading (AT) on market quality between 2001 and 2011. Despite representing the relatively early adoption of HFT, their findings were strikingly consistent across markets, indicating

that AT enhances liquidity and informational efficiency while also increasing volatility.

Contrastingly, Baron et al. (2012) find that only a subset of HFTs acts as liquidity providers, with the most profitable HFTs exclusively serving as liquidity takers. Furthering this debate, Brogaard et al. (2017) note a decline in market quality during high-volatility periods due to HFT engagement. Zhao and Wan (2018), in their study of China's commodity futures market, confirm HFTs' contributions to price discovery but limited informational benefits for institutional trades executed at millisecond intervals.

One group of academics researching HFT-related liquidity focuses on the effects of pre- and post-regulation order-to-trade (OTR) fee and penalty schemes imposed by market authorities. Friederich and Payne (2015), Jorgensen et al. (2018), and Malinova et al. (2018) acknowledge significant declines in OTR following the introduction of these regulations. Although spreads remain largely unaffected, retail investors often suffer due to the HFT-driven reduction in quoted depth and decreased turnover.

Returning to one of the central themes tick-size change effects, Van Ness et al. (2000) investigate the impact of sixteenth-dollar quoting increments introduced on AMEX, NASDAQ, and NYSE in 1997. Their results validate the widely accepted notion that smaller tick sizes narrow bid-ask spreads, yet order book depth decreased on AMEX and NYSE but increased on NASDAQ. Jiang et al. (2011) evaluate post-reform market quality for NASDAQ. They highlight the evidence of higher transaction costs and lower efficiency. In a broader context, Meling and Odegard (2017) argue that reducing tick sizes could boost exchanges' market share but reduce liquidity. The study diverges from the consensus, as HFTs were shown to favor larger tick sizes to enhance liquidity provision and maximize profits. The scholars attribute this finding to the strategic variability of HFT mechanisms.

In studies more closely related to the analysis, Verousis et al. (2018) conduct a pioneering literature survey, finding substantial empirical evidence supporting the link between minimum tick size regulations and the recent rise of HFT. Specifically, it has been documented that smaller minimum tick sizes enhance the effectiveness of HFT strategies, while larger tick sizes allow HFTs to earn greater profits due to increased trading activity. Generally, the majority of existing literature emphasizes a liquidity

enhancement following tick size reductions irrespective of discussions on HFT. Among others, Coughenour and Harris (2004), Chung et al. (2008), Barlet III and McCrary (2013), and Hagströmer and Norden (2013) are part of the first set of studies (i.e., the relationship between tick size and HFT activity), while Bacidore (1997), Ke et al. (2004), Aitken and Comerton-Forde (2005), and Anderson and Peng (2014) belong to the second set (i.e., the effects of tick size changes on liquidity).

O'Hara et al. (2019) center their related analysis on the biodiversity of the market (different trader types). They suggest that relative tick size affects HFT strategies. Larger relative tick sizes encourage HFT market makers to adopt more aggressive strategies by keeping limit orders in the book for longer and undercutting resting orders. This makes liquidity less “fleeting” compared to stocks with smaller tick sizes. As a result of such aggressiveness, HFT firms expand their market share and profit margins by leveraging competition. Hagstromer and Norden (2013) use OMX Stockholm data to compare market-making and opportunistic HFT strategies. They show that market-making HFT activity reduces intraday price volatility, depending on tick size changes. Kadan (2006) develops a model in which he proposes that decreases in tick size benefit dealers but may harm investors when the number of dealers is low.

Yamada (2022) contend that HFTs facilitate liquidity in stocks with larger relative tick sizes and alleged that increasing minimum tick sizes would amplify liquidity for substantial trades. Frino et al. (2015) remark that stock splits raise relative tick size, lowering OTR and increasing order resting times. Lastly, Pan et al. (2012) discuss that tick size reductions impair overall liquidity on the Hong Kong Exchange. In conceptual terms, they benefit retail investors through narrower bid-ask spreads, while disadvantaging institutional traders, thus hindering a broader Pareto improvement in market liquidity.

Unlike the typical focus on equities, Lee (2015) study the Korean index futures market to understand how HFT impacts market quality. He discovers that HFT neither boosts liquidity nor aids price discovery and efficiency; on the contrary it impedes them. This negative outcome may be linked to the view that the marginal utility of HFT is generally negative when a market is already liquid and has low latency. Alampieski and Lepone (2009) point out that tick size reductions tighten bid-ask spreads, reduce execution costs, and foster liquidity in the Sydney futures exchange. Furthermore,

Kurov and Zabolina (2005) estimate that minimum tick sizes in E-mini S&P 500 and E-mini NASDAQ 100 futures at Chicago Mercantile Exchange prevent bid-ask spreads from becoming competitive. They recommend that lowering the tick sizes for these contracts would be a beneficial policy.

Another example involving foreign exchange (FX) as a distinct asset type in the tick size debate is Mahmoodzadeh and Gencay's (2017) empirical research. They uncover that both spreads and depth contracted after decimalization in the interbank FX market. Despite this, HFTs, using sub-penny strategies, captured the rewards of spread improvement by front-running human traders. Chou and Chung (2006) identify similar patterns for ETFs in their study. Decimalization of S&P 500, NASDAQ 100, and Dow Jones index ETFs leads to a reduction in spreads, but at the cost of increased adverse selection. While trading activity for these instruments grow, market depth considerably drops post-tick size reduction.

2.4 HFT – Volatility Dynamics

During the sample period from 2019 to 2021, the daily average turnover reached approximately 346,000 contracts or TRY 4.91 billion (USD 0.68 billion) in terms of volume and notional value, respectively, for the chain. The average number of orders submitted per minute was 315. Additionally, the mean daily electronic order message traffic (i.e., engine load) for the closest expiry totalled 0.46 million transactions, peaking at 2.37 million on December 20, 2021.

HFT can be perceived as an emerging form of intermediary capable of both benefiting and disrupting the market in terms of volatility. A significant issue, focusing on whether these advanced investors act as key drivers or potential sources of unusual asset price volatility in times of market anxiety, has ignited widespread discourse. The divergent viewpoints on HFT may have arisen from the lack of consensus on its impact on volatility, where high-frequency traders might function as market makers, arbitrageurs, predators, or a hybrid of these roles (Carrion, 2013).

Previous investigations have meticulously analyzed the effects of HFT on volatility dynamics. In this vein, most early scientific work intend to address three cornerstone research questions: (i) whether HFT activity correlates with market volatility, (ii) how HFT behavior varies across low- and high-volatility periods, and (iii) whether linkages

exist between market crashes (turbulences) and HFT penetration. Contradictory findings in pertinent studies often prompt scholars to develop varying rationales to contend with the ongoing ambiguity of the issue.

To begin chronologically, Froot et al. (1992) assert that short-horizon traders, over-dependent on swift information, may contribute to price dislocations and excessive volatility in the market. They ascribe these effects to noise generation and disruption of price discovery processes. For examining HFT implications in simulated market conditions, Gsell (2008) designs an agent-based model featuring market participants with varying trading strategies. He concludes that HFT magnifies short-term volatility, especially under market stress, due to frequent order cancellations. In a comprehensive research, Zhang (2010) explores NYSE market using quarterly data spanning the period from 1985 to 2009. His results show that HFT exhibits a positive correlation with volatility, negatively impacting the market even after accounting for endogenous and exogenous factors, including stock fundamentals. The relationship is more evident for large-capitalization stocks and during times of heightened market uncertainty.

Jarrow and Protter (2012) illustrate that cumulatively, fast traders' arbitrage-focused strategies may displace prices from their intrinsic value by generating excess volatility. This may signal a dysfunctional role of HFTs as they capitalize on such market inefficiencies. On the contrary, Zervoudakis et al. (2012) argue that HFT diffusion bears no meaningful causal relationship with the intra-day-to-overnight volatility ratio. They reject the hypothesis that low-latency activity markedly intensifies volatility. Hagstromer and Norden (2013) investigate HFT practices through opportunistic and market-making strategies in Swedish (OMX30) equities from 2010 to 2012. Advocating for HFT, they determine that both strategies mitigate intraday return volatility by enhancing liquidity provision. Chaboud et al. (2014) also validate volatility lessening effect of HFT on the basis of reduced frequency of arbitrage opportunities. In a more event-specific analysis, Jiang et al. (2014) find that HFT elevates volatility in the US Treasury market around the time of macroeconomic news announcements. Egginton et al. (2016) identify that episodic spikes in low-latency quoting lead to upsurging volatility. In addition, Kelejian and Mukerji (2016) disclose that HFT significantly affects daily asset price

volatilities of 200 most traded S&P 500 shares over a relatively long period from 1985 through 2012. Interestingly, their findings suggest that the interaction between stock fundamentals and volatility has been undermined since the emergence of HFT.

The controversy on the subject persists in more recent studies as well. For instance, Brogaard et al. (2018) and Ke and Zhang (2020) state that HFTs do not cause any rise in volatility in a stable market environment when uncertainty is low. Ekinci and Ersan (2022) further depict that the variables HFT activity and volatility are not correlated in the Turkish equity market (BIST). The authors justify the result by claiming that the technological transformation of the exchange has only recently been implemented, and HFTs have yet to gain considerable penetration or influence in the market. An often-cited counterevidence for the discussion is presented by Shkilko and Sokolov (2020) with a novel approach. To account for speed disparities, they use unusual weather precipitation episodes, such as heavy rain or snow, as dummies in the model. Because the occurrence of such events temporarily worsens the latency capability—and thus the advantage— of HFTs' privileged microwave tower network, which is approximately 30% faster than the closest rival fiber-optic cable technology. The findings of the study highlight that volatility decreases under extreme precipitation conditions, implying a potential link between HFT and increased volatility.

Boehmer et al. (2021) conduct a thorough research to evaluate the influence of algorithmic trading (AT). Drawing on a dataset covering stocks from 42 global equity markets between 2001 and 2011, they demonstrate that, despite enhancing liquidity and informational efficiency, AT tends to accelerate short-term market fluctuations and volatility induced by AT often results in diminished liquidity. Consequently, it can be presumed that the market participants responsible for expanded volatility may also erode liquidity, or vice-versa. Breckenfelder (2024) constructs a model to discover the premises of HFT competition in the Swedish blue-chip stocks (OMX S30) during the period between 2009 and 2010. He reports that as competition heats up in the market environment, HFTs shift toward more speculative trading practices. This behavior, according to Breckenfelder, increases short-term (intraday) volatility but has no significant effect on inter-day volatility, which is consistent with the common HFT tendency to refrain from holding overnight positions.

In this context, a number of scholars attempt to detect substantial variations in HFT activity between low and high volatility periods. Early papers by Hasbrouck and Saar

(2013) and Hagstromer and Norden (2013) express that low-latency participation reduces short-term volatility. This effect remains valid even under stressed market conditions and during periods of high volatility. Following a contrarian perspective, Jain et al. (2016) imply that HFT represents a direct source of systemic risk after evaluating volatility shock propagation with any of order flow, correlation, and CoVAR metrics. Brogaard et al. (2018) scrutinize NASDAQ dataset to assess extreme price movements (EPM), which are defined as either (i) the 99.9th percentile of 10-second absolute midpoint quote returns, (ii) the residuals of a market and return model, or (iii) a local volatility ratio. They show that HFTs engage in transactions opposite to those of EPMs and supply liquidity to non-HFTs. Yet, they may still contribute to heightened volatility during times of market stress.

Kang et al. (2020) provide an in-depth analysis examining the relationship between HFT, order flow toxicity, and short-term volatility in Korean futures market. Their analysis reveal that during normal market conditions, HFT is negatively correlated with both order flow toxicity and short-term volatility. However, during periods of market stress, this relationship becomes positive, indicating that HFT may exacerbate volatility under such conditions. In tandem, Bazzana and Collini (2020) confirm the adverse impact of HFT, particularly aggressive strategies, on volatility in S&P 500 ETFs before and after the exogenous shocks such as US presidential election and Brexit. Alternatively, through observing post-earnings announcement drifts (PEAD), Ke and Zhang (2020) find a negative relationship between HFT and magnitude of PEAD, which stems from HFT's role in supplying liquidity. This effect is more pronounced when earnings surprises are extreme. As a result, Ke and Zhang propose that HFT might foster market efficiency during volatile periods. Ben Ammar and Hellara (2022) also document that higher HFT intensity is linked to lower stock price volatility under stable market conditions. Nonetheless, in times of intraday crashes, the escalation in HFT aggressive order submissions and cancellations leads to amplified volatility and excessive liquidity consumption. Simply put, during stressed market periods, high-frequency algorithms cause significant price fluctuations by executing rapid trades with each other, further destabilizing the market.

At last, with regard to evidence from Turkish markets, Dastan (2021) indicates that HFTs alleviate their trading volume and liquidity provision during EPMs in the BIST equity market. This behavior is explained by HFTs' superior forecast abilities

compared to non-HFT retail investors. Furthermore, Bodur (2022) investigates flash events affecting Turkish blue-chip stocks, TCELL and KCHOL, on February 22, 2017. Just before market close, both stocks saw nearly 10% price drops in a single second and then recovered very quickly. The study exposes that order flow toxicity metric, Volume-Synchronized Probability of Informed Trading (VPIN), displays a consistent upward movement before TCELL's sharp decline, clearly implying the influence of AT pressures throughout the event. In contrast, no meaningful precursor signals are present for KCHOL.



3. STRUCTURAL OVERVIEW OF MARKET PROPERTIES AND TRADING ECOSYSTEM

As Türkiye's sole authorized financial venue, Borsa Istanbul (BIST) offers a wide range of trading products, including equity, debt securities, derivatives, and precious metals, through fully automated, electronic limit order book markets. In line with global advancements in machine-based automated trading, BIST transformed its technological infrastructure via a multi-phase market-based program conducted between 2015 and 2018. Through a strategical partnership with NASDAQ, a global exchange leader, the partial deployment of the new infrastructure was carried out sequentially across markets: equity in Phase 1 (November 2015), derivatives in Phase 2 (March 2017), and debt securities and precious metals in Phase 3 (March 2018), respectively.

The new trading system, called “BISTECH”, significantly enhances the platform’s latency performance and FIX message traffic efficiency, providing a solid foundation for next-generation investors. Technically, the BISTECH platform can process 100,000 transactions per second, and the average processing time is designed to be less than 100 microseconds. As part of its development initiatives, the exchange has renovated its member colocation facilities ensuring a fast and secure communication network. These improvements uphold industry standards, making BIST attractive for HFTs. In 2023, HFT activity represented 29% of the total turnover in the BIST equity market and 46% in the BIST derivatives market (VIOP) (Borsa Istanbul 2023 Annual Report).

Supported by advanced technological transformation and steadily growing domestic retail interest, especially during the pandemic, the BIST equity market has emerged as one of the largest and most promising trading venues among emerging and developing countries. Its share turnover velocity remains exceptionally high, and the daily average traded value of shares has been steadily increasing throughout the past

decade¹. By the end of 2023, the total market capitalization of 539 listed public companies reached USD 417 billion, with a free float rate of 41% (Borsa Istanbul 2023 Annual Report). The BIST-30 spot index serves as the primary benchmark for tracking the price performance of the thirty selected large-cap blue-chip stocks in the market. Its constituents are periodically reviewed based on ranked market capitalization and turnover criteria. These stocks represent approximately 70% of the total trade value in the market.

Turkish derivatives trading officially commenced on 4 February 2005 with the establishment of the Futures and Options Exchange in Izmir. Meanwhile, Istanbul Stock Exchange (ISE) Derivatives Market began operations on 21 December 2012. On 5 August 2013, the two platforms were merged, centralizing all futures and options trading under Borsa Istanbul Derivatives Market (VIOP). Takasbank has acted as the central counterparty (CCP) for clearing and settlement in VIOP since 3 March 2014. Figure 3.1 illustrates the annual trend of the market's average daily traded value in millions of USD.

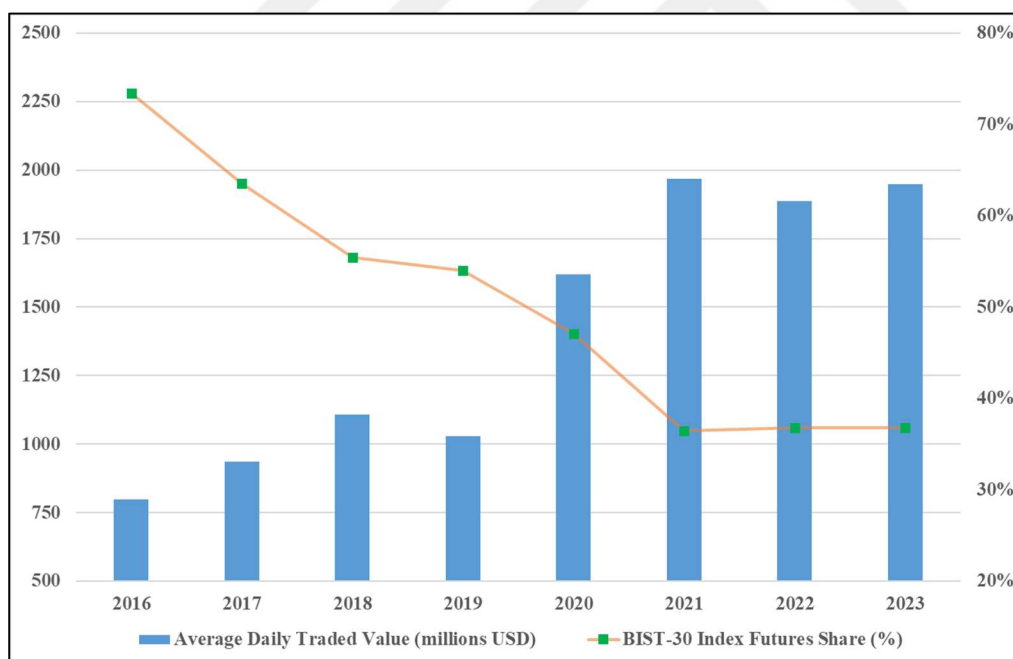


Figure 3.1 : Traded Value Progression in VIOP.

¹ According to the latest World Federation of Exchanges (WFE) annual statistics report, the BIST equity market recorded a 271% share turnover velocity in 2023, ranking 3rd globally after achieving 1st place the previous year with a 388% turnover velocity. Additionally, the average daily turnover value nearly reached USD 5.35 billion, placing BIST among the top 15 stock exchanges worldwide (WFE Annual Statistics Guide, 2023).

VIOP ranks among the most actively traded venues globally within its segment of emerging and frontier markets². The market includes a broad list of derivatives products from various underlying asset classes, but the BIST-30 index futures chain accounts for the solid portion of the traded value with single stock futures (Figure 3.1). Classified as one of the top twelve most traded stock index futures globally, these contracts have provided substantial liquidity to investors since their inception in 2005³. According to Table 3.1, roughly 2.6 trades are executed per second for BIST-30 index futures.

Table 3.1 : Global Market Statistics for Top Index Futures.

Panel A: Developed Markets						
Country	Index	Future	Avg Daily Volume	Median Daily Volume	Daily Nb Trades	Avg Nb Trades
USA	S&P 500	ESc1	1508599	1401168	586452	425
USA	NASDAQ 100	NQc1	484052	461403	494258	358
Japan	TOPIX	JT1c1	71052	61715	23216	17
Japan	N225	JN1c1	65399	57886	18767	14
UK	FTSE	FF1c1	110345	93140	57725	48
Germany	DAX 30	FDXc1	86970	84703	42706	36
France	CAC 40	FCEc1	82759	67905	50572	60
Switzerland	SMI	FSM1c1	41135	34868	13985	17
Panel B: Emerging Markets						
Country	Index	Future	Avg Daily Volume	Median Daily Volume	Daily Nb Trades	Avg Nb Trades
Brazil	Bovespa	INDc1	165937	157375	18737	35
Hong Kong	Hang Seng	HS1c1	150956	148211	95198	68
China	CSI 300	CIFc1	79568	78558	16010	49
Korea	KOSPI 200	MKS2c1	116615	100128	42840	102
India	NIFTY	NIFc1	11297787	10194225	12987	35
Russia	IRTS	RIRTSc1	501253	476349	34735	24
Türkiye	BIST 30	XU030c1	641853	596901	127651	158

Notes: Average and median daily volume show the number of contracts traded in the contracts with the nearest maturity over the period of the analysis (Jan 2, 2019 to Dec 31, 2021) while daily number of trades and average number of trades per minute (last two columns) refer to one-day data of Oct 13, 2023 (Source: Bloomberg Terminal, Global Stock Index Futures Data, accessed on August 2024, Bloomberg L.P.).

During the sample period from 2019 to 2021, the daily average turnover reached approximately 346,000 contracts or TRY 4.91 billion (USD 0.68 billion) in terms of

² As noted in the most recent WFE annual statistics report, the BIST index futures market generated a trade volume of 60.8 million contracts in 2023, accounting for a total of USD 143.6 billion in notional value. It is ranked as the 2nd most actively traded index futures market among 11 derivatives exchanges, following Eurex, within the regional group of Europe, the Middle East, and Africa (WFE Annual Statistics Guide, 2023).

³ The daily traded volume (value) record for VIOP was broken on August 26, 2022 (August 24, 2023) with 28.9 million contracts (TRY 112.5 billion or USD 4.32 billion). BIST-30 index futures composed nearly 41% (37%) of the total traded notional value of VIOP in 2023 (2022) in monetary terms (Borsa Istanbul 2022 and 2023 Annual Reports).

volume and notional value, respectively, for the chain. The average number of orders submitted per minute was 315. Additionally, the mean daily electronic order message traffic (i.e., engine load) for the closest expiry totaled 0.46 million transactions, peaking at 2.37 million on December 20, 2021.

The BIST-30 index futures are traded continuously on each working day from 09:30 to 18:15 during the regular session and from 19:00 to 23:00 in the after-hours session. Considering limited and transitory liquidity in the latter, the activity during this additional period is excluded from the analysis. On 27 July 2020, BIST's TRY-based price and return stock indices underwent a simplification process by being divided by 100 to simplify excessively high values, thereby enhancing operational efficiency. For example, the BIST-30 spot index close value of 134,849.352 on 24 July 2020 was recalibrated to 1348.49 for the opening on the next business go-live date. Following this adjustment, key amendments were introduced for BIST-30 index futures: (a) the underlying value factor was raised to 1 from 0.001; (b) contract size was reduced from 100 to 10; (c) price quotations were adjusted upwards by a factor of 10; and (d) the minimum tick size was increased from 0.025 to 0.25. The notional value of a single futures contract is calculated as 10 times the TRY value of the BIST-30 index level. Price quotations are expressed per unit of the underlying asset; for instance, a trade price of 5250.25 for a quantity of 10 results in a TRY 525,025 monetary position upon execution.

At any given time, three nearest contract expirations out of the bi-monthly terms—February, April, June, August, October, and December—are available for trading until the last business day of the respective contract month. These contracts are cash-settled on a T+0 basis, with same-day mark-to-market profit or loss distributions managed via the standard clearing procedure. The daily settlement price is determined as the weighted average price of all trades executed during the final 10 minutes of the normal session. The final settlement price, on the other hand, is calculated by combining two components: (i) the time-weighted average of the index values from the last 30 minutes of continuous auction trading in the equity market and (ii) the index closing price. These components are weighted at 80% and 20%, respectively, to arrive at the final value. Daily price fluctuations are capped within $\pm 10\%$ of the reference price, while position limits are set at 300,000 contracts per registered identity. Notably,

there is no officially designated market maker for the already liquid BIST-30 index futures.

The maximum order quantity for BIST-30 index futures is set at 2,000 contracts. For quantities exceeding this limit, up to a maximum of 4,000 contracts, the trade reporting procedure can be utilized during the normal session, provided the agreed price falls within the daily limits. Trade reports are exempt from vended price statistics (high, low, close, etc.) and settlement price calculations. However, they are included in volume and value metrics to reflect overall market activity.

The investor composition in VIOP highlights that domestic retail investors dominate liquidity-taking activities. At the end of 2023, 20,545 out of a total of 21,253 accounts with open positions belonged to this group⁴. However, exchange data reveals that 54% (33%) of the total traded volume across all derivatives in the preceding year was executed by a relatively small segment of institutional (foreign) accounts, most likely entities serving as liquidity providers (Borsa Istanbul 2022 Annual Report).

⁴ According to data provided by the clearinghouse Takasbank, the number of accounts with open positions in other investor groups is as follows: 577 for domestic institutions (2.71%), 78 for foreign retail investors (0.04%), and 53 for foreign institutions (0.02%) (<https://www.takasbank.com.tr/en/statistics/futures-and-options-market-1/futures-and-options-market>).



4. DATA WORKFLOW: ACQUISITION, PROCESSING AND TRANSFORMATION

The analysis involves two datasets, comprising tick-by-tick order and trade activity of the closest expiry BIST-30 index futures, directly sourced from the exchange. To put it differently, the entire trading activity of the most liquid contract month is examined for specific data entities. Due to insufficient trading in longer maturities, the sample is limited to expiration rolls closest to maturity.

Raw data files contain 342.4 million order and 95.3 million trade messages (in rows) covering 741 full trading days between January 2, 2019, and December 31, 2021. They are provided by fields (or columns) such as ID, date, interval, contract, high, low, close, best bid, best offer, total quote updates, trade count, volume, notional value, and, more importantly, HFT-related metrics, including HFT-quote updates, HFT-trade count, HFT-volume, and HFT-notional value. These four specialized low-latency participation gauges are calculated using designated HFT-flags for each order and/or trade message instance, which are derived from encrypted information about self-declared HFT accounts, provided under special consent by BIST.

By aggregating the message rows within each period, refined intraday (30-minute) and daily time series are generated for specific data fields, marking the conclusion of the process. A self-developed JAVA based program, called as HFT ProxyChecker, is used to complete the entire data transformation through a flow depicted in Figure 4.1. Static filtering conditions are applied along with given data objects HFT accounts (encrypted), included order update codes and included trade update codes. It should be emphasized that the application's computations for a trading day (i) have an average end-to-end runtime of approximately 15–20 minutes. Appendices section of the study encompasses a succinct note of computations for a specific trading day as an example. It is supplemented by reference Excel sheet screenshots and JAVA program codes.

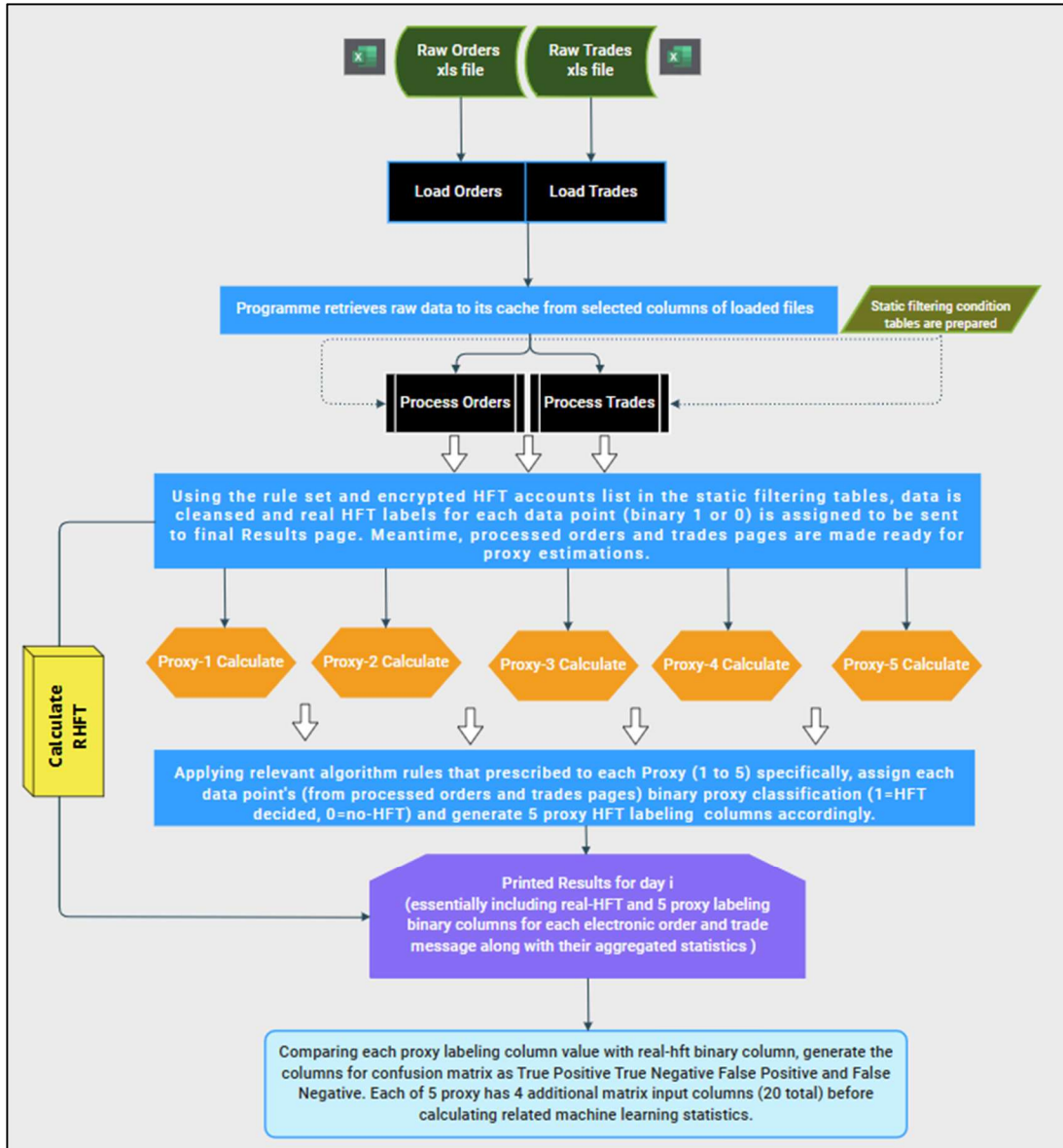


Figure 4.1 : HFT ProxyChecker Application Diagram.

VIOP operates as a pure limit order market without a hidden order mechanism. Incoming orders, identified by unique IDs and time stamps, are displayed to participants and data vendors through the ITCH channel. In tracking order lifespans, the relevant raw data files are restricted to encompass: (i) order update reasons, submission, user-modified (replaced), execution (trade), and user-cancelled (or inactivated); (ii) order methods, limit, market, and market-to-limit; and (iii) validity periods, daily, FOK, FAK, GTC, and GTD⁵. Because of system-enforced conditions, such as inadequate collateral, improper session, suspended instruments, or excess

⁵Additional explanatory information on order methods and validity periods can be found on the BIST website (<https://www.borsaistanbul.com/en/sayfa/2973/orders>).

position limits, directly rejected or deleted order messages by the PTRM are excluded from the observations. Each valid update sequence is assigned a millisecond time stamp and incremental code, creating a separate row in the data coverage tied to the order unifier ID given in the submission. Mixed-feature special order methods, such as stop and strategy orders, are also omitted.

Transactions outside regular market activity, such as trade reports or negotiated deals, are removed from the applicable raw trade series. Auto-generated matches from strategy (combination) orders are also subtracted from the computations using a similar approach. Since executed deals in the BIST data warehouse are logged with distinct buy and sell leg records (i.e., two trade IDs connected to a unique deal number) in the corresponding trade files, volume and notional value statistics are multiplied by 2 for each trade count that differs from certain exchange-presented statistics. This approach allows for a clear separation and appropriate weighing of HFT and non-HFT trade values.



5. THE PERFORMANCE OF SELECTED HIGH-FREQUENCY TRADING PROXIES: AN APPLICATION ON TURKISH INDEX FUTURES MARKET⁶

5.1 Background

In most markets, information regarding investor-based, HFT-labeled message traffic is either unavailable or inaccessible for research purposes. However, a few notable studies, including Hendershott and Riordan (2013), Brogaard et al. (2014), and Brogaard and Garriott (2019), benefit from source-assigned proprietary HFT tags. This approach to identifying HFT activity is commonly referred to as the “direct method.” Conversely, as noted by Hossain (2022), the majority of HFT research relies on an “indirect method,” which employs various proxies to trace the activities of HFTs. Key examples of this subset include Hendershott et al. (2011), Hasbrouck and Saar (2013), Kirilenko et al. (2017), Ben Ammar and Hellara (2022), and Ekinci and Ersan (2022), all of which develop quantitative HFT benchmarks on anonymous order and trade data lacking explicit flags.

Among the two methods, direct identification is often regarded as the more robust and reliable approach, as it is based on “real” data points. Nevertheless, Conrad et al. (2015) identify significant drawbacks to this method, including peculiar sampling biases and conflicts of interest, which can obscure genuine HFT activity. In contrast, the proxy identification approach (indirect method) is valued for its capacity to extract functional insights from HFT behavior embedded in trading activity, typically avoiding prejudgment biases associated with direct methods. However, this technique is not without criticism, primarily due to high variability in proxy definitions and sensitivity to cutoff thresholds, which lead to inconsistencies in reported results (Elizarov et al., 2017).

⁶ This section has been partially published in the paper titled “The Performance of Selected High-Frequency Trading Proxies: An Application on Turkish Index Futures Market” by Olgun, O., Ekinci, C. and Arıkan, R. (2024), in *Finance Research Letters*, 65, 105523.

To a greater extent, certain researchers, such as Shkilko and Sokolov (2020), focus on the effects of exogenous shocks to HFT activity instead of predicting these dynamics through proxies. Adopting a novel perspective, their research leverages unusual weather precipitation events as natural experiments in low-latency modeling. These disruptions to microwave networks temporarily eliminate speed advantages for the fastest traders, effectively reducing adverse selection and trading costs.

As a matter of fact, HFT is well-documented in equity markets. Nevertheless, it has also been studied, albeit less frequently, in other financial instruments such as derivatives (Lee, 2015; Zhao and Wan, 2018), bonds (Liu et al., 2014), foreign exchange markets (Manahov et al., 2014), and cryptocurrencies (Petukhina et al., 2021).

This research conducts an empirical analysis on time-stamped (microsecond precision) order and trade data of BIST-30 index futures, which enables the classification of market and limit orders in various stages (submission, modification, cancellation), as well as executed trades. More importantly, it allows for the identification of messages initiated by HFTs (with the HFT flag). The performance and consistency of widely applied HFT proxies are then evaluated by comparing them to exchange-tagged low-latency data (considered as real HFT, or RHFT). Additionally, the viability and predictive power of both direct and indirect methods are compared in certain models, based on their compatibility with stylized facts.

The study's contribution to the literature is threefold. First, it addresses a gap in the measurement of HFT activity by evaluating the performance of selected proxies, a crucial aspect for the validity and practicality of HFT studies. To date, no empirical inquiry has focused on this objective, aside from using a combination of indirect and direct HFT determinants. Second, by analyzing futures contracts, it extends the research beyond the dominant equity asset class, offering insights into the causalities of derivatives' unique characteristics. Third, building on Van Ness et al. (2015), the study introduces a modified proxy that incorporates observed trading habits, which may be better suited for uncharted futures products rather than being confined to widely applied equity market approaches.

5.2 Data and Variables

The dataset consists of tick-by-tick order and trade messages for the closest expiry BIST-30 index futures, spanning each of the 741 full trading days between January 2, 2019, and December 31, 2021. Irrelevant message activity is initially removed to prepare the data series. Following this, real HFT transaction messages are filtered out, and algorithms for selected proxies are executed to estimate HFT identification forecasts. This process generates binary identification values (real HFT or proxy) for each order and trade data point across the sample.

The average number of submitted orders per minute during the examined period is approximately 315. Time-dependent distributions of orders, trades, and electronic messages all exhibit U-shaped intraday patterns, as typically observed.

The HFT variables accompanying the econometric models are outlined below.

For day i ,

$$RHFT_i = \text{Nb of real HFT messages} / \text{Total nb of messages} \quad (5.1)$$

$$P1_i = \text{Nb of HFT messages by ARP proxy} / \text{Total nb of messages} \quad (5.2)$$

$$P2_i = \text{Nb of HFT messages by ARE proxy} / \text{Total nb of messages} \quad (5.3)$$

$$P3_i = \text{Nb of HFT messages by TSD proxy} / \text{Total nb of messages} \quad (5.4)$$

$$P4_i = \text{Nb of withdrawn limit orders} / \text{Nb of submitted limit orders} \quad (5.5)$$

$$P5_i = \text{Nb of withdrawn or modified limit orders} / \text{Total nb of limit orders} \quad (5.6)$$

ARP, *ARE* and *TSD* represent, AllRunsInProcess (Hasbrouck and Saar, 2013), AllRunsInExtended (Ersan and Ekinci, 2016) and Two-Step Detection (Ekinci and Ersan, 2018) algorithms, respectively. P4 corresponds to the Cancellation Rate proposed by Van Ness et al. (2015), with P5 serving as an extension to P4. Further details regarding the computation rules for the *RHFT* and proxy variables are specified in Table 5.1.

Table 5.1 : Explanatory Notes on the Construction of HFT Variables.

Variable	Numerator	Description
RHFT _i	Nb of real HFT messages	Sum of the electronic message traffic from the Exchange assigned self-declared HFT accounts.
P(1) _i	Nb of HFT messages denominated by ARP (<i>AllRunsInProcess</i>) proxy	Sum of the order submission, cancellation and execution messages within estimated “runs” with min. 10 message length constraint that complying RunsInProcess HFT criteria (i.e. two orders are linked if the cancelled order message followed by a same size and direction new order message, either submission, cancellation or execution, in 100ms) for day <i>i</i> and interval <i>f</i> (Hasbrouck and Saar, 2013)
P(2) _i	Nb of HFT messages denominated by ARE (<i>AllRunsInExtended</i>) proxy	Sum of the order submission, cancellation and modification messages within estimated “runs” with min. 10 message length constraint that complying RunsInExtended HFT criteria (i.e. two orders are linked with the same size if they have messages submitted within 1s.) for day <i>i</i> and interval <i>f</i> (Ersan and Ekinci, 2016)
P(3) _i	Nb of HFT messages denominated by TSD (<i>Two-Step Detection</i>) proxy	Sum of the order submission, cancellation and modification messages complying authors’ two-step detection’ mechanism within chronologically set submission – cancellation/execution sequences for day <i>i</i> and interval <i>f</i> (Ekinci and Ersan, 2018). {In the first step, HFT-label is assigned to messages if any of potential combinations (submission followed by modification; submission followed by cancellation; modification prior to cancellation and two successive modifications) in the sequence within the same second is observed. Second, add extra message(s) to prior HFT subset if there exists another unlabeled order message with same size of HFT-labeled ones in 1s assuming that it is also submitted by same high-frequency trader.}
P(4) _i	Nb of withdrawn limit orders	Sum of the limit order cancellation messages for day <i>i</i> and interval <i>f</i> (Van Ness et al., 2015)
P(5) _i	Nb of withdrawn or modified limit orders	Sum of the limit order cancellation and modification messages for day <i>i</i> and interval <i>f</i>

Notes: This table presents a brief overview of the computation procedure for the numerator value associated with RHFT and the proxy variables, as depicted in equations 5.1 to 5.6.

5.3 Methodology

Aligned with seminal multidisciplinary studies in the relevant literature, this research employs a three-component analytical framework to assess the eligibility of the aforementioned proxies. First, essential statistical techniques such as correlation analysis, hypothesis testing for sample means, and scatter plots featuring lines of equality are applied. Then, a confusion matrix, along with additional selected benchmarks, is developed for the variables, incorporating receiver operating characteristic (ROC) plots inspired by developments in the machine learning field. Dickinson et al. (2010) highlight

that such an integrative approach to direct and proxy methods facilitates a robust evaluation of clinical behavior in the given context.

In the final step, naive regression estimation is utilized to tackle the proxy aptitude challenge within the empirical setup. Using the model described in equation 5.7, the explanatory power of each HFT proxy $P(N)$, where N ranges from 1 to 5, is measured individually over the dependent variable $RHFT$ for day i .

$$RHFT_i = \alpha_i + \beta P(N)_i + \varepsilon_i \quad (5.7)$$

Mahnken et al. (2014) emphasize that the validity of a workable proxy measure should initially be confirmed by meeting the conditions of unbiasedness ($\alpha = 0 \cap \beta = 1$) and minimal error ($\varepsilon \sim 0$). Additionally, traditional OLS diagnostics, such as adjusted R^2 , F-statistics, and mean errors (MAE and RMSE), offer valuable insights into the robustness of the measure.

5.4 Empirical Results

5.4.1 Descriptive statistics

Tables 5.2 and 5.3 present summary statistics for the daily number of orders (both HFT and non-HFT) and trading volume in the market over the specified time period. The data indicate a rapid increase in the share of low-latency orders, consistent with global industry trends. Key contributors to this growth include BIST's recent technological upgrades, a significant rise in retail investor participation, and the heightened trading activity observed towards the Covid-19 pandemic. During 2020 and 2021, the majority of electronic order messages originated from HFT-designated accounts. While HFTs accounted for nearly a quarter of daily futures trading volume during the sample period, their order-to-trade (OTR) ratio was approximately 6.27⁷.

⁷ Transaction-based OTR is formulated by Eurex as [nb of HFT orders] / [nb of HFT trades] – 1 (<https://www.eurex.com/ex-en/rules-regs/order-to-trade-ratio>). Average daily HFT trades (denominator) equals 33,800 contracts during the period of analysis.

Table 5.2 : Summary Statistics for Avg. Daily Nb of Order Messages ($\times 10^3$).

	Years	Mean	Median	Max	Min	SD	Skewn.	Kurtos.	Obs.
Total	@all	462.1	401.9	2376.6	83.3	226.7	3.01	18.35	741
	2019	352.9	340.8	714.1	83.3	91.1	0.57	4.19	247
	2020	506.8	458.0	1324.1	198.3	196.2	1.50	5.85	249
	2021	526.8	441.3	2376.6	211.4	300.1	2.79	13.97	245
	HFT	@all	246.0	200.0	1921.6	39.5	179.3	3.89	26.97
	2019	160.1	148.1	426.9	39.5	60.1	1.00	4.71	247
	2020	256.2	226.2	781.9	66.5	132.1	1.63	6.06	249
	2021	322.2	242.8	1921.6	77.7	250.4	3.21	16.51	245

Notes: '@all' denotes the entire data ranging from 2019 to 2021. The abbreviations SD, Skewn., and Kurtos. correspond to standard deviation, skewness and kurtosis, while Obs. refers to the number of observations.

Furthermore, a joint evaluation of Tables 5.2 and 5.3 reveals that HFTs generated nearly half of the order flow, although their contribution to executed trades decreased to one-fourth, reflecting a not surprising HFT pattern.

Table 5.3 : Summary Statistics for Avg. Daily Trade Volume ($\times 10^3$ contracts).

	Mean	Median	Max	Min	SD	Skewn.	Kurtos.
Total Volume	644.8	597.1	1881.1	101.4	261.4	1.17	4.93
HFT Volume	168.0	147.1	739.9	20.1	93.2	1.81	7.87

Notes: Values in the table span the entire data range from 2019 to 2021, with a total of 741 daily observations. The abbreviations SD, Skewn., and Kurtos. correspond to standard deviation, skewness and kurtosis.

Table 5.4 provides the estimation statistics for the previously defined direct and proxy measures. Notably, P5 and P2 exhibit the closest matches with *RHFT*. To quantify the bilateral relationship between *RHFT* and the proxies, both daily interval time-series correlations and annual average cross-sectional correlations are computed across the full panel data. The former demonstrates how well a proxy tracks the direct measure, while the latter uncovers potential seasonal variations in the estimates.

Table 5.4 : Daily HFT Participation Estimates of Real HFT and Proxy Measures.

	Mean	Median	Max	Min	SD	Skewn.	Kurtos.
RHFT	0.5034	0.4995	0.8641	0.2119	0.0979	0.49	4.01
P1	0.0272	0.0241	0.1700	0.0002	0.0158	4.56	35.38
P2	0.4179	0.4339	0.6720	0.1856	0.0775	-0.51	2.87
P3	0.1833	0.1813	0.3932	0.0474	0.0588	0.23	2.69
P4	0.2016	0.1722	0.5282	0.0753	0.0742	1.10	3.38
P5	0.5442	0.5468	0.8530	0.2716	0.0953	0.13	2.94

Notes: Values in the table span the entire data range from 2019 to 2021, with a total of 741 daily observations. The abbreviations SD, Skewn., and Kurtos. correspond to standard deviation, skewness and kurtosis.

As shown in Table 5.5, the highest time-series correlation is observed for P5, as defined by equation 5.6. The rankings of the proxies are identical for the average cross-sectional correlations, although the correlations are somewhat lower, as expected (Schestag et al., 2016). In all cases, the null hypothesis of mean equality ($H_0: \mu_{RHFT} = \mu_{P(N)}$) is decisively rejected using t-and F-tests.

Table 5.5 : Pairwise Correlations of Real HFT vs. Proxies and Tests of Equality Results.

	d.f.	P1	P2	P3	P4	P5
Time-Series Correlations	741	0.4222	0.2268	0.3887	0.5541	0.8189
Average Cross-Correlations	247	0.3301	0.2613	0.3262	0.4680	0.7876
t-test value	1480	130.61 (0.0616)	18.64 (0.0096)	76.25 (0.0321)	66.83 (0.0303)	-8.15 (0.0097)
F-test value	(740, 740)	38.41 (0.0419)	1.59 (0.0678)	2.77 (0.0615)	1.74 (0.0677)	1.05 (0.0756)

Notes: Values in the table span the entire data range from 2019 to 2021, with a total of 741 daily observations. The Mean Squared and Mean Absolute Difference values are presented in brackets below the t-test and F-test statistics, respectively. P-values for all variable test scores are less than 0.0001. d.f. refers degrees of freedom.

5.4.2 Machine learning metrics

Table 5.6 displays the confusion matrix along with related machine learning parameter estimates. Sensitivity and precision ratios are prioritized, as they provide a clearer comparison by highlighting the relative performance of true classifications (e.g., RHFT and Proxy tags overlap in the same direction for both 1 and 0 in any electronic message). Based on these, P4 and P1 emerge as the consistently worst performers. While slight differences exist among other proxies, P5 and P2 deliver superior results.

Table 5.6 : Machine Learning (ML) Parameters.

		P1	P2	P3	P4	P5
True – Positives	(a)	3.03	38.97	21.14	3.68	31.71
False –Positives	(b)	64.91	23.59	37.63	46.90	41.21
False –Negatives	(c)	30.91	19.90	37.73	46.66	18.63
True – Negatives	(d)	1.15	17.54	3.50	2.76	8.45
Accuracy		67.94	62.56	58.77	50.58	72.92
Sensitivity / TPR		8.39	66.23	35.95	7.32	61.93
False Alarm / FPR		1.84	40.58	7.96	5.57	31.51
PPV / Precision		70.34	69.60	85.44	54.91	77.44
NPV		67.58	54.37	49.89	50.12	68.49

Notes: The values in the table represent average daily percentages over the entire data period from 2019 to 2021, encompassing 741 daily observations. The confusion matrix parameters—True Positives (a), False Positives (b), False Negatives (c), and True Negatives (d)—are calculated for each day based on the following conditions: P(N) = HFT & RHFT = True for (a); P(N) = HFT & RHFT = False for (b); P(N) = Non-HFT & RHFT = True for (c); and P(N) = Non-HFT & RHFT = False for (d).

Scholars analyzing proxy qualification often consider the upper-left quartile of ROC plots as a key criterion in this context. From the diagrams shown in Figure 5.1, it is evident that the highest density (high sensitivity vs. low 1-specificity) is achieved by P5 and P2, while the underperformance of P4 and P1 is consistent with the machine learning ratio estimates.

5.4.3 Regression analysis

This step analyzes the bilateral relationship and interactions between direct and proxy measures through the estimation of equation 5.7. The OLS model outputs in Table 5.7 facilitate the evaluation of both the applicability of sound proxy assumptions and the derived model diagnostics, as previously discussed. Upon alignment of these factors, none of the selected proxies meet the unbiasedness condition, which represents the desired theoretical outcome. However, P5 and P4 yield the closest parameter matches for the qualification of $\alpha = 0 \cap \beta = 1$.

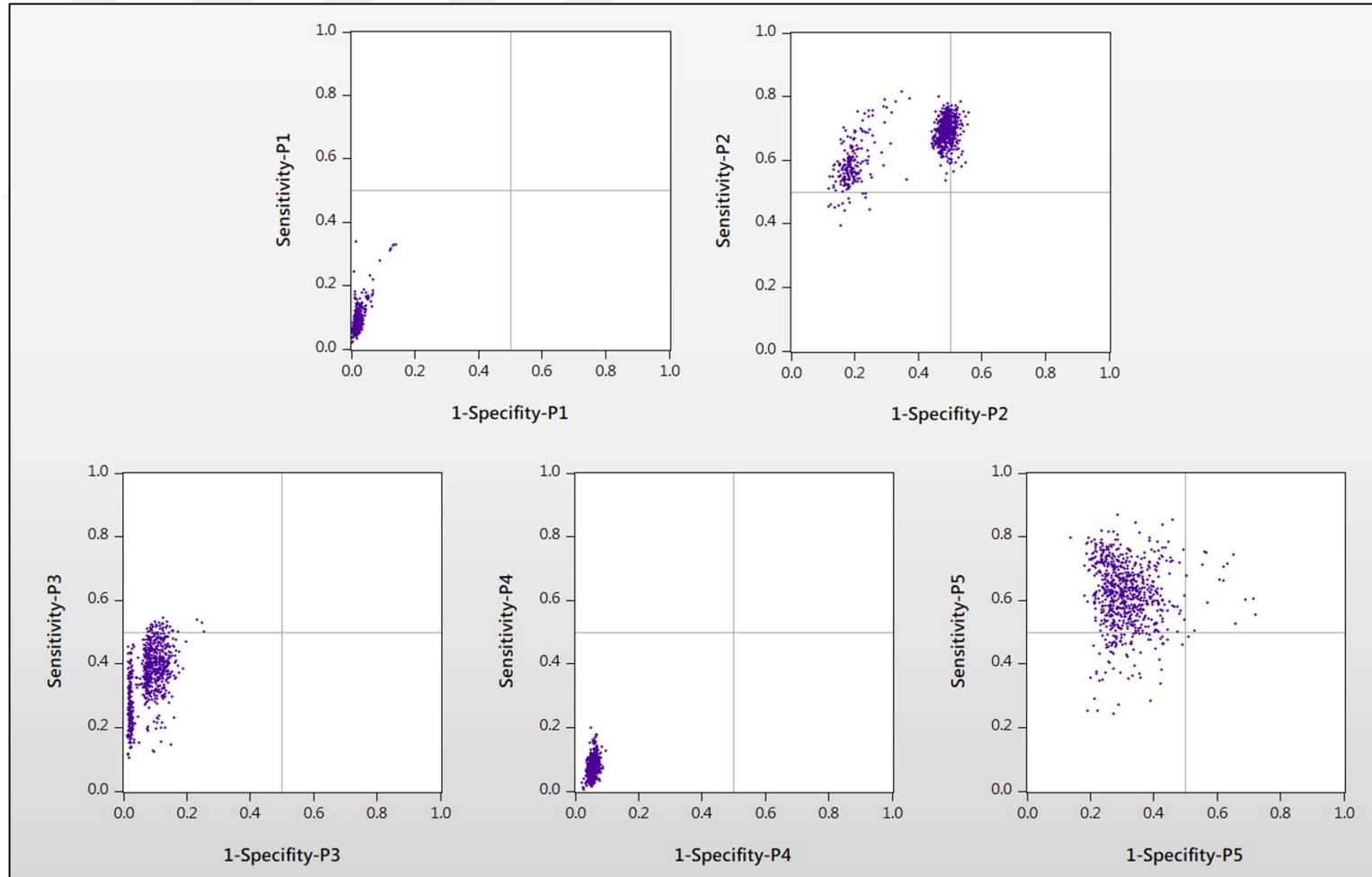


Figure 5.1 : ROC Plots of Proxy Measures.

Table 5.7 : OLS Regression Results.

	Md.1a	Md.1b	Md.1c	Md.1d	Md.1e
Constant (α)	0.4321 (0.0065)	0.3836 (0.0192)	0.3847 (0.0109)	0.3559 (0.0087)	0.0455 (0.0112)
P1	2.6164 (0.2066)				
P2		0.2866 (0.0453)			
P3			0.6474 (0.0564)		
P4				0.7311 (0.0404)	
P5					0.8411 (0.0217)
AIC	-2.001	-1.857	-1.968	-2.171	-2.915
Adj.R-sq	0.177	0.050	0.149	0.307	0.671
F-Value	160.32	40.07	131.49	327.35	1504.92
MAE	.0680	.0756	.0723	.0650	.0452
RMSE	.0887	.0953	.0902	.0814	.0561

Notes: The table presents the results of univariate regressions where the dependent variable is real HFT, and the independent variable is one of the five HFT proxies (P1 to P5) in each of the models, Md.1a to Md.1e originated from equation 4.7. The dataset comprises daily HFT ratios derived from orders in BIST 30 index futures contracts with the nearest maturity, spanning the period from 2019 to 2021 (741 days). Standard errors are reported in brackets below the coefficients, and all p-values are below 0.0001. The abbreviations AIC, Adj.R-sq, MAE and RMSE correspond to statistical measures namely Akaike Information Criterion, Adjusted R-squared, Mean Absolute Error and Root Mean Squared Error, respectively.

The diagnostics in the bottom panel highlight P5 as the dominant proxy among the alternatives, followed by P4 in terms of similar scales. Interestingly, the more sophisticated algorithm-based proxies (P1, P2, and P3) demonstrate relatively poor performance in this segment.

5.5 Conclusion

In summary, the study identifies P5 and P4 (P5 and P2) as the most effective proxies within the group based on evaluation metrics like correlation and tests of equality, while P2 and P3 (P4 and P1) exhibit the weakest outcomes in the *RHFT* course analysis. Moreover, P5 and P4 demonstrate the closest alignment with the unbiasedness criterion, yielding superior results in regression estimation.

The findings reveal that the proposed extension (P5) and the Cancellation Rate proxy (P4) from Van Ness et al. (2015) demonstrate stronger performance compared to other measures based on the evaluated metrics. Given that P5's message counts represent an extension of P4's framework, the similarity between these proxies is to be expected. The

relative inefficacy of certain sophisticated proxies (e.g., P1, P2, and P3) can be attributed to limitations inherent in Exchange-supplied (predetermined) HFT data. These datasets, often self-declared, are susceptible to Type-1 and/or Type-2 errors, which may distort HFT forecasts. Accordingly, some HFT-flagged accounts may not engage in low-latency activity, while others may remain hidden without notifying the Exchange. Additionally, these proxies were initially designed for equity markets, which may differ substantially from futures markets. Variations in factors such as traders' financial literacy, margin requirements, and short sale restrictions may influence proxy performance, emphasizing the importance of considering market-specific characteristics.

Cumulatively, the indirect method for HFT determination through proxies retains its value despite the noted limitations. These action-oriented, data-driven techniques can, in certain cases, yield more reliable low-latency footprints than Exchange-provided data, which often depends on deliberately declared member information and is challenging to access. To enhance the perspective and generalize the findings, further research should replicate this analysis in comparable futures markets or apply it to other asset classes, such as equities.



6. IMPACT OF REVERSE DECIMALIZATION ON HFT-DRIVEN LIQUIDITY IN TURKISH INDEX FUTURES MARKET: A COMPARATIVE TICK SIZE ANALYSIS

6.1 Scope

Since the inception of the HFT phenomenon in the financial industry, numerous researchers have sought to uncover the proprietary and concealed characteristics of this trading style. A particular focus has been on its significant impact on key market quality indicators, especially liquidity. Nevertheless, a consensus remains elusive, as scholars have reported divergent evidence on whether HFT enhances or undermines liquidity from the investors' standpoint.

On the other hand, a vast body of research has been attempting to determine the conditional effects of traded instruments' minimum tick size changes (or decimalization) over a number of market microstructure concepts like transaction costs, trading strategies, investor decomposition and competition. The most oft-cited assumption regarding the potential aftermath of decimalization is that lower tick size reduces transaction costs by essentially narrowing the bid-ask spread for unsophisticated investors. However, this adjustment may concurrently diminish incentives market makers' incentives, leading to reduced quoting and weaker market depth. Therefore, predicting the net effect on liquidity and identifying the optimal tick size pose significant challenges when considering overall market quality from the Exchange perspective.

The causality between HFT presence and tick size factors underlines divergent interpretations of absolute (minimum) or relative concepts within the market microstructure framework. A reduction in minimum tick size is likely to encourage HFT participation due to (i) simplified front-running opportunities, (ii) heightened adverse selection costs, and (iii) faster quote adjustments stemming from narrower bid-ask spreads. In addition, decreasing the minimum tick size quintessentially raises its

substitute parameter, relative tick size⁸. This dynamic is likely to attract more HFTs seeking to capitalize on enhanced liquidity incentives.

Combining the aforementioned research topics with an empirical perspective, this study primarily intends to shed light on the effects of the recent tick-size increase (or reverse decimalization)⁹ on HFT-driven liquidity in the Turkish index futures market. Within this framework, the dynamics of the hypothesized relationship between HFT activity (dependent variable) and instrument tick size (explanatory variable) are closely analyzed.

To evaluate the overall market liquidity impact of this trading rule change, decomposed non-HFT and HFT activities are modeled both ex-ante and ex-post, capturing the full spectrum of investor types. It is important to note that most previous research has predominantly focused on equity markets, while this study explores derivatives instruments like futures, which possess distinct characteristics such as inherent leverage, low transaction costs, high liquidity, short-selling options, and roll-over routines. Ultimately, this research seeks to contribute to multiple fundamental streams of literature by connecting the themes of HFT, tick-size effects, and futures markets, offering insights into the interplay among these domains.

The study addresses the rare occurrence of a minimum price tick increase (multiplied by 10), which necessitates in-depth analysis to assess its implications. It overcomes the HFT identification challenge by using a unique dataset¹⁰ that directly sorts low-latency transactions, eliminating the need for less reliable proxy methods. The research fills gaps in the literature on product types and venues by analyzing less prominent futures contracts traded in emerging markets, which offers new insights and promotes

⁸ Relative tick size is defined as the lowest price tick (or minimum tick size) of traded instrument divided by its price.

⁹ On 27 July 2020, Borsa Istanbul (BIST) TRY based price and return stock indices were updated through dividing them by 100 with the approval of Exchange authority in order to simplify impractical high values with the aim of increasing operational efficiency. (eg. 24 July 2020 old BIST-30 spot index close value of 134,849.652 was revalued and listed as 1348.49 at opening on next business go-live date 27 July 2020). As result, for BIST-30 index futures: (a) underlying value factor was increased to 1 from 0.001; (b) contract size was decreased from 100 to 10; (c) price quotations were multiplied by 10; and (d) minimum tick size (price tick) was increased from 0.025 to 0.25 subsequent to implemented rule amendments. (eg. old futures quotations as 135.600 – 135.625 – 135.650 were requoted as 1356.00 – 1356.25 – 1356.50 as of changes took effect)

¹⁰ Including tick by tick order and trade electronic message fingerprints of self-declared HFT accounts (encrypted- identity hidden form) as provided under special consent by Borsa Istanbul.

the diversification of future evidence. The findings contribute to the ongoing local debate¹¹ on whether an order-to-trade (OTR) fee or penalty should be imposed on ultra-speed traders with regard to market welfare.

The remainder of the study is organized as follows: Section 2 presents the data samples and elaborates the variables, hypotheses, and empirical models used in the research. Section 3 reports the findings, and Section 4 concludes.

6.2 Data and Models

The analysis involves two datasets, comprising millisecond stringed tick-by-tick chronological trade and electronic message information for the nearest maturity BIST-30 index futures, obtained directly from exchange sources. These raw data files are processed to create 30-minute and daily interval samples. Covering 328 regular full trading days between December 27, 2019, and April 21, 2021, the sample range is strategically chosen to ensure symmetrical representation of pre- and post-periods surrounding the minimum tick size change on July 27, 2020, while mitigating bias from extended bullish or bearish trends. An aggregated price change of 0.008% across the dataset supports the absence of any directional trend.

In alignment with the study's research objectives, a multi-step longitudinal analytical framework is adopted, as explained below. To address potential endogeneity issues previously identified in related HFT-liquidity studies (e.g., Hendershott and Riordan, 2013; Ben Ammar et al., 2020), an instrumental variable (IV) approach is utilized, and the two-stage least squares (2SLS) method is employed to compute regression specifications unless the Wu-Hausman test (Hausman, 1978) preliminarily fails to affirm this concern for any likely candidate regressor(s).¹² The IV is required to correlate with the original predictor while remaining independent of the error terms. Drawing on Ekinçi and Ersan (2022), n-period lagged explanatory measure(s) are selected as IVs to meet these criteria¹³. Moreover, given the presumed natural

¹¹ Further details can be found in the country regulatory and supervisory authority Capital Market Board of Türkiye Chairman İbrahim Ömer Gönül's Sabah newspaper interview on 20.01.2023. <https://spk.gov.tr/data/63ca543b8f95db1b14de5f04/Sabah%20Gazetesi%20R%C3%BCB1.pdf>

¹² If the test does not significantly reject the null hypothesis of exogeneity (i.e. p-value > 0.05).

¹³ In cases where the 2SLS regression approach is necessarily applied, post-estimation diagnostics, such as the Sargan-Hansen (Sargan, 1958 and Hansen, 1982) test and weak instruments test, will be conducted to verify that the instruments are both valid and relevant in effectively addressing endogeneity.

distribution characteristics of the data, including serial correlation, heteroskedasticity, and multicollinearity, p-values of coefficients in the models are calculated using the Driscoll-Kraay extension of Newey-West HAC robust standard errors (Newey and West, 1987; Driscoll and Kraay, 1998), as recommended by Hasbrouck and Saar (2013). The stationarity condition of the variables is also ensured prior to executing model analyses.

6.2.1 Model 1

HFTs' liquidity tendencies regarding the minimum tick size change event are first scrutinized employing the updated version of Chae et al.'s (2013) regression method, which underscores the time series interactions between asset returns and net-buy volume by HFTs. A significant and negative contemporaneous relationship between these variables would imply that HFTs enhance liquidity in the market or vice versa. In this model, the HFT net-buy volume is replaced with HFT active volume as an explanatory variable (AV). Within this framework, low-latency trades are categorized as active unless the execution is triggered by an associated non-HFT counter-order bearing a later time-stamp. The liquidity provision (taking) behavior of HFTs should be connected with negative (positive) active volume coefficient finding hereby as well. It is noteworthy that this innovative independent variable may be more suited to short-sell-allowed futures data than the authors' stock trading-derived net-buy volume alternative, as it more comprehensively reflects the liquidity attributes of derivative HFTs. Specifically, the model shown below, Eq. 6.1, is executed in the process:

$$r_{i,t} = c_i + \beta AV_{i,t} + \theta TUR_{i,t} + \varepsilon_{i,t} \quad (6.1)$$

r and AV denote futures log return and prior explained HFT active volume during the observed period. As a control variable, TUR represents total traded turnover value and c bears fixed-effects. Each sample day trading session (t) is layered by eighteen 30-minute intervals (i). The magnitude and significance variability of before and after change period samples' AV coefficient findings give insight regarding the extent of the impact.

6.2.2 Model 2

Next, the following specification Eq. 6.2 is estimated, as outlined by Ekinçi and Ersan (2022), across pre- and post-tick size terms to assess whether the enforced regulatory change alters the relevant market dynamics. This approach also facilitates the investigation of the interactions between HFT activity and liquidity, with particular emphasis on spread measures as the response variable.

$$ES_{i,t} = p_1 HQ_{i,t} + p_2 TV_{i,t} + p_3 VOL_{i,t} + \varepsilon_{i,t} \quad (6.2)$$

ES is recognized as the dependent variable measuring effective spread.¹⁴ The primary explanatory variable, HQ quantifies the total number of low-latency electronic order messages, including submission, modification, cancellation, or, execution update reason types. Control variables for trade volume (TV) and volatility (VOL) are incorporated. VOL is calculated by dividing the difference between period high and low prices by half of their sum. In accordance with the analytical framework, the 30-minute intraday sample is applied to estimate the model.

Zero-mean standardization is conducted on each non-dummy regressor, consistent with the existing literature (e.g., Brogaard et al., 2017).¹⁵ Such transformation enables capturing fixed effects and eliminates the need for an intercept term.

6.2.3 Model 3

In the latter part of the analysis, the relative tick size variable (REL_{index}) is examined to identify any connection with HFT liquidity flow, and whether this potential relationship exhibits significant deviations following the tick size rule change. To evaluate market low-latency participation, the HFT order magnitude (HO_{index}) is used as the dependent variable, defined as the total of HFT order submissions and modifications divided by a constant of 1,000 for each sample day (t). The daily time-series regression model, based on Frino et al. (2015), is presented as follows:

$$HO_{index,t} = c_{index,t} + k_1 HO_{index,t-1} + k_2 REL_{index,t} + k_3 C_{j,t} + k_4 d_t + \varepsilon_t \quad (6.3)$$

¹⁴ Effective spread (ES) is determined by doubling the absolute difference between executed trade price and the midpoint price.

¹⁵ By subtracting the mean and dividing by the standard deviation over the relevant period.

REL_{index} denotes the lowest (absolute) valid tick size divided by the volume-weighted average prices (VWAP). C stands for the vector of market quality control variables: TUR_{index} (Eq.6.1) and VL_{index} – defined as the high-low ratio (VOL , Eq.6.2) – on a given day t . d is an indicator variable set to 1 for trading days after the rule change (ex-post) and 0 for before (ex-ante), while c signifies the constant term. The analysis uses a daily interval time-series approach and encompasses data from a total of 328 trading days.

6.2.4 Robustness check

The final aspect of the empirical approach focuses on the robustness-checking process. For this evaluation, gold futures (USD/Ounce) are introduced as a control instrument due to their dominant HFT activity and stability in tick size across the sample period. Consequently, the original regression model in Eq.6.3 is modified as:

$$HO_{gold,t} = c_{gold,t} + n_1 HO_{gold,t-1} + n_2 REL_{gold,t} + n_3 A_{j,t} + \varepsilon_t \quad (6.4)$$

HO_{gold} , c_{gold} and REL_{gold} are computed identically to their definitions in the unrestricted version of Eq.6.3. A represents the vector of predefined control variables, TUR_{gold} and VL_{gold} , for the control instrument (gold futures) in a similar manner. The estimation procedure and data interval of treatment model are replicated to attain statistical conformity for the robustness check.

6.3 Results and Discussion

6.3.1 Descriptive statistics

Table 6.1 provides key insights into BIST-30 index futures market activity during the sample period, including daily average values for both pre-change and post-change phases. In the second phase, the average daily number of electronic order quote updates reaches approximately half a million. During a typical trading day within the sample period, 169,140 orders are submitted on average, of which 71,248 are ultimately executed. Order modification, as essential as order entry, is a crucial and frequently employed strategy among HFTs to facilitate liquidity actions. The traded notional value in USD ranges from a minimum of 540.5 million USD to a maximum of 3,380.4 million USD, with a daily mean of roughly 1,451 million USD. Upon verifying the data period selected for its no-trend pricing behavior, the market shows neutrality between the initial and final observations, resulting in a zero overall return;

meanwhile, the average daily log returns increase from 0.009% to 0.016%. Broadly, HFT accounts participate in one-of-two order quote update and one-of-four trade execution messages, both of which have moderately grown over time in ex-ante and ex-post comparisons. The increasing penetration of low-latency trading in the market stems from Borsa Istanbul's recent technological advancements such as enhanced colocation facilities and a streamlined trading system.

Table 6.1 : Trading Patterns in BIST-30 Index Futures.

		Overall	Ex-Ante	Ex-Post	% Diff.
Number of Quote Updates	(Market)	480.7	438.3	517.1	+17.98
	(HFT %)	50.53	47.43	53.92	+6.49
Number of Order Submissions	(Market)	169.1	164.4	172.2	+4.74
	(HFT %)	38.59	34.27	42.64	+8.37
Number of Order Modifications	(Market)	146.9	130.3	154.3	+18.41
	(HFT %)	80.93	79.27	87.30	+8.03
Number of Trade Counts	(Market)	71.21	66.41	72.84	+9.68
	(HFT %)	24.13	20.92	27.47	+6.55
Trading Volume (contracts)	(Market)	750.1	784.2	728.1	-7.15
	(HFT %)	23.21	19.71	26.71	+7.00
Traded Notional Value (10 ³ USD)	(Market)	1,450.5	1,531.3	1,395.6	-8.86
	(HFT %)	23.86	21.08	26.41	+5.33
Avg. Trading Prices	(Level)	1,352	1,312*	1,460	–
Term Price Change (%)		0.179	0.160	0.190	–
Observations (days)		328	144	184	–

Notes: This table presents median values of given daily market activity parameters. Ex-ante and Ex-post change terms are determined considering enforced minimum tick-size rule change on July 27, 2020 (i.e. from 0.025 to 0.25) within the entire sample period between December 27, 2019 and April 21, 2021. HFT percentage values are calculated according to Exchange-provided identification information for self-declared low-latency accounts. Quote updates consist total of electronic order messages for submission, modification, cancellation or execution. Number of quote updates, number of order submissions, number of trades, trading volume (contracts) and traded notional value (10³ USD) is divided by 1,000. Trade related statistics are given as two-rounded buy sell logic except number of trade count. Values are shown for the most liquid, closest expiry, of traded chain, quoted as XU030c1.

*To allow for a proper comparison, the variable is adjusted by multiplying it by 10 in the ex-ante period, considering the tick-size change. The first and the last (adjusted) closing price pairs for the ex-ante and ex-post periods are provided as: 1,399 – 1,358 and 1,358 – 1,398, respectively.

The summary statistics for variables considered within the subsequent analysis is illustrated in Table 6.2. Panels A and B highlight the strengthened presence of HFT, evidenced by both ex-ante and ex-post mean values of the *AV*, *HQ*, and *HO*. Spreads, represented by *ES*, widened by 15% following the tick size adjustment. Conversely, while security prices remained largely stable, the *REL* measure narrowed by approximately 10% in the latter period, likely due to declining trading volumes, which, as the denominator, led to higher VWAPs. Retail non-HFT interest to the market that peaked during the COVID period has been gradually declining since then, possibly

contributing to lower turnover, as this investor group accounts for an extensive portion of trading activity.

Table 6.2 : Overview of Summary Statistics for Model Variables.

Panel A: Ex-Ante Term						
Measure	Symbol	Mean	SD	Min	Max	Obs.
Log Return (%)	r	-0.0004	0.2257	-0.4760	0.4499	2592
HFT Active Volume	AV	4,990	3,329	1,085	13,213	2592
Traded Value	TUR	76,421	42,482	22,780	174,592	2592
Effective Spread (bp.)	ES*	4.440	3.503	0.000	11.302	2592
HFT Quote Updates	HQ	13,050	8,840	2,980	35,873	2592
Trade Volume	TV	38,859	21,756	11,638	89,130	2592
Volatility (%)	VOL	0.4270	0.2715	0.1416	1.1586	2592
HFT Order Magnitude	HO	192.80	105.14	73.56	462.54	144
Relative Tick Size (bp.)	REL	1.965	0.210	1.682	2.357	144
Daily Turnover	TUR	1,618,203	482,632	904,779	2,632,076	144
Daily Volatility (%)	VL	2.165	1.258	0.796	5.751	144
Panel B: Ex-Post Term						
Measure	Symbol	Mean	SD	Min	Max	Obs.
Log Return (%)	r	0.0020	0.2456	-0.4886	0.4682	3311
HFT Active Volume	AV	5,192	3,307	1,164	13,168	3311
Traded Value	TUR	70,298	39,693	21,060	167,147	3311
Effective Spread (bp.)	ES	5.086	3.826	0.000	12.304	3311
HFT Quote Updates	HQ	15,437	9,516	3,042	38,317	3311
Trade Volume	TV	36,778	20,132	11,628	84,764	3311
Volatility (%)	VOL	0.4581	0.2253	0.1813	1.0211	3311
HFT Order Magnitude	HO	222.49	85.61	103.44	423.76	184
Relative Tick Size (bp.)	REL	1.750	0.197	1.487	2.042	184
Daily Turnover	TUR	1,411,498	339,308	770,686	2,184,879	184
Daily Volatility (%)	VL	2.236	0.953	1.018	4.483	184
Panel C: Full-Sample						
Measure	Symbol	Mean	SD	Min	Max	Obs.
Log Return (%)	r	0.0009	0.2370	-0.4886	0.4682	5903
HFT Active Volume	AV	5,103	3,317	1,085	13,213	5903
Traded Value	TUR	72,987	41,050	21,060	174,592	5903
Effective Spread (bp.)	ES	4.802	3.701	0.000	12.304	5903
HFT Quote Updates	HQ	14,389	9,300	2,980	38,317	5903
Trade Volume	TV	37,692	20,884	11,628	89,130	5903
Volatility (%)	VOL	0.4445	0.2471	0.1416	1.1586	5903
HFT Order Magnitude	HO	209.46	95.68	73.56	462.54	328
Relative Tick Size (bp.)	REL	1.845	0.229	1.487	2.357	328
Daily Turnover	TUR	1,502,246	449,055	770,686	2,632,076	328
Daily Volatility (%)	VL	2.204	1.096	0.796	5.751	328

Notes: This table presents descriptive statistics for variables employed in Models 1, 2, and 3 at specified intervals (30-minute or daily) across ex-ante, ex-post, and overall data periods, displayed in Panels A, B, and C, respectively. Definitions of each liquidity, HFT and control measure are detailed in Eq.6.1, Eq.6.2 and Eq.6.3 as presented throughout the study. Ex-ante and Ex-post change terms are determined considering enforced minimum tick-size rule change on July 27, 2020 (i.e. from 0.025 to 0.25) within the entire sample period between December 27, 2019 and April 21, 2021. Traded Value (TUR) is divided by 1,000. "Obs.", "bp." and "SD" denote observations, basis points and standard deviation, in that sequence.

*To allow for a proper comparison, the variable is adjusted by multiplying it by 10 in the ex-ante period, considering the tick-size change.

To expand on the preliminary analysis, Table 6.3 displays the Kendall correlation values for individual variable pairs, arranged by model. Coefficient estimates reveal weak bilateral interactions between the dependent variables (r and ES) and the main explanatory HFT variables (AV and HQ) in Models 1 and 2. For Model 3, however, the direction of the relationship between HO (response) and REL (predictor) reverses from positive to negative after the tick-size rule change. Despite this, certain HFT and control variables demonstrate strong correlations in Models 2 and 3,¹⁶ in line with existing literature.

Table 6.3 : Correlation Analysis.

Panel-A.Model-1 Variables	r	AV	
AV	(.0199) .0331		
TUR	(.0367) (.0197)	.1410 .1055	
Panel-B.Model-2 Variables	ES	HQ	TV
HQ	.2794 .2462		
TV	.3967 .3513	.7456 .7606	
VOL	.4753 .4514	.6846 .6224	.8325 .8017
Panel-C.Model-3 Variables	HO	REL	TUR
REL	.4503 (.3156)		
TUR	.6090 .7327	.0166 (.3448)	
VL	.5670 .4189	.2708 .0964	.6902 .5804

Notes: This table presents pairwise Kendall correlation coefficients for variables employed in Models 1, 2, and 3 at specified intervals (30-minute or daily) across ex-ante and ex-post data periods, displayed in Panels A, B, and C, respectively. Definitions of each liquidity, HFT and control measure are detailed in Eq.6.1, Eq.6.2 and Eq.6.3 as presented throughout the study. Ex-ante and Ex-post change terms are determined considering enforced minimum tick-size rule change on July 27, 2020 (i.e. from 0.025 to 0.25) within the entire sample period between December 27, 2019 and April 21, 2021. In the table, each cell displays pre-change values in the top row and post-change values in the bottom row. Negative values are shown in brackets for clarity. Natural logarithmic transformations are applied to the measures AV, TUR, HQ and TV. All variables are winsorized at the 0.05th and 99.5th percentiles to mitigate the impact of outliers and ensure data consistency.

In general, the time-dependent distributions of either orders, trades or electronic messages all exhibit U-shaped patterns during the day. Yet, there are no substantial differences in the relevant intraday trading measures between pre- and post-amendment intervals, aside from a few exceptions.

¹⁶ Such as HQ vs. TV , HQ vs. VOL , and TV vs. VOL or HO vs. TUR , HO vs. VL and TUR vs. VL .

6.3.2 Model findings

6.3.2.1 Model 1

Table 6.4 presents the return-volume regression estimates (Eq. 6.1) examining the liquidity provision tendencies of HFT accounts. While no significant relationship between variables was found in the ex-ante period (Panel A), the positive and significant coefficient for active volume (AV) in Panel B indicates a negative impact of HFT on market liquidity following the minimum tick-size change. This effect appears to reflect a demand for liquidity, rather than a supply, in the ex-post period, likely due to the increased rewards associated with voluntary market-making practices amidst larger achievable spreads and heightened low-latency participation. The findings contrast with those of Chae et al. (2013), who report that algorithms consistently inject liquidity into the (warrant) derivatives market through strategies such as scalping, high-level arbitrage, or momentum trading.

Table 6.4 : Returns and HFTs Active Volume.

	Panel-A: Ex-Ante	Panel-B: Ex-Post
constant	0.0017 (1.53)	0.0026** (2.52)
AV	-0.0205 (-0.40)	0.0236** (2.05)
TUR	-0.0191** (-2.03)	-0.0192** (-2.09)
Obs.	2,592	3,311
AR2	0.03	0.02
F-Statistics	4.73	4.62
Breusch-Godfrey LM Test	0.56	0.33
Breusch-Pagan-Godfrey Test	685***	759***
Wu-Hausman Test	1.21	3.90

Notes: This table presents the OLS estimation results for the regression model specified in Eq.6.1, covering a total of 5,903 30-minute intervals across the ex-ante and ex-post data periods, as shown in Panels A and B. Ex-ante and Ex-post change terms are determined considering enforced minimum tick-size rule change on July 27, 2020 (i.e. from 0.025 to 0.25) within the entire sample period between December 27, 2019 and April 21, 2021. The definitions of the variables are thoroughly provided in the study. With the exception of r , natural logarithmic transformations are applied to the measures. All variables are winsorized at the 0.05th and 99.5th percentiles to mitigate the impact of outliers and ensure data consistency. AV and TUR are demonstrated by multiplying each by 100. The stationarity of the variables is verified. Obs. and AR2 denote number of observations and adjusted R-squared value along with selected diagnostics given as Breusch-Godfrey LM test for serial correlation, Breusch-Pagan-Godfrey test for heteroskedasticity and Wu-Hausman tests for endogeneity. T-statistics are given in brackets and significance levels are computed by referring Newey-West HAC robust standard errors. P-values are denoted as * < 0.10, ** < 0.05 and *** < 0.01.

The Wu-Hausman test results indicate no significant evidence of endogeneity for the combination of any likely candidate regressors (AV and TUR) in the OLS estimation under the null hypothesis of exogenous variables. Each independent variable in the structural equation is individually validated using an appropriate procedure to ensure they are exogenous and free from endogeneity. For brevity, these results are not shown here but are available upon request.

6.3.2.2 Model 2

From a different liquidity perspective, the relevant Eq.6.2 spread calculations for both ex-ante and ex-post data samples are provided in the panels of Table 6.5 respectively. Aligned with the findings of the prior model (1), the main explanatory variable HQ is statistically significant only in the ex-post term. The positive sign of the associated coefficient indicates a detrimental influence of HFT engagement on the dependent variable, market quality indicator ES . As a result, it can be argued that a increase in HFT order activity leads to a widening of effective spreads, or alternatively, that as effective spreads rise, this type of traders seems to appear more on the market, particularly in the later period. One could infer that, the observed decline in market trading volumes over the post-change period, despite the growth in order message traffic, may independently contribute to larger effective spreads, regardless of a direct HFT effect. The derived findings are consistent with Lee (2015), who examines futures markets, yet they contradict the views of equity market researchers like Hendershott et al. (2011), who claim that HFT tends to narrow relevant spread measures. Furthermore, the control variable volatility (VOL) is found to be significant and positive for both samples, as highlighted by Olgun et al. (2024). On the other hand, no meaningful relationship was detected for trading volume (TV) within the estimated vector throughout the investigation.

6.3.2.3 Model 3

In the subsequent stage of the empirical analysis, the impact of the exogenous variable, relative tick size (REL_{index}), on HFT market dynamics is assessed. Eq.6.3 is run for this purpose, and the resulting estimates are detailed in left panel (A) of Table 6.6. The results show that relative tick size (REL_{index}) serves as a predictor of HFTs order transactions (HO_{index}), with a positive relationship. This insight supports prior evidence, including that of Yamada (2022), which propose that larger relative tick sizes elevate momentum and promote trading willingness in the market by expanding profit margins. Nevertheless, the significant dummy coefficient (d) finding reveals notable variation in the association between the periods before and after the minimum tick-size rule change. As expected, the market quality parameters, TUR_{index} and VL_{index} (i.e. turnover and volatility), are both significant and positive (e.g. Ben Ammar et al., 2020).

Table 6.5 : Spread Dynamics and HFT.

	Panel-A: Ex-Ante	Panel-B: Ex-Post
HQ	0.0831 (0.33)	0.5807** (2.08)
TV	0.1755 (0.79)	-0.2680 (-0.97)
VOL	4.3218*** (7.80)	5.1988*** (4.98)
Obs.	2,592	3,311
AR2	0.23	0.20
Breusch-Godfrey LM Test	0.39	0.92
Breusch-Pagan-Godfrey Test	71.54***	91.59***
Wu-Hausman Test	33.92***	35.52***
Sargan-Hansen Test (p-value)	0.11	0.17
Weak Instruments Test	161	62.8

Notes: This table presents the 2SLS estimation results for the regression model specified in Eq.6.2, covering a total of 5,903 30-minute intervals across the ex-ante and ex-post data periods, as shown in Panels A and B. Ex-ante and Ex-post change terms are determined considering enforced minimum tick-size rule change on July 27, 2020 (i.e. from 0.025 to 0.25) within the entire sample period between December 27, 2019 and April 21, 2021. The definitions of the variables are thoroughly provided in the study. With the exception of ES and VOL, natural logarithmic transformations are applied to the measures. All variables are standardized by subtracting the mean and dividing by the standard deviation over the relevant period. ES is expressed in basis points (bp.), and VOL is scaled by a factor of 100. The stationarity of the variables is verified. Obs. and AR2 denote number of observations and adjusted R-squared value along with selected diagnostics given as Breusch-Godfrey LM test for serial correlation, Breusch-Pagan-Godfrey test for heteroskedasticity and Wu-Hausman tests for endogeneity. T-statistics are given in brackets and significance levels are computed by referring Newey-West HAC robust standard errors. P-values are denoted as * < 0.10, ** < 0.05 and *** < 0.01.

The Wu-Hausman test results indicate a significant evidence of endogeneity for the combination of any likely candidate regressors (HQ and TV) in the OLS estimation under the null hypothesis of exogenous variables. Each independent variable in the structural equation is individually validated using an appropriate procedure to ensure they are exogenous and free from endogeneity. For brevity, these results are not shown here but are available upon request. As instrumental variables (IVs), n-period lagged endogenous regressors (HQ and TV) are specified. Post-estimation, it is confirmed that the instruments are valid and relevant in addressing endogeneity, using the Sargan-Hansen test (null hypothesis: instruments are valid and uncorrelated with error term) and the weak instruments test (F-statistics above 10), respectively.

6.3.3 Robustness check

The robustness examination is shown in the right panel (B) of Table 6.6. The analysis reveals that the primary explanatory variable, relative tick size (REL_{gold}) is significant for gold futures. In contrast, the sign of the parameter is now negative, unlike the treatment security, index futures, coefficient (REL_{index}). This might be attributed to the presence of (mostly HFT-oriented) official market maker participants involved in the Exchange's revenue-sharing campaign, which specifically targets less-liquid futures contracts, such as gold futures and single-stock futures, rather than heavily traded BIST-30 index futures. For these instruments, the competitive landscape is entirely dissimilar in this regard. Due to the distinct strategies, tactics, or inventory approaches employed, it is not surprising, therefore, that sharp variations exist in the behaviors of voluntary and obligated HFT market makers (HO_{index} and HO_{gold}). Additionally, the

control variables turnover (TUR_{gold}) and volatility (VL_{gold}) are calculated as significant with positive signs, similar to the treatment model (3).

Table 6.6 : Relative Tick Size and HFT Activity.

	Panel-A: Index Fut. (Treatment Var.)	Panel-B: Gold Fut. (Control Var.)
constant	−0.0055*** (−3.63)	0.0159* (1.66)
HO_{t-1}	0.4838*** (12.0)	0.4613*** (8.79)
REL	0.0064*** (3.52)	−0.3520** (−2.01)
TUR	8.91E-09*** (6.19)	1.40E-07*** (7.74)
VL	0.0504* (1.92)	0.3855* (1.70)
d	0.0033*** (5.09)	
Obs.	328	328
AR2	0.70	0.75
F-Statistics	155	249
Breusch-Godfrey LM Test	0.07	0.24
Breusch-Pagan-Godfrey Test	2.48**	6.88***
Wu-Hausman Test	3.32	3.27

Notes: This table presents the OLS estimation results for the regression models specified in Eq.6.3 and Eq.6.4. The analysis spans 328 daily observations, from December 27, 2019, to April 21, 2021, examining both treatment and control variables for index and gold futures, displayed in Panels A and B. The definitions of the variables are thoroughly provided in the study. Natural logarithmic transformations are applied to the measures HO and TUR. All non-dummy variables are winsorized at the 0.05th and 99.5th percentiles to mitigate the impact of outliers and ensure data consistency. HO is scaled by a factor of 100. The stationarity of the variables is verified. Obs. and AR2 denote number of observations and adjusted R-squared value along with selected diagnostics given as Breusch-Godfrey LM test for serial correlation, Breusch-Pagan-Godfrey test for heteroskedasticity and Wu-Hausman tests for endogeneity. T-statistics are given in brackets and significance levels are computed by referring Newey-West HAC robust standard errors. P-values are denoted as * < 0.10, ** < 0.05 and *** < 0.01.

The Wu-Hausman test results indicate no significant evidence of endogeneity for the combination of any likely candidate regressors (REL and TUR) in the OLS estimation under the null hypothesis of exogenous variables. Each independent variable in the structural equation is individually validated using an appropriate procedure to ensure they are exogenous and free from endogeneity. For brevity, these results are not shown here but are available upon request.

6.4 Final Remarks

This research scrutinizes the interdependency between minimum (absolute) and relative tick size parameters and HFT-driven liquidity in the Turkish BIST-30 index futures market, with a particular focus on the recently implemented tick size increase (i.e. reverse decimalization). The findings are as follows: HFTs' liquidity-consuming behavior negatively affects market quality by absorbing tradeable volumes. From other perspective, HFT presence exacerbates intraday effective spreads—a key liquidity measure—at the cost of retail investors. Post tick-size rule change, these adverse

liquidity impacts have notably intensified, suggesting that the minimum tick-size parameter may have a contributory role in shaping algorithmic strategies. It is also worth noting that the recent steady upward trend in electronic message traffic for the instrument may inherently encourage some degree of HFT involvement, irrespective of the minimum tick-size change.

Moreover, it is reported that relative tick size influences HFT order propensity, serving as a marked exogenous factor. This relationship is verified across both treatment and control (robustness) models, despite the emergence of opposing effects. Dummy variable results further underscore a significant discrepancy between ex-ante and ex-post periods.

The remarks of this study shed light on the ongoing debate surrounding the economic implications of HFT. The analysis is further extended by focusing specifically on potential trading determinants, such as tick size, that are pertinent to this investor type. By studying futures contracts, the research broadens the range of product type in the literature, moving beyond the conventionally dominant equity asset class and advancing related scientific exploration in the field of derivatives.

7. DO HIGH-FREQUENCY TRADERS FUEL MARKET HAYWIRE WHEN VOLATILITY PEAKS? EVIDENCE FROM THE TURKISH INDEX FUTURES MARKET

7.1 Context

The linkage between high-frequency trading (HFT) and volatility in both normal and stressed market environments has been rigorously examined through academic, practical, and regulatory lenses. A major concern—whether these sophisticated investors are the primary drivers, or perhaps even the cause, of abnormal asset price volatility during times of market fear—has sparked considerable discussion. Relevant studies often yield conflicting results, leading scholars to adopt various rationales in response to the persistent ambiguity surrounding the issue.

HFT proponents including Gsell (2008), Hagstromer and Norden (2013), Jovanovic and Menkveld (2016), Brogaard et al. (2018), and Ke and Zhang (2020), argue that no significant causal relationship exists between fast (algorithmic) trading and rising volatility. These studies allege that enhanced latency aids in efficient price discovery by mitigating information asymmetry between slow liquidity providers and front-runners, which in turn lowers inventory costs for market makers. In contrast, critics such as Sanchez Serrano (2020), Yang and Zhu (2020), and Breckenfelder (2024) claim that HFT exacerbates volatility and trading costs due to intensified adverse selection. The high correlation among HFT strategies, driven by interconnectedness and synchronization of fast agents, may trigger overreaction and herding behavior in trading venues. As markets heat up, particularly during bearish trends, liquidity provision from these investors is quickly depleted, amplifying the systemic risk of a potential ‘crash,’ according to HFT critics. Boehmer et al. (2021), through extensive research, analyzed 42 equity markets worldwide to assess the impact of algorithmic trading (AT). Their results, consistent across diverse trading environments, suggest that, on average, AT increases short-term volatility, although it improves liquidity and informational efficiency. It is also documented that higher volatility driven by AT

leads to lower liquidity. Hence, the same traders who cause volatility may also deplete liquidity, or vice versa.

In this regard, numerous authors seek to discern meaningful differences in HFT activity across low and high volatility periods. Based on approximated excess volatility sample periods in these studies¹⁷, the findings point to a lack of consensus, with polarized views on HFT's time-specific net welfare effect. Brogaard et al. (2018) maintain that HFTs provide liquidity to non-HFTs in response to extreme price movements (EPMs) by absorbing trade imbalances in the NASDAQ market. Kang et al. (2020) reveal a multidimensional positive (negative) relationship between HFT and short-term volatility during high (low) toxicity periods in the Korean futures market. Furthermore, Bazzana and Collini (2020) and Ben Ammar and Hellara (2022) investigate the relationship between HFT and volatility in the S&P 500 ETF and CAC40 stock markets by focusing exogenous shocks such as the US presidential election and Brexit. The researchers contend that low-latency algorithms elevate volatility during market crashes by transmitting excessive amount of ultra-speed orders to exchange systems. On the contrary, Foucault et al. (2016) and Brogaard et al. (2018) assert that HFT does not intensify price volatility during market turbulences. Dastan (2021) demonstrates that HFTs diminish their trading activity and liquidity supply during EPM periods in the Borsa Istanbul (BIST) equity market. This tendency is attributed to HFTs' ability to anticipate market conditions ahead of unsophisticated or retail investors. Ekinici and Ersan (2022) further maintain that no significant interaction exists between HFT and volatility in BIST equity market, where technological infrastructure has only recently been established and HFTs have not yet achieved sufficient penetration or influence. Moreover, Bodur (2022) scrutinize flash events impacting Turkish blue-chip stocks TCELL and KCHOL on February 22, 2017. Prior to market close, both stocks experienced nearly 10% price drops within a second before recovering swiftly. The study remarks that the Volume-Synchronized Probability of Informed Trading (VPIN) metric, an indicator of order flow toxicity for

¹⁷ Intuitively defined by following various methods including: (i) observing around before and after exogenous shocks like US May 2010 Flash Crash (Bazzana and Collini 2020), Korean Credit Rating Upgrade (Kang et al. 2020), Post-Earning Announcement Drifts (PEAD) (Ke and Zhang 2020); (ii) estimating tail events within the sample during which market return is at lowest/highest percentiles (Jain et al. 2016); (iii) using residuals of market and return model (Brogaard et al. 2018); and (iv) picking top x% of the highest realized (local) volatility data points (Boehmer et al. 2021).

informed trading, displayed a sustained upward trend preceding TCELL's sudden dip, suggesting algorithmic trading-induced pressures, whereas no significant early warning signals were detected for KCHOL.

The study intends to spotlight HFT activity in the BIST index futures market by centering its bilateral impact on volatility under both normal and extreme conditions. The analysis benefits from a unique dataset¹⁸, which allows bypassing concerns related to indirect HFT identification tools, measures, or proxies (Elizarov et al., 2017). The findings aim to deepen the existing microstructure framework by providing not only supplementary but also extensive and diversified international evidence from a rarely studied asset class, as opposed to the commonly analyzed equity products. As noted by Ersan and Ekinici (2016) and Olgun et al. (2024), scientific work on HFT in emerging markets is considerably less compared to developed markets. Additionally, the BIST index futures market holds a steady global ranking among its peers¹⁹, demonstrating high growth potential, which warrants further examination from a different perspective. Notably, the BIST equity market's extremely high share turnover velocity and subsequent volatility (ranked second and first globally, respectively; see Table 6.1) underscore the practical relevance of this study, particularly in terms of the hedging capabilities of index futures from a risk management viewpoint²⁰.

Toward this end, the study intends to deliver valuable implications into both theoretical and practical realms across various fields. Initially, it enriches the understanding of HFT behavior by addressing perspectives from the relatively underexplored grounds of emerging market and derivatives instruments (i.e., futures). Moreover, the results are expected to provide useful guidance for policymakers in the development of the

¹⁸Including tick by tick order and trade electronic message fingerprints of self-declared HFT accounts (encrypted- identity hidden form) as provided under special consent by Borsa Istanbul.

¹⁹ As noted by most recent World Federation Exchanges (WFE) annual statistics report, BIST index futures market generated 60.8 million contracts trade volume in 2023 by accounting a total of USD 143.6 billion notional value. It is ranked as the 2nd most actively traded index futures market among 11 derivatives exchanges, following Eurex, within its regional group Europe-Middle East-Africa (WFE Annual Statistics Guide, 2023)

²⁰ Referring to the WFE 2023 annual statistics, BIST equity market the average daily turnover value is reached almost USD 5.35 billion which is ranked as the 3rd most actively traded equity market among 48 stock exchanges, following Euronext and CBOE Europe, within its regional group Europe-Middle East-Africa (WFE Annual Statistics Guide, 2023).

HFT rulebook draft being worked on²¹. The steady HFT dominance in the current decade underpins the trade-off between enriched liquidity and increased systemic risks for market participants. Existing literature emphasize that regulatory decisions in this field are delicate and must be prudently evaluated prior to implementation, taking into account different HFT strategies and their segmentation, to avoid foregoing its potential benefits (Sanchez Serrano, 2020; Boehmer et al., 2021; Ben Ammar and Hellara, 2022).

Table 7.1 : Selected Trading Dynamics in Major Global Stock Exchanges.

Panel-A: Dispersion of Broad Stock Market Indices				
Exchange	Country	Region	Name of Index	HighLow
Borsa Istanbul	Türkiye	EMEA	BIST 100	0,6369
The Egyptian Exchange	Egypt	EMEA	EGX 30	0,5531
Tehran Stock Exchange	Iran	EMEA	TEPIX	0,4940
Athens Stock Exchange	Greece	EMEA	ASE	0,3591
Hong Kong Exchanges	Hong Kong	Asia-Pacific	HKEX Lar. Cap	0,3551
Warsaw Stock Exchange	Poland	EMEA	WIG	0,3433
Shenzhen Stock Exchange	China	Asia-Pacific	SZI	0,2816
Japan Exchange Group	Japan	Asia-Pacific	TOPIX	0,2616
Zagreb Stock Exchange	Croatia	EMEA	CROBEX	0,2484
NASDAQ OMX Helsinki	Finland	EMEA	OMXH	0,2385
BME Spanish Exchanges	Spain	EMEA	IGBM	0,2087
Saudi Exchange (Tadawul)	Saudi Arabia	EMEA	TASI	0,1814
Nasdaq OMX Stockholm	Sweden	EMEA	OMXS	0,1806
Deutsche Boerse AG	Germany	EMEA	CDAX	0,1314
Panel-B: Turnover Velocity of Domestic Shares				
Exchange	Country	Region	Velocity (%)	Change (%)
Shenzhen Stock Exchange	China	Asia-Pacific	395,4	+ 0,8
Borsa Istanbul	Türkiye	EMEA	323,0	+ 52,0
Taipei Exchange	Taiwan	Asia-Pacific	283,1	- 45,1
Shanghai Stock Exchange	China	Asia-Pacific	192,3	- 14,6
Korea Exchange	South Korea	Asia-Pacific	172,4	- 13,1
B3 - Brasil Bolsa Balcão	Brazil	Americas	121,0	- 41,7
Japan Exchange Group	Japan	Asia-Pacific	103,6	- 5,3
TMX Group	Canada	Americas	60,2	- 21,7
Hong Kong Exchanges	Hong Kong	Asia-Pacific	57,3	- 5,0
ASX Australian Exchange	Australia	Asia-Pacific	50,2	- 14,7
Deutsche Boerse AG	Germany	EMEA	47,3	- 20,5
National Stock Exch. of India	India	Asia-Pacific	44,8	- 5,0
BME Spanish Exchanges	Spain	EMEA	40,9	- 13,6
SIX Swiss Exchange	Switzerland	EMEA	37,6	- 9,6
Euronext	Germany	EMEA	36,8	- 10,3
Athens Stock Exchange	Greece	EMEA	35,7	+ 4,1

Source: WFE 2023 Annual Statistics Report.

Notes: EMEA represents Europe-Middle East-Africa. HighLow dispersion (volatility) measure in Panel-A is calculated for indices as $(2023 \text{ High} - 2023 \text{ Low}) / (0.5 \times (2023 \text{ High} - 2023 \text{ Low}))$. Velocity (%) measure in Panel-B denotes 2023 Average values whereas Change differs $(2023 \text{ Average} - 2022 \text{ Average})$ velocities.

²¹ Further details can be found in the country regulatory and supervisory authority Capital Market Board of Türkiye Chairman İbrahim Ömer Gönül's Sabah newspaper interview on 20.01.2023. <https://spk.gov.tr/data/63ca543b8f95db1b14de5f04/Sabah%20Gazetesi%20R%C3%B6portaj%C4%B1.pdf>

7.2 Empirical Framework

7.2.1 Sample

The study utilizes comprehensive tick-level order and trade data for the BIST-30 index futures, directly retrieved from exchange records. Spanning 328 full trading days from December 30, 2019, to April 22, 2021, it ensures a robust time frame that mitigates biases and maintains structural consistency by excluding prolonged one-sided market conditions. The sample confirms no persistent dominance of bull or bear trends lasting beyond three months.

7.2.2 Variables

A distinct premise of this study entails employing an outright low latency variable, $HFTQ$, shown below, for day t , to track market participation directly, circumventing the reliance on proxy estimations deduced from specific order and/or trade measures.²²

$$HFTQ_t = QH_t / TQ_t \quad (7.1)$$

QH and TQ represent the number of exchange-identified HFT accounts and total electronic order messages, respectively, including updates related to submissions, modifications, executions, and cancellations.

For robustness checks, an additional trade-linked low-latency measure, $HFTT$, is computed as follows:

$$HFTT_t = VH_t / TV_t \quad (7.2)$$

where VH and TV denote the traded volumes attributed to HFT-designated accounts and the overall market for the period, respectively.

The key volatility variable, VOL , is derived from the price range of the logarithmic differences between the highest (*High*) and lowest (*Low*) prices on a given trading day (t) (Eq.7.3). This measure is less influenced by microstructure noise, which can exaggerate volatility estimates, and captures a wider array of price movements (Brandt and Diebold, 2006). Two commonly used volatility measures, VRV and VCN , are

²² Hossain (2022) points out that a majority of HFT studies adopt various proxies to identify actions of these specialized investors. Olgun et al. (2024) investigate the performance of widely applied HFT proxies in the literature.

further computed as described in Eq.7.4, and Eq.7.5, respectively, to expand the analysis for robustness validation.

$$VOL_t = \text{Log}(\text{High}_t) - \text{Log}(\text{Low}_t) \quad (7.3)$$

$$VRV_t = \text{Variance from midquote returns} \quad (7.4)$$

$$VCN_t = \text{Conditional variance from GARCH (1,1) by using period returns} \quad (7.5)$$

In order to investigate whether the effect of HFT on market quality varies between extreme and stable volatility periods, a dummy variable, *EXT*, is incorporated. This regressor is set to one when absolute returns surpass the 90th percentile, in a given interval, and zero in all other cases.²³

Control variables, such as total traded volume (*TV*), inverse settlement price (*IP*), and effective spread (*EP*), are integrated to the analysis alongside a vector (*X*).²⁴ Following the generation of time-series for each of the aforementioned variables (comprising all criteria and input measures, excluding dummy *EXT*) at the given intervals, these are winsorized at the 1st and 99th percentiles to preserve data consistency by addressing outliers.

7.2.3 Model formation

A multi-dimensional analysis framework is implemented to rigorously examine the core propositions of the study. In the first stage, the model in Eq. 7.6 is applied to assess the relationship between HFT (as the predictor) and volatility (as the response) under normal market conditions.

$$\text{Model 1: } VL_t = d_0 + d_1 VL_{t-1} + d_2 HF_t + d_3 X_t + \varepsilon_t \quad (7.6)$$

VL and *HF* signify the formulated measures for volatility (*VOL*, *VRV*, or *VCN*) and high-frequency trading (*HFTQ* or *HFTT*), respectively. *X* represents a set of predefined control variables (*TV*, *IP*, and *EP*) for the observed data, and d_0 is constant term. Lagged volatility, VL_{t-1} , is used as an instrumental variable to ensure consistent results for exogenous changes. This method mitigates potential endogeneity by

²³ The statistical cut-off point for identifying *EXT* is selected intuitively after scanning the related work.

²⁴ To prevent biases arising from potential multicollinearity, *TV* is dropped from control vector *X* when using *HFTT* as the predictor variable during robustness checks.

addressing the correlation between the predictor and the error term (Blundell and Bond, 1998).

Second, a revised version of the specification proposed by Ben Ammar and Hellara (2022) is conducted to explore potential variations in HFT behavior during high volatility periods versus calmer market conditions. By introducing the antecedent factor EXT into the unrestricted model (Eq. 7.6) for extreme volatility intervals, Eq. 7.7 is developed and calculated, as presented below:

$$\text{Model 2: } VL_t = m_0 + m_1 VL_{t-1} + m_2 EXT_t + m_3 HF_t \times EXT_t + m_4 HF_t \times (1 - EXT_t) + m_5 X_t + \varepsilon_t \quad (7.7)$$

where multiplicative terms $HF_t \times EXT_t$ and $HF_t \times (1 - EXT_t)$ represent HFT activity under high and normal volatility conditions, respectively. This formulation strengthens interpretability by unveiling non-linear relationships and conditional effects among variables, based on binary dummy values assigned to distinct volatility levels. For a deeper understanding of HFT patterns, an additional dimension is added to the interaction parameter EXT , which enables a more nuanced distinction between high-volatility periods with price increases and those with price declines. Consequently, Eq. 7.7 is repeated two more times, with EXT assigned a value of one only under the following conditions: the interval's absolute return exceeds the 90th percentile, and the price change is positive (in the first run) or negative (in the second run). If these criteria are not met, the dummy variable EXT is assigned a value of zero.

To avoid spurious results and misleading conclusions arising from endogeneity-driven challenges, the Maximum Likelihood (MLH) procedure is adopted for the regression models 1 and 2 (Eq. 7.6 and Eq. 7.7). As supported by Moral-Benito (2013) and Hsiao et al. (2002), this method is more efficient and generally produces lower variance and higher precision in parameter estimates compared to the Generalized Method of Moments (GMM) alternative (Blundell and Bond, 1998). When the cross-sectional dimension of the dynamic panel is relatively small in a finite large sample, the GMM estimator may suffer from convergence issues, over-identification, and instability, which could introduce bias into the results. Additionally, the simplicity of the likelihood function in hypothesis testing makes MLH less susceptible to specification errors when selecting appropriate instruments.

A series of diagnostic checks are performed on the regressions in Eq. 7.6 and Eq. 7.7, encompassing the Breusch-Godfrey LM test for serial correlation, the Breusch-Pagan-Godfrey test for heteroskedasticity, and the calculation of variance inflation factors (VIF) to detect multicollinearity. Moreover, the Durbin–Wu–Hausman test (Hausman, 1978) and the Wald test are carried out. The former test checks the presence of uncorrelated regressors and error terms, while the latter appraises the joint significance and interaction effects across variables.

7.3 Findings and Interpretations

7.3.1 Summary statistics

Table 7.2 highlights the summary statistics of the variables used and the associated market activity metrics across the sample period. On average, 482.4 thousand incoming order messages result in 34.5 thousand trades (equivalent to a notional turnover value of 1515 million USD) during a typical BIST-30 index futures trading session. This translates to approximately 800 order messages and 5 trades being recorded per minute for the contract.²⁵

Market dynamics indicate the escalating prominence of low-latency activities, with high-frequency trading accounting for nearly half of order quote updates (*HFTQ*) and a quarter of executed trades (*HFTT*). This growth is supported by Borsa Istanbul's advancements, such as enhanced colocation services. Intraday distributions exhibit anticipated U-shaped patterns for variables associated with order messages (*QH* and *TQ*) and trades (*VH* and *TV*).

The primary volatility measure (*VOL*) shows a median value of 0.83%, ranging from 0.24% to 5.19%. Out of 328 trading days, 33 are marked as highly volatile (*EXT* = 1) based on the outlined criteria. Among these, 17 days exhibited positive returns, while

²⁵ To facilitate a more comprehensive assessment of the metrics, recent benchmark figures for per-minute order message and trade counts from several actively traded global index futures are rounded as follows:

1. S&P 500 E-Mini Futures (CME) 10.000 and 500; 2. KOSPI 200 Futures (Korea Exchange) 8.500 and 400; 3. DAX Futures (Eurex) 8.000 and 400; 4. Nikkei 225 Futures (Osaka Exchange) 7.000 and 350; 5. Hang Seng Futures (HKEX) 6.500 and 300, etc... (Source: Bloomberg Terminal, Global Stock Index Futures Data, accessed on September 2024, Bloomberg L.P.).

16 days recorded losses, underlining a degree of balance. Excess volatility²⁶ during sell-offs (1.39%) was marginally higher than during upward price movements (0.80%). High-frequency trading activity, measured by *HFTQ* and *HFTT*, did not vary significantly between volatile and normal market conditions.

Table 7.2 : Descriptive Statistics for Variables.

Measure	Symbol	Median	Min	Max	SD
Order Messages					
HFTs	QH	244.2	66.5	781.9	131.6
Total	TQ	482.4	172.7	1324.1	193.5
HFT measure-1	HFTQ	0.506 (0.514)	0.275 (0.388)	0.864 (0.653)	0.078 (0.069)
Trades					
HFTs Volume	VH	174.3	57.2	570.1	83.7
Total Volume	TV	751.5	227.1	1881.1	256.0
Total Trade Count	TH	34.5	10.2	131.0	16.4
HFT measure-2	HFTT	0.240 (0.249)	0.125 (0.143)	0.511 (0.409)	0.057 (0.058)
Dispersion/Variability					
Volatility measure-1	VOL	0.834 (1.737)	0.238 (1.045)	5.186 (5.186)	0.633 (0.902)
Volatility measure-2	VRV	1.544 (4.263)	0.101 (2.115)	38.72 (38.72)	4.500 (8.287)
Volatility measure-3	VCN	2.491 (3.689)	1.524 (1.589)	26.68 (26.68)	2.564 (4.573)
Prices					
Log Returns	r	0.179	-10.243	5.093	1.741
Effective Spread	EP	1.389	0.005	1.759	0.306
Level	1/IP	1351	1004	1713	172

Notes: The table reports descriptive statistics regarding the employed HFT and volatility measures for closest expiry (i.e. most liquid) BIST-30 index futures (XU030c1), based on daily data. The full sample spans 328 regular trading days, covering the period from December 30, 2019, to April 22, 2021; whereas the estimations pertain to 33 (percentiled) extreme volatility days, where EXT=1, are reported separately in brackets if applicable. The definitions of the above specified measures are detailed in the study.

Order message types consist submission, modification, cancellation or execution particularly. Order message and trade volume metrics QH, TQ, VH, TV and TH are divided by 1000. Volatility variables VOL, VRV and VCN; and log return (r); and effective spread (EP) values are displayed as percentages (%) subsequent to multiply by 100. Trade related statistics are given as two-rounded buy-sell logic. "Min", "Max" and "SD" indicate minimum, maximum and standard deviation values of the specified parameters, respectively.

Kendall correlations are calculated for daily HFT, volatility, and control measures²⁷, as detailed in Table 7.3, to support conceptual insights. The findings reveal no

²⁶ Excess volatility measure is calculated as the difference between mean daily volatilities (VOL) of extreme and normal days considering direction of the market prices. In this regard, the mean volatility is 2.32% and 1.76% for extreme negative and extreme positive days. During normal conditions, by leaving aside each of these outliers, the same parameter is reported as a 0.93% (without extreme negative) and 0.96% (without extreme positive) respectively.

²⁷ Due to the rejection of the normality assumption for the series based on significant Jarque-Bera test results, a nonparametric correlation measure, Kendall's tau, is utilized to ensure robustness and validity.

meaningful bidirectional relationship between any paired HFT and volatility variables, with the highest observed correlation being 0.263 between *HFTQ* and *VCN*, present in neither stable nor extreme market periods. This aligns with Ekinçi and Ersan (2022), who found no correlation between HFT participation and volatility in blue-chip stocks within the BIST equity market. Nevertheless, the lack of direct relationships among control variable pairs verifies the absence of multicollinearity concerns.

Table 7.3 : Correlation Matrix.

		1	2	3	4	5	6	7
HFT Variables								
HFTQ	1							
HFTT	2	0.501 (0.43)						
Volatility Variables								
VOL	3	0.159 (0.13)	0.170 (0.10)					
VRV	4	0.197 (0.27)	0.168 (0.06)	0.600 (0.38)				
VCN	5	0.263 (0.28)	0.214 (0.07)	0.244 (0.40)	0.321 (0.24)			
Control Variables								
Traded Volume (TV)	6	0.162 (0.31)	0.108 (0.07)	0.497 (0.47)	0.574 (0.27)	0.325 (0.45)		
Effective Spread (EP)	7	0.052 (0.03)	-0.008 (-0.04)	0.394 (0.14)	0.302 (0.06)	0.047 (0.08)	0.309 (0.33)	
Inv.Setl. Price (IP)	8	-0.123 (0.18)	-0.288 (-0.25)	0.083 (0.12)	0.100 (0.36)	0.098 (0.25)	0.174 (0.28)	0.078 (0.13)

Notes: The table reports Kendall correlation coefficients regarding the employed HFT, volatility and control variables based on daily data. The full sample spans 328 regular trading days, covering the period from December 30, 2019, to April 22, 2021; whereas the estimations pertain to 33 (percentiled) extreme volatility days, where EXT=1, are reported separately in brackets if applicable. The definitions of the above specified measures are detailed in the study.

7.3.2 Models results

The Durbin–Wu–Hausman test (Hausman, 1978) is first applied to determine whether the OLS estimation is inappropriate due to potential endogeneity bias in the regressors of Eq.7.6 and Eq.7.7. Significant F-values from the test lead to the rejection of the null hypothesis of exogeneity. Therefore, a more robust and suitable computational approach, MLH estimation, is employed for regression models 1 and 2.

Table 7.4 illustrates the findings of Model 1 and Model 2, which assess the HFT impact on volatility under normal and stressful market conditions, respectively, throughout the sample period. Based on the coefficient estimates for *HFTQ* and *HFTQ x (1-EXT)*, both models demonstrate a significant positive association between HFT and volatility

in relatively stable market environments. The alignment of calculated parameters across the specified models corroborates this result. However, the insignificance of the interaction term, $HFTQ \times EXT$, in Model 2 indicates that the behavior of low-latency traders should not be attributed to amplified volatility during abnormal price fluctuations, whether in severe market downturns or upturns. This suggests that while HFT activity appears to intentionally elevate volatility under certain conditions, it should not be considered responsible for the magnitude or velocity of sharp price swings during periods of excessive market overreaction driven by greed or fear. These findings conflict with Ben Ammar and Hellara (2022), which could stem from structural disparities between equity and futures markets.

Table 7.4 : Volatility and HFT Dynamics.

	Panel-A: Model 1	Panel-B: Model 2
constant	-0.0398*** (-3.20)	-0.0386*** (-2.51)
VOL_{t-1}	0.1319** (2.17)	0.1538*** (2.82)
HFTQ	0.0059** (2.40)	
TV	2.88e-09*** (6.54)	3.68e-09*** (4.58)
EP	0.7015*** (9.48)	0.5306*** (6.69)
IP	0.1201*** (3.26)	0.1144*** (3.37)
EXT		0.0025 (0.25)
HFTQ × (1 – EXT)		0.0056* (1.69)
HFTQ × EXT		0.0144 (0.69)
AR2	0.4791	0.5778
F-Statistics	60.95***	64.67***
Breusch-Godfrey LM Test	1.61	1.28
Breusch-Pagan-Godfrey Test	0.77	1.83
Wald Test		3.15*

Notes: The table reports parameter estimates of the regression models specified by Eq.7.6 and Eq.7.7 in Panel-A and Panel-B respectively, through maximum-likelihood (MLH) procedure. The full sample spans 328 regular trading days, covering the period from December 30, 2019, to April 22, 2021. VL and HF measures in the models are represented by VOL and HFTQ as dependent and main explanatory variables respectively. The definitions of the variables are detailed in the study. All the variables are standardized around the means. AR2 denotes adjusted R-squared value along with selected diagnostics given as Breusch-Godfrey LM test for serial correlation, Breusch-Pagan-Godfrey test for heteroskedasticity and Wald test for equality of interaction terms in Model-2. T-statistics are given in brackets and significance levels are computed by referring Newey-West HAC robust standard errors. P-values are denoted as * < 0.10, ** < 0.05 and *** < 0.01.

Endogeneity of regressors for the OLS estimation are verified preliminarily by significant F-values of Wu-Hausman test under the null hypothesis of exogenous variables. It is later confirmed, using the same test, that the proposed models are free from endogeneity. To maintain brevity, Wu-Hausman test results and variance inflation factors (VIF) of control variable pairs are not presented here; however they are available upon request.

In the subsequent step, the analysis is broadened by recalculating the estimates of the extended model (Eq.7.7) using Model 2a and Model 2b, which concentrate exclusively on positive and negative daily return series, respectively, as presented in Table 7.5. The objective is to consider whether the relationship between HFT and volatility deviates with the direction of market price movements. It is observed that while HFTs tend to increase volatility, particularly in declining markets under normal conditions ($HFTQ \times (1-EXT)$ in Panel B), no significant interaction emerges between the explanatory and response variables during low volatile bullish or extreme sample intervals ($HFTQ \times (1-EXT)$ in Panel A and/or $HFTQ \times EXT$ in both Panel A and Panel B).

Two explanations clarify this outcome. First, HFT algorithms might implement conservative portfolio adjustment maneuvers preemptively before the bearish market deteriorates and limit their trading activity as a result when investor sentiment becomes highly unstable and tensions intensify (Dastan, 2021). Second, from a different perspective, the commonly employed liquidity provision- oriented HFT strategies may pursue asymmetric trajectories over the course of various market periods, which could potentially offset their effects on volatility through mutual opposition (Ersan and Ekinci, 2016).

The time series data exhibit serial correlation and leptokurtosis at the outset of the analysis. To address these issues, the Driscoll-Kraay extension of Newey-West HAC robust standard errors (Driscoll and Kraay, 1998) is applied across all model specifications to generate more reliable t-statistics and p-values. Diagnostic checks, as defined in Section 7.2.3, were conducted to validate the robustness of the estimates, and the corresponding results are provided in the accompanying tables. VIFs were further evaluated within the regression specifications to confirm the absence of multicollinearity for the VL and HF predictors. The maximum recorded VIF value is 1.881, indicating no significant multicollinearity in these measures. Additionally, the inequality of the interaction terms $HFTQ \times EXT$ and $HFTQ \times (1-EXT)$ was confirmed using Wald test statistics across the relevant model configurations, with the exception of Model 2a in Table 7.5, which accounts for shifts in HFT effects on volatility between normal and extreme market conditions.

Table 7.5 : Volatility - HFT Relationship based on Market Price Behavior.

	Panel-A: Model 2a ($r > 0$)	Panel-B: Model 2b ($r < 0$)
constant	-0.0159 (-0.98)	-0.0599*** (-2.75)
VOL_{t-1}	0.1924*** (3.23)	0.1219* (1.77)
TV	2.64e-09*** (2.65)	4.70e-09*** (4.42)
EP	0.5188*** (4.88)	0.5289*** (5.92)
IP	0.0579 (1.25)	0.1644*** (2.41)
EXT	0.0002 (0.05)	0.0152 (0.60)
HFTQ $\times$ (1 - EXT)	-0.0012 (-0.27)	0.0134*** (3.54)
HFTQ $\times$ EXT	0.0092 (0.87)	0.0007 (0.01)
AR2	0.4928	0.6395
F-Statistics	25.29***	38.50***
Breusch-Godfrey LM Test	0.63	1.17
Breusch-Pagan-Godfrey Test	0.54	1.46
Wald Test	1.12	3.54***

Notes: The table reports parameter estimates of the regression model specified by Eq.7.7 through maximum-likelihood (MLH) procedure using positive or negative daily return sample variants in Panel-A and Panel-B respectively. The full sample spans 328 regular trading days, covering the period from December 30, 2019, to April 22, 2021. VL and HF measures in the models are represented by VOL and HFTQ as dependent and main explanatory variables respectively. The definitions of the variables are detailed in the study. All the variables are standardized around the means. AR2 denotes adjusted R-squared value along with selected diagnostics given as Breusch-Godfrey LM test for serial correlation, Breusch-Pagan-Godfrey test for heteroskedasticity and Wald test for equality of interaction terms. T-statistics are given in brackets and significance levels are computed by referring Newey-West HAC robust standard errors. P-values are denoted as * < 0.10, ** < 0.05 and *** < 0.01.

Endogeneity of regressors for the OLS estimation are verified preliminarily by significant F-values of Wu-Hausman test under the null hypothesis of exogenous variables. It is later confirmed, using the same test, that the proposed models are free from endogeneity. To maintain brevity, Wu-Hausman test results and variance inflation factors (VIF) of control variable pairs are not presented here; however they are available upon request.

7.3.3 Robustness assessment

The regressions from Models 1 and 2 were replicated using alternative measures for HFT (*HF*) and volatility (*VL*) to ensure robustness. The findings, displayed in Table 7.6, remain consistent and effectively support the implications previously drawn in the study. A minor contradiction exists in Model 4y concerning the insignificant (significant) parameter estimate of $VL_{t-1}(HF \times EXT)$, in particular.

Table 7.6 : Robustness Results of Alternative Models.

	Model 3x	Model 3y	Model 4x	Model 4y	Model 5x	Model 5y
constant	−0.0559*** (−4.77)	−0.0437*** (−2.90)	−0.1784*** (−3.95)	−0.6081*** (−3.18)	−0.1809*** (−3.26)	−0.1437* (−1.78)
VL_{t−1}	0.0922** (1.89)	0.2315*** (3.59)	0.1363*** (2.66)	0.1870 (1.20)	0.5972*** (17.66)	0.5869*** (17.89)
HF	0.0187*** (4.40)		0.0251*** (2.99)		0.0417** (2.14)	
TV			1.35e-08*** (3.64)	2.73e-08*** (3.24)	1.09e-08*** (3.77)	1.08e-08*** (2.76)
EP	0.7094*** (9.40)	0.7700*** (7.74)	0.6169*** (3.35)	2.7394** (2.06)	−0.4574 (−1.43)	−0.4886 (−1.17)
IP	0.1667*** (4.74)	0.1381*** (3.04)	0.4956*** (3.85)	1.7099*** (3.11)	0.4934*** (3.17)	0.3990* (1.73)
EXT		0.0032 (0.53)		−0.1206 (−1.23)		−0.0838 (−0.96)
HF × (1 − EXT)		0.0186*** (3.33)		0.0657*** (2.94)		0.0281** (2.55)
HF × EXT		0.0336 (1.17)		0.3433* (1.66)		0.1946 (1.18)
AR2	0.5101	0.5405	0.4852	0.5107	0.5513	0.5606
F-Statistics	66.85***	64.93***	60.51***	49.62***	78.87***	60.42***
Breusch-Godfrey LM Test	1.30	2.35	1.36	1.14	1.33	2.11
Breusch-Pagan-Godfrey Test	1.35	4.78***	8.65***	10.45***	1.55	8.61***
Wald Test		2.56**		1.76*		1.99*

Notes: The table reports parameter estimates of the regression models specified by Eq.7.6 and Eq.7.7 through maximum-likelihood (MLH) procedure. The full sample spans 328 regular trading days, covering the period from December 30, 2019, to April 22, 2021. Within the robustness checking framework, VL and HF coefficients in the models are represented by VOL and HFTT in Model 3x and Model 3y; VRV and HFTQ in Model 4x and Model 4y; and VCN and HFTQ in Model 5x and Model 5y as dependent and key explanatory alternative variables respectively. The definitions of the variables are detailed in the study. All the variables are standardized around the means. AR2 denotes adjusted R-squared value along with selected diagnostics given as Breusch-Godfrey LM test for serial correlation, Breusch-Pagan-Godfrey test for heteroskedasticity and Wald test for equality of interaction terms in Model 3y, Model 4y and Model 5y. T-statistics are given in brackets and significance levels are computed by referring Newey-West HAC robust standard errors. P-values are denoted as * < 0.10, ** < 0.05 and *** < 0.01.

7.4 Concluding Implications

The study investigates whether the recent surge in HFT activity influences, and potentially drives, volatility in the Turkish (BIST) index futures market. For a rigorous understanding of low-latency trading dynamics, the analysis is extended through an empirical investigation across multiple data samples, each differentiated by distinct market conditions such as stable versus extreme volatility phases and bull versus bear trend cycles.

The BIST index futures market consistently ranks highly on a global scale, thereby meriting further investigation from a novel perspective. In addition, the high share turnover velocity and leading volatility in the BIST equity market underscore the practical importance of the study particularly from a risk management standpoint. The findings contribute not only additional evidence to the market microstructure framework but also document a comprehensive range of international data from the relatively underrepresented asset class of futures, moving beyond the primary focus on equity products.

The results highlight a significant positive relationship between HFT and volatility throughout the full sample period. In other words, as low-latency participation in the market increases, volatility rises over the course of a normal trading day, reflecting a natural, and to some extent desirable, consequence of HFT-induced liquidity enhancement. The interaction becomes particularly prominent during negative return phases but diminishes somewhat within clusters of positive returns. Yet, no discernible HFT impact on market volatility is observed when momentum peaks amid extreme price fluctuations. Rule-making authorities should reconsider this implication in their decision-making process, as HFT does not worsen market quality by amplifying volatility, and inherently systematic risks, during exceptional market conditions (e.g., stressful events or crashes). A preemptive, one size fits all regulatory policy for HFT may not be ideal and necessitates a prudent and holistic review to safeguard the marginal benefits for market stakeholders before introducing any related legislative actions.



8. CONCLUSIVE REMARKS

The thesis aims to reveal the dynamics of high-frequency trading (HFT) in the BIST-30 index futures market with a comprehensive approach. A self-contained, theme-based framework was adopted for the drafting with three individual chapters addressing the primary research objective and sharing interconnected aspects. One distinctive strength of the research is to have an opportunity to work on an exceptional dataset, acquired with special authorization from Borsa Istanbul, featuring HFT-flags designated from encrypted information of self-reported low latency accounts. This unique attribute facilitates precise identification of HFTs and overcomes common reliability issues of indirect measurement prevalent in the literature. Reflecting collectively, the study enriches the literature by vertically expanding the product scope beyond the traditionally dominant equity asset class. Furthermore, its derived propositions reinforce the theoretical underpinnings of this field while offering valuable practical implications from both business and regulatory perspectives.

In the first chapter (Section 5), it is reported that HFT proxies demonstrate varying levels of effectiveness in identifying low-latency activity within the Turkish index futures market. No single proxy consistently outperforms the others across the evaluated metric segments, indicating the absence of a dominant approach. Nevertheless, the proposed new proxy, extended modification rate (P5), derived from Van Ness's (2015) measure (P4), delivers solid results in the valuation procedure.

Yet, it should be underlined that, the relative underperformance of certain sophisticated proxies is attributed to Type-1 and Type-2 errors inherent in the predetermined HFT dataset. Besides, proxies designed for equity markets, may not align perfectly with futures markets dynamics, as divergences in investor profiles, margin requirements, and short-selling allowance could hinder their applicability and underscore the necessity of envisioning market-specific traits. Despite these limitations, the research implies that the indirect method of determining HFT activity through proxies remains a valuable approach for gaining insights into high-frequency trading mechanics. As far as can be determined, no prior study has offered a

quantitative analysis of HFT proxy performance using an approach akin to that of the first chapter.

The second chapter (Section 6) investigates causal dynamics between HFT and liquidity, with particular attention to the very infrequently observed event of a minimum price tick increase, also referred to as reverse decimalization. Conceptually, the study aspires to advance core areas of literature by integrating the themes of liquidity, HFT, tick-size effects and futures markets. Such an approach sheds light on variables' multifaceted interconnections.

Overall, the evidence indicates HFT activity undermines liquidity in the Turkish index futures market by depleting tradeable volumes and widening effective spreads. The adverse effect becomes more pronounced after the introduction of the minimum tick-size regulation change. Additionally, the relative tick size of futures contracts emerges as a pivotal factor influencing the order behavior of HFT participants. The gradual rise in electronic message traffic may inevitably foster some level of HFT activity independent of the scrutinized rule amendment. While the results support Lee's (2015) analysis of futures markets, they contrast with the conclusions of equity market researchers, including Hendershott et al. (2011), who assert liquidity improvement effects of HFT.

And the final chapter (Section 7) focuses on whether HFT influences volatility in stable versus stressful market conditions. The BIST equity market's strikingly high share turnover velocity and ensuing volatility accentuate the applied significance of the research in the context of the risk management benefits offered by index futures. The findings remark that HFT activity generally drives volatility in the BIST futures market. The effect is most noticeable during negative returns but vanishes under extreme conditions. Superior trend forecasting skills and counteracting strategic moves undertaken by this type of investor may provide justification for the explored pattern. Hence, HFTs should not be held responsible for pumping up volatility or deepening stress when the market faces turbulent periods. The results oppose those of Ben Ammar and Hellara (2022) yet concur with Dastan (2021), likely reflecting structural variations between equity and futures markets.

Reassessing the broad picture, the policy implications suggested by the study are grounded in the distinct HFT segments operating within the market. HFT-associated

systematic risk exposures hampering the trading environment through adverse selection, distortions in price discovery, herding tendencies or ghost liquidity should be prudently evaluated. In his seminal work, Serrano (2020) outlined that the appropriate set of policies, including auction mechanisms, circuit breakers, order-to-trade penalties and capital requirements can be utilized to mitigate HFT induced vulnerabilities to some extent. However, he emphasized the critical challenge of ensuring participants stay active when exiting the market might appear more favorable and be the rational choice.

Therefore, to address the major shifts in HFT's impact on market quality, a regulatory framework must be rigorously designed to uphold the substantial benefits of these activities. The relevant legislative strategies must balance the mitigation HFT risks with the preservation of its advantages. Further analysis and research is essential to elucidate how various HFT strategies affect systematic risk factors, requiring diligent and continuous monitoring by policymakers in this domain .

The fundamental limitation of the thesis lies in its dependence on voluntarily self-reported HFT data provided by the exchange. Such reliance introduces the potential for statistical inaccuracies, including both Type-1 and Type-2 errors. Specifically, certain accounts labeled as HFT might not engage in genuine low-latency activities, while some high-frequency traders may deliberately avoid disclosing their status to the exchange. Consequently, portrayals of HFT behavior could lack precision or completeness, which might compromise the practicability of the findings. Therefore, this drawback requires the results to be interpreted carefully within the framework of the acknowledged data constraints.

To enhance the robustness and versatility of future analyses, researchers could widen the scope of the study across both horizontal and vertical dimensions. Horizontal dimension might involve replicating the analysis in comparable index futures markets, enabling cross-market validation of the findings. Vertical dimension, on the other hand, could examine other asset classes, such as equities, commodities, or fixed-income securities, to uncover potential variations in HFT dynamics across different financial instruments. These approaches could provide a more comprehensive understanding of HFT behavior and its implications, which contribute to a broader and more nuanced viewpoint on the subject.



REFERENCES

- Aitken, M. & Comerton-Forde, C.** (2005). Do reductions in tick size affect market liquidity. *Account Finance* 45, 171–184.
- Aitken, M., Harris, F., McNish, T., Aspris, A. & Foley, S.** (2012). High frequency trading—assessing the impact on market efficiency and integrity. UK Government Office for Science. *Foresight Driver Review* 28.
- Alampieski, K. & Lepone, A.** (2009). Impact of a tick size reduction on liquidity: evidence from the Sydney Futures Exchange. *Accounting and Finance* 49(1), 1–20.
- Aldridge, I.** (2014). High-frequency runs and flash-crash predictability. *Journal of Portfolio Management* 40(3), 113–123.
- Anderson, H.D. & Peng, Y.** (2014). From cents to half-cents and its liquidity impact. *Pacific Accounting Review* 26(3), 160–176.
- Arnoldi, J.** (2016). Computer algorithms, market manipulation and the institutionalization of high frequency trading. *Theory, Culture & Society* 33(1), 29–52.
- Bacidore, J.M.** (1997). The impact of decimalization on market quality: An empirical investigation of the Toronto stock exchange. *Journal of Financial Intermediation* 6(2), 92–120.
- Baldauf, M. & Mollner, J.** (2020). High-frequency trading and market performance. *Journal of Finance* 75, 1495–1526.
- Barker, W. & Pomeranets, A.** (2011). The growth of high-frequency trading: implications for financial stability. Bank of Canada. *Financial System Review*, 47–52.
Retrieved from: <https://www.bankofcanada.ca/wp-content/uploads/2011/12/fsr-0611-barker.pdf>
- Barlett III, R.P. & McCrary, J.** (2013). Shall we haggle in pennies at the speed of light or in nickels in the dark? How minimum price variation regulates high frequency trading and dark liquidity. Working Paper. *Berkeley School of Law*.
- Baron, M., Brogaard, J. & Kirilenko, A.** (2012). The trading profits of high frequency traders. *8th annual Central Bank workshop on the microstructure of financial markets* 25. Available at SSRN 2106158.
- Baron, M., Brogaard, J., Hagstromer, B. & Kirilenko, A.** (2019). Risk and return in high-frequency trading. *Journal of Financial and Quantitative Analysis* 54, 993–1024.
- Bazzana, F. & Collini, A.** (2020). How does HFT activity impact market volatility and the bid-ask spread after an exogenous shock? An empirical analysis on S&P 500 ETF. *North American Journal of Economics and Finance* 54, 101240.

- Ben Ammar, I. & Hellara, S.** (2022). High-frequency trading, stock volatility, and intraday crashes. *Quarterly Review of Economics and Finance* 84, 337–344.
- Ben Ammar, I., Hellara, S. & Ghadhab, I.** (2020). High-frequency trading and stock liquidity: An intraday analysis. *Research in International Business and Finance* 53, 101235.
- Biais, B. & Foucault, T.** (2014). HFT and market quality. *Bankers, Markets and Investors* 128, 5–19.
- Biais, B., Foucault, T. & Moinas, S.** (2015). Equilibrium fast trading. *Journal of Financial Economics* 116, 292–313.
- Bland, J. & Altman, D.** (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 8, 135–160.
- Blocher, J., Cooper, R., Seddon, J. & Van Vliet, B.** (2016). Phantom liquidity and high-frequency quoting. *Journal of Trading* 11(3), 6–15.
- Blundell, R. & Bond, S.** (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87(1), 115–143.
- Bodur, M.** (2022). Informed trading around extreme events in Borsa Istanbul. Master's thesis, *Kadir Has University, School of Graduate Studies*. Council of Higher Education Thesis Center 743203.
- Boehmer, E., Fong, K. & Wu, J.** (2012). Algorithmic trading and changes in firms' equity capital. FIRN Research Paper. Available at SSRN 2050856.
- Boehmer, E., Fong, K. & Wu, J.** (2021). Algorithmic trading and market quality: International evidence. *Journal of Financial and Quantitative Analysis* 56 (8), 2659–2688.
- Boehmer, E., Li, D. & Saar, G.** (2018). The competitive landscape of high-frequency trading firms. *Review of Financial Studies* 31 (6), 2227–2276.
- Bollen, N. & Whaley, R.** (2015). Futures market volatility: what has changed ? *Journal of Futures Markets* 35 (5), 426–454.
- Borsa Istanbul (BIST)** (2022). Annual Report.
Retrieved from: <https://borsaistanbul.com/files/2022-annual-integrated-report.pdf>
- Borsa Istanbul (BIST)** (2023). Annual Report.
Retrieved from: <https://borsaistanbul.com/files/2023-annual-integrated-report.pdf>
- Bouveret, A., Guillaumie, C., Roqueiro, C., Winkler, C. & Nauhaus, S.** (2014). High-frequency trading activity in EU equity markets. *European Securities and Markets Authority, Economic Report* 1, 1–31.
- Brandt, M. & Diebold, F.** (2006). A no-arbitrage approach to range based estimation of return covariances and correlations. *Journal of Business* 79, 61–73.
- Breckenfelder, J.** (2024). Competition among high-frequency traders and market quality. *Journal of Economic Dynamics and Control* 166, 104922.
- Brogaard, J.** (2010). High frequency trading and its impact on market quality. Working Paper. *Northwestern University*.
Retrieved from: <https://conference.nber.org/confer/2010/MMf10/Brogaard.pdf>

- Brogaard, J., Carrion, A., Riordan, R., Shkilko, A. & Sokolov, K.** (2018). High-frequency trading and extreme price movements. *Journal of Financial Economics* 128(2), 253–265.
- Brogaard, J. & Garriott, C.** (2019). High-frequency trading competition. *Journal of Financial and Quantitative Analysis* 54, 1469–1497.
- Brogaard, J., Hendershott, T. & Riordan, R.** (2014). High-frequency trading and price discovery. *Review of Financial Studies* 27(8), 2267–2306.
- Brogaard, J., Hendershott, T. & Riordan, R.** (2017). High frequency trading and the 2008 short-sale ban. *Journal of Financial Economics* 124(1), 22–42.
- Brogaard, J., Hendershott, T. & Riordan, R.** (2019). Price discovery without trading: Evidence from limit orders. *Journal of Finance* 74, 1621–1658.
- Budish, E., Cramton, P. & Shim, J.** (2015). The high-frequency trading arms race: frequent batch auctions as a market design response. *Quarterly Journal of Economics* 130, 1547–1621.
- Cao, C., Hansch, O. & Wang, X.** (2009). The information content of an open limit-order book. *Journal of Futures Markets* 29(1), 16–41.
- Carrion, A.** (2013). Very fast money: High-frequency trading on Nasdaq. *Journal of Financial Markets* 16, 680–711.
- Chaboud, A., Chiquoine, B., Hjalmarsson, E. & Vega, C.** (2014). Rise of the machines: algorithmic trading in the foreign exchange market. *Journal of Finance* 69(5), 2045–2084.
- Chae, J., Khil, J. & Lee, E.** (2013). Who makes markets? Liquidity providers versus algorithmic traders. *Journal of Futures Markets* 33(5), 397–420.
- Chakabarty, B., Huang, J. & Jain, P.** (2018). Exchange competition with levelled speed. Working Paper. Available at SSRN 3197088.
- Chakabarty, B., Jain, P., Shkilko, A. & Sokolov, K.** (2020). Unfiltered market access and liquidity: evidence from Sec rule 15c3-5. *Management Science* 67(2), 1183–1198.
- Chou, R. & Chung, H.** (2006). Decimalization, trading costs, and information transmission between ETFs and index futures. *Journal of Futures Markets* 26 (2), 131–151.
- Chung, K., Chuwonganant, C. & Jiang, J.** (2008). The dynamics of quote adjustments. *Journal of Banking & Finance* 32(11), 2390–2400.
- Comerton-Forde, C., Malinova, K. & Park, A.** (2018). Regulation dark trading: Order flow segmentation and market quality. *Journal of Financial Economics* 130, 347–366.
- Conrad, J., Wahal, S. & Xiang, J.** (2015). High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics* 116, 271–291.
- Coughenour, F. & Harris, L.** (2004). Specialist profits and the minimum price increment. Working Paper. *University of Delaware*. Available at SSRN 537785.

- Cvitanic, J. & Kirilenko, A.** (2010). High frequency traders and asset prices. Working Paper. *California Institute of Technology*. Available at SSRN 1569067.
- Dastan, I.** (2021). Investigating high-frequency trading around the extreme price movements in Borsa Istanbul. Master's thesis, *Kadir Has University, School of Graduate Studies*. Council of Higher Education Thesis Center 662690.
- Degryse, H., Winne, R., Gresse, C. & Payne, R.** (2021). Cross-venue liquidity provision: High frequency trading and ghost liquidity. Working Paper. Available at SSRN 3356695.
- Dickinson, H., Hrisos, S., Eccles, P., Francis, J. & Johnston, M.** (2010). Statistical considerations in a systematic review of proxy measures of clinical behavior. *Implementation Science* 5.
- Ding, S., Hanna, J. & Hendershott, T.** (2014). How slow is the NBBO? A comparison with direct exchange feeds. *Financial Review* 49, 313–332.
- Driscoll, J. & Kraay, A.** (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics* 80(4), 549–560.
- Easley, D., Prado, M. & O'Hara, M.** (2011). The microstructure of the “Flash Crash”: Flow toxicity, liquidity crashes, and the probability of informed trading. *Journal of Portfolio Management* 37, 119–128.
- Easley, D., Prado, M. & O'Hara, M.** (2012). The volume clock: Insights into the high-frequency paradigm. *Journal of Portfolio Management* 39, 19–29.
- Egginton, F., Van Ness, B. & Van Ness, R.** (2016). Quote stuffing. *Financial Management* 45, 583–608.
- Ekinci, C. & Ersan, O.** (2018). A new approach for detecting high-frequency trading from order and trade data. *Finance Research Letters* 24, 888–896.
- Ekinci, C. & Ersan, O.** (2022). High-frequency trading and market quality: The case of a “slightly exposed” market. *International Review of Financial Analysis* 79, 102004.
- Elizarov, M., Ivanyuk, V. & Soloviev, V.** (2017). Identification of high-frequency traders using fuzzy logic methods. *Tenth International Conference Management of Large-Scale System Development (MLSD)*, 1–4.
- Ersan, O., Dalgic, N., Ekinci, C. & Bodur, M.** (2021). High-frequency trading and its impact on market liquidity: A review of literature. *Alanya Akademik Bakis* 5(1), 345–368.
- Ersan, O. & Ekinci, C.** (2016). Algorithmic and high-frequency trading in Borsa Istanbul. *Borsa Istanbul Review* 16, 233–248.
- Foucault, T., Hombert, J. & Rosu, I.** (2016). News trading and speed. *Journal of Finance* 71(1), 335–381.
- Foucault, T., Kadan, O. & Kandel, E.** (2013). Liquidity cycles and make/take fees in electronic markets. *Journal of Finance* 68(1), 299–341.
- Foucault, T., Kozhan, R. & Tham, W.** (2017). Toxic arbitrage. *Review of Financial Studies* 30, 1053–1094.

- Friederich, S. & Payne, R.** (2011). Computer based trading, liquidity and trading costs. UK Government Office for Science. *Foresight Driver Review DR5*.
- Friederich, S. & Payne, R.** (2015). Order-to-trade ratios and market liquidity. *Journal of Banking & Finance* 50, 214–223.
- Frino, A., Mollica, V., Monaco, E. & Palumbo, R.** (2017). The effect of algorithmic trading on market liquidity: Evidence around earnings announcements on Borsa Italiana. *Pacific Basin Finance Journal* 45, 82–90.
- Frino, A., Mollica, V. & Zhang, S.** (2015). The impact of tick size on high frequency trading: The case for splits. Working Paper. *Macquarie University*. Available at SSRN 2607391.
- Froot, K., Scharfstein, D. & Stein, J.** (1992). Herd on the street: Informational inefficiencies in a market with short-term speculation. *Journal of Finance* 47, 1461–1484.
- Gsell, M.** (2008). Assessing the Impact of Algorithmic Trading on Markets: A Simulation Approach. *ECIS Proceedings* 225.
- Hagstromer, B. & Norden, L.** (2013). The diversity of high-frequency traders. *Journal of Financial Markets* 16(4), 741–770.
- Haldane, A.** (2011). The race to zero. Bank of England. *International Economic Association Sixteenth World Congress*.
- Hansen, L.** (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029 – 1054.
- Harris, L.** (2013). What to do about high-frequency trading. *Financial Analysts Journal* 69, 6–9.
- Hasbrouck, J. & Saar, G.** (2009). Technology and liquidity provision: the blurring of traditional definitions. *Journal of Financial Markets* 12(2), 143–172.
- Hasbrouck, J. & Saar, G.** (2013). Low-latency trading. *Journal of Financial Markets* 16(4), 646–679.
- Hausman, J.** (1978). Specification tests in econometrics. *Econometrica* 46, 1251–1271.
- Hautsch, N. & Horvath, A.** (2019) How effective are trading pauses? *Journal of Financial Economics* 131, 378–403.
- Hendershott, T., Jones, M. & Menkveld, A.** (2011). Does algorithmic trading improve liquidity? *Journal of Finance* 66(1), 1–33.
- Hendershott, T. & Mouton, P.** (2011). Automation, speed, and stock market quality: the NYSE’s hybrid. *Journal of Financial Markets* 14(4), 568–604.
- Hendershott, T. & Riordan, R.** (2013). Algorithmic trading and the market for liquidity. *Journal of Financial and Quantitative Analysis* 48, 1001–1024.
- Hirschey, N.** (2018). Do high-frequency traders anticipate buying and selling pressure? Working Paper. *London Business School*.
- Hoffmann, P.** (2014). A dynamic limit order market with fast and slow traders. *Journal of Financial Economics* 113(1), 156–169.

- Hossain, S.** (2022). High-frequency trading (HFT) and market quality research: An evaluation of the alternative (HFT) proxies. *Journal of Risk and Financial Management* 15 (2), 54.
- Hrisos, S., Eccles, P., Francis, J., Dickinson, H., Kaner, E., Beyer, F. & Johnston, M.** (2009). Are there valid proxy measures of clinical behavior? A systematic review. *Statistical Methods in Medical Research* 4.
- Hsiao, C., Pesaran, M. & Tahmiscioglu, K.** (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109(1), 107–150.
- Hu, E.** (2018). Intentional access delays, market quality, and price discovery: evidence from IEX becoming an exchange. Working Paper. *Securities and Exchange Commission*.
- Huh, Y.** (2014). Machines vs. machines: high frequency trading and hard information. Finance and economics discussion series 34. *Board of Governors of the Federal Reserve System*.
- Jain, K., Jain, P. & McInish, T.** (2016). Does high-frequency trading increase systemic risk? *Journal of Financial Markets* 31, 1–24.
- Jarnecic, E. & Snape, M.** (2014). The provision of liquidity by high-frequency participants. *Financial Review* 49, 371–394.
- Jarrow, R. & Protter, P.** (2012). A dysfunctional role of high frequency trading in electronic markets. *International Journal of Theoretical and Applied Finance*, 15, 1250022.
- Jiang, C., Kim, J. & Wood, R.** (2011). A comparison of volatility and bid–ask spread for NASDAQ and NYSE after decimalization. *Applied Economics* 43(10), 1227–1239.
- Jiang, G., Lo, I. & Valente, G.** (2014). High-frequency trading around macroeconomic news announcements: evidence from the U.S. Treasury market. Working Paper. *Bank of Canada*.
- Johnson, N., Zhao, G., Hunsader, E., Qi, H., Johnson, N., Meng, J. & Tivnan, B.** (2013). Abrupt rise of new machine ecology beyond human response time. *Scientific Reports* 3, 2627.
- Jorgensen, K., Skjeltorp, J. & Odegaard, B.** (2018). Throttling hyperactive robots – Order-to-trade ratios at the Oslo Stock Exchange. *Journal of Financial Markets* 37, 1–16.
- Jovanovic, B. & Menkveld, J.** (2016). Middlemen in Limit Order Markets. Available at SSRN 1624329.
- Kadan, O.** (2006). So who gains from a small tick size? *Journal of Financial Intermediation* 15(1), 32–66.
- Kang, J., Kwon, Y. & Kim, W.** (2020). Flow toxicity of high-frequency trading and its impact on price volatility: Evidence from the KOSPI 200 futures market. *Journal of Futures Markets* 40, 164–191.
- Ke, C., Jiang, H. & Huang, S.** (2004). The impact of tick size on intraday stock price behavior: evidence from the Taiwan Stock Exchange. *Pacific-Basin Finance Journal* 12 (1), 19–39.

- Ke, Y. & Zhang, Y.** (2020). Does high-frequency trading reduce market underreaction to earnings news? *Finance Research Letters* 34, 101239.
- Kelejian, H. & Mukerji, P.** (2016). Does high frequency algorithmic trading matter for non-AT investors? *Research in International Business and Finance* 37, 78–92.
- Kirilenko, A. & Lo, A.** (2013). Moore’s law versus Murphy’s Law: algorithmic trading and its discontents. *Journal of Economic Perspective* 27(2), 51–72.
- Kirilenko, A., Kyle, A., Samadi, M. & Tuzun, T.** (2017). The flash crash: High-frequency trading in an electronic market. *Journal of Finance* 72, 967–998.
- Kurov, A. & Zabolina, T.** (2005). Is it time to reduce the minimum tick sizes of the E-mini futures? *Journal of Futures Markets* 25(1), 79–104.
- Lafarguette, R.** (2016). High-frequency trading, market volatility and regulation: the role of high-frequency quoting and dark pools. Working Paper. *European Central Bank Macprudential Bulletin* 2.
- Lee, E.** (2015). High frequency trading in the Korean index futures market. *Journal of Futures Markets* 35(1), 31–51.
- Linton, O. & Mahmoodzadeh, S.** (2018). Implications of high-frequency trading for security markets. *Annual Review of Economics* 10, 237–259.
- Linton, O. & O’Hara, M.** (2012). The impact of computer trading on liquidity, price efficiency/discovery and transaction costs. UK Government Office for Science. *Foresight Driver Review WP2*.
- Liu, X., Lo, I., Nguyen, M. & Valente, G.** (2014). High-frequency trading and treasury bond returns. *23rd European Financial Management Association Annual Meetings*, EFMA 2014.
- Mahmoodzadeh, S. & Gencay, R.** (2017). Human vs. high-frequency traders, penny jumping, and tick size. *Journal of Banking & Finance* 85, 69–82.
- Mahnken, D., Chen, X., Brown, A., Vidoni, E., Billinger, S. & Gajewski, B.** (2014). Evaluating variables as unbiased proxies for other measures: Assessing the step test exercise prescription as a proxy for the maximal, high-intensity peak oxygen consumption in older adults. *International Journal of Statistics and Probability* 3.
- Malceniece, L., Malceniaks, K. & Putnins, T.** (2019). High frequency trading and comovement in financial markets. *Journal of Financial Economics* 134, 381–399.
- Malinova, K., Park, A. & Riordan, R.** (2018). Do retail investors suffer from high frequency traders? Available at SSRN 2183806.
- Manahov, V., Hudson, R. & Gebka, B.** (2014). Does high frequency trading affect technical analysis and market efficiency? And if so, how? *Journal of International Financial Markets, Institutions and Money* 28, 131–157.
- Meling, T. & Odegaard, B.** (2017). Tick size wars, high frequency trading, and market quality. Working Paper in Economics, *University of Bergen Department of Economics* 5(17).

- Menkveld, A.** (2013). High-frequency trading and the new market makers. *Journal of Financial Markets* 16(4), 712–740.
- Menkveld, A.** (2016). The economics of high-frequency trading: taking stock. *Annual Review of Financial Economics* 8, 1–24.
- Menkveld, A. & Yueshen, B.** (2018). The Flash Crash: a cautionary tale about highly fragmented markets. *Management Science* 65, 4470–4488.
- Menkveld, A. & Zoican, M.** (2017). Need for speed? Exchange latency and liquidity. *Review of Financial Studies* 30, 1188–1228.
- Moral-Benito, E.** (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business & Economic Statistics* 31, 451–472.
- Myers, B. & Gerig, A.** (2014). Simulating the synchronizing behavior of high-frequency trading in multiple markets. *Financial Econometrics and Empirical Market Microstructure*, pp. 207–213. Springer.
- Newey, K. & West, D.** (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–708.
- O’Hara, M.** (2015). High-frequency market microstructure. *Journal of Financial Economics* 116, 257–270.
- O’Hara, M., Saar, G. & Zhong, Z.** (2019). Relative tick size and the trading environment. *Review of Asset Pricing Studies* 9(1), 47–90.
- Olgun, O., Ekinci, C. & Arıkan, R.** (2024). The performance of selected high-frequency trading proxies: An application on Turkish index futures market. *Finance Research Letters* 65, 105523.
- Pagnotta, E. & Philippon, T.** (2018). Competing on speed. *Econometrica* 86, 1067–1115.
- Pan, W., Song, F. & Tao, L.** (2012). The effects of a tick-size reduction on the liquidity in a pure limit order market: evidence from Hong Kong. *Applied Economics Letters* 16, 1639–1642.
- Petukhina, A., Reule, G. & Hardle, W.** (2021). Rise of the machines? Intraday high-frequency trading patterns of cryptocurrencies efficiency. *European Journal of Finance* 27, 8–30.
- Sanchez Serrano, A.** (2020). High-frequency trading and systemic risk: A structured review of findings and policies. *Review of Economics* 71(3), 169–195.
- Sargan, J.** (1958). The estimation of economic relationships using instrumental variables. *Econometrica* 26, 393 – 415.
- Schestag, R., Schuster, P. & Uhrig-Homburg, M.** (2016). Measuring liquidity in bond markets. *Review of Financial Studies* 29, 1170–1219.
- Shkilko, A. & Sokolov, K.** (2020). Every cloud has a silver lining: Fast trading, microwave connectivity, and trading costs. *Journal of Finance* 75, 2899–2927.
- Van Kervel, V.** (2015). Competition for order flow with fast and slow traders. *Review of Financial Studies* 28(7), 2094–2127.

- Van Ness, B., Van Ness, R. & Pruitt, W.** (2000). The impact of the reduction in tick increments in major US markets on spreads, depth, and volatility. *Review of Quantitative Finance and Accounting* 15, 153–167.
- Van Ness, B., Van Ness, R. & Watson, E.** (2015). Canceling liquidity. *Journal of Financial Research* 38, 3–33.
- Verousis, T., Perotti, P. & Sermpinis, G.** (2018). One size fits all? High frequency trading, tick size changes and the implications for exchanges: market quality and market structure considerations. *Review of Quantitative Finance and Accounting* 50, 353–392.
- World Federation of Exchanges (WFE)** (2023). Annual Statistics Guide.
Retrieved from: <https://www.world-exchanges.org/our-work/articles/2023-annual-statistics-guide>
- Yamada, M.** (2022). Profitability and liquidity provision of HFTs during large price shocks: Does relative tick size matter? *Finance Research Letters* 46, 102308.
- Yang, L. & Zhu, H.** (2020). Back-running: seeking and hiding fundamental information in order flows. *Review of Financial Studies* 33(4), 1484–1533.
- Yao, C. & Ye, M.** (2018). Why trading speed matters: A tale of queue rationing under price controls. *Review of Financial Studies* 31, 2157–2183.
- Ye, M., Yao, C. & Gai, J.** (2013). The externalities of high frequency trading. Research Paper. *WBS Finance Group*.
- Zaharudin, K., Young, M. & Hsu, W.** (2021). High-frequency trading: Definition, implications, and controversies. *Journal of Economic Surveys* 36, 75–107.
- Zervoudakis, F., Lawrance, D., Gontikas, G. & Al, M.** (2012). Perspectives on high-frequency trading. Working Paper. *University College London*.
- Zhang, F.** (2010). High-frequency trading, stock volatility, and price discovery. Working Paper. *Yale School of Management*. Available at SSRN 1691679.
- Zhao, Y. & Wan, D.** (2018). Institutional high frequency trading and price discovery: evidence from an emerging commodity futures market. *Journal of Futures Markets* 38(2), 243–270.



APPENDICES

APPENDIX A: Computational Procedures for Data Preparation

APPENDIX B: Program Java Codes

APPENDIX C: Declaration of Disclosures



APPENDIX A : Computational Procedures for Data Preparation

The date March 15, 2019, was selected from the sample to exemplify and simulate data preparation computations, throughout the self-developed program HFT ProxyChecker, for one day in the study.

Computational Flow:

- (a) Raw (unfiltered) order and trade messages are individually uploaded to Proxy Checker for selected date 15.03.2019. (Figures A.1 and A.2)
- (b) Applying the preset acceptance rules (Figure A.3), these messages are sorted/cleaned both vertical and horizontal directions using order and trade data filters in the cached memory. (i.e. only receive messages with valid update reasons; exclude irrelevant columns and sort the messages according to update time-stamps)
- (c) Depending on the Exchange-provided HFT accounts and users information (confidential-encrypted) (Figure A.4), each approved order and trade message's binary Real-HFT values are assigned in the associated column of cached orders and cached trades tables in relevant sheets (Figures A.5 and A.6).
- (d) Following the algorithm rules of each applied proxy (Figure A.7), the binary HFT results of those are computed and reported automatically by the Proxy Checker software for included tick-by electronic order messages (Figure A.8).
- (e) By making a vis-à-vis comparison between Real-HFT and Proxies' binary estimations, machine learning confusion matrix parameters (true positives, true negatives, false positives and false negatives) are generated and related statistics such as accuracy, sensitivity or precision are calculated (Figure A.9).
- (f) Aggregating the binary values from preceding steps, real and proxy hft variables in the study are obtained through equations from 4.1 to 4.6 respectively. Findings for the day is printed by the programme for variables and other selected parameters (Figure A.10).
- (g) Loop from (a) to (f) is iterated with the next trading date until the last covered day (i.e. 31.12.2021) in the data span.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1	Orders Data Cleaning Process (before cache)																														
2																															
3	1. Remove (not relevant) other than below shown 19 included columns from Raw Orders data (first sheet)																														
4																															
5	Included 19 Arrayed Columns = "Binary Order ID"- "Order Date"- "Order Entry Time Stamp"- "Contract Code"- "Depository Participant Code"- "Depository Account Number"- "Order Status"- "Buy/Sell"- "Order Validity"- "Order Update Time Stamp"- "Order Update Reason"- "Update no"																														
6	"Price" - "Quantity" - "Remain Quantity" - "Order Notional Amount" - "Best Bid" - "Best Offer" - "User"																														
7																															
8	2. Exclude messages having not applicable Order Status 4 ("Deleted by the System").																														
9																															
10	3. Exclude messages other than acceptable Order Update Reason codes 1 ("Canceled By User"), 3 ("Trade"), 5 ("Modified By User") and 6 ("Insert/New"). (not applicable codes 4-Inactivate, 8-ConvertedbySystem, 12-TriggerredbySystem, 52-Pause)																														
11																															
12	4. Exclude previous days long orders having Order Entry Date < Current Date ("15.03.2019")																														
13																															
14	5. Make ascending sort according to Order Update Time Stamp																														
15																															
16																															
17																															
18	Trades Data Cleaning Process (before cache)																														
19																															
20	1. Remove (not relevant) other than below shown 16 included columns from Raw Trades data (second sheet)																														
21																															
22	Included 16 Arrayed Columns = "Deal Number"- "Trade Date"- "Trade Time Stamp"- "Contract Code"- "Deal Type"- "Direction"- "Deal Status"- "Depository Participant Code"- "Depository Account Number"																														
23	"Clearing Trade Number" - "Active/Passive" - "Price" - "Volume" - "Traded Notional Value" - "Binary Order ID" - "User"																														
24																															
25	2. Exclude trades having same depository account in both sides (wash sales).																														
26																															
27	3. Exclude trades other than acceptable Deal Type codes 1 ("deal-source auto"), 7 ("combo_combo") and 43 ("combo_single"). (not involve negotiated deals)																														
28																															
29	4. Only get active trades where Deal Status shows A (not involve rectification, amendment or cancel)																														
30																															
31	5. Make ascending sort according to Trade Time Stamp																														
32																															
	<	>	Raw Orders	Raw Trades	Order & Trade Msg Filters	HFT Accounts & Users	Cached Orders	Cached Trades	Proxy Rules (conditions)	Calculated Proxies (order msg)	Machine Learning Parameters	Printed (Aggregated) Output	⊕																		

Figure A.3: Data Cleaning Process (XLSX Screenshot).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	HFT Accounts List														HFT Users List										
2																									
3	Depository Participant Code (I)	Depository Account Number (II)																							
4	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value		Total of 582 accounts from 22 participant firms and 294 users from 19 participants is given by the Exchange in the encrypted form.																					
5	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
6	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
7	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
8	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value		We determine Real-HFT binary value (TRUE or FALSE) of cached order and/or trade messages by looking any of (a) participant and account number combination (unique after concatenated) or (b) user code (unique) values exists in these lists																					
9	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
10	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
11	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
12	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
13	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
14	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
15	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
16	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
17	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
18	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
19	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
20	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
21	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
22	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
23	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
24	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
25	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
26	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
27	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
28	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
29	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
30	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
31	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
32	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
33	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
34	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
35	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
36	Encrypted Alphabetic Code	Encrypted Alpha Numeric Value																							
	< >	Raw Orders	Raw Trades	Order & Trade Msg Filters	HFT Accounts & Users	Cached Orders	Cached Trades	Proxy Rules (conditions)	Calculated Proxies (order msg)	Machine Learning Parameters	Printed (Aggregated) Output	+													

Total of 582 accounts from 22 participant firms and 294 users from 19 participants is given by the Exchange in the encrypted form.

We determine Real-HFT binary value (TRUE or FALSE) of cached order and/or trade messages by looking any of (a) participant and account number combination (unique after concatenated) or (b) user code (unique) values exists in these lists

Cached Orders Sheet:
Column T
Cached Trades Sheet:
Column Q

Figure A.4: HFT Accounts and Users (XLSX Screenshot).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	Binary Order ID	Order Date	Order Entry Time Stamp	Contract Code	Depository Par	Depository Account	Order Status	Buy/Sell	Order Validity	Order Update Time Stamp	Order Update Reason	Update no	Price	Quantity	Remain Quan	Order Notional Amount	Best Bid	Best Offer	User	Real-HFT	Nb of Submitted Limit Orders	Check (Proxy-4)		
2	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	B	GTC	2019-03-15 09:30:00	6 2	127,725	34	34	434265	130,125	130,325	Encrypted Al	FALSE	GTC-6					
3	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	B	GTC	2019-03-15 09:30:00	6 1	127,725	34	34	434265	130,125	130,325	Encrypted Al	FALSE	GTC-6					
4	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	B	GTC	2019-03-15 09:30:00	6 2	127,45	1	1	12745	130,125	130,325	Encrypted Al	FALSE	GTC-6					
5	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	B	GTC	2019-03-15 09:30:00	6 1	127,45	1	1	12745	130,125	130,325	Encrypted Al	FALSE	GTC-6					
6	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	S	DAY	2019-03-15 09:30:00	6 2	131,45	1	1	13145	130,125	130,325	Encrypted Al	FALSE	DAY-6					
7	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	S	DAY	2019-03-15 09:30:00	6 1	131,45	1	1	13145	130,125	130,325	Encrypted Al	FALSE	DAY-6					
8	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	S	DAY	2019-03-15 09:30:00	6 2	131	1	1	13100	130,125	130,325	Encrypted Al	FALSE	DAY-6					
9	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	S	DAY	2019-03-15 09:30:00	6 1	131	1	1	13100	130,125	130,325	Encrypted Al	FALSE	DAY-6					
10	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	B	DAY	2019-03-15 09:30:00	6 2	128,7	1	1	12870	130,125	130,325	Encrypted Al	FALSE	DAY-6					
11	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	B	DAY	2019-03-15 09:30:00	6 1	128,7	1	1	12870	130,125	130,325	Encrypted Al	FALSE	DAY-6					
12	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	S	DAY	2019-03-15 09:30:00	6 2	130,8	1	1	13080	130,125	130,325	Encrypted Al	FALSE	DAY-6					
13	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	S	DAY	2019-03-15 09:30:00	6 1	130,8	1	1	13080	130,125	130,325	Encrypted Al	FALSE	DAY-6					
14	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	B	DAY	2019-03-15 09:30:00	6 2	128,6	1	1	12860	130,125	130,325	Encrypted Al	FALSE	DAY-6					
15	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	B	DAY	2019-03-15 09:30:00	6 1	128,6	1	1	12860	130,125	130,325	Encrypted Al	FALSE	DAY-6					
16	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	S	DAY	2019-03-15 09:30:00	6 2	132,7	1	1	13270	130,125	130,325	Encrypted Al	FALSE	DAY-6					
17	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	S	DAY	2019-03-15 09:30:00	6 1	132,7	1	1	13270	130,125	130,325	Encrypted Al	FALSE	DAY-6					
18	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	S	DAY	2019-03-15 09:30:00	6 2	133,3	1	1	13330	130,125	130,325	Encrypted Al	FALSE	DAY-6					
19	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	S	DAY	2019-03-15 09:30:00	6 1	133,3	1	1	13330	130,125	130,325	Encrypted Al	FALSE	DAY-6					
20	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	S	DAY	2019-03-15 09:30:00	6 2	133,925	1	1	13392,5	130,125	130,325	Encrypted Al	FALSE	DAY-6					
21	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	S	DAY	2019-03-15 09:30:00	6 1	133,925	1	1	13392,5	130,125	130,325	Encrypted Al	FALSE	DAY-6					
22	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	S	DAY	2019-03-15 09:30:00	6 2	131,925	6	6	79155	130,125	130,325	Encrypted Al	FALSE	DAY-6					
23	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	S	DAY	2019-03-15 09:30:00	6 1	131,925	6	6	79155	130,125	130,325	Encrypted Al	FALSE	DAY-6					
24	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	S	DAY	2019-03-15 09:30:00	6 2	131,1	5	5	65550	130,125	130,325	Encrypted Al	FALSE	DAY-6					
25	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	S	DAY	2019-03-15 09:30:00	6 1	131,1	5	5	65550	130,125	130,325	Encrypted Al	FALSE	DAY-6					
26	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	S	DAY	2019-03-15 09:30:00	6 2	131,075	5	5	65537,5	130,125	130,325	Encrypted Al	FALSE	DAY-6					
27	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	S	DAY	2019-03-15 09:30:00	6 1	131,075	5	5	65537,5	130,125	130,325	Encrypted Al	FALSE	DAY-6					
28	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	S	DAY	2019-03-15 09:30:00	6 2	131,1	10	10	131100	130,125	130,325	Encrypted Al	FALSE	DAY-6					
29	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	S	DAY	2019-03-15 09:30:00	6 1	131,1	10	10	131100	130,125	130,325	Encrypted Al	FALSE	DAY-6					
30	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	B	DAY	2019-03-15 09:30:00	6 2	129,225	8	8	103380	130,125	130,325	Encrypted Al	FALSE	DAY-6					
31	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	B	DAY	2019-03-15 09:30:00	6 1	129,225	8	8	103380	130,125	130,325	Encrypted Al	FALSE	DAY-6					
32	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	B	DAY	2019-03-15 09:30:00	6 2	129,3	1	1	12930	130,125	130,325	Encrypted Al	FALSE	DAY-6					
33	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	B	DAY	2019-03-15 09:30:00	6 1	129,3	1	1	12930	130,125	130,325	Encrypted Al	FALSE	DAY-6					
34	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	S	GTC	2019-03-15 09:30:00	6 2	131,075	2	2	26215	130,125	130,325	Encrypted Al	FALSE	GTC-6					
35	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	2	S	GTC	2019-03-15 09:30:00	6 1	131,075	2	2	26215	130,125	130,325	Encrypted Al	FALSE	GTC-6					
36	7112919399	2019-03-15 09:30:00	F_XU030041	Encrypted Alph	Encrypted Alpha Num	1	B	DAY	2019-03-15 09:30:00	6 2	128	1	1	12800	130,125	130,325	Encrypted Al	FALSE	DAY-6					
Raw Orders Raw Trades Order & Trade Msg Filters HFT Accounts & Users Cached Orders Cached Trades Proxy Rules (conditions) Calculated Proxies (order msg) Machine Learning Parameters Printed (Aggregated) Output ⊕																								

Figure A.5: Cached Orders (XLSX Screenshot).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	Deal Number	Trade Date	Trade Time	Contract Code	Deal Type	Direction	Deal Status	Depository Participant Co	Depository Account Number	Clearing Trade Number	Active/Passive	Price	Volume	Traded Notional Value	Binary Order ID	User	Real-HFT							
2	9DF04420001	2019-03-1	09:30:00	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487745	A	130,325	10	130325	7112919399	Encrypted Alf	FALSE							
3	9DF04420001	2019-03-1	09:30:00	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487744	P	130,325	10	130325	7112523575	Encrypted Alf	FALSE							
4	9DF04420001	2019-03-1	09:30:00	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487747	A	130,325	10	130325	7112919399	Encrypted Alf	FALSE							
5	9DF04420001	2019-03-1	09:30:00	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487746	P	130,325	10	130325	7112523575	Encrypted Alf	FALSE							
6	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487749	A	130,250	1	13025	7112919399	Encrypted Alf	TRUE							
7	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487748	P	130,250	1	13025	7112919399	Encrypted Alf	FALSE							
8	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487751	A	130,300	1	13030	7112919399	Encrypted Alf	FALSE							
9	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487750	P	130,300	1	13030	7112919399	Encrypted Alf	FALSE							
10	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487753	A	130,275	1	13027,5	7112919399	Encrypted Alf	TRUE							
11	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487752	P	130,275	1	13027,5	7112919399	Encrypted Alf	FALSE							
12	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487755	P	130,300	1	13030	7112919399	Encrypted Alf	FALSE							
13	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487754	A	130,300	1	13030	7112919399	Encrypted Alf	FALSE							
14	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487757	P	130,300	1	13030	7112919399	Encrypted Alf	FALSE							
15	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487756	A	130,300	1	13030	7112919399	Encrypted Alf	FALSE							
16	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487759	A	130,350	1	13035	7112919399	Encrypted Alf	FALSE							
17	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487758	P	130,350	1	13035	7112523575	Encrypted Alf	FALSE							
18	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487761	A	130,350	2	26070	7112919399	Encrypted Alf	FALSE							
19	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487760	P	130,350	2	26070	7112523575	Encrypted Alf	FALSE							
20	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487763	A	130,350	1	13035	7112919399	Encrypted Alf	FALSE							
21	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487762	P	130,350	1	13035	7112523575	Encrypted Alf	FALSE							
22	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487765	A	130,350	13	169455	7112523575	Encrypted Alf	FALSE							
23	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487764	P	130,350	13	169455	7112523575	Encrypted Alf	FALSE							
24	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487767	A	130,350	1	13035	7112523575	Encrypted Alf	FALSE							
25	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487768	A	130,350	8	104280	7112523575	Encrypted Alf	FALSE							
26	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487766	P	130,350	9	117315	7112919399	Encrypted Alf	FALSE							
27	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487770	A	130,350	1	13035	7112523575	Encrypted Alf	FALSE							
28	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487769	P	130,350	1	13035	7112919399	Encrypted Alf	FALSE							
29	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487772	A	130,375	1	13037,5	7112523575	Encrypted Alf	FALSE							
30	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487771	P	130,375	1	13037,5	7112523575	Encrypted Alf	FALSE							
31	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487774	P	130,350	1	13035	7112919399	Encrypted Alf	FALSE							
32	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487773	A	130,350	1	13035	7112919399	Encrypted Alf	FALSE							
33	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487776	P	130,350	1	13035	7112919399	Encrypted Alf	FALSE							
34	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487775	A	130,350	1	13035	7112919399	Encrypted Alf	FALSE							
35	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	B	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487778	P	130,350	1	13035	7112919399	Encrypted Alf	FALSE							
36	9DF04420001	2019-03-1	09:30:01	F_XU030041	1	S	A	Encrypted Alphabetic Co	Encrypted Alpha Numeric	80487777	A	130,350	1	13035	7112919399	Encrypted Alf	FALSE							
	<	Raw Orders	Raw Trades	Order & Trade Msg Filters		HFT Accounts & Users	Cached Orders	Cached Trades		Proxy Rules (conditions)		Calculated Proxies (order msg)		Machine Learning Parameters		Printed (Aggregated) Output		⊕						

Figure A.6: Cached Trades (XLSX Screenshot).

	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT
1	<u>Proxy-1/Hasbrouck and Saar's (2013) AllRunsInProcess Conditions:</u>											<u>Proxy-3/Ekinci and Ersan's (2018) Two-Step Detection Conditions:</u>											<u>Proxy-5/Extended Withdrawn or Modified Limit Orders Conditions:</u>										
2	n=cached orders row count											n=cached orders row count											n=cached orders row count										
3	FOR row i (i=2,...,n) and row k (k = i - 1),											FOR row i (i=2,...,n) and row k (k = i - 1),											FOR row m (m=1,...,n)										
4	Proxy-1 = TRUE if any of below conditions realized											Proxy-3 = TRUE if any of below conditions realized											Proxy-5 = TRUE if below condition realized										
5																																	
6	cond-1: FOR row i order update time stamp - row i-k order update time stamp < "100ms" AND											cond-1: FOR row i order update time stamp - row i-k order update time stamp < "100ms" AND											FOR row j order update reason = "1" OR "5" AND row j order validity NOT "IMMEDIATE"										
7	row i order update reason = "6"											row i order update reason = "5" OR "1"																					
8	IF row i -k order update reason ="1" AND,											IF row i -k order update reason ="6" AND,																					
9	row i direction ("B" or "S") = row i-k direction ("B" or "S") AND,											row i direction ("B" or "S") = row i-k direction ("B" or "S") AND,																					
10	row i quantity = row i-k quantity											row i quantity = row i-k quantity																					
11	cond-2: FOR row i order update time stamp - row i-k order update time stamp < "100ms" AND											cond-2: FOR row i order update time stamp - row i-k order update time stamp < "100ms" AND																					
12	row i order update reason = "3"											row i order update reason = "1"																					
13	IF row i -k order update reason ="1" AND,											IF row i -k order update reason ="5" AND,																					
14	row i direction ("B" or "S") = row i-k direction ("B" or "S") AND,											row i direction ("B" or "S") = row i-k direction ("B" or "S") AND,																					
15	row i quantity - row i remaining quantity = row i-k quantity											row i quantity - row i remaining quantity = row i-k quantity																					
16																																	
17	cond-3: FOR row i order update time stamp - row i-k order update time stamp < "100ms" AND											cond-3: FOR row i order update time stamp - row i-k order update time stamp < "100ms" AND																					
18	row i order update reason = "1"											row i order update reason = "5"																					
19	IF row i -k order update reason ="6" AND,											IF row i -k order update reason ="5" AND,																					
20	row i direction ("B" or "S") = row i-k direction ("B" or "S") AND,											row i direction ("B" or "S") = row i-k direction ("B" or "S") AND,																					
21	row i quantity = row i-k quantity											row i quantity = row i-k quantity																					
22	ELSE row i (i=2,...,n)											ELSE row i (i=2,...,n)																					
23	Proxy-1 = FALSE											Proxy-3 = FALSE																					
24																																	
25	<u>Proxy-2/Ersan and Ekinci's (2016) AllRunsInExtended Conditions:</u>											<u>Proxy-4/Van Ness et al.'s (2015) Withdrawn Limit Orders Conditions:</u>																					
26	n=cached orders row count											n=cached orders row count																					
27	FOR row i (i=2,...,n) and row k (k = i - 1),											FOR row j (j=1,...,n)																					
28	Proxy-2 = TRUE if below condition realized											Proxy-4 = TRUE if below condition realized																					
29												FOR row j order update reason = "1" AND row j order validity NOT "IMMEDIATE"																					
30	FOR row i order update time stamp - row i-k order update time stamp < "100ms" AND																																
31	row i quantity = row i-k quantity																																
32																																	
33	ELSE row i (i=2,...,n)											ELSE row j (j=1,...,n)																					
34	Proxy-2 = FALSE											Proxy-4 = FALSE																					
35																																	
36																																	
37																																	
38																																	
39																																	
40																																	

Figure A.7: Proxy Rules (XLSX Screenshot).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Binary Order ID	Real-HFT	Proxy-1	Proxy-2	Proxy-3	Proxy-4	Proxy-5									
2	7112919399314812816_80_170_126_0_13765_15518_0_1	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE									
3	7112919399314812816_80_170_126_0_13765_15518_0_1	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
4	7112919399314812870_80_170_126_0_13765_15518_0_1	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE									
5	7112919399314812870_80_170_126_0_13765_15518_0_1	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
6	7112919399314812991_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
7	7112919399314812991_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
8	7112919399314812992_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
9	7112919399314812992_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
10	7112919399314812993_80_170_126_0_13765_15518_0_1	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
11	7112919399314812993_80_170_126_0_13765_15518_0_1	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
12	7112919399314812994_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
13	7112919399314812994_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
14	7112919399314812995_80_170_126_0_13765_15518_0_1	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
15	7112919399314812995_80_170_126_0_13765_15518_0_1	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
16	7112919399314812996_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
17	7112919399314812996_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
18	7112919399314812997_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
19	7112919399314812997_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
20	7112919399314812998_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
21	7112919399314812998_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
22	7112919399314813009_80_170_126_0_13765_15518_0_2	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE									
23	7112919399314813009_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
24	7112919399314813011_80_170_126_0_13765_15518_0_2	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE									
25	7112919399314813011_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
26	7112919399314813012_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
27	7112919399314813012_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
28	7112919399314813013_80_170_126_0_13765_15518_0_2	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE									
29	7112919399314813013_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
30	7112919399314813014_80_170_126_0_13765_15518_0_1	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE									
31	7112919399314813014_80_170_126_0_13765_15518_0_1	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
32	7112919399314813029_80_170_126_0_13765_15518_0_1	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
33	7112919399314813029_80_170_126_0_13765_15518_0_1	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
34	7112919399314813050_80_170_126_0_13765_15518_0_2	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE									
35	7112919399314813050_80_170_126_0_13765_15518_0_2	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
36	7112919399314813269_80_170_126_0_13765_15518_0_1	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE									
Raw Orders		Raw Trades	Order & Trade Msg Filters	HFT Accounts & Users	Cached Orders	Cached Trades	Proxy Rules (conditions)	Calculated Proxies (order msg)		Machine Learning Parameters			Printed (Aggregated) Output			

Figure A.8: Calculated RHFT and Proxies (XLSX Screenshot).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
1	Binary Order	Real-HFT/Proxy-1	Real-HFT/Proxy-2	Real-HFT/Proxy-3	Real-HFT/Proxy-4	Real-HFT/Proxy-5	Proxy-1 TPs	Proxy-1 FPs	Proxy-1 TNs	Proxy-1 FNs	Proxy-2 TPs	Proxy-2 FPs	Proxy-2 TNs	Proxy-2 FNs	Proxy-3 TPs	Proxy-3 FPs	Proxy-3 TNs	Proxy-3 FNs	Proxy-4 TPs	Proxy-4 FPs	Proxy-4 TNs	Proxy-4 FNs	Proxy-5 TPs	Proxy-5 FPs	Proxy-5 TNs	Proxy-5 FNs	
2	711291935	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	
3	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
4	711291935	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0
5	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
6	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
7	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
8	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
9	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
10	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
11	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
12	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
13	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
14	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
15	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
16	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
17	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
18	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
19	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
20	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
21	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
22	711291935	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0
23	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
24	711291935	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0
25	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
26	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
27	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
28	711291935	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0
29	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
30	711291935	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0
31	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
32	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
33	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
34	711291935	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0
35	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
36	711291935	FALSE-FALSE	FALSE-TRUE	FALSE-FALSE	FALSE-FALSE	FALSE-FALSE	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
	< >	Raw Orders	Raw Trades	Order & Trade Msg Filters	HFT Accounts & Users	Cached Orders	Cached Trades	Proxy Rules (conditions)	Calculated Proxies (order msg)	Machine Learning Parameters					Printed (Aggregated) Output					⊕							

Figure A.9: Machine Learning Parameters (XLSX Screenshot).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1																					
2	Date	Instrument	Total Order Msgs	Real HFT Ords	Proxy-1 HFTs	Proxy-2 HFTs	Proxy-3 HFTs	Proxy-4 HFTs	Proxy-5 HFTs	Real HFT %	Proxy-1 %	Proxy-2 %	Proxy-3 %	Proxy-4 %	Proxy-5 %	Total Trd Vol	Total Trd Value	Real HFT Trd Vol	Real HFT Trd Value %	Real HFT Trd Value %	Real HFT Trd Value %
3	2019-03-15	F_XU0300419	243886	138290	2970	128334	78660	9580	137825	0,5670272	0,0121778	0,526204866	0,3225277	0,1659104	0,7312175	228996	2992824745	52296	0,228370801	683451095	0,22836322
4																					
5																					
6																					
7	Confusion Matrix Parameters																				
8																					
9		Proxy-1	Proxy-2	Proxy-3	Proxy-4	Proxy-5															
10	TPs	1833	93969	69003	4424	113187															
11	%	0,75%	38,53%	28,29%	1,81%	46,41%															
12	FPS	1137	34365	9657	5156	24638															
13	%	0,47%	14,09%	3,96%	2,11%	10,10%															
14	TNs	104459	71231	95939	100440	80958															
15	%	42,83%	29,21%	39,34%	41,18%	33,20%															
16	FNs	136457	44321	69287	133866	25103															
17	%	55,95%	18,17%	28,41%	54,89%	10,29%															
18	Total	243886	243886	243886	243886	243886															
19		100,00%	100,00%	100,00%	100,00%	100,00%															
20																					
21																					
22																					
23																					
24																					
25																					
26																					
27																					
28																					
29																					
30																					
31																					
32																					
33																					
34																					
35																					
36																					
	<	>	Raw Orders	Raw Trades	Order & Trade Msg Filters	HFT Accounts & Users	Cached Orders	Cached Trades	Proxy Rules (conditions)	Calculated Proxies (order msg)	Machine Learning Parameters	Printed (Aggregated) Output	+	:	<	>					

Figure A.10: Printed Aggregated Output (XLSX Screenshot).

APPENDIX B : Program JAVA Codes

Main.Java²⁸

```
package com.hftproxychecker;
import com.hftproxychecker.entity.Order;
import com.hftproxychecker.entity.Trade;
import com.hftproxychecker.proxychecker.PeriodUtil;
import com.hftproxychecker.proxychecker.ProxyChecker;
import com.hftproxychecker.proxychecker.algorithms.Algorithm;
import com.hftproxychecker.proxychecker.writer.CombinedExcelWriterProvider;
import com.hftproxychecker.proxychecker.writer.SimpleExcelWriterProvider;
import com.hftproxychecker.proxychecker.writer.WriterProvider;
import com.hftproxychecker.reader.DefaultFileNames;
import com.hftproxychecker.util.TimeSpan;
import lombok.extern.slf4j.Slf4j;
import java.io.File;
import java.io.IOException;
import java.io.InputStream;
import java.lang.reflect.InvocationTargetException;
import java.util.*;
@SuppressWarnings("resource")
@Slf4j
public class Main {
    private static final String FILE_NAME_TITLE_ORDER = "orderFile";
    private static final String FILE_NAME_TITLE_TRADE = "tradeFile";

    private static final String FILE_NAME_TITLE_COMBINED_OUTPUT =
"combinedOutputFile";
    private static final String
FILE_NAME_TITLE_COMBINED_OUTPUT_FILE_DIR =
"combinedOutputFileDirectory";
```

²⁸ The codes for entities Order, Trade, ProxyChecker and TimeSpan are available upon request.

```

    private static final String
FILE_NAME_TITLE_COMBINED_OUTPUT_ORDER_SHEET_NAME =
"combinedOutputFileOrderSheetName";

    private static final String FILE_NAME_TITLE_ML_EXT_FILE =
"machineLearningExtensionFile";

    private static final String
FILE_NAME_TITLE_ML_EXT_FILE_PREFERRED_SHEET =
"machineLearningExtensionFilePreferredSheetName";

    private static final String FILE_NAME_TITLE_HFT_ACC = "hftAccountFile";
    private static final String FILE_NAME_TITLE_HFT_USER = "hftUserFile";
    private static final String FILE_NAME_TITLE_CHR_ORDER =
"includedOrdersChangeReasonsFile";

    private static final String FILE_NAME_TITLE_CHR_TRADE =
"includedTradesChangeReasonsFile";

    private static final String FILE_NAME_TITLE_ALGO_NAMES =
"includedAlgoNames";

    private static final String FILE_NAME_TITLE_ORDER_PROCESSED =
"processedOrderFileName";

    private static final String FILE_NAME_TITLE_TRADE_PROCESSED =
"processedTradeFileName";

    private static final String IS_OUTPUT_COMBINED = "combinedOutput";
    private static final String IS_WRITE_PROCESSED_ITEMS =
"writeProcessedItems";

    private static final String EXCEL_PROGRAM_PATH = "excelProgramPath";
    private static final String OPEN_EXCEL_PROGRAM = "openExcelProgram";
    private static final String COPY_TO_CLIPBOARD = "copyToClipboard";
    private static final String ALL_SUMMARIES_EXCEL_FILE =
"allSummariesExcelFile";

    private static final String ALL_SUMMARIES_EXCEL_FILE_SHEET =
"allSummariesExcelFileSheetName";


    private static final String PERIOD = "period";


    private static final String COMMAND = "command";


    public static void main(String[] args) {
        long startTime = System.nanoTime();

```

```

    if (!parseGivenArguments(args)) {
        System.exit(1);
    }

    String orderFiles = getArgValue(FILE_NAME_TITLE_ORDER);
    String tradeFiles = getArgValue(FILE_NAME_TITLE_TRADE);
    String combinedOutputFiles =
getArgValue(FILE_NAME_TITLE_COMBINED_OUTPUT);
    String combinedOutputFilesDirectory =
getArgValue(FILE_NAME_TITLE_COMBINED_OUTPUT_FILE_DIR);
    String combinedOutputFileOrderSheetName =
getArgValue(FILE_NAME_TITLE_COMBINED_OUTPUT_ORDER_SHEET_NAME);
    String mlExtensionFileName =
getArgValue(FILE_NAME_TITLE_ML_EXT_FILE);
    String mlExtensionFilePreferredSheetName =
getArgValue(FILE_NAME_TITLE_ML_EXT_FILE_PREFERRED_SHEET);
    String hftAccounts = getArgValue(FILE_NAME_TITLE_HFT_ACC);
    String hftUsers = getArgValue(FILE_NAME_TITLE_HFT_USER);
    String includedOrdersChangeReasons =
getArgValue(FILE_NAME_TITLE_CHR_ORDER);
    String includedTradesChangeReasons =
getArgValue(FILE_NAME_TITLE_CHR_TRADE);
    String processedOrderFileName =
getArgValue(FILE_NAME_TITLE_ORDER_PROCESSED);
    String processedTradeFileName =
getArgValue(FILE_NAME_TITLE_TRADE_PROCESSED);
    String outputCombined = getArgValue(IS_OUTPUT_COMBINED);
    String writeProcessedItems =
getArgValue(IS_WRITE_PROCESSED_ITEMS);
    String excelProgramPath = getArgValue(EXCEL_PROGRAM_PATH);
    String openExcelProgram = getArgValue(OPEN_EXCEL_PROGRAM);
    String copyToClipboard = getArgValue(COPY_TO_CLIPBOARD);
    String allSummariesExcelFile =
getArgValue(ALL_SUMMARIES_EXCEL_FILE);
    String allSummariesExcelFileSheetName =
getArgValue(ALL_SUMMARIES_EXCEL_FILE_SHEET);
    String period = getArgValue(PERIOD);
    String command = getArgValue(COMMAND);

```

```

PeriodUtil.setPeriod(period);

if ("splitTradeFile".equalsIgnoreCase(command)) {
    final Set<String> tradeFileNameSet =
getMultipleFileNames(FILE_NAME_TITLE_TRADE, tradeFiles);
    ProxyChecker.splitTradeFile(tradeFileNameSet.toArray(new String[0]));
} else if ("createMachineLearningExtension".equalsIgnoreCase(command)) {
    if (combinedOutputFiles != null && !combinedOutputFiles.isEmpty()) {
        final Set<String> fileNames =
getMultipleFileNames(FILE_NAME_TITLE_COMBINED_OUTPUT,
combinedOutputFiles);

        ProxyChecker.createMachineLearningExtension(fileNames.toArray(new
String[0]), combinedOutputFileOrderSheetName, mlExtensionFileName,
mlExtensionFilePreferredSheetName);

    } else if (combinedOutputFilesDirectory != null &&
!combinedOutputFilesDirectory.isEmpty()) {

ProxyChecker.createMachineLearningExtensionByDirectory(combinedOutputFilesD
irectory, combinedOutputFileOrderSheetName, mlExtensionFileName,
mlExtensionFilePreferredSheetName);

    }
} else {
    final Set<String> orderFileNameSet =
getMultipleFileNames(FILE_NAME_TITLE_ORDER, orderFiles);

    final Set<String> tradeFileNameSet =
getMultipleFileNames(FILE_NAME_TITLE_TRADE, tradeFiles);
    checkFileExistence(FILE_NAME_TITLE_TRADE, tradeFiles);
    checkFileExistence(FILE_NAME_TITLE_HFT_ACC, hftAccounts);
    checkFileExistence(FILE_NAME_TITLE_HFT_USER, hftUsers);
    checkFileExistence(FILE_NAME_TITLE_CHR_ORDER,
includedOrdersChangeReasons);

    checkFileExistence(FILE_NAME_TITLE_CHR_TRADE,
includedTradesChangeReasons);

    WriterProvider writerProvider;

    if (outputCombined == null || outputCombined.isEmpty() ||
isTrue(outputCombined)) {
        writerProvider = new CombinedExcelWriterProvider();
    }
}

```

```

    } else {
        writerProvider = new SimpleExcelWriterProvider();
    }

    ProxyChecker.runAll(
        includedOrdersChangeReasons,
        includedTradesChangeReasons,
        hftAccounts,
        hftUsers,
        orderFileNameSet.toArray(new String[0]),
        tradeFileNameSet.toArray(new String[0]),
        processedOrderFileName,
        processedTradeFileName,
        getOrderAlgorithms(),
        getTradeAlgorithms(),
        writerProvider,
        isTrue(writeProcessedItems));

    log.info("Completed total elapsed time > " + new
    TimeSpan(System.nanoTime() - startTime));

    if (isTrue(copyToClipboard)) {
        try {
            String clipboardData = ProxyChecker.getClipboardData();
            try {
                clipboardData =
ExcelRowCopyUtil.getRowAsString(CombinedExcelWriterProvider.getFileName(),
2);

            } catch (Exception e) {
                log.error("ExcelRowCopyUtil failed" + e);
            }
            ClipboardUtil.copyToClipboard(clipboardData);
        } catch (Exception e) {
            log.error("Failed to copy to clipboard", e);
        }
    }
}

```

```

    try {
        AllSummaryWriterUtil.append(allSummariesExcelFile,
allSummariesExcelFileSheetName);
    } catch (Exception e) {
        log.error("Failed to append print summary to allSummariesExcelFile=" +
allSummariesExcelFile, e);
    }

    if (isTrue(openExcelProgram)) {
        try {

            if (excelProgramPath == null) {
                String programFilesDir = System.getenv("PROGRAMFILES");
                if (programFilesDir == null || programFilesDir.isEmpty()) {
                    programFilesDir = "C:\\Program Files\\";
                }

                for (int i = 30; i > 0; i--) {
                    excelProgramPath = programFilesDir + "\\Microsoft
Office\\Office" + i + "\\EXCEL.EXE";
                    File excel = new File(excelProgramPath);
                    if (excel.isFile()) {
                        break;
                    }
                }
                File excel = new File(excelProgramPath);
                if (excel.isFile()) {
                    Runtime.getRuntime().exec(excelProgramPath + " " +
CombinedExcelWriterProvider.getFileName());
                } else {
                    log.error("Failed to find excel.exe");
                }
            } catch (IOException e) {
                log.error("Failed to run excel", e);
            }
        }
    }
}

```

```

    }
    }
}
}

```

```

private static Set<String> getMultipleFileNames(String fileType, String
fileNames) {
    final Set<String> nameSet = new HashSet<>();
    for (String orderFile : fileNames.split(",")) {
        if (orderFile != null && !orderFile.trim().isEmpty()) {
            checkFileExistence(fileType, orderFile);
            nameSet.add(orderFile);
        }
    }

    if (nameSet.isEmpty()) {
        log.error(fileType + " is not given!");
        System.exit(1);
    }
    return nameSet;
}

```

```

private static void checkFileExistence(String fileType, String fileName) {
    if (fileName == null || fileName.isEmpty()) {
        log.error(fileType + " is not given!");
        System.exit(1);
    }
}

```

```

File file = new File(fileName);
if (!file.isFile()) {
    InputStream is;
    try {
        is = Main.class.getClassLoader().getResourceAsStream(fileName);
    } catch (Exception e) {
        log.error(e.toString(), e);
    }
}

```

```

        is = null;
    }

    if (is == null) {
        log.error(fileType + " is not exists!");
        System.exit(1);
    }
}

private final static Map<String, Argument> defaultArguments = new
HashMap<>();

private static void addPossibleArgument(String name, String value) {
    defaultArguments.put(name, new Argument(name, value));
}

private static boolean isTrue(String booleanValue) {
    return booleanValue.equalsIgnoreCase("yes") ||
booleanValue.equalsIgnoreCase("true") || booleanValue.equalsIgnoreCase("y");
}

static {
    addPossibleArgument(FILE_NAME_TITLE_ORDER, "");
    addPossibleArgument(FILE_NAME_TITLE_TRADE, "");
    addPossibleArgument(FILE_NAME_TITLE_COMBINED_OUTPUT, "");

    addPossibleArgument(FILE_NAME_TITLE_ML_EXT_FILE, "ml_extension.xlsx");

    addPossibleArgument(FILE_NAME_TITLE_ML_EXT_FILE_PREFERRED_SHEET, "");

    addPossibleArgument(FILE_NAME_TITLE_COMBINED_OUTPUT_FILE_DIR, "");

    addPossibleArgument(FILE_NAME_TITLE_COMBINED_OUTPUT_ORDER_SHEET_NAME, "Orders");

    addPossibleArgument(FILE_NAME_TITLE_HFT_ACC,
DefaultFileNames.FILE_NAME_HFT_ACC);
}

```

```

        addPossibleArgument(FILE_NAME_TITLE_HFT_USER,
DefaultFileNames.FILE_NAME_HFT_USER);

addPossibleArgument(FILE_NAME_TITLE_CHR_ORDER,DefaultFileNames.FIL
E_NAME_CHR_ORDER);

addPossibleArgument(FILE_NAME_TITLE_CHR_TRADE,DefaultFileNames.FIL
E_NAME_CHR_TRADE);

addPossibleArgument(FILE_NAME_TITLE_ALGO_NAMES,"realhft,proxy1,proxy
2,proxy3,proxy4,proxy5,proxy6");

addPossibleArgument(FILE_NAME_TITLE_ORDER_PROCESSED,"Orders");

addPossibleArgument(FILE_NAME_TITLE_TRADE_PROCESSED,"Trades");
        addPossibleArgument(IS_OUTPUT_COMBINED,"yes");
        addPossibleArgument(IS_WRITE_PROCESSED_ITEMS,"no");
        addPossibleArgument(OPEN_EXCEL_PROGRAM, "no");
        addPossibleArgument(COPY_TO_CLIPBOARD, "yes");
        addPossibleArgument(ALL_SUMMARIES_EXCEL_FILE,
"allSummaries.xlsx");
        addPossibleArgument(ALL_SUMMARIES_EXCEL_FILE_SHEET,
"Summaries");
        addPossibleArgument(PERIOD, "");
        addPossibleArgument(COMMAND, "processProxies");
    }

    private final static Map<String, Argument> givenArguments = new
HashMap<>();

    private static boolean parseGivenArguments(String[] args) {
        log.info("Command line=" + Arrays.toString(args));

        if (args == null || args.length == 0) {
            log.info("Invalid command line parameters. Possible parameters:");
            for (Argument argument : defaultArguments.values()) {
                log.info(argument.toString());
            }
            return false;
        }
    }

```

```

    }

    givenArguments.clear();
    for (String arg : args) {
        String[] splits = arg.split("=");
        if (splits.length != 2) {
            continue;
        }
        String name = splits[0].trim();
        String value = splits[1].trim();
        givenArguments.put(name, new Argument(name, value));
    }

    log.info("Given valid arguments:");
    for (Argument argument : givenArguments.values()) {
        log.info(argument.toString());
    }
    return true;
}

private static String getDefaultValueForArg(String name) {
    String value = null;

    for (Argument argument : defaultArguments.values()) {
        if (argument.equals(name)) {
            value = argument.getValue();
            break;
        }
    }
    log.info(name + "=" + value + " (default value)");
    return value;
}

private static String getArgValue(String name) {
    String value = null;

```

```

    for (Argument argument : givenArguments.values()) {
        if (argument.equals(name)) {
            value = argument.getValue();
            break;
        }
    }
    log.info(name + "=" + value);
    if (value == null) {
        value = getDefaultValueForArg(name);
    }
    return value;
}

@SuppressWarnings("unchecked")
private static List<Algorithm<Order, Order>> getOrderAlgorithms() {
    List<Algorithm<Order, Order>> algorithms = new ArrayList<>();
    String includedAlgoNames =
getArgValue(FILE_NAME_TITLE_ALGO_NAMES);
    String[] algoNames = includedAlgoNames.toLowerCase().split(",");
    for (String algo : algoNames) {
        String className = "com.hftproxychecker.proxychecker.algorithms." + algo
+ ".OrderProcessor" ;
        try {
            algorithms.add((Algorithm<Order, Order>)
Class.forName(className).getDeclaredConstructor().newInstance());
        } catch (ClassNotFoundException e) {
            log.error("Order processor class not found for algo = " + algo + "
className = " + className);
        } catch (InstantiationException e) {
            log.error("Order processor class not initialized for algo = " + algo + "
className = " + className);
        } catch (IllegalAccessException e) {
            log.error("IllegalAccessException: Order processor class for algo = " +
algo + " className = " + className);
        } catch (InvocationTargetException e) {
            log.error("InvocationTargetException: Order processor class for algo = " +
algo + " className = " + className);
        } catch (NoSuchMethodException e) {

```

```

        log.error("NoSuchMethodException: Order processor class for algo = " +
algo + " className = " + className);
    }
}
return algorithms;
}
@SuppressWarnings("unchecked")
private static List<Algorithm<Trade, Trade>> getTradeAlgorithms() {
    List<Algorithm<Trade, Trade>> algorithms = new ArrayList<>();
    String includedAlgoNames =
getArgValue(FILE_NAME_TITLE_ALGO_NAMES);
    String[] algoNames = includedAlgoNames.toLowerCase().split(",");
    for (String algo : algoNames) {
        String className = "com.hftproxychecker.proxychecker.algorithms." + algo
+ ".TradeProcessor" ;
        try {
            algorithms.add((Algorithm<Trade, Trade>)
Class.forName(className).getDeclaredConstructor().newInstance());
        } catch (ClassNotFoundException e) {
            log.error("Trade processor class not found for algo = " + algo + "
className = " + className);
        } catch (InstantiationException e) {
            log.error("Trade processor class not initialized for algo = " + algo + "
className = " + className);
        } catch (IllegalAccessException e) {
            log.error("IllegalAccessException: Trade processor class for algo = " +
algo + " className = " + className);
        } catch (InvocationTargetException e) {
            log.error("InvocationTargetException: Trade processor class for algo = " +
algo + " className = " + className);
        } catch (NoSuchMethodException e) {
            log.error("NoSuchMethodException: Trade processor class for algo = " +
algo + " className = " + className);
        }
    }
    return algorithms;
}
}

```

APPENDIX C : Declaration of Disclosures

Author Onur Olgun is in the employ of Borsa Istanbul. The views expressed in this thesis are those of the author and do not necessarily represent the official views of the Borsa Istanbul.

The data in the thesis is obtained from Borsa Istanbul by special consent. The author does not have permission to share data.

OpenAI ChatGPT-4 was utilized by the author in the thesis drafting process to improve readability, refine flow, and ensure grammatical correctness.





CURRICULUM VITAE

Name Surname : Onur OLGUN

EDUCATION :

- **B.Sc.** : 2005, Hacettepe University, Faculty of Economic and Administrative Sciences, Business Administration
- **M.Sc.** : 2007, University of Southampton, Management Faculty, International Financial Markets
- **M.A** : 2010, Izmir University of Economics, Graduate School, Financial Economics

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2006-2010: Research Assistant, Izmir University of Economics, Dep. of Finance
- 2010-2013: Assistant Expert, Turkish Derivatives Exchange, Market Operations
- 2013-2021: Senior Market Specialist, Borsa Istanbul, Derivatives Market (VIOP)
- 2021-2021: Senior Business Analyst, XYZ Technologies
- 2021-Present: Lead Business Analyst, Borsa Istanbul, System Development (IT)

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Olgun, O., Ekinci, C. & Arıkan, R.** (2024). The performance of selected high-frequency trading proxies: An application on Turkish index futures market. *Finance Research Letters* 65, 105523.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- **Olgun, O. & Yetkiner, H.** (2011). Determination of optimal hedging strategy for index futures: Evidence from Turkey. *Emerging Markets Finance and Trade* 47(6), 68-79.
- **Baklaci, H., Olgun, O. & Can, E.** (2011). Noise traders: a new approach to understand the phantom of stock markets. *Applied Economics Letters* 18(11), 1035-1041.
- **Okan, B., Olgun, O. & Takmaz, S.** (2009). Volume and volatility: A case of ISE-30 index futures. *International Res. Journal of Finance and Economics* 32(10), 93-103.
- **Aksoy, G. & Olgun, O.** (2009). Optimal Hedge oranı tahminlemesi üzerine ampirik bir çalışma: VOB örneği. *İktisat İşletme ve Finans* 24(274), 33-53.