KADIR HAS UNIVERSITY
SCHOOL OF GRADUATE STUDIES
DEPARTMENT OF FINANCIAL ENGINEERING

INFORMED TRADING AROUND EXTREME EVENTS IN BORSA ISTANBUL

MEHMET BODUR

MASTER OF SCIENCE THESIS

ISTANBUL, JANUARY, 2022

Mehmet Bodur

Master of Science Thesis

2022

INFORMED TRADING AROUND EXTREME EVENTS IN BORSA ISTANBUL

MEHMET BODUR

A thesis submitted to
the School of Graduate Studies of Kadir Has University
in partial fulfilment of the requirements for the degree of
Master of Science in
Financial Engineering

Istanbul, January, 2022

APPROVAL

This thesis titled INFORMED TRADING AROUND EXTREME EVENTS IN BORSA ISTANBUL submitted by MEHMET BODUR, in partial fulfilment of the requirements for the degree of Master of Science in Financial Engineering is approved by

Asst. Prof. Oğuz Ersan (Advisor) ....
Kadir Has University

Assoc. Prof. Cumhur Ekinci ....
Istanbul Technical University

Asst. Prof. Gamze Öztürk Danışman ....
Kadir Has University

I confirm that the signatures above belong to the aforementioned faculty members.

....
Prof. Dr. Mehmet Timur Aydemir
Director of the School of Graduate Studies
Date of Approval: 11/01/2022

DECLARATION ON RESEARCH ETHICS AND PUBLISHING METHODS

I, MEHMET BODUR; hereby declare

- that this Master of Science Thesis that I have submitted is entirely my own work and I have cited and referenced all material and results that are not my own in accordance with the rules;
- that this Master of Science Thesis does not contain any material from any research submitted or accepted to obtain a degree or diploma at another educational institution;
- and that I commit and undertake to follow the "Kadir Has University Academic Codes of Conduct" prepared in accordance with the "Higher Education Council Codes of Conduct".

In addition, I acknowledge that any claim of irregularity that may arise in relation to this work will result in a disciplinary action in accordance with the university legislation.

Mehmet Bodur

....
11/01/2022

To all uninformed traders...

ACKNOWLEDGMENTS

I am deeply thankful to Asst. Prof. Oğuz Ersan for his extensive contribution to my academic performance during my education at Kadir Has University. Also, I am more than grateful to him for his enormous generosity to let me participate in his research project. By working on his side, I did improve my academic-oriented research perspective. Additionally, I thank him for his guidance, contribution, and evaluation on this study which helped me to cover the topic in a holistic view.

I would like to thank Assoc. Prof. Cumhur Ekinci for letting me learn from his extensive years of experience in academic research. I wish to thank his understanding, patience, and confidence in me during our co-operated research practices. Also, I will be deeply thankful for his contribution to the revision of the thesis. In addition, Asst. Prof. Gamze Öztürk Danışman's recognition of and participation in the defense jury will always be appreciated from my side. Thanks to her suggestions, I have enhanced some of the cardinal parts of the thesis. I also thank Asst. Prof. Murat Tiniç for letting me participate to work with him on an academic research paper. With his guidance during the process, I experienced a unique opportunity in the context of academic research. Next, I wish to thank Assoc. Prof. Mehmet Horasanlı for his lectures and positive effects on my mindset on academia and industry. I will always be pleased for attending his classes. I thank Prof. Dr. Volkan Şevket Ediger to help me on built my academic-wise way of thinking. His perspective gives me the opportunity to conduct better research efforts.

Last but not least, I wish to thank Prof. Dr. Ayşe Hümeyra Bilge for accepting me into the graduate program at Kadir Has University with a merit-based scholarship. Because of her assessment and approval, I had the chance to follow unique education in Kadir Has University where I enjoyed it from the beginning. Therefore, finally, I wish to thank the School of Graduate Studies for letting me experience this academic journey which broadened my perspective, both academically and professionally.

INFORMED TRADING AROUND EXTREME EVENTS IN BORSA ISTANBUL

ABSTRACT

On February 22, 2017, two Turkish blue-chip stocks Koç Holding (KCHOL) and Turkcell (TCELL) experienced a simultaneous flash event resulting in a sudden price crash during the continuous auction in Borsa Istanbul Equity Market, right before the market closing time. Both stocks experienced a nearby 10% fall before subsequent price recovery. KCHOL (TCELL) falls as much as %9.86 (%10.77) between 17:45:00 – 17:45:01 time period for an approximately 1-second interval. Before the respective event, order flow toxicity for informed trading proxy VPIN – Volume-Synchronized Probability of Informed Trading shows consecutive increasing behavior even before the sudden crash for TCELL whereas no concrete in advance reaction for KCHOL. VPIN levels for KCHOL (TCELL) increase (decrease) in the course of the post-event interval. Such a difference may be interpreted as increasing (decreasing) order flow toxicity for KCHOL (TCELL) trade balance. Univariate and multivariate regressions' implied empirical findings result in the statistically significant predictive power of VPIN for TCELL on impending VWAP - Volume Weighted Average Price pattern. However, for KCHOL, no reliable explanatory role of VPIN after considering control variables. Such indefinite results may imply different algorithmic trading strategy execution for KCHOL and TCELL with respect to the event and post-event periods.

Keywords: Borsa Istanbul, Order Flow Toxicity, Informed Trading, Flash Crash, Market Microstructure, Volume-Synchronized Probability of Informed Trading

BORSA İSTANBUL'DA AŞIRI OLAYLAR ETRAFINDAKİ BİLGİYE DAYALI İŞLEM

ÖZET

Şubat 22, 2017 tarihinde, yüksek işlem hacmine sahip iki adet Türk hisse senedi olan Koç Holding (KCHOL) ve Turkcell (TCELL) işlemlerinde, Borsa İstanbul Pay Piyasası'nda seans kapanışından kısa süre önce sürekli işlem saatleri içerisinde ani fiyat çöküşü yaşanmıştır. Her iki hisse senedi de %10' yakın düşüş göstermiş, ardından fiyatlar tekrar eski seviyelerine doğru toparlanmıştır. Saat 17:45:00 – 17:45:01 aralığındaki yaklaşık 1-saniyelik zaman diliminde KCHOL (TCELL) yaklaşık %9.86 (%10.77) düşüş göstermiştir. Olaydan önce, bilgiye dayalı alım-satım göstergesi olan VPIN – Volume-Synchronized Probability of Informed Trading TCELL için düzenli bir şekilde yükseliş davranışı göstermiş iken KCHOL için ise belirli bir tepki vermemiştir. KCHOL (TCELL) için VPIN seviyesi olay sonrası aralıkta yükseliş (düşüş) göstermiştir. Böyle bir farklılık KCHOL (TCELL) işlem dengesi için artan (düşen) toksik emir akışı olarak yorumlanabilir. TCELL için tek ve çok değişkenli regresyonların işaret ettiği ampirik bulgular, VPIN için bir sonraki HAOF - Hacim Ağırlıklı Ortalama Fiyat patikası üzerinde istatistiki geçerlilikte açıklayıcılığı ortaya koymaktadır. Fakat, KCHOL için kontrol değişkenleri dikkate alındıktan sonra VPIN için güvenilir bir açıklayıcılıktan bahsedilememektedir. Böylesine kesin olmayan sonuçlar, KCHOL ve TCELL için olay ve olay sonrası dönemlerde uygulanmış farklı algoritmik alım-satım stratejisi uygulandığına işaret edebilmektedir.

Anahtar Sözcükler: Borsa İstanbul, Emir Akış Toksikliği, Bilgiye Dayalı Alım-Satım, Ani Çöküş, Piyasa Mikroyapısı, Volume-Synchronized Probability of Informed Trading

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$\rho_{kchol,tcell}$	1-year period correlation between KCHOL and TCELL
ϵ	Error Term
σ^4	Fourth Moment
μ	Mean
R^2	R-Squared
$S.E.$	Standard Error
σ	Standard Deviation
σ^3	Third Moment
σ^2	Variance

LIST OF ABBREVIATIONS

AMEX	American Stock Exchange
BIST	Borsa Istanbul
CMB	Capital Markets Board of Turkey
EMH	Efficient Market Hypothesis
HHI	Herfindahl-Hirschman Index
HFT	High-Frequency Trading
HML	High Minus Low
LSE	London Stock Exchange
MLE	Maximum Likelihood Estimation
NYSE	New York Stock Exchange
PIN	Probability of Informed Trading
PDP	Public Disclosure Platform
QMJ	Quality Minus Junk
SMB	Small Minus Big
UMD	Up Minus Down
VPIN	Volume-Synchronized Probability of Informed Trading

1. INTRODUCTION

Financial markets tend to experience rare but instrumental pricing behaviors. Due to the widespread application of systematic trading activities, especially after high-frequency trading (HFT) participation[1] these price movements have started to cause even intense sudden swings. Therefore, flash crash and jump events are occurred instantaneously, according to high-frequency quotation ability. Observation of any imbalances on market quality proxies[2] may be related to a subsequent extreme event. The market quality around such abnormal events deserves to be investigated carefully due to the difficulties of observing rapid price movement outputs. During such market conditions, liquidity provision between HFT and market maker counterparties becomes even more crucial on behalf of the market efficiency balance. Order flow toxicity as a proxy for information asymmetry is yielded as a result of an unbalanced quotation dynamic between informed and uninformed traders. Accordingly, informed traders (HFTs) who have information superiority, demand liquidity based on positive adverse selection whereas uninformed traders (non-HFT market makers) provide liquidity at a loss (Easley et al., 2012a). During periods of unprecedentedly intense trading environment, HFT initiated order submission in the direction of extreme movement may cause order flow toxicity (Kang et al., 2020). Therefore, detecting such toxicity prior to flash events should be highly appreciated in order

[1]Ersan et al. (2021) document the level of HFT participation in financial markets. Accordingly, Brogaard (2010) shows as high as 74% of HFT firms' share in NASDAQ between 2008-2009 sample period. Popper (2012) observes 51% of HFT share in the U.S. in 2012. Similarly, Boehmer et al. (2018) document 46% of HFT firms' trading portion in the Canadian stock market in 2010-2011. Whereas, in Europe, similar HFT participation has also been estimated by several studies focusing on European markets (Grant, 2010; Haldane, 2010; Hagströmer and Norden, 2013).

[2]Ersan et al. (2021) review the literature with concerning HFT impact on market quality proxies such as liquidity (Hasbrouck and Saar, 2013; Brogaard et al., 2014; Brogaard et al., 2017), volatility (Brogaard, 2010; Zhang, 2010; Chaboud et al., 2014), price discovery (Carrion, 2013; Menkveld, 2013; Conrad et al. 2015), market downturn (Kirilenko et al., 2017; Brogaard et al., 2017; Madhavan, 2012; McInish et al., 2014).

to predict following extreme price volatility. Even though algorithmic trading (AT) participation contributes to efficient price discovery better than human trading, its various effects[3] on other participants are also valid, especially in the course of intraday extreme price movements. Investigation of the cause and effect relationship between order flow toxicity and flash events directly leads to the concept of informed trading. Even if not all rapid speed crashes and jumps are originated from exogenous news arrival events (Johnson et al., 2013), still an information asymmetry may cause intraday sudden price variations. Thus, informed trading activities may be based on an adverse selection environment due to asymmetry in information accessing and/or dissemination.

Considering the late-2015 Borsa Istanbul BISTECH[4] first phase transformation on the equity market and following that second phase of transformation in 2017 on the derivatives market, intraday flash trading swings have started to be observed. Thus, paying regard to the post-BISTECH transformation era in Borsa Istanbul (BIST), consecutive intraday flash crash and jump events have become more interesting and valuable to be investigated in order to uncover their fundamental rationales. As a result, in this study, two specific intraday flash equity market events are investigated which occurred on February 22, 2017, on Koç Holding and Turkcell equity stocks, KCHOL and TCELL, respectively. These respective events present sudden price drops in the course of a 1-second time interval between 17:45:00 – 17:45:01 during the continuous auction in BIST Equity Market right before the market close time at 18:00.00. Both Turkish blue-chip equity stocks realized an almost 10% sudden drop in the course of the 1-second time period, simultaneously. However, KCHOL concluded the day with a flattish change against the previous day's close by c.-0.78%

[3]Impact on non-HFT participants as crowding out effect (Malinova et al., 2013; Jones, 2013; Hoffmann, 2014), adverse selection effect (Cartea and Penalva, 2012; Biais et al., 2015; Egginton et al., 2016), welfare effect (Boehmer et al., 2015; Stiglitz, 2014; Budish et al., 2015), HFT competition (Baron et al., 2019; Brogaard and Garriott, 2019).

[4]BISTECH is the technological transformation protocol implemented by Borsa Istanbul and cooperation with NASDAQ, to build a common trading platform for trading and post-trade operations on Borsa Istanbul markets with the aim of higher order processing capacity and lower latency. More detailed information is available at https://www.borsaistanbul.com/en/sayfa/2521/bistech-technology.

at 15.30 TRY/share thanks to the following recovery from 13.85 TRY/share during the crash. In parallel, TCELL concluded the day at 11.84 TRY/share with c.0.34% due to post-event recovery up from 10.66 TRY/share. The then BIST Chairman Himmet Karadağ stated[5] that the series of crash events on the respective shares have taken place due to sequential algorithmic order submission and also he mentioned the fact that the responsible parties' total loss over TRY 100K. Considering the respective simultaneous flash crashes without no relevant firm-specific announcement raises many questions on the main rationale behind such events. After two days from the event, on February 24, 2017, CMB[6] imposed a series of penal sanctions against three individuals due to their fictitious pricing activities on the event. Events' simultaneous occurrence deserves attention to be investigated. The uniqueness of the synchronized flash crash events on two blue-chip stocks, KCHOL and TCELL, comes from the questioning of the presence of informational asymmetry or algorithmic trading error on both stocks. Saying that simultaneous order submission at the same time on two stocks indicates an algorithmic trading error rather than information asymmetry caused by informed trading activity. Since two firms, Koç Holding and Turkcell, pursue their operations in completely irrelevant businesses with neither partnership nor shareholding relationship, therefore concurrent abnormal trading events contradict the possibility of pair informed trading activity (Koç Holding and Turkcell equity stock return correlation, $\rho_{kchol,tcell}$ is approximately c.0.50 in the course of 16.02.2016 – 16.02.2017 period which is the 1-year pre-event period with neglecting 1-week pre-event period due to preventing overlapping issue). Alternatively, considering their relatively high beta characteristics, no macro announcement related to market downturn had even been released either. However, even in this environment, intraday data for both stocks demonstrate sudden, abnormal, and biased (buy vs sell) order flow activity during the respective time. Thus, paying attention to the pre-and post-event periods raises the possibility of capturing

[5] https://www.bloomberght.com/haberler/haber/1988313-bistkaradag-turkcell-ve-koc-hisselerindeki-hareket-algoritma-kaynakli.

[6] Capital Markets Board of Turkey is the regulatory and supervisory state agency of Turkish financial markets. The CMB is mainly responsible for maintaining the stability of the financial markets. CMB has been making regulations for organizing the Turkish capital markets and developing capital market instruments.

an order imbalance on order flow structure to detect possible toxicity on order flow in the course of the pre-event window. All in all, measuring the level of order flow toxicity prior to the event regarding information asymmetry should unfold the fact that whether the presence and absence of informed trading according to the event.

This thesis is motivated by investigation of two simultaneous but distinct flash crash events regarding to order flow toxicity as a proxy for information asymmetry to present a valid observation related to pre-event market deterioration. To measure the level of order flow toxicity on both stocks, volume focused informed trading metric, VPIN - Volume-Synchronized Probability of Informed Trading, developed by Easley et al. (2012a) is employed. By considering the order imbalances biased to either buy or sell, VPIN is able to identify the probability of informed trading due to the presence of significant order flow imbalance portion by comparing bulk volumes for buy and sell order submissions. By contrast with the original informed trading metric PIN - Probability of Informed Trading developed by Easley et al. (1996), VPIN is the suitable measure for order flow toxicity in a high-frequency environment (Easley et al., 2012a). Thus, in this study, trade-by-trade data respectively for flash events on KCHOL and TCELL stocks is much appropriate for VPIN measurement. Additionally, by measuring the VPIN metric, the predictive power of such metric on following volatility is also investigated during the study. Therefore, apart from investigating the presence of possible order flow toxicity during the pre-event period (under the assumption of VPIN metric's validity and accuracy), the study also aims to examine the usefulness of VPIN on signaling the post-event extreme event volatility as in the original study by Easley et al. (2012a). In order to observe post-event market condition as well, regarding quality, several market quality proxies are also computed in addition to volatility. Easley et al. (2011) observe a high level of toxic order flow prior to May 6, 2010, flash crash event. As a result, the presence of possible early signals on market quality deterioration may be detected prior to the event, which would be a valuable success considering the early warning mechanism modelling chance. Hence, an alternative tool for circuit breaker market precaution may be developed to ignore and/or minimize possible extreme market movements. As a result, this thesis mainly aims to detect twofold: i) possible order flow toxicity detection in the course of pre-event period via VPIN and observing its pre-event level and, ii) testing the predictive power of VPIN on impending price level via considering subsequently deteriorated market quality proxies for the pre- and post-event periods. By this means, this study eventually targets to investigate the order flow toxicity condition in advance by VPIN measure together with conducting a model to test its forecasting ability of subsequent flash event.

Apart from suitability with intraday market data, the main rationale why employing the bulk volume approach informed trading measurement, VPIN, stemmed from several reasons. Easley et al. (2012a) developed the VPIN metric as a measure of toxic order flow regarding adverse selection between informed and uninformed traders. High liquidity supply by uninformed traders to informed ones due to private information causes market disturbance (Andersen and Bondarenko, 2013) and therefore, high VPIN refers to high toxic order flow structure. Bethel et al. (2012) show advance signal prior to the flash crash event with a high VPIN level. In addition, Easley et al. (2012a) indicate the liquidity provision by uninformed traders to informed ones at a loss due to adverse selection issues. In parallel with this study's motivation of investigating the pre-event market condition, Easley et al. (2011) refer to toxic order flow structure before May 6, 2010, flash crash event on E-mini S&P 500 futures. Unbalanced and abnormal volume observations during the pre-event period lead to further liquidity problems due to market makers' subsequent absence in the course of the order flow toxicity stage. Since market makers (non-HFTs) tend to generate profit under balanced order flow dynamics, when order flows become unbalanced due to adverse selection, then market makers tend to liquidate their positions and leave the market in order to ignore severe losses against informationally advantageous informed traders. As a result, illiquidity condition occurs as Kirilenko et al. (2017) also argue. In parallel, Wu et al. (2013) demonstrate the strong predictive power of VPIN for lack of liquidity caused volatility. In addition, Wei et al. (2013) observe a relationship between VPIN as a toxicity and intraday volatility as Wu et al. (2013). Contrary, Andersen and Bondarenko (2014b) show no predictive power of bulk volume classification methodology of VPIN on future volatility.

The main findings of this thesis suggest indefinite results for VPIN - Volume-Synchronized Probability of Informed Trading measure as a proxy for order flow toxicity presence, on impending VWAP - Volume Weighted Average Price in the course of the flash event on February 22, 2017, for KCHOL and TCELL blue-chip Turkish stocks. On the one hand, VPIN estimation shows no statistically significant implications for predictive power on VWAP, after controlling various market quality proxy variables for KCHOL in general. At time t-4 and t-5 with respect to the event, only VPIN 1-75-50 (and partially VPIN 1-25-50) estimate shows significant predictive power on impending VWAP as a proxy price level during the event. On the other hand, for TCELL, empirical findings explicitly suggest vice versa. For both VPIN measures and lagged time periods, VPIN as a proxy for toxic order flow ables to explain the impending weighted average price behavior with a significance at 99% confidence level. Even stressing the validity of VPIN via considering additional control variables as a robustness test, VPIN measures keep result quite robust outputs. Under the acceptance of VPIN's accuracy as a proxy for order flow toxicity with respect to informed trading, such discrepancies for KCHOL and TCELL naively suggest different trading strategy activities on the respective event period. Since the main motivation of this study is based on pre-event toxic order flow presence, empirical results demonstrate that prior to flash crashes, order flow balance seems to be unbalanced towards sell-initiated orders. Additionally, results indicate that the lower the sample length, the higher the sensitivity of VPIN measures to price changes. Therefore, 1-25-50 parameter-based VPIN estimation shows a more rapid and consequent response to sudden price changes in the course of flash events.

Contribution of this study to the literature along with the following respects. Informed trading literature contains an extensive number of varying studies on the context of different markets. However, emerging markets related literature becomes relatively shallow due to a lack of data, trading volume, and attention from academia and industry. Therefore, lower HFT participation in emerging markets (Haldane, 2012; Ekinci and Ersan, 2018; Ersan and Ekinci, 2016) rises the deficiency in literature. Also, to the best of one's knowledge, this study is the first one on a specific intraday flash event on an emerging market, Borsa Istanbul, within the scope of informed trading. The post-BISTECH period gives an extensive research environment to research scholars in order to raise appropriate and well-pointed research questions, together with market microstructure data. For that matter, this study should contribute to the existing literature with specific event study empirical findings in an emerging market, Turkey. Additionally, constructing such a lead-lag relationship between VPIN and WVAP enables us to observe and accordingly display the effectiveness of VPIN's predictive power on impending price change in an environment of flash events. Thus, the empirical findings and concluding remarks of this thesis should add additional empirical observation proof to the ongoing debate on VPIN's effectiveness in literature.

The rest of the study is organized as follows: In section 2, the literature is reviewed in a broader sense. Then in Section 3, the sample data and methodology of the study is presented. Thereafter, the remaining part of Section 3, empirical findings are introduced. Lastly, in Section 4, concluding remarks of the study and further extension for possible future studies are discussed.

2. LITERATURE REVIEW

Investigating informed trading requires approaching financial markets' efficiency and accordingly, the condition of adverse selection due to information asymmetry between different traders. Due to growing systematic trading implementation in financial markets, conditions related to market efficiency may be measured by market reaction and price incorporation of additional informational arrival (Ersan et al., 2021). By this means, intraday extreme market movements are analyzed through a direct relationship with market efficiency and information asymmetry. Since order flow balance is motivated by efficient market conditions without adverse selection due to information asymmetry, market efficiency and its assumptions are also needed to be remembered.

2.1 Efficient Market Hypothesis

In 1970, the definitive study and discussion on the Efficient Market Hypothesis (EMH) were argued by Fama (1970) with the definition of "a market in which prices always fully reflect available information is called efficient". The theory suggests that when information arises, the information itself spreads almost simultaneously and it becomes already incorporated into the market prices of the related asset without significant latency. Therefore, neither chart analysis nor fundamental security valuation could predict and yield an abnormal return premium over the market. As a result, Fama (1970) supports no undervalued nor overvalued asset pricing in the market and no chance to build a profitable trading strategy to beat the market by mimicking past price patterns. In parallel, the EMH is highly associated with the concept of random walk[7] and therefore, it should be investigated accordingly.

[7]The Random Walk Theory explicitly characterizes price time series without any pattern but behaving stochastic swings around the historical mean developed and acknowledged by Bachelier (1900). In his Ph.D. thesis, Bachelier (1900) develops the Brownian

Regnault (1863) argues the observation that the security price variance is directly related and proportional to the square root of time. The efficient markets statement is clearly mentioned by Gibson (1889) and discusses the value of public stocks regarding the best knowledge on them, i.e. the information flow itself. Early stages of the 20th century, random walk theory becomes the motivation on market price behaviors and dynamics with the sole purpose of understanding the enigma behind the asset pricing rationale. After discussing random walk and Brownian motion by Bachelier (1900), another two important studies were conducted in the same year by Pearson (1905) and Einstein (1905) where Pearson (1905) introduced the random walk term and Einstein (1905) developed the Brownian motion-based equations without knowledge of Bachelier (1900). In terms of risk-reward and trading performance, Keynes (1923) states that investors do not get a reward because of absolute competitive advantage over the market in terms of knowing the future, but instead, investors are rewarded due to their risk baring appetite as a consequence of the EMH. Regarding random walk, MacCauley (1925) points out the similarities between stock market fluctuations and throwing dice. Similary, Working (1934) argues stock returns and their lottery-like behaviors. In terms of beating the market, studies show no significant overperformance by professionals (Cowles, 1944; Working, 1949). While after, supporting the random walk theorem, Harry (1959) observes random walk like stock price series. In parallel, Osborne (1959) references the Bachelier (1900) and Regnault (1863) implications by showing that the logarithm of stock prices follows Brownian motion and the existence of the square root of time argument, indeed.

---

motion for security prices and their consecutive price drift. Thus, again, random walk theory backs the EMH with today's newly disseminated information that is significantly irrelevant from tomorrow's pricing pattern. Tomorrow's pricing behavior is determined by only tomorrow's information flow. Before Fama (1970), the random walk and market efficiency relationship is conducted by Samuelson (1965) and differently from Bachelier (1900) in his study Samuelson (1965) focuses on martingale rather than random walk in order to provide the EMH. Fama (1970) and Samuelson (1965) arguments on EMH are accepted as contradictory by their implications where Fama (1970) develops the practical implications whereas Samuelson (1965) develops the implications for the purpose of policy-making matters.

Contrary to the above arguments and the EMH condition, Cootner (1962) concludes no random walk situation in the stock market. Partly similar to Cootner (1962) argument, Granger and Morgenstern (1963) separate market prices into two: short-run and long-run. By applying spectral analysis, findings demonstrate that short-run price movements show random walk behavior whereas long-run movements do not. Alexander (1964) and Steiger (1964) both conclude that stock prices do not follow a random walk-like price pattern. Then, Fama (1965) states the efficient market term for the first time and Samuelson (1965) provides the first economic argument for efficient markets with the concept of martingale different from random walk theory. Roberts (1967) raises the efficient market hypothesis term and distinguishes the difference between the weak and strong forms of the EMH. Accordingly, Fama (1970) coins the available information reflection into the prices and efficiency hypothesis. Related to not fully efficient market camp, Scholes (1972) argues partially efficient market condition due to minor post-event price drift observation from some indicators. Similar to the post-event drift argument raised by Scholes (1972), in parallel, Ball (1978) demonstrates sticky excess return premiums after public announcements of quarterly earnings. Regarding information gathering and reflecting into prices, Grossman and Stiglitz (1980) show the imperfection of an informationally efficient market. Due to the cost of information access, no compensation would be earned in exchange for obtaining the information itself. Therefore, Grossman and Stiglitz (1980) argue an incentive and its existence in the market equilibrium which results informationally inefficient market. By discovering the overreaction in the stock market De Bondt and Thaler (1985) prove the weak-form market inefficiency and accordingly, raises the behavioral finance concept. Also, there are other studies rejecting the EMH due to weekly stock returns' behavior (Lo and MacKinlay, 1988; Lehmann, 1990). Further studies show also supportive conclusions with prior findings, such as Jegadeesh (1990) documents predictable behavior of asset returns and therefore, rejects the random walk.

On the other hand, Chopra et al. (1992) observe stock market overreaction similar to De Bondt and Thaler (1985). Besides informationally efficient markets, studies show market anomalies concerning factor analysis and find abnormal return results trading strategies (Jegadeesh and Titman, 1993; Lakonishok et al., 1994; Chan et al., 1996). In contrast, Shiller (2000) shows no possibility of market explanation through historical movements of earnings and dividends. Regarding both arguments on market anomalies on yielding abnormal returns, Schwert (2003) implies the cause-effect relationship between market anomalies and efficiency with the argument of market anomaly related published studies' weakening the profitability of those trading strategies due to causing a crowding-out effect. By absolute contradictory findings compared to Granger and Morgenstern (1963), Wilson and Marashdeh (2007) show no EMH related stock price behaviors in the short run but consistent EMH like movements in the long run. Malkiel (2003) shows more efficiency and fewer predictability conclusions contrary to the academic papers which suggest partial predictive power of market anomalies. Through investigating the stationary for developed and emerging markets together, Lee et al. (2010) find no conclusive efficient stock market.

The EMH hypothesis simply rejects the sustainable consecutive excess return performances in the long run. Contrary, there are significant market anomalies[8] which yield a sustainable long-term abnormal return over the market and/or benchmark index itself. In return, both the academic literature and asset management industry cover multiple market anomaly factors which violate the EMH and its fundamental assumptions. For that reason, in order to segregate different market forms related to efficiency levels, Weak, Semi-Strong, and Strong From of the EMH are necessary definitions for the information incorporation into market prices.

[8]The most well-known anomalies are Value and Size factors are proposed by Fama and French (1993) and denoted by HML and SMB, respectively. Value factor implies that undervalued (overvalued) stocks tend to outperform (underperform) the market. Whereas the Size factor shows that small (big) market capitalization stocks tend to outperform (underperform) the market. The Beta factor is developed by Fama and MacBeth (1973) regarding the relationship between single stock and the overall market. On the other hand, the Momentum factor was raised by Carhart (1997) and denoted by UMD. The factor implies that past winner (loser) stocks tend to keep yielding positive (negative) return performances. The Quality factor is developed by Asness et al. (2014) and denoted by QMJ. The factor suggests that high (low) balance sheet quality stocks tend to outperform (underperform) the market.

2.1.1 Weak form

The premature form of the EMH assumes (supports) the reflection of all publicly available information but not yet publicly available information. This means, already disseminated public information announcements have been incorporated into prices, and the current market prices (should) reflect all publicly announced and available information. However, in the Weak Form of the EMH, not released information and its price impact are not acknowledged yet, and therefore, fair market prices do not represent information that is not publicly available. As a result, focusing on historical prices does (should) not represent and/or reflect future possible price patterns in order to follow similar return performances.

2.1.2 Semi-strong form

The second form of the EMH, the Semi-Strong Form rejects the possibility of excess return generation by fundamental analysis, in addition to the Weak Form. Thus, the Semi-Strong Form carries assumptions from the Weak Form where no such an effect of historical (technical) price trends on the market prices, together with fundamental data predictive power. In the equity stock market, public companies regularly release their quarterly financial performance by disclosing quarterly financial statements. In the foreign exchange currency market, country-specific macroeconomic parameters and their scheduled data announcements are accepted and used as a proxy for fundamental information.

2.1.3 Strong form

Lastly, the most comprehensive form of the EMH is the Strong Form which implies no outperformance opportunity over the market regarding historical prices, publicly available fundamental information, and also, not publicly disclosed private information. As saying private information, relevant information, and/or dataset which matters for the market price but only acknowledged by insiders and/or superior information accessing advantage holders. The Strong Form argues the strictest form of the EMH with no violation tolerance and therefore, it assumes that even privately owned information is already incorporated into market prices.

The EMH assumptions and forms are rest against the information availability and its latency in market price incorporation. As discussed in the Weak Form, private information – without publicly disclosed and/or acknowledged information – is not adjusted into fair market prices since no dissemination action is taken by the main source. Examination of different informationally valuable news dissemination plays a unique role in the quest for market efficiency. Therefore, besides information itself, the way of dissemination and time-varying trading effects are also equally important. As a result, no pre-scheduled information flows cause violations of the EMH assumptions in terms of the rapid information incorporation perspective.

2.2 Information Asymmetry

In financial markets, participants whether short- or long-term focused mindset, do follow information flow in order to catch the respective market reaction to such flows. Intermediaries who disseminate the recent informationally important news to relevant data providers and/or terminals play a vital role in the information flow framework. Bloomberg LP, Thomson Reuters, Morningstar, FactSet Research Systems Inc., etc. are such data providers for global financial markets. Accordingly, firm-specific and/or macroeconomic datasets are provided to market participants by these data providers. Therefore, the unique data disseminated from these intermediaries is monitored, accessed, and processed by (mostly) institutional investors.

Schaub (2018) examines the role of Thomson Reuters' First Call quarterly earnings announcements with respect to investigating the market reaction to those disclosures. Observations show that the instant price and volume reactions are not strong. However, the significance of the observations is comprised of the post-earnings announcement drift. The market anomaly of such a drift occurs right after an announcement and shows medium- to long-term persistence. Schaub (2018) shows the market response to earnings announcements with the delayed dissemination timeliness. Results show that a significant part of the drift aggregated around the announcement day. Thus, specific information formation results in an abnormal drift and causes market anomaly. Similar to earnings announcements, unscheduled informational datasets, disseminated by data providers yield abnormal return performances. Womack (1996) analyzes stock recommendations issued by sell-side analysts. For buy (sell) analyst recommendations' average post-event drift is 2.4% (-9.1%) and considerably short-lived (long-lived). According to the EMH, sell-recommended analyst issuance should not have caused an anomaly where an excess return drift happens. The literature covers and gives quite an attention to analyst recommendations and their implied market reactions. For an emerging market, Turkey, Tinic et al. (2021) study the effect of Turkish sell-side analyst recommendations, disseminated by Matriks Data Terminal[9] platform. Results demonstrate 35 (-45) basis points of abnormal returns around the announcement day, on average for upgraded (downgraded) recommendations, respectively. Tinic et al. (2021) also examine the comparison between local and foreign brokerage firms' informativeness from the perspective of Turkish investors. Observations show larger attention to foreign brokerage firms' analyst recommendations than local ones'. Similarly, Barber et al. (2001) document the most (least) favorable stock recommendations by consensus and their consequent daily return performances. Without considering the trading transaction costs in the study, observations show more than 400 basis points of annualized gross return performance by simply following that recommendation consensus. In parallel, Easley et al. (1998) investigate the informational role of sell-side analyst coverage on NYSE. Contrary, results are not supportive to accept sell-side analysts' coverage as an information flow proxy and to build an information-based trading strategy. Differently, Aktas et al. (2007) yield the information-based trading around M&A announcements on Euronext Paris stock market for 1995-2000 period. Results show some contradictory observations against the informed trading literature with

[9]Matriks Data Terminal is a subscription-based trading and data platform for specifically Turkish capital markets. Detailed information on the platform can be reached at https://www.matriksdata.com/website/.

evidence of information leakages around sample M&A announcements. Similar to Aktas et al. (2007), an informed trading investigation is conducted by Christophe et al. (2010) by focusing on information-based trading appearance around analyst downgrade recommendations on NASDAQ for the 2000-2001 period. Observations on short-selling activity show significant abnormal short-sale trading activity even in the three days before downgrade recommendations become publicly available, by controlling scheduled earnings announcements.

Other than sell-side analyst informativeness on stock prices and their post-event behaviors, trade data itself may also be used as an information proxy. Hasbrouck (1991a) estimates the impact and information asymmetry through trade and quote revisions. Results show that large size trades cause spreads to widen and accordingly, trade orders during larger spread conditions have greater impacts. As a result, information asymmetry becomes more apparent and significant for relatively small market capitalization firms. As a continuation, Hasbrouck (1991b) studies on NYSE stocks and asymmetrically informed components of trade data. By investigating intraday price patterns, trade orders are apparently more informative for relatively smaller firms. DellaVigna and Pollet (2009) focus on earnings announcements that take place on Friday, before the weekend. Therefore, through investigating the day of the week anomaly, observations show approximately 15% (%70) lower (higher) immediate (delayed) market response for Friday announcements. Garfinkel (1994) documents the post-event insider trading activity around quarterly earnings announcements. Observations show more prudent insider trading activity during the post-event period than pre-event. By investigating cross-country sample data, Griffin et al. (2008) show the market reaction to earnings and takeover announcements. Results indicate that the difference between countries exists because of insider trading and press freedom factors. Whereas, Bhattacharya et al. (2000) focus on an emerging market, Mexico, and document no abnormal trading parameters in the event window for corporate news announcements for July 1994 to June 1997 period. By working on earnings announcements, Ke and Zhang (2019) result in lower post-earnings announcement drift with higher HFT activity, due to the HFT role of liquidity supplying.

Even though the information itself is disseminated and becomes publicly available for the overall market, there is continuous unbalance between different market participants with respect to information access. By considering annual reports, Brown and Hillegeist (2007) show a negatively associated correlation between the quality of annual reports and information asymmetry. Also, results demonstrate that besides the level of quality, investor relations activities are also negatively correlated with information asymmetry. Thus, empirical findings prove the transparency versus opaqueness of firms and their corresponding information asymmetry on stock trading activities. Similarly, Healy and Palepu (2001) document the effect of financial reporting and disclosures on information asymmetry in relation to stock performance. As discussed previously, the informativeness of scheduled and unscheduled announcements disclosed by corporates may show a varying effect on trading activity. Chae (2005) results in decreasing cumulative trading volume prior to scheduled announcements. However, for unscheduled announcements, trading volume increases dramatically and demonstrates the fact that the presence of information asymmetry. Brown et al. (2004) study on conference calls and their adding value on the market quality. Empirical findings show that voluntary conference calls reduce the level of information asymmetry among investors. Results also demonstrate the difference between information asymmetry for firms with regularly holding conference calls and for one-time callers.

Due to market fairness and transparency, emerging markets are accepted more volatile and ambiguous markets in the perspective of market participants. Accordingly, information asymmetry may be more prudent in those regional markets. Chan et al. (2008) investigate the effect of information asymmetry on the Chinese stock market focusing on local A- and foreign B-share markets. Results support the presence of adverse selection situations through the bid-ask spread component. Similarly, by analyzing foreign trading activity on BIST, Cinier and Karagozoglu (2008) show significant evidence on the explanatory power of trading activities on informed trading. Results show that foreign trading activities, in BIST, are indeed associated with informed trading by implementing the predictive power of Tobin's Q and raising the information asymmetry imbalance. Kyle (1985) investigates the relationship between the informativeness of insider trading and its incorporation into prices.

Besides response to newly disseminated publicly available information and investigating abnormal trading activity around such an announcement, insider trading outputs measure considerably significant anomalies on information symmetry. Cohen et al. (2012) simply decompose insider trading activities in order to identify opportunistic insider trading actions. Results show that a trading strategy that follows those opportunistic insider trading activities, indeed yields an abnormal return of 82 (180) basis points, per month for value-weighted (equal-weighted) portfolio balancing. Focusing on 21 countries with 2,189 firms, Durnev and Nain (2007) show the success of strict regulations on insider trading activities regarding reduced private information-based trading. For the U.S. stock markets NYSE, AMEX, and NASDAQ, Lakonishok and Lee (2001) examine insider traders' trading strategies. Results show that insider traders are more like contrarian investors and the informativeness of that trader profile comes from their purchases, not selling activities. Accordingly, by differentiating informed and uninformed traders, Wang (1993) shows higher price volatility due to information asymmetry between informed and uninformed traders. Hence, uninformed or less-informed traders accordingly show price chasers with positive feedback (momentum) behavior. Whereas, Hughes et al. (2007) study the relationship between private information as a proxy for information asymmetry and risk premiums. Results document that higher information asymmetry yields higher risk premiums, and accordingly higher cost of capital.

2.3 Extreme Market Movements

Unexpected pricing behaviors cause sudden abnormal market dynamics due to severe price swings. Such behaviors are cardinal market events for toxic order flow characteristics and therefore, information asymmetry suspicion. Because of the higher implementation of algorithmic and/or systematic trading applications via electronic trading platforms, subsequent extreme market movements have occurred at a considerably rapid speed. Relationship with the concept of order flow toxicity and informed trading comes from order imbalances prior to the event itself by observing unbalanced order tendency toward either buy or sell direction. As a result, examining such abnormal market conditions around unexpected events should be seen as an explanatory investigation in order to present pre-and post-event market behaviors. Contradictory to the EMH, literature on extreme market movements for both U.S. and European stock markets actually show statistically significant abnormal return behavior following an extreme market movement. In U.S. stock markets, Brown et al. (1988) observe positive abnormal return in the course of a 60-day post-event period following a greater than 2.5% price change in equity stock, in both positive and negative change. Similarly, Corrado and Jordan (1997) work with a 10% price change threshold rather than 2.5%, and observe positive (negative) abnormal returns following the positive (negative) events. Bremer and Sweeney (1991) investigate statistically significant price reversals following more than 10% price declines. Alternatively, in European market stock markets, similar observations are also valid as in the U.S. markets. Mehdian et al. (2008) report positive abnormal return drift in the Turkish stock market following positive and negative events.

Since the study is very much related to the high-frequency trading (HFT) concept with instantaneously order submission capability, its effect on market quality and relation with information asymmetry deserves to be mentioned. Ersan et al. (2021) study different cause and effect relationships of HFT activities and market quality with an extensive literature review. As a result, the literature demonstrates various and inconclusive empirical results on HFT and market quality proxies such as liquidity, depth, bid-ask spread, volatility, so on and so forth. For the Turkish stock market, similar studies (Akyildirim et al., 2015; Ekinci et al., 2019) also conduct parallel research and observe inline results regarding investigating the relationship between algorithmic-based order flow and trading activity. Differently, Zhang (2010) finds that high-frequency actually could result in hazardous effects on the market, specifically on volatility. Empirical findings point out a positive correlation between HFT activity and stock price volatility even after controlling for firm-specific fundamental factors. Consequently, relatively advantageous market participants benefit through positioning on an asset regarding the information they accessed priorly. On the other hand, relatively slower market participants in terms of accessing the specific publicly available information and/or news suffer from tighter profit margins relative to ones with superior information access. However, private information or "not yet" publicly available information-based trading activities cause a problem on the market efficiency. A market participant who yields a financial benefit via not publicly available information is classified as an informed trader. However, it certainly deteriorates the competition to some degree and causes adverse selection situations.

2.3.1 Algorithmic events

Easley et al. (2011) study the May 6, 2010 flash crash on U.S. markets. Empirical findings show increasing toxicity before the event by applying the VPIN measure as a proxy for informed trading activity. Easley et al. (2011) suggest forming a VPIN contract in order to monitor such an event and manage possible risks. Huang and Wang (2009) show a positive correlation between illiquidity and asymmetric return. On the other hand, by examining emerging markets, Morck et al. (2000) propose more co-movements in relatively poor countries with lower economies than rich economies due to lower market size and noise trading effects. In terms of price volatility, Lee and Liu (2011) use price informativeness and document a U-shaped relationship between price-based informativeness and idiosyncratic return volatility. Kang et al. (2020) investigate the KOSPI 200 futures market during normal but distressed market periods, and document that VPIN is indeed helping to detect order flow toxicity regarding predicting short-term volatility.

By focusing on flash jumps, Christensen (2014) argues that the accurate price variation due to flash jumps is quite overstated and indeed, is lower than what academic literature demonstrates. Around extreme price movement periods, Brogaard et al. (2016) show that HFT profile traders actually provide liquidity to the market during those distressed market movements. However, the dilemma occurs when more than one extreme events' presence where high-frequency traders' liquidity demand overtakes their supply. Similarly, Golub et al. (2012) analyze mini flash crashes on U.S. equity stock markets and result the fact that mini flash crashes are the effect, not the cause. Whereas, Kirilenko et al. (2017) study the May 6, 2010 flash crash and the behavior of HFT investors. Results show no behavioral change when prices experience a flash crash during the event. Additionally, Braun et al. (2018) examine the characteristics of intraday ultra-extreme flash events in U.S. equities between the 2007-2008 period and demonstrate that all investor types regardless of algorithmic computation power may cause extreme price movements. Braun et al. (2018) show that at least 60% of mini flash crashes are due to a single market order.

2.3.2 Fat finger events

Apart from systematic trading errors, differently from algorithmic trading and its direct and/or indirect effect on extreme market movements context, unexpected flash crashes/jumps may be occurred due to trader error which is called fat-finger. By simply ordering inaccurate order parameters when submitting the real-time order, that error may cause a sudden price crash/jump due to submitting more than what it should be. Jin et al. (2019) document that traders who traded during the fat-finger event show risk-taking reduction behavior even though the event happens with no fundamental information. Contrary, traders who did not trade during the event show no significant behavioral change. In 2014, on Tokyo Stock Exchange, approximately USD 711 billion worth of accidental stock orders were voided before the execution. Again, in Japan, the year 2009, Swiss bank UBS mistakenly send Yen 30 trillion worth of fixed income order instead of 30 million. Later, the order was placed out and avoided such a catastrophe. In 2001, a Lehman Brothers trader mistakenly executed an order 100 times larger than the actual intention. LSE – London Stock Exchange fined Lehman Brothers GBP 20 thousand. In 2012 and

2013, two fat-finger events happened in Melbourne and Sydney, respectively.

2.3.3 Extreme events in Borsa Istanbul

BIST has been experiencing rare but material intraday flash events as well. On May 22, 2014, İş Bankası stock experienced a sudden jump around 15:50:38 local time. The price of one share of İş Bankası goes from TRY 5.53 to 6.01 suddenly, and then it comes back to the prior level with sudden price recovery. A total of 2,945,023 shares were purchased in a single minute. On April 18, 2017, Akbank shares experienced a similar flash jump at 16:16:26 sharp, in local time. One share of Akbank goes from TRY 9.24 to 10 with a rapid speed. Borsa Istanbul Chairman at that time disclosed about the event and said “no fat-finger but the algorithmic trading error”. A total of 2,096,824 shares were purchased by an error. On the other hand, on May 26, 2017, Ulker shares experienced 15,754 number of short-selling order pressure which resulted in a freefall on price, from TRY 20.74 to 18.75 around 10:46:53 local time. Borsa Istanbul Chairman at that time said “no algorithmic trading, just a normal order submission”. Therefore, it can be accepted as a fat finger since no rational investor would execute such a block order on continuous auction. On September 9, 2019, Pegasus shares experienced a flash crash with a price fall from TRY 70 to 60.45 at 12:58:58 by total of 133,453 shares of selling orders.

Figure 2.1 displays the four respective flash events that occurred in BIST in different time periods. The respective charts show intraday pricing in the course of each flash event during the continues auction for ISCTR – İş Bankası between 15:30 – 16:30, AKBNK - Akbank between 16:00 – 16:28, ULKER – Ulker between 10:00 – 11:00, and PGSUS – Pegasus between 12:45 – 14:05, from the upper left corner to lower right bottom, respectively. As it can be seen, consecutive rapid recovery in pricing following all the aforementioned flash events in Borsa Istanbul. Even though these flash jumps and crashes did not create a systematic risk to the market back then, they are still deserved to be investigated separately in order to compare

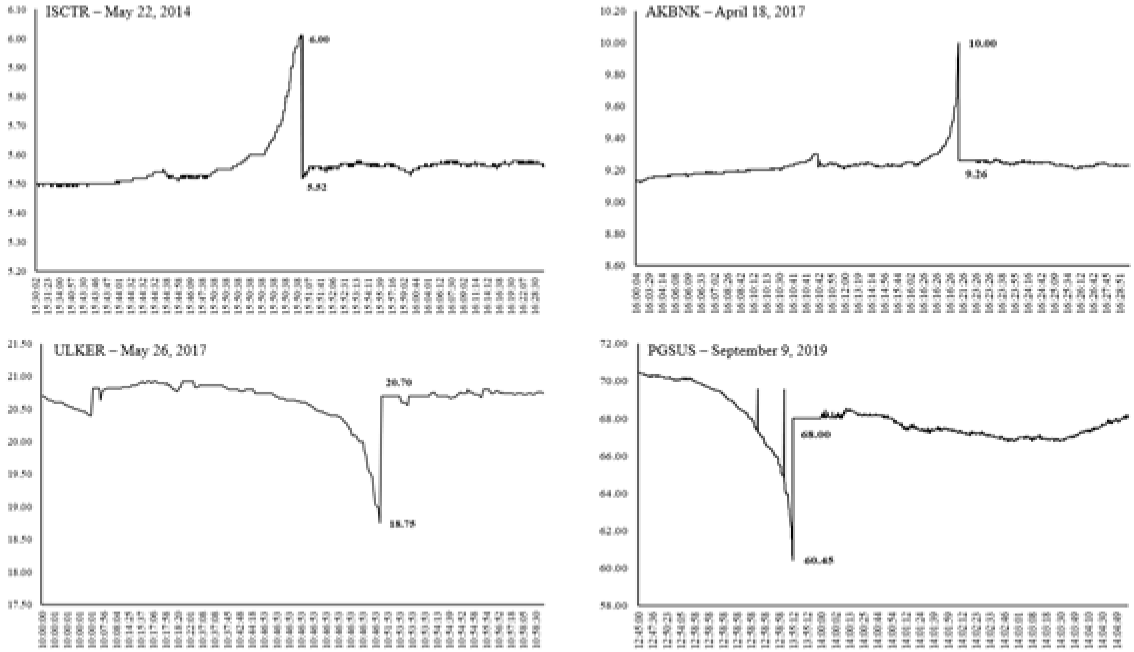

Figure 2.1 Extreme Market Movements in Borsa Istanbul

their unique features in terms of market microstructure perspective. The KCHOL-TCELL synchronized flash event(s) makes itself quite different than above flash events in BIST by its simultaneous nature on two "irrelevant" and considerably low correlated equity stocks. The beauty in this research and its respective event study lies in the questioning of toxic order flow patterns in a broken (as BIST Chairman stated briefly at that time) algorithmic trading strategy. Figure 2.1 shows the most debated and popular extreme events in BIST apart from the rest of them - including mini flash crashes/jumps. Considering the relatively lower market share of high-frequency-oriented trading strategies implemented on BIST equity stocks, there are limited events relevant to the respective examination motivation.

2.4 Order Flow Toxicity

In order to monitor order flow structure and detect possible order imbalances with respect to bias through buy or sell initiated trading activities, order flow toxicity measurement represents an adverse selection environment between informed and uninformed market participants. Such an environment may be realized in the course of liquidity provided by market makers (uninformed traders) to relatively or absolutely more informationally superior counterparties (informed traders) at a loss (Easley et al., 2012a). The importance of order flow toxicity as a proxy for information asymmetry is twofold: i) causing an efficient market by creating asymmetric information-based trading activities, and ii) liquidity provision. As Abad and Yagüe (2012) stated, Easeley et al. (2012a) developed the VPIN measurement in order to capture informed trading by estimating the probability based on order imbalance (via volume).

2.5 Informed Trading

The literature has used several proxies as an indication of informed traders' degree such as insider and institutional ownership level (Stoll, 1978; Brennan and Subrahmanyam, 1995; Grullon and Wang, 2001), market capitalization and trading volume (Chari et al.,1988), abnormal volatility (Krishnaswami and Subramaniam, 1999), etc. Still, identification of an informed trader and consequently related trading activities remains considerably challenging due to untraceable implementations by informed traders, e.g., selling the company's stocks by insiders through the course of multiple quarters prior to bad news (Ke et al., 2003) which would highly affect the stock. Therefore, asymmetric information issues negatively affect the rest of the market participants who are labeled as uninformed or noise traders. Since, informed traders are able to reach, analyze, and interpret the data in advance, the execution price yields considerably effective profit margins. On the other hand, uninformed traders are suffered from adverse selection during extreme events related to information-based unexpected trading activities. As a consequence, a fully efficient market environment gets damaged and liquidity shocks are observed in an informed trading environment. In parallel with informed trading and liquidity shocks, Collin-Dufrense and Fos (2015) suggest that informed trading activities exist in market conditions where liquidity is sufficiently high. In today's financial markets with heavily technology-based trading environments and infrastructures, market participants tend to search the accurate and reliable pieces of information by following unorthodox methodologies. Through implementing complex but efficient algorithms by sophisticated traders/investors, intraday market orders are monitored in order to detect an unusual positioning. Contrary to the perception of local investors' informationally superiority over foreign investors, it is possible that foreign investors could be relatively more informed than local investors (Yang, 2003). Therefore, rather than publicly available and/or soon-to-be-released scheduled data, privately owned information forms a basis for informed trading and adverse selection.

2.5.1 PIN

Easley et al. (1996) suggest the initial informed trading measure known as PIN – Probability of Informed Trading and also show an extended version of the model developed by Easley and O'Hara (1987, 1992a, 1992b). Through this model, the probabilistic measure is computed based on informed and uninformed investors' trading activities on with and without information event periods. As a result, PIN has been the widely accepted estimation for the fraction of informed trading to the overall trading in the market. The main research question of this study is to investigate intraday extreme market movements in BIST connecting with informed trading suspicion. Due to the BISTECH transformation in BIST in order to build more effective technological infrastructure, high-speed trading activities and their implementation on BIST became more appropriate and rational. Therefore, high computing power-based algorithmic trading volume in the BIST equity stock market started to show observable increase even though it is still relatively low when compared with other markets (Ersan and Ekinci, 2016; Ekinci and Ersan, 2018). PIN, widely implemented in the relevant literature, has become prone to essential computation problems especially with dramatic increase in trading activity in last two decades. This may have led to significant biases in many studies using the PIN measure with limited attention paid on these problems. Several studies suggest remedial solutions for the computational problems (Lin and Ke, 2011; Yan and Zhang, 2012; Ersan and Alıcı, 2016). Moreover, given today's financial markets and the corresponding financial data, original PIN model and its assumptions are also being challenged in the literature. Addressing these issues, model extensions and new measures are suggested (e.g., Duarte and Young, 2009; Ersan, 2016). Nevertheless, PIN studies have been numerous and within diverse topics. One essential question inquired in PIN studies has been whether PIN is priced in financial markets or not. There exist inconsistent results and the question remains to be an open question. While Easley et al. (2002) and Hwang et al. (2013) document that PIN is a priced factor; Duarte and Young (2009), Mohanram and Rajgopal (2009), and Lai et al. (2014) disagree. The conflicting results signal for the fact that the role of computational errors might be significant.

2.5.2 VPIN

A new (at least compared to the original PIN) method to accurately estimate the order flow toxicity has been developed by Easley et al. (2012a). It should be fair enough to say that VPIN is an adjusted version of and inspired by the originally informed trading estimation measurement PIN. Unobservable parameters and their MLE – Maximum Likelihood Estimation procedure via fitting on Poisson Distribution yield multiple approximations and therefore, misleadings in the original PIN measurement. However, Easley et al. (2012a) overcome the PIN's difficulties via not requiring non-observable parameter estimation. Different from the original measurement PIN, VPIN is based on volume time in an attempt to catch the arrival of new information via volume order imbalances. Therefore, in a high-frequency market structure, VPIN should become more effective than PIN in terms of the level of success for measuring order flow toxicity. Abad and Yagüe (2012) explicitly demonstrate the innovation of VPIN with comparing to the original informed trading proxy PIN. With focusing on the Spanish stock market, empirical findings display the significant ability of VPIN on detecting adverse selection risk indeed. Also, the respective results imply the fact that determining the number of volume buckets in VPIN yields the main difference. Similarly, Paparizos et al. (2016) employ the VPIN metric on another European market, Athens Stock Market in order to detect its predictive power on impending volatility. Considering the Greek sovereign debt crisis period, empirical findings result in positive correlation between VPIN and following price volatility. Therefore, the study argues the significant role of VPIN on future order imbalances. By covering several international equity markets, Low et al. (2018) implement VPIN measure and observe rising VPIN level which foreshadows impending high-level volatility. As a policy recommendation, the study offers the implication of VPIN as a risk monitoring tool. In another study, probably the most related one with this thesis in terms of investigation motivations, Abad et al. (2018) employs VPIN proxy in order to investigate i) VPIN's reliability on being the order flow toxicity proxy, and ii) VPIN's ability on being an early warning signal. Empirical findings imply no statistically significant role of VPIN as the order flow toxicity proxy since most of the volatility peaks alarmed by VPIN are not toxic at all. Additionally, Abad et al. (2018) suggest that VPIN usage as an alternative for circuit breaker mechanism would yield costly and unnecessary halts. However, the literature also covers additional studies other than equity market and/or price-related volatility. Kitamura (2017) implements VPIN on a foreign exchange market, specifically on the USD/JPY parity. Observations show a reliable measure of VPIN on predicting the flash crash on the parity. Other than firm-specific and/or intraday trading anomalies, various exogenous factors may also affect the trading behavior in financial markets. Linking this fact with the informed trading concept, other studies investigate the relationship between VPIN and geopolitical risk initiated market turmoils (Silva and Volkova, 2018; Tinic, 2019[10]). Silva and Volkova (2018) analyze the invasion of the Crimean region in 2014. By employing VPIN on the Russian RTS equity index, significant increasing behavior is observed prior to the invasion pricing. During the event, a significant and negative relationship between VPIN and price change is also argued. The study implies the additional support of VPIN on undesired events' likelihood monitoring.

[10]Tinic (2019) employs PIN measure, not VPIN, however due to its investigation on political turmoils in Turkey and their effects on Borsa Istanbul volatility, the study uses PIN in order to predict such an event.

2.5.3 PIN vs VPIN

Both VPIN and PIN focus on trading volume in buy and sell directions and they are based on order imbalances. Such an imbalance is accepted as an adverse selection detection. In contrast to PIN, VPIN can be used with high-frequency data. Since volume-buckets are independent of each other in high-frequency data, MLE estimation cannot be employed based on the product of each bulk periods' likelihood values with the assumption of independence. PIN maximizes the likelihood in which buy and sell order flows are assumed to follow Poisson Distribution whereas Normal Distribution is used in VPIN. As a result, as Easley et al. (2012a) mention, the resulting time-series of observations follow Normal Distribution (at least closer to normal) and as a result, lower heteroskedasticity may have occurred than it would be in a clock-time as in PIN. As Abad and Yagüe (2012) discuss, VPIN is able to deal with high-frequency data idiosyncrasy. Again, Abad and Yagüe (2012) show the main differences between PIN and VPIN on a table[11] as an illustration purpose. The table explicitly introduces the main four innovations of VPIN differently than the original PIN measurement. Those innovations can be classified as information, time, classification, and size. VPIN innovate these adjusted methods in order to match with the market microstructure data.

[11]The table compares Easley et al. (1996) and Easley et al. (2012a) studies with respect to four main differences: information, time, classification, and size. Easley et al. (1996) have assumptions of fundamental information, clock-time, itemized classification, and no explicit role for trade intensity for information, time, classification, and size factors, respectively. Whereas, Easley et al. (2012a) argue broader definition of information, volume-time, bulk classification, and trade intensity matter for the corresponding factors as well, respectively.

3. EMPIRICAL STUDY, DATA AND ESTIMATES

This thesis is solely based on trade-by-trade data on KCHOL and TCELL over the February 21, 2017, through February 23, 2017, period. In order to detect order flow toxicity in the course of the flash event on February 22, 2017, adjusted and modified informed trading measure VPIN is employed in the empirical investigation. Due to the high-frequency trading framework following the BISTECH transformation, algorithmic trading focused sophisticated institutional investors' exposure to the Turkish equity market has enabled them to materialize sudden market movements. The respective flash crash duality on KCHOL and TCELL deserves to be investigated in the context of order flow toxicity via VPIN - to examine the presence of informed trading-based order imbalances prior to the event by VPIN as a proxy for toxic order flow sequence. In this section, the empirical investigation will be introduced and findings will be discussed accordingly.

3.1 Sample Data

Sample data regarding the investigation covers the period between February 21, 2017, and February 23, 2017, with [t-1;t+1] event window with respect to February 22, 2017, event day. The event window is selected with a 1-day pre-and post-event period due to the sufficiency of a 3-day event window with market microstructure data. Even though the existence of validity on approximately 1-month (30-day) on observing practicability trade data for the PIN estimation (Easley et al., 1997a), the sample data covers a much narrower period in order to neglect exogenous factors out of the essential toxic order flow interval. Similar to Easley et al. (1997a), Akay et al. (2012) use 20-day in PIN estimation as the ability of a c.1-month estimation horizon for the PIN. Thus, even though 1-month trading data should be well enough to cover observations for informed trading metrics (Abad and Yagüe, 2012), instead of as argued above, this thesis focuses on much more relevant trade data with the purpose of zooming the event relevant trade data. In this study, price data is gathered from Matriks IQ which is the latest version of Matriks Data Platform, introduced earlier. In Matriks IQ, historical trade data (no passive order quotes but only executed orders) is available. The extracted raw data consists of such classifications[12] as Trade, Symbol, Time, Price, Size, Volume, and ID as presented below Table 3.1 for KCHOL and TCELL, respectively, as illustrative purposes. As it can be seen, the ID column assigns binary identification to each trade data, either 0 or 1 where 0 (1) are assigned for buy (sell) executed orders. This classification is in parallel with Van Ness et al. (2017) as mentioned earlier and enables to ignore trade classification algorithm in order to differentiate buy and sell orders. However, as the raw data itself contains such segregation, no further classification via algorithm employment as in the original VPIN estimation by Easley et al. (2012a) is needed for buy and sell initiated order identification. Additionally, Andersen and Bondarenko (2015) also show supportive results for the accuracy of by default buy and sell identification rather than an algorithm classification as in VPIN. Andersen and Bondarenko (2015) yield no useful role of the VPIN metric in capturing order flow toxicity, which is the main motivation of this study. Indeed, Andersen and Bondarenko (2015) argue nonlinear VPIN transformation applied to the actual signed order flow results in no independent explanatory power of impending market imbalances. Therefore, by default order direction (buy or sell) identification in the data set is used in this study to enhance the accuracy level of order flow toxicity estimation.

Additionally, Matriks IQ does not provide the name of brokerage firms for buy and sell sides on the respective data. However, currently, it contains the brokerage firm identification on the most recent data. Nevertheless, this very study is not concerning such data classification regarding the informed trading context. Lastly,

[12]Trade represents the number executed trade data of the respective stock in a day, Symbol represents the ticker of the respective stock, Time represents the clock time of the executed trade, Price represents the executed price for each trade, Size represents the number of shares executed for each trade, Volume represents the executed volume level in TRY for each trade, and ID represents the binary classification for each trade whether buy or sell with 0 or 1, respectively.

Table 3.1 Sample Data: KCHOL and TCELL

Trade	Symbol	Time	Price	Size	Volume	ID	Trade	Symbol	Time	Price	Size	Volume	ID
1	KCHOL	09:55:07	15.42	30	463	0	**10**	KCHOL	10:00:03	15.43	1,294	19,966	0
2	KCHOL	09:55:07	15.42	16	247	0	**11**	KCHOL	10:00:04	15.42	165	2,544	1
3	KCHOL	09:55:07	15.42	165	2,544	0	**12**	KCHOL	10:00:29	15.43	36	555	0
4	KCHOL	09:55:07	15.42	165	2,544	0	**13**	KCHOL	10:00:29	15.43	34	525	0
5	KCHOL	09:55:07	15.42	24	370	0	**14**	KCHOL	10:00:29	15.43	997	15,384	0
6	KCHOL	09:55:07	15.42	131	2,020	0	**15**	KCHOL	10:00:29	15.43	997	15,384	0
7	KCHOL	10:00:01	15.42	156	2,406	1	**16**	KCHOL	10:01:12	15.44	300	4,632	0
8	KCHOL	10:00:01	15.42	175	2,699	1	**17**	KCHOL	10:01:12	15.44	54	834	0
9	KCHOL	10:00:03	15.43	165	2,546	0	**18**	KCHOL	10:01:12	15.44	64	988	0
Trade	**Symbol**	**Time**	**Price**	**Size**	**Volume**	**ID**	**Trade**	**Symbol**	**Time**	**Price**	**Size**	**Volume**	**ID**
1	TCELL	10:00:01	11.65	1	12	1	**10**	TCELL	10:01:12	11.65	906	10,555	0
2	TCELL	10:00:03	11.65	1	12	0	**11**	TCELL	10:01:12	11.65	873	10,170	0
3	TCELL	10:00:05	11.65	92	1,072	0	**12**	TCELL	10:01:12	11.65	1,000	11,650	0
4	TCELL	10:00:27	11.64	250	2,910	1	**13**	TCELL	10:01:12	11.65	586	6,827	0
5	TCELL	10:00:27	11.64	500	5,820	1	**14**	TCELL	10:01:58	11.65	500	5,825	1
6	TCELL	10:00:27	11.64	620	7,217	1	**15**	TCELL	10:03:42	11.64	628	7,310	1
7	TCELL	10:00:27	11.64	1,370	15,947	1	**16**	TCELL	10:04:11	11.64	1,000	11,640	1
8	TCELL	10:01:12	11.65	4,906	57,155	0	**17**	TCELL	10:06:36	11.65	280	3,262	0
9	TCELL	10:01:12	11.65	1,594	18,570	0	**18**	TCELL	10:07:46	11.65	43	501	0

regarding the ID column, due to the call auction periods in 2017 BIST sessions, Matriks IQ differentiates call auctions (opening, midday, closing call auctions) by assigning ID classification number 128 other than 0 and 1. Continuous auction starts at 10:00:00 and lasts until 13:00:00 before 1-hour midday break until 14:00:00. Then, again, continuous auction continuous up until 18:00:00. Therefore, call auctions in a single trading day are occurred between 09:55:00 - 09:59:59, 13:55:00 - 13:59:59, and 18:00:00 - 18:04:59 for morning, midday, and closing call auctions, respectively. Accordingly, executed trade data within such call auction periods are normally classified with the ID number 128. However, the extracted raw data from Matriks IQ simply assign those orders 0 or 1 rather than 128. In this study, these ambiguously classified orders are accepted as 0 or 1 and considered accordingly in the analysis. Their relatively negligible volumes in total event window volume (%7.52 and %6.74 for KCHOL and TCELL, respectively) should be vanished and no possible misleading during the computation process should be materialized. Correspondingly, all orders which arrive in a call auction are excluded inline to the literature (Choe et al., 1999; Nyholm, 2002; Griffin et al., 2003; Comerton-Forde and Rydge, 2006; Hasbrouck, 2018; Tinic and Salih, 2020) in order to focus on only continuous auction data. Figures 3.1 and 3.2 display the sample data as well.

3.2 Methodology

In this study, intraday data is used rather than daily data due to the motivation of focusing on the respective intraday short-term flash events as illustrated in Figures 3.1 and 3.2. Different from PIN, the VPIN metric was developed by considering volume information rather than price information. Therefore, VPIN assumes the value of information by focusing on intraday volume imbalances. As a result, VPIN captures volume information rather than price information (Easley et al., 2011). Clark (1973), and Ane and Geman (2000) show empirical findings supporting the suitability of the volume-time approach rather than the clock-time approach. In parallel, Easley et al. (2012a) show closer price return distribution to Normal Distribution under the volume-time approach together with less heteroskedasticity

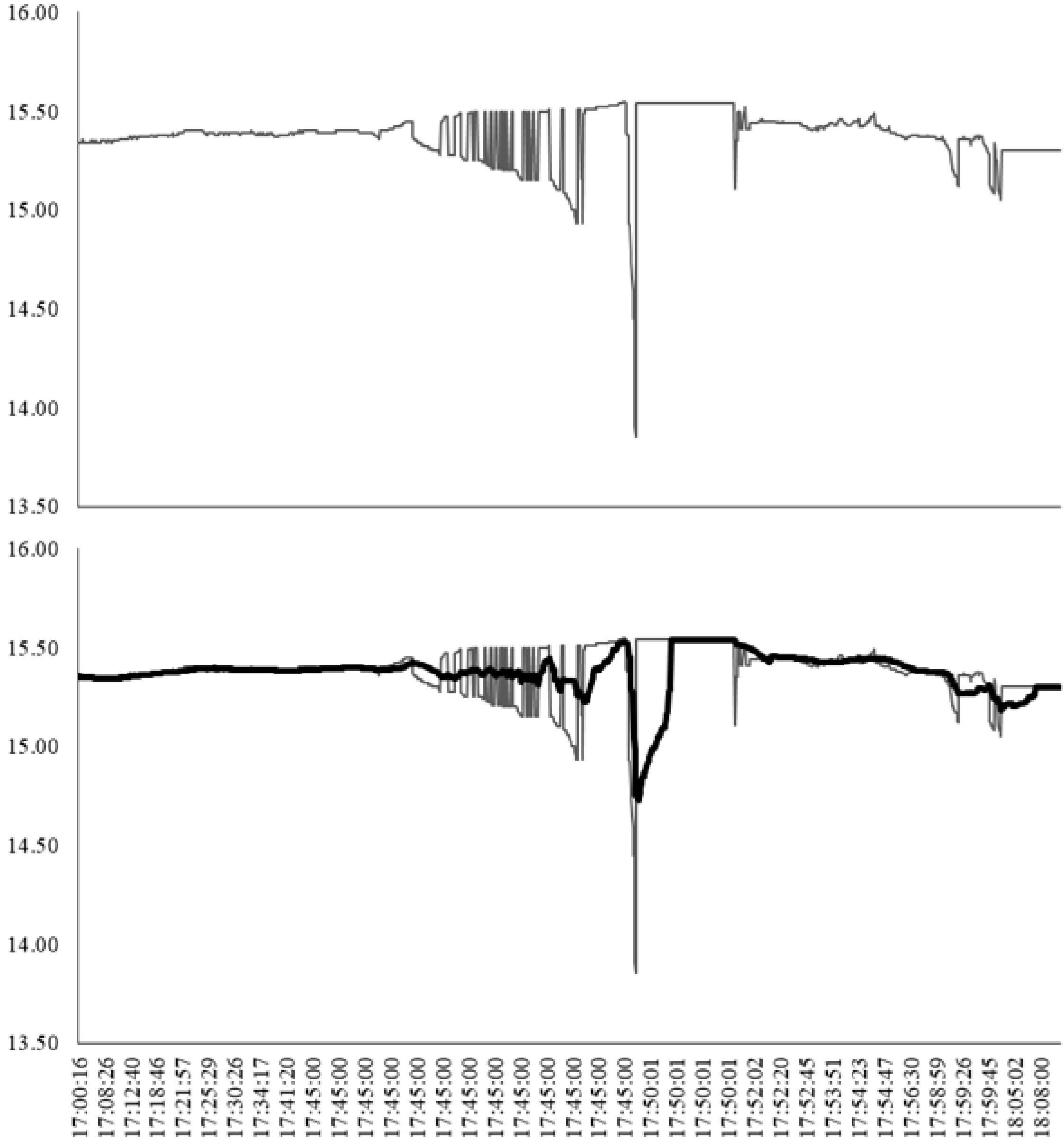

Figure 3.1 Price and VWAP: KCHOL

The above figure shows the tade-by-trade price data (light line) and 50-period VWAP - Volume Weighted Average Price (bold line) level for KCHOL equity stock on February 22, 2017 during the flash crash.

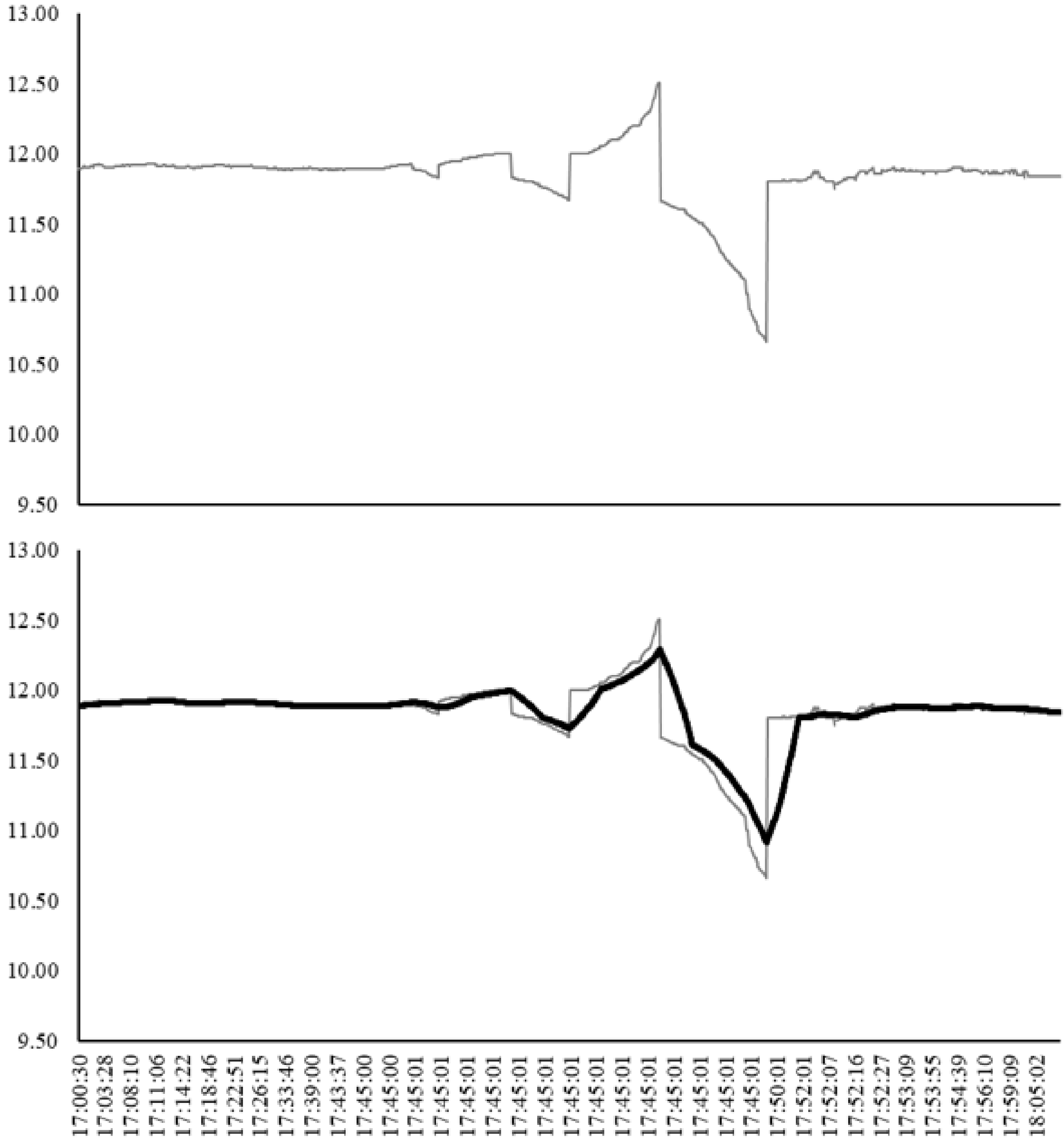

Figure 3.2 Price and VWAP: TCELL

The above figure shows the tade-by-trade price data (light line) and 50-period VWAP - Volume Weighted Average Price (bold line) level for TCELL equity stock on February 22, 2017 during the flash crash.

than the clock-time approach. Another underlying motivation for employing the VPIN measure is based on volume bucket trade clustering. Since VPIN argues the volume-time information aspect, each volume bucket is combined by equal volume size which means equal information. Additionally, the PIN metric was developed around the Poisson Distribution assumption for buy and sell order flows. Whereas, VPIN assumes normally distributed buy and sell order flows which are applicable for dependent volume buckets without Maximum Likelihood Estimation. Another reason for employing VPIN emerges due to the fact that PIN's clock-time estimation when estimating different buckets in order to estimate PIN with segregated intraday data. However, VPIN uses volume-time bucketing by segregating intraday with respect to equal volume size buckets. Most importantly, VPIN is able to be run via using high-frequency data as in this study. Therefore, VPIN ables to detect volume-based order flow toxicity better than PIN due to controlling the volume-synchronized order flow data. Literature shows the significant usefulness of VPIN on liquidity during extreme events where it can be used as an advanced signal (Easley et al., 2012a; Bethel et al., 2012; Wu et al., 2013; Abad and Yagüe, 2012). Yildiz et al. (2013) show a negative (positive) correlation between trade intensity (return volatility) and VPIN. Abad and Yagüe (2012) compare PIN to VPIN and document that the significant difference comes from the number of buckets used in VPIN. Andersen and Bondarenko (2014a) argue VPIN's poor predictive power on short-term volatility. Similarly, Andersen and Bondarenko (2014b) document no predictive power of VPIN neither on crashes nor volatility.

3.2.1 VPIN classification

VPIN metric can be computed by different approaches with respect to trade classification. These approaches are BV-VPIN – Bulk-Volume Volume-Synchronized Probability of Informed Trading, TR-VPIN – Tick-Rule Volume-Synchronized Probability of Informed Trading, and LR-VPIN – Lee-Ready Volume Synchronized Probability of Informed Trading. All above approaches calculate VPIN metric by different classification in order to classify data whether buy or sell initiated. BV-VPIN approach splits aggregated volume data in each bar as a buyer- or seller-initiated volume. Whereas, the TR-VPIN approach assign buy/sell initiation classification trade by trade. Literature has no conclusive and specific choice on the approach selection. While Easley et al. (2011a), and Andersen and Bondarenko (2014) rely on the TR-VPIN, other studies such as Abad and Yagüe (2012), Easley et al. (2012a), and Yıldız et al. (2013) employ the BV-VPIN approach. Easley et al. (2012a) argue that the BV-VPIN approach is superior to the TR-VPIN measure in terms of detecting informed trading accurately. In similar, Chakrabarty et al. (2012) compare BV-VPIN and TR-VPIN classifications, and empirical results show the better ability of BV-VPIN on toxic events in the classification of large and high-frequent trade data. However, on the other hand, there are other studies that show contrary empirical findings to BV-VPIN superiority. Andersen and Bondarenko (2014a) show the vice versa and argue that TR-VPIN performs better than BV-VPIN by examining E-mini S&P 500 future contracts. Subsequently, Andersen and Bondarenko (2014b) observe the BV-VPIN classification's lack of ability on detecting flash events. By investigating German DAX equity stocks, Poeppe et al. (2014) conduct a comparison between BV-VPIN and TR-VPIN approaches.

Results show robust VPIN measure by TR-VPIN on signaling the following crash. In this thesis, as introduced above, the sample data extracted from the Matriks IQ has its own by default buy and sell classification dummy variables for each trade. As a result, no need to employ further classification algorithms. Thus, again as Van Ness et al. (2017), this study's computations are free from any biases due to trade classification algorithms. However, in some sense, the TR-VPIN approach shows similarities with the sample data in terms of classifying data as trade by trade.

3.2.2 VPIN parameters

VPIN has three main parameters that need to be determined, i, N, and T where they represent the time bar in a minute, estimation period for average order imbalance computation, and a number of volume buckets per day, respectively. In the original

VPIN study, Easley et al. (2012a) assign 1, 50, and 50 values for the respective parameters, simply denoted as 1-50-50. Accordingly, in this investigation, the same parameter values are used initially. Therefore, the VPIN measure can be seen as a function of these parameters as illustrated below in Equation 3.1. For robustness purposes, different parameters have been also used following the original estimation.

$$VPIN = f(i, N, T) \tag{3.1}$$

As in the original study, this thesis accepts T=50 as a constant parameter without any further changing via robustness. Also, due to the ability to classify each trade whether buy or sell, no further alteration is necessary for i which is accepted as another constant parameter as i=1. The original VPIN procedure initiates the estimation by trade aggregation in time bars. In each 1-minute time bar, trades (in volume) are respectively aggregated by adding all trades in the respective time bar. Accordingly, raw data become 1-minute classified sample data. Then, the volume bucket is structured in the second step. Each volume bucket simply represents an information content itself in order to measure the order imbalance. In the original VPIN measure, Easley et al. (2012a) measure each volume bucket by dividing the average daily volume by 50. Therefore, in this study, 1-year pre-event daily volume averages for KCHOL and TCELL have been calculated and they are divided into 50 in order to estimate the volume per bucket. The reason for running a look-back methodology to estimate daily volume average comes from the motivation to minimize any overlapping error between the daily volume average period and event window. Thus, the pre-event period is selected for this respective motivation. Volume buckets are filled by adding subsequent trade volume up until filling the respective bucket volume threshold. The excess volume in a volume bucket is given to the next volume bucket and added into it. As mentioned earlier, the data set in this study contains the identification of each trade's direction whether buy or sell. Therefore, in the course of the volume bucket phase, no additional classification algorithm is employed in order to distinguish each trade in terms of the buy or sell direction. Hence, each volume bucket and buy/sell aggregation is estimated quite easily – probably easier than the original VPIN measurement by Easley et al. (2012a). Afterward, following the estimation of buy and sell volume buckets, order imbalance computation has to be performed. By calculating a straightforward absolute difference between buy and sell volume in each bucket, order imbalance is defined. Table 3.2 shows order imbalance for the first eighteen buckets together with displaying each bucket's time interval as an illustrative purpose. As it can be observed, according to trade volume and intensity, each volume's time interval may vary. And at the final step, in order to estimate the VPIN, the sample length of volume buckets needs to be defined. The respective parameter establishes the number of buckets in order to measure VPIN. Therefore, in a simplistic definition, VPIN is the average of order imbalances in the selected sample length of buckets by dividing the summation of order imbalances by the production of volume bucket size multiplied by the priorly defined sample length.

3.2.3 VPIN estimation

Since the data set contains buy and sell identification by default, no further classification efforts are needed in this study which should consequently improve the accuracy of the measure. Below, in Equation 3.2, VPIN estimation is explicitly presented in plain vanilla form as,

$$VPIN_j = \frac{OI_{j;j+49}}{N} \tag{3.2}$$

where, j represents the number of rolling VPIN measure, OI represents Order Imbalance which measures buy and sell order flow imbalance per bucket. Additionally, N represents the number of a volume-time buckets per day. In order to estimate OI measure, volume-time buckets' buy and sell levels are compared in absolute value terms with the assumption of order flow toxicity as in Equation 3.3,

$$OIj = \sum_{t=1}^{N} |W_t^B - W_t^S| \tag{3.3}$$

where, W_t^B and W_t^S represent trading volume for buy and sell orders at time t, respectively and V represents equal number of volume for each volume bucket. In addition, V is estimated by considering the average daily volume of W for KCHOL and TCELL, with respect to a number of buckets per day, T,

$$V_j = \frac{W_j}{T} \tag{3.4}$$

where, T is 50 and constant throughout the thesis investigation as the number of volume buckets per day. By looking back to a 1-year period (approx. 252 trading days) in order to estimate the daily average volume, below Equation 3.5 as,

$$W_i = \frac{\sum_{j=1}^{252} Volume_{t-j}^B + Volume_{t-j}^S}{252}, \; for \; i = KCHOL, TCELL \tag{3.5}$$

and as a result, KCHOL (TCELL) respective volume per bucket is TRY 972,105 (TRY 957,900), and accordingly,

$$V_j = \begin{cases} TRY \; 972,105, & if \; j = KCHOL \\ TRY \; 957,900, & if \; j = TCELL \end{cases} \tag{3.6}$$

where, $Volume_{t-j}^B$ and $Volume_{t-j}^S$ represent daily volume buy and sell per day in the course of [t-252;t-1] pre-event window. Therefore, V becomes a constant average volume size for each volume-time bucket. As a result, VPIN simply groups sequential order flows (trade data in this study) with equal volume bucket size, W, in order to identify an order flow toxicity with controlling abnormal volume imbalance. As originally proposed by Easley et al. (2014), in this study W has been determined at daily average volume for KCHOL and TCELL by estimating the daily average trading volume with respect to 1-year pre-event period. Estimated volume bucket sizes for KCHOL and TCELL are TRY 972,105 and TRY 957,900, respectively as illustrated in Equation 3.5.

KCHOL and TCELL have 153 and 182 volume buckets between the 21-23 February 2017 period, respectively. The reason for this inequality comes from more trading volume and order submission on TCELL than KCHOL as a more liquid blue-chip stock. Flash crashes on February 22, 2017, can be observed in the course of a 1-second period between 17:45:00 – 17:45:01 time period. As can be seen in Figures 3.1 and 3.2, both stocks show considerably different pricing behavior around the flash crash. However, at the same time, they both experience transitory upward move prior to the event, especially on TCELL. Whereas, KCHOL shows consecutive price variations right before the event. Such a move may mimic the possibility of algorithmic trading error and/or systematic trading strategy caused behavior rather than a sudden fat-finger event. As a result, pre-event price volatility on both stocks indicates a possibility of toxic order flow presence and accordingly, deteriorated liquidity prior to the event. In Tables 3.6 and 3.7, descriptive statistics are displayed for the pre-event, event, and post-event periods. As of the number of observations for KCHOL, 37, 53, and 64 with respect to the pre-event, event, and post-event days. On the other, for TCELL, 50, 85, and 47 with respect to the pre-event, event, and post-event days. These observations are sub-period numbers of volume buckets for each period. For KCHOL (TCELL), the summation of such sub-period bucket observations is exactly equal to the above-mentioned figure of 153 (182) of volume buckets between 21-23 February 2017.

However, since this study focuses on VPIN, volume buckets are computed based on volume-time specification. Therefore, this study is able to segregate order flows accurately with respect to the level of toxicity. Figures 3.3 and 3.4 show respective VPIN levels for KCHOL and TCELL. The process of estimating respective VPIN levels follow a rolling-based estimation. This means, for VPIN #1 computation, volume bucket #1 to bucket #50 are used, a total of 50 buckets. Next, for VPIN #2 computation, volume bucket #2 to bucket #51 are used, again a total of 50 buckets. And the rest follows the same rolling criteria. In the end, KCHOL (TCELL) has a total of 106 (136) different VPIN measures until the end of February 23, 2017,

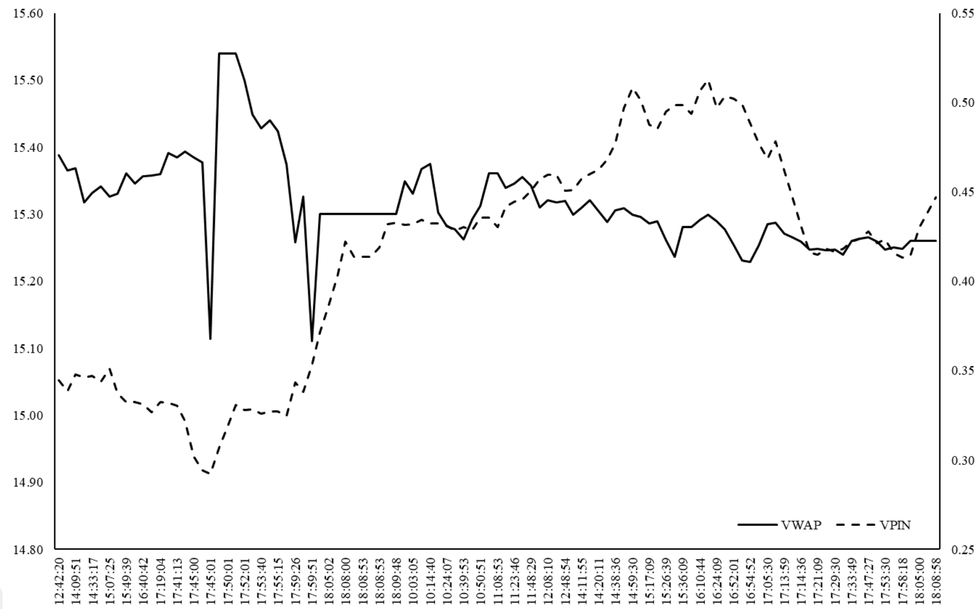

Figure 3.3 VPIN 1-50-50: KCHOL

The above figure shows the 50-period VWAP (Volume Weighted Average Price) and VPIN 1-50-50 (dashed line) for KCHOL equity stock on February 22, 2017 during the event window.

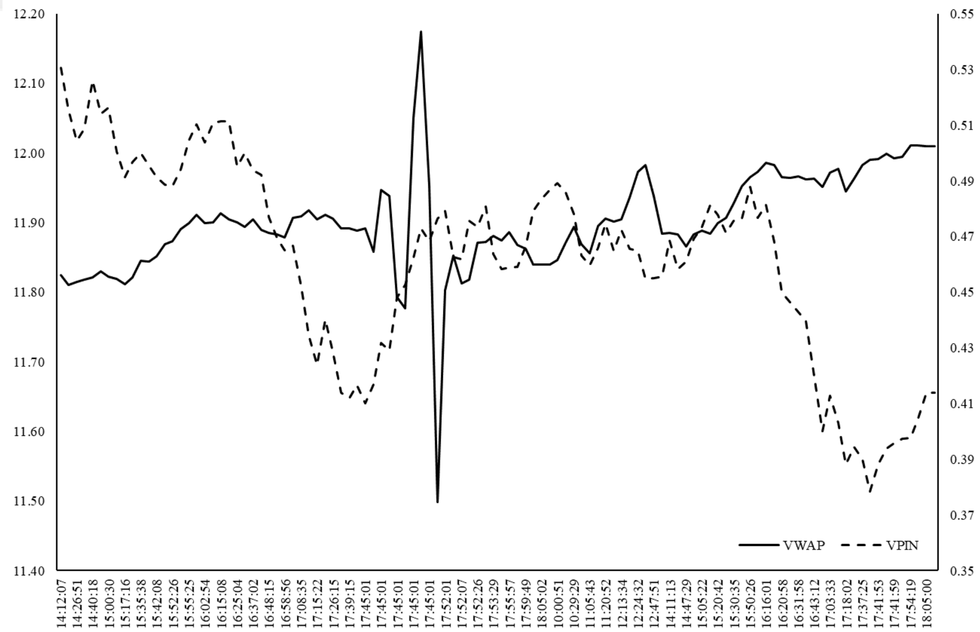

Figure 3.4 VPIN 1-50-50: TCELL

The above figure shows the 50-period VWAP (Volume Weighted Average Price) and VPIN 1-50-50 (dashed line) for TCELL equity stock on February 22, 2017 during the event window.

Table 3.2 Sample Volume Buckets: KCHOL and TCELL

Bucket	Time	Buy	Sell	OI	Duration	Bucket	Time	Buy	Sell	OI	Duration
1	10:13:27	644,181	327,924	316,257	00:13:27	**10**	11:51:09	241,811	730,294	488,482	00:08:16
2	10:20:24	478,191	493,914	15,723	00:06:57	**11**	12:03:34	603,764	368,341	235,423	00:12:25
3	10:33:54	447,894	524,211	76,317	00:13:30	**12**	12:12:33	581,128	390,977	190,150	00:08:59
4	10:46:06	545,786	426,319	119,466	00:12:12	**13**	12:30:32	401,673	570,432	168,758	00:17:59
5	11:00:08	635,780	336,325	299,454	00:14:02	**14**	12:46:56	599,389	372,716	226,674	00:16:24
6	11:14:44	452,667	519,438	66,771	00:14:36	**15**	12:54:13	151,293	820,812	669,518	00:07:17
7	11:20:27	100,609	871,496	770,887	00:05:43	**16**	12:54:13	0	972,105	972,105	00:00:00
8	11:32:13	612,627	359,478	253,150	00:11:46	**17**	14:04:22	296,773	675,332	378,559	00:10:09
9	11:42:53	375,059	597,046	221,987	00:10:40	**18**	14:16:15	431,331	540,774	109,442	00:11:53
Bucket	**Time**	**Buy**	**Sell**	**OI**	**Duration**	**Bucket**	**Time**	**Buy**	**Sell**	**OI**	**Duration**
1	10:23:46	540,407	417,493	122,915	00:23:46	**10**	12:35:54	659,916	297,984	361,932	00:18:30
2	11:03:38	407,356	550,544	143,188	00:39:52	**11**	12:42:35	616,708	341,192	275,515	00:06:41
3	11:14:09	711,513	246,387	465,126	00:10:31	**12**	12:59:56	222,341	735,559	513,217	00:17:21
4	11:20:25	321,775	636,125	314,351	00:06:16	**13**	12:59:56	0	957,900	957,900	00:00:00
5	11:34:08	514,797	443,103	71,695	00:13:43	**14**	12:59:56	0	957,900	957,900	00:00:00
6	11:54:09	549,833	408,067	141,767	00:20:01	**15**	14:00:02	26,304	931,596	905,292	00:00:06
7	11:59:25	913,533	44,367	869,167	00:05:16	**16**	14:07:11	547,959	409,941	138,017	00:07:09
8	12:06:35	345,650	612,250	266,601	00:07:10	**17**	14:12:50	939,309	18,591	920,717	00:05:39
9	12:17:24	755,207	202,693	552,514	00:10:49	**18**	14:34:08	368,634	589,266	220,632	00:21:18

continuous auction. Figures 3.3 and 3.4 show sudden price fall at the crash event however, also document rapid price recovery over respective weighted-average price levels for both KCHOL and TCELL. Following the day of the event day, stocks tend to keep their regular trading activity around pre-event levels without any mini flash crashes and/or consecutive abnormal activities. This post-event behavior shows how traders on KCHOL and TCELL behave considering the fact that experiencing such an event. Therefore, risk-taking reduction behavior cannot be mentioned for the market, specifically on KCHOL and TCELL stock traders following the event.

3.2.4 Regression analysis

During each volume bucket estimation, in this study, different variables are also estimated with respect to each bucket. They are time duration (TIME) between two buckets, maximum (MAX) and minimum (MIN) price levels falling in each bucket, Volume-Weighted Average Price (VWAP), return volatility (VOL), skewness (SKEW), kurtosis (KURT) as a proxy for third and fourth moments, buy:sell order volume ratio (BUYSELL), average order size (AVGOS), and (ILLIQ) illiquidity factor for each bucket calculated as Amihud (2002). These variables represent unique trading order flow behavior and accordingly, toxicity with respect to each volume bucket. Other variables different than VPIN measure are calculated as below,

$$TIME_i = Time\_VolumeBucket_i - Time\_VolumeBucket_{i-1} \tag{3.7}$$

$$MAX_i = Maximum[Price_j; Price_{j+N}] \tag{3.8}$$

$$MIN_i = Minimum[Price_j; Price_{j+N}] \tag{3.9}$$

$$VWAP_i = \frac{\sum_{j=1}^{n} Volume_j}{Size_j} \tag{3.10}$$

$$VOL_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(Return_j - \mu)^2} \tag{3.11}$$

$$SKEW_i = \frac{1}{(n-1)\sigma^3} \sum_{j=1}^{n} (Return_j - \mu)^3 \tag{3.12}$$

$$KURT_i = \frac{1}{(n-1)\sigma^4} \sum_{j=1}^{n} (Return_j - \mu)^4 \tag{3.13}$$

$$ILLIQ_i = \frac{1}{n} \sum_{j=1}^{n} \frac{|Return_j|}{Volume_j} \tag{3.14}$$

$$BUYSELL_i = \frac{\sum_{j=1}^{n} BuyVolume_j}{\sum_{j=1}^{n} SellVolume_j} \tag{3.15}$$

$$AVGOS_i = \frac{\sum_{j=1}^{n} Size_j}{n} \tag{3.16}$$

As mentioned above, each variable is calculated via using observations (trade data) within each volume bucket. Therefore, each variable has its own respective sample data regarding the respective bucket. After estimating and observing VPIN levels for KCHOL and TCELL, the linear relationship between respective bucket intervals' VPIN and intraday variables is conducted. By building an OLS - Ordinary Least Squares linear regression model, the equation tries to explain VWAP behavior by observing the corresponding VPIN levels. By considering the lead-lag relationship between VWAP and VPIN variables, the main purpose of such an econometric model is to investigate the ability of laggard VPIN measure's predictive power on VWAP with [t-1;t-5] period VPIN measures as Equation 3.17 below,

$$VWAP_{i,t} = \gamma_{0,t} + \gamma_{1,t} VPIN_{i,t-j} + \epsilon_{i,t} \tag{3.17}$$

where, $VWAP_{i,t}$ is the i^{th} VWAP at time t, and $VPIN_{i,t-j}$ is the i^{th} VPIN at time t-j. As a result, Equation 3.17 is the univariate model on VWAP by considering VPIN solely as a single independent variable. Whereas, Equation 3.18 considers additional variables as control variables with stressing the ability of VPIN. As a result,

$$VWAP_{i,t} = \upsilon_{0,t} + \upsilon_{1,t} VPIN_{i,t-j} + \upsilon_{2,t} X_{i,t-j} + \nu_{i,t} \tag{3.18}$$

where, $VWAP_{i,t}$ is the i^{th} VWAP at time t, and $VPIN_{i,t-j}$ is the i^{th} VPIN at time t-j, and $X_{i,t-j}$ is the i^{th} control variable set at time t-j. Different than Equation 3.17, Equation 3.18 is the multivariate model on VWAP by considering additional control variable together with VPIN as an independent variable. Control variables are priorly introduced market quality proxy measures as TIME, MAX, MIN, VOL, SKEW, KURT, ILLIQ, BUYSELL, and AVGOS.

Table 3.3 shows results for KCHOL and TCELL univariate regression models' outputs for [t-1;t-5]. As it can be seen, both models' VPIN measures convey statistically significant predictive power ability on subsequent VWAP levels for each t time lag. However, the VPIN metric for TCELL seems to be relatively more solid than the KCHOL. All else equal, VPIN 1-50-50 shows great explanatory power on latter pricing behavior as a proxy for order flow toxicity in the course of the pre-event period. This result is no surprise or at least as expected/hoped. On the other hand, Tables 3.4 and 3.5 show multivariate regressions' outputs for VPIN via controlling additional variables on VWAP predictability. In one sense, multivariate regression outputs in Tables 3.4 and 3.5 regarding VPIN show similarities with univariate regression in terms of KCHOL and TCELL comparison. The predictive power of VPIN 1-50-50 for KCHOL weakens after controlling such control variables. However, in TCELL, even after considering such control variables, VPIN is able to maintain its efficiency as a proxy for the abnormal order imbalance estimator. One other empirical finding is with respect to VPIN's period. Without no differentiation, for both stocks, it can be interpreted that control variables' significance decreases around time t-3, t-4, and t-5 periods. Therefore, VPIN 1-50-50 performs much better via modelling for VWAP at time t+1 and t+2 with VPIN at time t.

Table 3.3 Univariate Regression: VPIN 1-50-50

Panel A: KCHOL					
Variable	**t-1**	**t-2**	**t-3**	**t-4**	**t-5**
INTERCEPT	15.55*** (380.28)	15.56*** (386.24)	15.56*** (383.94)	15.57*** (387.08)	15.56*** (384.39)
VPIN 1-50-50	-0.57*** (-5.85)	-0.59*** (-6.09)	-0.60*** (-6.23)	-0.61*** (-6.35)	-0.61*** (-6.27)
R^2	0.2510	0.2683	0.2794	0.2894	0.2864
S.E.	0.0610	0.0605	0.0601	0.0599	0.0604
N	104	103	102	101	100
Panel B: TCELL					
Variable	**t-1**	**t-2**	**t-3**	**t-4**	**t-5**
INTERCEPT	12.35*** (209.56)	12.36*** (209.65)	12.36*** (204.94)	12.35*** (199.22)	12.35*** (195.84)
VPIN 1-50-50	-0.98*** (-7.95)	-0.99*** (-8.09)	-0.99*** (-7.83)	-0.97*** (-7.49)	-0.97*** (-7.36)
R^2	0.3239	0.3334	0.3205	0.3031	0.2973
S.E.	0.0621	0.0616	0.0622	0.0631	0.0632
N	134	133	132	131	130

Parentheses represent t-test values with *, **, and *** for significance at 10%, 5%, and 1%, respectively.

Table 3.4 Multivariate Regression: VPIN 1-50-50 for KCHOL

Variable	**t-1**	**t-2**	**t-3**	**t-4**	**t-5**
INTERCEPT	1.20 (1.55)	5.15*** (3.06)	7.91*** (4.44)	10.44*** (5.17)	10.42*** (4.96)
VPIN 1-50-50	0.14 (1.38)	0.01 (0.12)	-0.07 (-0.61)	-0.24* (-1.81)	-0.31** (-2.25)
TIME	0.63 (1.05)	0.45 (0.69)	0.76 (1.11)	0.64 (0.83)	0.76 (0.96)
MAX	0.93*** (9.13)	0.67*** (6.06)	0.54*** (4.60)	0.37*** (2.81)	0.36*** (2.63)
MIN	-0.01 (-0.28)	-0.00 (-0.08)	-0.05 (-1.42)	-0.05 (-1.15)	-0.03 (-0.79)
VOL	-2.99 (-1.16)	6.16*** (2.22)	6.87** (2.34)	1.30 (0.40)	-0.44 (-0.13)
SKEW	0.01*** (3.31)	0.00 (0.57)	-0.00 (-0.48)	0.01 (1.30)	0.01** (2.43)
KURT	-0.00 (-1.32)	-0.00*** (-2.96)	-0.00 (-1.35)	0.00 (0.17)	-0.00 (-1.47)
ILLIQ	-1,176.6 (-1.00)	256.53 (0.20)	192.41 (0.14)	998.65 (0.66)	382.77 (0.25)
BUYSELL	0.00 (0.45)	0.01 (1.12)	0.00 (0.25)	0.00 (0.17)	-0.00 (-0.45)
AVGOS	0.00 (1.36)	0.00 (0.92)	0.00 (1.20)	0.00 (1.34)	0.00 (0.99)
R^2	0.6325	0.5735	0.5284	0.4111	0.3889
S.E.	0.0448	0.0484	0.0510	0.0573	0.0587
N	134	133	132	131	130

Parentheses represent t-test values with *, **, and *** for significance at 10%, 5%, and 1%, respectively.

Table 3.5 Multivariate Regression: VPIN 1-50-50 for TCELL

Variable	**t-1**	**t-2**	**t-3**	**t-4**	**t-5**
INTERCEPT	5.70*** (7.49)	11.03*** (10.22)	12.82*** (10.91)	12.03*** (10.63)	8.01*** (7.04)
VPIN 1-50-50	-0.47*** (-4.56)	-0.89*** (-6.19)	-1.02*** (-6.49)	-1.01*** (-6.68)	-0.69*** (-4.51)
TIME	0.58 (1.03)	-0.17 (-0.22)	-0.40 (-0.47)	-0.43 (-0.53)	0.27 (0.33)
MAX	0.40** (2.16)	-1.14*** (-4.85)	-0.70*** (-2.72)	0.14 (0.58)	-0.38 (-1.52)
MIN	0.18 (1.07)	1.25*** (5.25)	0.66** (2.55)	-0.11 (-0.44)	0.74*** (2.87)
VOL	-24.60* (-1.69)	87.88*** (4.26)	50.66** (2.26)	-65.36*** (-3.04)	4.61 (0.21)
SKEW	0.02*** (9.08)	0.01*** (3.46)	0.00* (-1.35)	-0.01** (-2.10)	0.01*** (4.46)
KURT	0.00*** (4.34)	0.00** (2.61)	0.00 (1.12)	0.00 (0.08)	0.00** (82.59)
ILLIQ	-4,297.84** (-2.24)	4,072.52 (1.50)	1,917.58 (0.65)	6,447.29** (2.17)	6,350.95* (2.12)
BUYSELL	-0.00 (-1.52)	0.00 (0.15)	0.00 (0.97)	0.00 (0.76)	-0.00 (-0.63)
AVGOS	-0.00 (-1.31)	-0.00** (-2.09)	-0.00* (-1.76)	-0.00* (-1.70)	-0.00 (-0.99)
R^2	0.7496	0.4999	0.4109	0.4593	0.4522
S.E.	0.0391	0.0552	0.0601	0.0576	0.0579
N	104	103	102	101	100

Parentheses represent t-test values with *, **, and *** for significance at 10%, 5%, and 1%, respectively.

Table 3.6 Descriptive Statistics: KCHOL

Panel A: Pre-Event	**Min**	**Q1**	**Mean**	**Median**	**Q3**	**Max**	**SD**	**Skew.**	**Kurt.**
Price, TRY	15.42	15.45	15.47	15.47	15.50	15.53	0.16	-5.83	36.99
VWAP, TRY	14.06	15.44	15.43	15.47	15.49	15.53	0.02	0.68	18.13
Buy Volume, TRY ('000)	0.00	362.43	480.75	478.19	612.63	972.11	1.06	2.97	11.97
Sell Volume, TRY ('000)	0.00	359.48	491.36	493.91	609.67	972.11	0.53	0.87	1.88
Size	1.00	61.25	1,247.36	398.50	745.25	13,863.00	128.36	4.11	18.79
VPIN 1-50-50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-	-
Panel B: Event	**Min**	**Q1**	**Mean**	**Median**	**Q3**	**Max**	**SD**	**Skew.**	**Kurt.**
Price, TRY	13.90	15.30	15.33	15.37	15.39	15.54	0.02	0.96	24.42
VWAP, TRY	15.11	15.32	15.36	15.37	15.39	15.54	0.01	1.93	15.33
Buy Volume, TRY ('000)	0.00	270.84	446.75	482.93	599.53	972.11	3.41	6.80	50.15
Sell Volume, TRY ('000)	0.00	372.58	525.36	489.18	701.26	972.11	0.47	0.72	1.11
Size	1.00	256.00	3,483.91	1,000.00	2,141.00	25,000.00	140.83	6.17	43.12
VPIN 1-50-50	0.00	0.00	0.11	0.00	0.33	0.42	0.01	1.42	1.30
Panel C: Post-Event	**Min**	**Q1**	**Mean**	**Median**	**Q3**	**Max**	**SD**	**Skew.**	**Kurt.**
Price, TRY	15.21	15.26	15.29	15.28	15.32	15.37	0.00	-0.20	0.73
VWAP, TRY	15.23	15.26	15.29	15.29	15.31	15.37	0.00	0.00	2.35
Buy Volume, TRY ('000)	0.00	237.35	427.04	426.86	628.32	869.87	1.37	3.44	17.67
Sell Volume, TRY ('000)	102.24	343.78	541.20	545.25	726.44	972.11	1.00	2.12	6.52
Size	1.00	333.25	1,459.72	539.50	1,867.50	8,890.00	60.07	5.56	36.05
VPIN 1-50-50	0.39	0.41	0.44	0.43	0.47	0.49	0.01	-0.60	0.79

Table 3.7 Descriptive Statistics: TCELL

Panel A: Pre-Event	**Min**	**Q1**	**Mean**	**Median**	**Q3**	**Max**	**SD**	**Skew.**	**Kurt.**
Price, TRY	11.57	11.63	11.70	11.69	11.76	11.80	0.00	-0.34	0.21
VWAP, TRY	11.58	11.64	11.69	11.69	11.76	11.80	0.00	-0.66	2.05
Buy Volume, TRY ('000)	0.00	462.29	606.64	611.09	867.62	957.90	2.82	6.45	45.93
Sell Volume, TRY ('000)	0.00	90.28	351.26	346.81	495.61	957.90	10.09	4.07	20.02
Size	1.00	260.50	1,630.48	1,085.50	1,932.75	13,000.00	183.41	5.74	36.57
VPIN 1-50-50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-	-
Panel B: Event	**Min**	**Q1**	**Mean**	**Median**	**Q3**	**Max**	**SD**	**Skew.**	**Kurt.**
Price, TRY	11.60	11.82	11.86	11.86	11.89	12.25	0.01	-4.65	39.05
VWAP, TRY	11.50	11.82	11.86	11.85	11.89	12.17	0.01	-1.38	19.59
Buy Volume, TRY ('000)	0.00	358.43	541.50	518.99	795.55	957.90	2.92	5.69	39.11
Sell Volume, TRY ('000)	0.00	162.35	416.40	438.91	599.47	957.90	2.42	3.98	21.73
Size	1.00	548.25	2,059.72	1,311.50	2,416.75	20,582.00	138.43	4.82	24.51
VPIN 1-50-50	0.00	0.00	0.35	0.48	0.51	0.53	0.01	-0.05	0.37
Panel C: Post-Event	**Min**	**Q1**	**Mean**	**Median**	**Q3**	**Max**	**SD**	**Skew.**	**Kurt.**
Price, TRY	12.01	12.01	12.01	12.01	12.01	12.01	0.00	-0.15	-0.52
VWAP, TRY	12.01	12.01	12.01	12.01	12.01	12.01	0.00	-0.84	1.77
Buy Volume, TRY ('000)	0.00	463.02	572.84	578.67	718.98	930.19	0.67	2.42	9.83
Sell Volume, TRY ('000)	27.71	238.92	385.06	379.23	494.88	957.90	1.36	2.42	8.47
Size	1.00	232.00	1,227.06	689.50	1,879.00	10,000.00	20.70	3.21	10.39
VPIN 1-50-50	0.42	0.43	0.45	0.45	0.46	0.48	0.01	0.10	-0.41

3.3 Robustness

In order to enhance the results, and test the results via additional tests, robustness implications are implemented. Robustness tests are twofold – by estimating additional VPIN measure via selecting different parameters other than 1-50-50, and by running additional regression analysis via adding different independent variables into the regression.

3.3.1 Parameters

Other than the original 1-50-50 as firstly initiated by Easley et al. (2012a) in the course of VPIN estimation, 1-25-50 and 1-75-50 parameter structures are used in additionally with the purpose of measuring VPIN via different sample lengths. Rather than examining many other parameter formations, only sample length variation is focused in the parameter robustness section. The reason comes from the order imbalance and accordingly possible adverse selection occurrence by the volume itself. By focusing on the recent order imbalances via shorter sample length, VPIN may be enhanced by detecting the recent high-frequency order imbalance. Since flash events happen suddenly, the recent order flow imbalance should be more carefully monitored. In Figures 3.5 to 3.8, 1-25-50 and 1-75-50 parameter-based VPIN estimations' illustration can be observed. In order to examine shorter and longer sample lengths than the original 50 value, equal distanced 25 and 75 sample lengths parameters are additionally employed in the robustness. Both VPIN measures show a response to subsequent price changes during the flash event period. However, it can be said that 1-25-50 VPIN both for KCHOL and TCELL show more robust and quite rapid response to each flash event than 1-75-50. The reason is quite understandable since the lower the sample length, the faster the price adjustment in VPIN measurement by its formulation. Therefore, higher frequency trade data employed VPIN measurements may signal a possible order imbalance quicker than lower frequency data. Tables 3.8 to 3.11 show the outputs of respective VPIN based multivariate regressions.

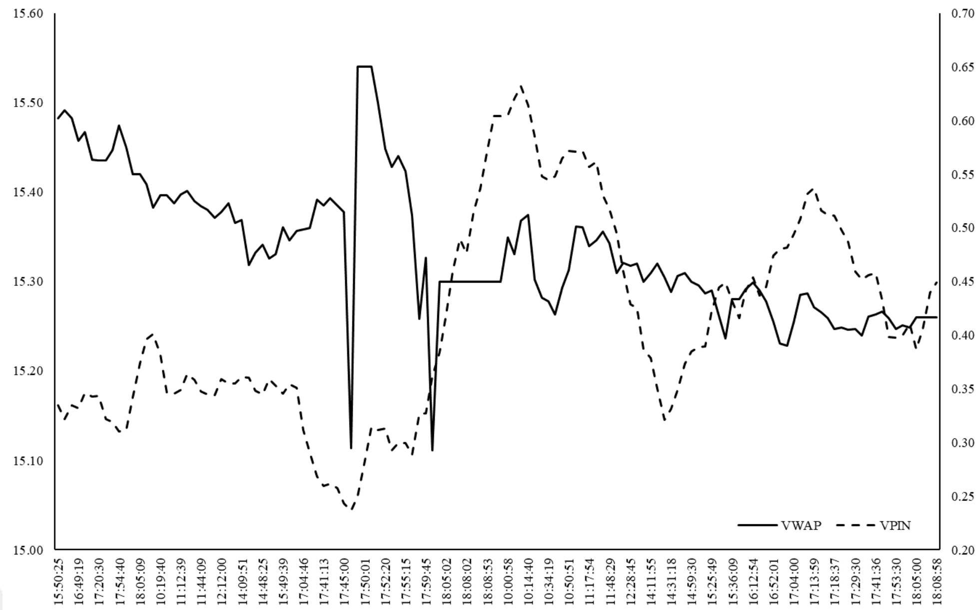

Figure 3.5 VPIN 1-25-50: KCHOL

The above figure shows the 50-period VWAP (Volume Weighted Average Price) and VPIN 1-25-50 (dashed line) for KCHOL equity stock on February 22, 2017 during the event window.

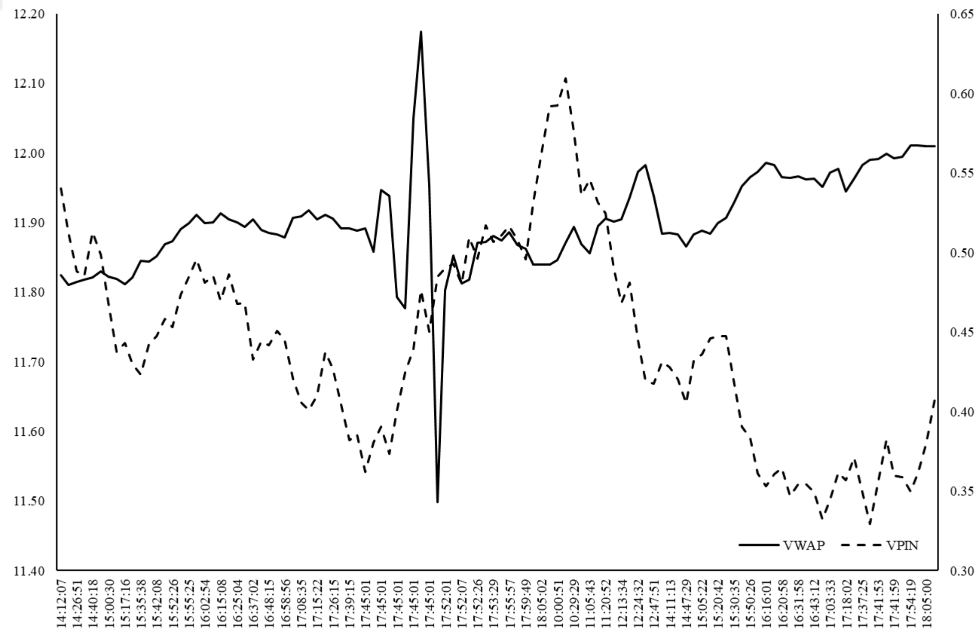

Figure 3.6 VPIN 1-25-50: TCELL

The above figure shows the 50-period VWAP (Volume Weighted Average Price) and VPIN 1-25-50 (dashed line) for TCELL equity stock on February 22, 2017 during the event window.

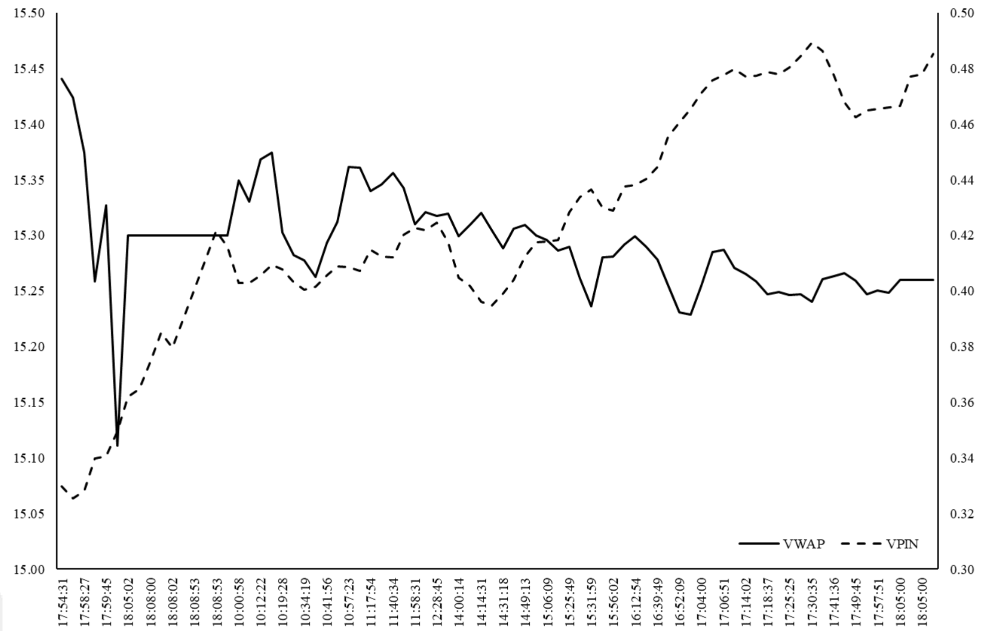

Figure 3.7 VPIN 1-75-50: KCHOL

The above figure shows the 50-period VWAP (Volume Weighted Average Price) and VPIN 1-75-50 (dashed line) for KCHOL equity stock on February 22, 2017 during the event window.

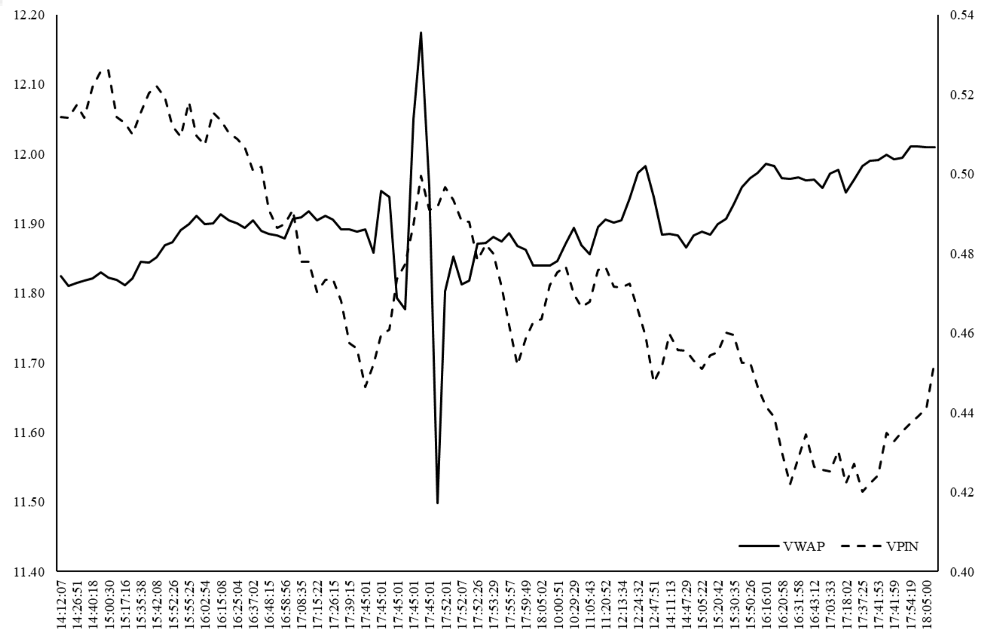

Figure 3.8 VPIN 1-75-50: TCELL

The above figure shows the 50-period VWAP (Volume Weighted Average Price) and VPIN 1-75-50 (dashed line) for TCELL equity stock on February 22, 2017 during the event window.

Table 3.8 Multivariate Regression: VPIN 1-25-50 for KCHOL

Variable	**t-1**	**t-2**	**t-3**	**t-4**	**t-5**
INTERCEPT	1.57* (1.67)	3.44*** (3.33)	4.52*** (4.03)	5.89*** (4.63)	6.18*** (4.74)
VPIN 1-25-50	0.00 (0.09)	0.01 (0.11)	0.02 (0.30)	-0.06 (-0.87)	-0.09 (-1.35)
TIME	0.52 (0.99)	0.50 (0.87)	0.77 (1.23)	0.61 (0.87)	0.70 (0.98)
MAX	0.89*** (13.83)	0.76*** (10.74)	0.74*** (9.59)	0.65*** (7.41)	0.62*** (6.93)
MIN	0.00 (0.10)	0.01 (0.35)	0.04* (-1.03)	-0.03 (-0.82)	-0.02 (-0.54)
VOL	-3.81 (-1.63)	4.88* (1.90)	5.27* (1.91)	-1.04 (-0.34)	-2.34 (-0.74)
SKEW	0.01*** (3.65)	0.00 (0.84)	-0.00 (-0.05)	0.01** (2.15)	0.01*** (-1.98)
KURT	-0.00* (-1.65)	-0.00*** (-3.64)	-0.00** (-2.17)	-0.00 (-0.47)	-0.00** (-1.98)
ILLIQ	-1,151.4 (-1.09)	25.18 (0.02)	-290.76 (-0.23)	577.87 (0.41)	144.79 (0.10)
BUYSELL	0.00 (0.62)	0.01 (1.44)	0.00 (0.40)	0.00 (0.20)	-0.00 (-0.49)
AVGOS	0.00 (1.23)	0.00 (0.93)	0.00 (1.10)	0.00* (1.71)	0.00 (1.58)
R^2	0.7373	0.6764	0.6162	0.5069	0.4826
S.E.	0.0412	0.0452	0.0487	0.0548	0.0556
N	129	128	127	126	125

Parentheses represent t-test values with *, **, and *** for significance at 10%, 5%, and 1%, respectively.

Table 3.9 Multivariate Regression: VPIN 1-25-50 for TCELL

Variable	**t-1**	**t-2**	**t-3**	**t-4**	**t-5**
INTERCEPT	4.43*** (7.37)	7.82*** (8.83)	8.92*** (9.23)	8.59*** (9.40)	6.10*** (6.90)
VPIN 1-25-50	-0.29*** (-4.57)	-0.51*** (-5.41)	-0.55*** (-5.35)	-0.56*** (-5.75)	-0.38*** (-4.02)
TIME	0.91* (1.68)	0.57 (0.71)	0.42 (0.49)	0.41 (0.51)	0.84 (1.07)
MAX	0.53*** (3.35)	-0.78*** (-3.37)	-0.36 (-1.41)	0.39 (1.62)	-0.12 (-0.53)
MIN	0.11 (0.66)	1.14*** (4.79)	0.63** (2.41)	-0.09 (-0.38)	0.63** (2.58)
VOL	-31.51** (-2.22)	69.67*** (3.34)	33.27 (1.46)	-74.98*** (-3.49)	-3.76 (-0.18)
SKEW	0.02*** (10.04)	0.01*** (4.63)	0.00 (0.06)	-0.00 (-0.69)	0.02*** (5.32)
KURT	0.00*** (4.69)	0.00*** (3.42)	0.00** (2.21)	0.00 (1.06)	0.00*** (2.93)
ILLIQ	-4,787.41** (-2.57)	3,420.43 (1.25)	1,633.48 (0.55)	6,213.17** (2.11)	5,108.28* (1.79)
BUYSELL	-0.00* (-1.71)	-0.00 (-0.17)	0.00 (0.49)	0.00 (0.19)	-0.00 (-0.89)
AVGOS	-0.00 (-0.21)	-0.00 (-0.12)	0.00 (0.28)	0.00 (0.31)	0.00 (0.16)
R^2	0.8033	0.5677	0.4785	0.5296	0.5551
S.E.	0.0389	0.0572	0.0623	0.0588	0.0568
N	159	158	157	156	155

Parentheses represent t-test values with *, **, and *** for significance at 10%, 5%, and 1%, respectively.

Table 3.10 Multivariate Regression: VPIN 1-75-50 for KCHOL

Variable	**t-1**	**t-2**	**t-3**	**t-4**	**t-5**
INTERCEPT	2.47 (1.54)	9.89*** (4.65)	13.75*** (5.64)	18.69*** (8.10)	18.83*** (8.42)
VPIN 1-75-50	0.04 (0.32)	-0.21 (-1.30)	-0.18 (-0.94)	-0.56*** (-3.09)	-0.51*** (-2.91)
TIME	0.51 (0.999)	0.74 (1.09)	0.88 (1.13)	0.63 (0.86)	0.77 (1.07)
MAX	0.55*** (4.06)	0.09 (0.48)	0.15 (0.729)	-0.14 (-0.74)	-0.03 (-0.14)
MIN	0.29*** (3.09)	0.27** (2.21)	-0.04 (-0.30)	-0.06 (-0.47)	-0.19 (-1.47)
VOL	1.53 (0.37)	-0.64 (-0.12)	-4.41 (-0.70)	-9.07 (-1.53)	-11.27* (-1.96)
SKEW	0.01*** (4.31)	-0.00 (-0.19)	0.00 (1.01)	0.01 (1.64)	0.01* (1.87)
KURT	-0.00** (-2.03)	-0.00 (-0.24)	0.00 (0.61)	-0.00 (-1.02)	-0.00* (-2.05)
ILLIQ	-12,939.00** (-2.22)	-8,689.70 (-1.13)	-3,007.70 (-0.34)	-5,216.80 (-0.62)	2,757.24 (0.34)
BUYSELL	0.00 (0.91)	0.06* (1.91)	0.03 (0.91)	0.00 (0.99)	0.00 (0.17)
AVGOS	0.00 (0.31)	0.00 (0.92)	0.00* (1.86)	0.00 (1.28)	0.00** (2.01)
R^2	0.6958	0.4064	0.1837	0.2758	0.3231
S.E.	0.0256	0.0337	0.0386	0.0365	0.0354
N	79	78	77	76	75

Parentheses represent t-test values with *, **, and *** for significance at 10%, 5%, and 1%, respectively.

Table 3.11 Multivariate Regression: VPIN 1-75-50 for TCELL

Variable	**t-1**	**t-2**	**t-3**	**t-4**	**t-5**
INTERCEPT	5.92*** (7.16)	11.96*** (9.65)	14.02*** (10.41)	12.94*** (10.08)	8.06*** (6.01)
VPIN 1-75-50	-0.57*** (-3.62)	-1.14*** (-4.87)	-1.26*** (-4.97)	-1.12*** (-4.68)	-0.72*** (-2.85)
TIME	0.35 (0.47)	-1.07 (-0.99)	-1.62 (-1.37)	-1.94* (-1.74)	-0.29 (-0.25)
MAX	0.33* (1.97)	-1.21*** (-4.83)	-0.74*** (-2.70)	0.10 (0.38)	-0.46* (-1.68)
MIN	0.19 (1.09)	1.26*** (4.80)	0.61** (2.15)	-0.14 (-0.50)	0.81*** (2.81)
VOL	-19.47 (-1.34)	91.91*** (4.23)	47.10** (2.00)	-74.40*** (-3.35)	6.17 (0.27)
SKEW	0.02*** (9.57)	0.01*** (2.97)	-0.01** (-2.20)	-0.01*** (-3.30)	0.01*** (3.90)
KURT	0.00*** (3.06)	0.00* (1.94)	0.00 (1.72)	0.00 (0.82)	0.00** (2.23)
ILLIQ	-4,363.60** (-2.18)	4,146.58 (1.39)	1,676.40 (0.52)	6,706.66** (2.09)	7,128.99** (2.13)
BUYSELL	0.00 (0.60)	0.00 (0.37)	0.00 (0.13)	-0.00 (-0.31)	0.00 (0.19)
AVGOS	-0.00 (-0.58)	-0.00* (-1.86)	-0.00* (-1.77)	-0.00* (-1.89)	-0.00 (-0.67)
R^2	0.7625	0.4727	0.3790	0.4467	0.3961
S.E.	0.0381	0.0567	0.0614	0.0579	0.0604
N	109	108	107	106	105

Parentheses represent t-test values with *, **, and *** for significance at 10%, 5%, and 1%, respectively.

3.3.2 Dummy interaction

The respective data sample for this thesis covers [t-1;t+1] event window as introduced earlier. Due to the fact that the respective flash event happens for a very short time, approximately within 1-second. Therefore, previous regression structures for the whole sample data including pre-and post-event periods may result in biased and/or inaccurate outputs with respect to VPIN and VWAP relationship. Therefore, in order to vanish such possible negativity, a considerably enhanced regression model is structured via employing a dummy variable and dummy interaction variable with respect to the flash event. As the event for both KCHOL and TCELL, occurs between 17:45:00 - 17:45:01 specifically, the pre-event period dummy variable should be notated prior to the respective event period. Therefore, with the aim of considering such order flow toxicity before the event, a dummy variable is assigned with 1 for such trades prior to the event window of [17:30:00 - 18:00:00]. As a result, the new regression model should be benefited from identifying the VPIN-VWAP relationship prior to the event, and accordingly, VPIN's predictive power on impending VWAP as an early warning signal would be resulted with higher conviction. For the rest of the data, the dummy variable is assigned as 0.

However, to conduct an enhanced investigation on VPIN's early warning signal as a proxy for impending order flow toxicity, one should consider the sole pre-event effect[13] as well. The above regressions simply cover the whole sample data observations for the pre-, event, and post-event periods. By applying such a dummy interaction effect for the event period, the specific lead-lag relationship between VPIN and VWAP has been differentiated from the whole observation set. Therefore, in a separate robustness check, a similar analysis should be conducted for the

[13]Since one of the main motivations of this thesis is to investigate the pre-event order flow toxicity via VPIN measure, such analysis needs to be conducted. Besides capturing the event period VPIN ability as predictive power, its usefulness in the course of the pre-event period should be also well appreciated as a proxy for an early warning signal. As a result, if possible, such a market risk management mechanism would be employed by the BIST other than sudden circuit breaker

pre-event period in order to detect any early warning signal ability for VPIN. In order to investigate such analysis, a similar method will be used. By assigning 1 to the pre-event period observations (and 0 for the rest of the observations), the additional effect of VPIN for the pre-event period before the flash crash as below Equation 3.19,

$$VWAP_{i,t} = \xi_{0,t} + \xi_{1,t-j}VPIN_{i,t-j} + \xi_{2,t-j}F_{i,t-j} + \xi_{3,t-j}X_{i,t-j} + \psi_{i,t} \qquad (3.19)$$

where, $F_{i,t-j}$ represents the set of dummy variables with respect to the analogy mentioned above according to being in the pre-event period or not, thus,

$$F_{i,t} = \Big\{DUMMY_{i,t-j}\ , VPIN^{*}DUMMY_{i,t-j}\Big\} \qquad (3.20)$$

where, $DUMMY_{i,t-j}$ and $VPIN^{*}DUMMY_{i,t-j}$ represents dummy variable whether 0 or 1, and dummy interaction with respect to VPIN. The dummy variable is assigned as 1 if the respective bucket lies in the pre-event period, 0 otherwise, as,

$$Dummy_{t,j} = \begin{cases} 1, & if\ j = pre - event \\ 0, & otherwise \end{cases} \qquad (3.21)$$

accordingly, if a bucket is out of the pre-event interval, then no dummy interaction effect will be observed since both DUMMY and VPIN*DUMMY variables eventually become zero. On the other hand, if a bucket is in the pre-event interval, then the additional marginal effect of being in the pre-event period will be observed on the respective coefficients of intercept and VPIN.

Tables 3.12 and 3.13 show multivariate regression outputs considering the pre-event period as a dummy. Respective VPIN 1-50-50 significance now should be interpreted as the early signal ability prior to the event as an accurate order flow toxicity. Due to using dummy interaction for VPIN, the summation of coefficients of VPIN and interaction term should be interpreted. For the pre-event period, an additional contribution to VPIN comes from the interaction term when the dummy variable is 1. Therefore, results simply yield the effect of VPIN 1-50-50 on impending VWAP specifically for the pre-event period. VPIN and interaction terms show statistically insignificant results for KCHOL. In parallel, interpreting the level of VPIN and interaction term coefficient summation should be irrelevant since neither is statistically significant. However, for TCELL, VPIN shows a quite significant in the pre-event period. By summing each VPIN and interaction term coefficients, respective levels show negative signs and statistically significant results. As a result, VPIN has explanatory power on impending VWAP for TCELL even before the pre-event and since the crash happens via sell-initiated orders, increasing VPIN prior to the crash implies being an early warning signal with an upward reaction. As a result, VPIN for TCELL does not differ from the event period VPIN role on the following price level. VPIN shows quite significant explanatory power on following VWAP for both the pre-event and event periods. Results indicate the difference between two blue-chip Turkish stocks' different features or additional explanatory variables on the flash event. While VPIN successfully plays the role of being an early warning mechanism for impending sudden price crash for TCELL, no such significance can be mentioned for KCHOL prior to the event - even the fact that the event is the same. Additionally, in order to compare with the initial regression structure - considering the whole observations without any dummy variables - with the dummy variable-based regression, KCHOL VPIN 1-50-50 coefficients become even more positive sign and insignificant. This means, contrary to what was expected prior to the study, VPIN for KCHOL shows positive relation with VWAP prior to the event. Although, since the crash is a sell-initiated event, the rational expectation should be a significantly negative sign of the VPIN coefficient. This is not in line with the outputs for KCHOL. However, for TCELL, results are in parallel with such an expectation indeed. A negative and statistically significant lead-lag relationship between VPIN and VWAP implies an increase in VPIN level when VWAP starts to fall - which is the main expectation from VPIN as an early warning signal role.

Table 3.12 Multivariate Dummy Regression: VPIN 1-50-50 for KCHOL

Variable	t-1	t-2	t-3	t-4	t-5
INTERCEPT	0.73 (0.46)	4.57*** (2.66)	7.05*** (3.94)	9.52*** (4.65)	9.84*** (4.54)
VPIN 1-50-50	0.24* (1.92)	0.14 (0.98)	0.11 (0.75)	-0.06 (-0.37)	-0.20 (-1.17)
DUMMY	0.46 (0.79)	0.41 (0.66)	0.80 (1.24)	0.81 (1.10)	0.63 (0.83)
VPINDUMMY	-1.27 (-0.75)	-1.11 (-0.61)	-2.22 (-1.17)	-2.23 (-1.04)	-1.78 (-0.80)
TIME	0.14 (0.21)	-0.13 (-0.18)	-0.12 (-0.15)	-0.24 (-0.27)	0.21 (0.24)
MAX	0.96*** (9.22)	0.71*** (6.27)	0.59*** (5.07)	0.43*** (3.23)	0.40*** (2.82)
MIN	-0.01 (-0.42)	-0.01 (-0.24)	-0.06* (-1.66)	-0.05 (-1.35)	-0.04 (-0.90)
VOL	-0.63 (-1.01)	6.63** (2.37)	7.51** (2.59)	1.94 (0.59)	-0.10 (-0.03)
SKEW	0.01*** (3.30)	0.00 (0.57)	-0.00 (-0.49)	0.01 (1.32)	0.01** (2.43)
KURT	-0.00 (-1.12)	-0.00*** (-2.72)	-0.00 (-1.06)	0.00 (0.45)	-0.00 (-1.30)
ILLIQ	-1,227.50 (-1.04)	189.16 (0.15)	99.26 (0.07)	902.35 (0.60)	327.74 (0.21)
BUYSELL	0.00 (0.31)	0.00 (0.98)	0.00 (0.03)	0.00 (-0.03)	-0.00 (-0.58)
AVGOS	0.00 (1.49)	0.00 (1.09)	0.00 (1.44)	0.00 (1.56)	0.00 (1.09)
R^2	0.6410	0.5852	0.5557	0.4386	0.4000
S.E	0.0447	0.0482	0.0500	0.0566	0.0588
N	104	103	102	101	100

Parentheses represent t-test values with *, **, and *** for significance at 10%, 5%, and 1%, respectively.

Table 3.13 Multivariate Dummy Regression: VPIN 1-50-50 for TCELL

Variable	**t-1**	**t-2**	**t-3**	**t-4**	**t-5**
INTERCEPT	5.74*** (7.36)	11.15*** (10.12)	12.97*** (10.84)	12.20*** (10.59)	8.01*** (6.88)
VPIN 1-50-50	-0.36** (-2.26)	-0.70*** (-3.08)	-0.78*** (-3.12)	-0.80*** (-3.29)	-0.56** (-2.24)
DUMMY	0.05 (0.44)	0.07 (0.47)	0.10 (0.55)	0.07 (0.41)	0.09 (0.55)
VPINDUMMY	-0.12 (-0.51)	-0.19 (-0.57)	-0.24 (-0.67)	-0.18 (-0.53)	-0.21 (-0.58)
TIME	0.50 (0.88)	-0.34 (-0.42)	-0.60 (-0.69)	-0.62 (-0.74)	0.20 (0.24)
MAX	0.33* (1.94)	-1.20*** (-4.98)	-0.77*** (-2.95)	0.08 (0.33)	-0.41 (-1.58)
MIN	0.20 (1.18)	1.29*** (5.34)	0.71*** (2.71)	-0.07 (-0.29)	0.76*** (2.91)
VOL	-23.50 (-1.60)	-0.32*** (4.35)	53.66** (2.38)	-62.56*** (-2.90)	5.69 (0.26)
SKEW	0.02*** (9.00)	0.01*** (3.40)	-0.00 (-1.41)	-0.01** (-2.17)	0.01*** (4.39)
KURT	0.00*** (4.25)	0.00** (2.55)	0.00 (1.09)	0.00 (0.04)	0.00** (2.59)
ILLIQ	-4,130.60** (-2.13)	4,374.92 (1.60)	2,288.42 (0.77)	6,660.32** (2.24)	6,421.60** (2.12)
BUYSELL	-0.00 (-1.37)	0.00 (0.32)	0.00 (1.17)	0.00 (0.94)	-0.00 (-0.50)
AVGOS	-0.00 (-1.40)	-0.00** (-2.25)	-0.00* (-1.95)	-0.00* (-1.90)	-0.00 (-1.05)
R^2	0.7512	0.5063	0.4206	0.4676	0.4542
S.E	0.0393	0.0553	0.0600	0.0576	0.0583
N	134	133	132	131	130

Parentheses represent t-test values with *, **, and *** for significance at 10%, 5%, and 1%, respectively.

3.3.3 Herfindahl-Hirschman Index

In order to yield more reliable and robust results, additional control variable considerations should be materialized other than market quality proxies above. Since the VPIN measure actually estimates the level of biased order flow environment, one more similar measure would be beneficial to stress the level of effectiveness of VPIN. Therefore, in a separate robustness test, HHI - Herfindahl-Hirschman Index is used. The index[14] owes its name to the two economists' independent studies, Hirschman (1945) and Herfindahl (1997). Even though both studies work on the level of market concentration, the specific topic for each study is varied. The index is widely used for calculating industries' monopolistic conditions by measuring the level of concentration by different participants. Therefore, equality between different firms and their respective market shares are analyzed in order to measure the HHI which results in the level of concentration. Accordingly, HHI simply considers a market share of each competitor in a market to estimate the level of concentration as a proxy for lack of competition as,

$$HHI = s_1^2 + s_2^2 + s_3^2 + ... + s_n^2 = \sum_{j=1}^{n} s_j^2 \tag{3.22}$$

where, HHI is the single level of Herfindahl-Hirschman Index as a proxy for market concentration in a market where n companies are operating and s_j^2 is squared of market share of a company j^{th} in the respective market. Due to using market share expression as fractions of the market, each market share level variable gets value from 0 to 1 (i.e., $0 < s_j \leq 1$), and therefore HHI becomes also somewhere between 0 and 1 (i.e., $0 < HHI \leq 1$) accordingly. In this study, HHI is applied in similar to its

[14]Even though Hirschman (1945) develops his work before Herfindahl's original work as an unpublished doctoral dissertation in 1950, the index is called Herfindahl-Hirschman Index and/or Herfindahl Index. Since Herfindahl (1997) works on market competition in the steel industry which is linked to market concentration investigation, the respective literature acknowledges the study as the reference. Hirschman (1945) works on the concentration of a country's foreign trade by measuring the concentration in the export-import pattern. On the other hand, Herfindahl (1997) focuses on the concentration with respect to steel market competition via market shares.

role in estimating the level of market concentration by considering volume bucket concentration. Accordingly, high HHI up to 1 suggests the market concentration due to a single participant's relatively higher market share than the rest of them. In parallel, low HHI down to 0 implies fairly distributed market share among competitors without concentration. The main motivation for using the HHI as a robustness test is to capture the order flow toxicity by estimating each trade volume's fraction in each volume bucket. Literature generally uses the HHI as a proxy for market fragmentation in order to detect cross-exchange volume concentration (Bethel et al., 2012; Madhavan, 2012; Ibikunle et al., 2020). However, differently from the wider implication of the index, in this thesis, microstructure application of the HHI is conducted to detect possible intraday order flow anomalies in a single stock trading. Similarly, Akins et al. (2012) also employ the HHI methodology to estimate institutional investors' concentration as an additional control variable on impending return. However, in this thesis, different than companies' market share, venues' volume share, and investor types' share, trade volume share in a single bucket is estimated as uniquely as,

$$OFS_i = \frac{TradeVolume_i}{BucketVolume} \tag{3.23}$$

where, OFS_i is the Order Flow Share of i^{th} trade in a bucket, $TradeVolume_i$ is the traded volume of the i^{th} trade in a specific bucket, and $BucketVolume$ is the respective volume threshold for each bucket - TRY 972,105 for KCHOL and TRY 957,900 for TCELL. The respective index is computed for each volume bucket interval with respect to volume-time consideration. By using each order flow in a bucket, buy and sell orders' concentration, thus, the level of abnormal imbalance can be estimated. As a result, with the same analogy as applied on firms in an industry, in this study, buy and sell order flows' imbalances are computed via implementing the HHI method.

The main robustness motivation by applying the index is to stress[15] VPIN a bit more in order to see its predictive power ability after controlling such a control variable[16] measure, therefore HHI is explicitly measured as,

$$HHI_{i,t} = OFS_{i,1}^2 + OFS_{i,2}^2 + OFS_{i,3}^2 + ... + OFS_{i,N}^2 = \sum_{i=1}^{N} OFS_i^2 \tag{3.24}$$

where, $HHI_{i,t}$ represents the Herfindahl-Hirschman Index and $OFS_{i,t}^2$ represents the order flow share, which indicates the portion of an order flow (whether buy or sell) with respect to total order flow volume in a volume bucket interval. Accordingly, the linear relationship between VPIN 1-50-50 and VWAP can be estimated even after controlling the order flow concentration measure together with HHI as,

$$VWAP_{i,t} = \beta_{0,t} + \beta_{1,t-j} VPIN_{i,t-j} + \beta_{2,t-j} X_{i,t-j} + \beta_{3,t-j} HHI_{i,t-j} + \epsilon_{i,t} \tag{3.25}$$

where, $VWAP_{i,t}$ is the volume-weighted average price for bucket i^{th} at time t, $VPIN_{i,t-j}$ is the VPIN measure for bucket i^{th} at time t, $X_{i,t-j}$ and $HHI_{i,t-j}$ are control variables and HHI for i^{th} bucket at time t. Tables 3.14 and 3.15 show multivariate regression outputs. Results imply that VPIN for KCHOL has quite limited explanatory power on the following price level. This is mostly in parallel with the previous results for KCHOL where low VPIN predictive power on VWAP presence.

[15]Due to similar roles with VPIN as being the measure of estimating such order flow toxicity, possible multicollinearity problem may be the subject for HHI. Therefore, before adding HHI into multivariate regression as an additional control variable, multicollinearity tests should be applied. VIF - Variance of Inflation Factor tests the significant correlation among the independent variables in multivariate regression. As a result, HHI has a VIF value below 5 which is the possible multicollinearity suspicion threshold. Additionally, quite a low correlation between HHI and VPIN level of c.0.08 implies no multicollinearity problem thus, no obstacle in order to use HHI as a control variable.

[16]Since HHI and VPIN measures show some similarities with respect to estimating abnormal behaviors in a market, their usage in the same regression as an independent variable could be questioned. However, since no significant VIF and correlation exist, there is no statistically reliable obstacle against using the HHI as a control variable in a regression where VPIN is also an independent variable. Akins et al. (2012) also employs the HHI together with PIN in the same regression in order to explain the impending price return level.

On the other hand, TCELL trading data show negative signs and statistically significant VPIN for the impending VWAP level. Getting such a significance level at 1% for VPIN implies quite a robust predictive power role on VWAP even after controlling the explanatory contribution of HHI. Such an interpretation should be highly appreciated since HHI is actually used as an alternative for or at least a similar method with VPIN as a proxy[17] for order flow imbalance in the literature. Bethel et al. (2012) focus on May 6, 2010, the flash event via using VPIN and HHI as an early warning signals. In order to measure how concentrated exchanges are, HHI is calculated by trade volumes executed by different stock exchanges. As a result, HHI is promoted as an early warning signal for a more gradual normalization intervention by the exchange rather than a circuit breaker. In another study, Madhavan (2012) calculates the HHI in a similar way as in Bethel et al. (2012), by considering volume shares from different exchanges. Madhavan (2012) focuses on the market structure by detecting venue concentration. By doing so, the study captures the level of abnormal venue volume flow. Differently from this thesis motivation, both studies consider cross-exchange volume concentration instead. In this thesis, a higher frequency trading activity-based order flow anomaly tries to be detected. Therefore, quite a unique HHI method and implementation relationship[18] with VPIN 1-50-50 is used in the thesis in order to stress the level of VPIN predictive power.

[17]Even though VPIN and HHI seem relatively similar in a way of estimating a concentration, and yields possible hazardous implication of HHI together with VPIN in a regression, there are also significant objections. Even though previous studies use HHI as an alternative early warning signal prior to a flash event, its ability to capture such an event does not imply an absolute informed trading behavior. Literature uses HHI for cross-venue market fragmentation purposes by comparing each venue's volume share. Possible abnormal volume from a single venue may be occurred due to latency arbitrage as well rather than informed trading. Temporarily mispriced bid-ask quotation could simply drive possible abnormal order flow to the respective trading venue.

[18]As mentioned earlier, the correlation between VPIN and HHI is estimated at 0.08. The respective level of correlation is for VPIN 1-50-50 and HHI measures. However, the correlation of VPIN 1-25-50 and HHI is as high as 0.32. The reason is stemmed from VPIN 1-25-50 measure's ability to weigh the recently occurred abnormal order flows, as HHI. They both are able to react to such order flow anomalies. However, via VPIN 1-50-50, the recent order flow is weighted by 50 rather than 25 in VPIN 1-25-50, and as a result, it does not react as solid as HHI. Accordingly, their correlation becomes lower. Thus, using HHI as a control variable does not undermine the significance of the regression model with VPIN 1-50-50.

Table 3.14 Robust Multivariate Regression: VPIN 1-50-50 for KCHOL

Variable	**t-1**	**t-2**	**t-3**	**t-4**	**t-5**
INTERCEPT	1.47 (0.91)	5.81*** (3.36)	8.83*** (4.85)	11.80*** (5.79)	11.39*** (5.39)
VPIN 1-50-50	0.14 (1.30)	-0.00 (-0.04)	-0.10 (-0.82)	-0.28** (-2.13)	-0.33** (-2.42)
TIME	0.56 (0.91)	0.26 (0.40)	0.50 (0.73)	0.27 (0.35)	0.46 (0.59)
MAX	0.91*** (8.71)	0.63*** (5.61)	0.48*** (4.10)	0.29** (2.22)	0.31** (2.24)
MIN	-0.01 (-0.33)	-0.01 (-0.22)	-0.06 (-1.61)	-0.06 (-1.40)	-0.04 (-1.01)
VOL	-3.01 (-1.17)	6.10** (2.21)	6.77** (2.34)	1.17 (0.37)	-0.63 (-0.19)
SKEW	0.01*** (3.28)	0.00 (0.54)	-0.00 (-0.53)	0.01 (1.28)	0.01** (2.44)
KURT	-0.00 (-1.28)	-0.00*** (-2.88)	-0.00 (-1.26)	0.00 (0.32)	-0.00 (-1.34)
ILLIQ	-1171.14 (-0.99)	266.69 (0.21)	209.63 (0.16)	1024.64 (0.70)	385.74 (0.25)
BUYSELL	0.00 (1.39)	0.00 (0.95)	0.00 (0.02)	-0.00 (-0.12)	-0.00 (-0.69)
AVGOS	0.00 (1.39)	0.00* (1.73)	0.00** (2.24)	0.00*** (2.77)	0.00** (2.25)
HHI	-0.00 (-0.64)	-0.00 (-1.50)	-0.00* (-1.91)	-0.00** (-2.49)	-0.00** (-2.09)
R^2	0.6342	0.5838	0.5468	0.4494	0.4177
S.E.	0.0449	0.0480	0.0503	0.0557	0.0576
N	104	103	102	101	100

Parentheses represent t-test values with *, **, and *** for significance at 10%, 5%, and 1%, respectively.

Table 3.15 Robust Multivariate Regression: VPIN 1-50-50 for TCELL

Variable	**t-1**	**t-2**	**t-3**	**t-4**	**t-5**
INTERCEPT	5.73*** (7.51)	11.08*** (10.22)	12.86*** (10.89)	12.04*** (10.59)	7.95*** (6.99)
VPIN 1-50-50	-0.47*** (-4.58)	-0.90*** (-6.20)	-1.02*** (-6.49)	-1.01*** (-6.65)	-0.68*** (-4.45)
TIME	0.64 (0.14)	-0.08 (-0.11)	-0.32 (-0.37)	-0.41 (-0.49)	0.12 (0.15)
MAX	0.37** (2.21)	-1.12*** (-4.75)	-0.68*** (-2.65)	0.15 (0.58)	-0.40 (-1.58)
MIN	0.17 (0.98)	1.23*** (5.12)	0.64** (2.46)	-0.12 (-0.45)	0.76*** (2.95)
VOL	-25.69* (-1.75)	86.46*** (4.16)	49.41** (2.19)	-65.78*** (-3.03)	7.18 (0.33)
SKEW	0.02*** (9.09)	0.01*** (3.49)	-0.00 (-1.31)	-0.01** (-2.06)	0.01*** (4.33)
KURT	0.00*** (4.24)	0.00** (2.53)	0.00 (1.05)	0.00 (0.07)	0.00*** (2.67)
ILLIQ	-4373.97** (-2.27)	3695.01 (1.43)	1823.15 (0.61)	6454.45** (2.16)	6319.70** (2.11)
BUYSELL	-0.00 (-1.57)	0.00 (0.10)	0.00 (0.93)	0.00 (0.74)	-0.00 (-0.53)
AVGOS	-0.00 (-1.48)	-0.00** (-2.13)	-0.00* (-1.77)	-0.00 (-1.53)	-0.00 (-0.25)
HHI	0.00 (0.72)	0.00 (0.69)	0.00 (0.55)	0.00 (0.19)	-0.00 (-1.10)
R^2	0.7507	0.5018	0.4124	0.4595	0.4577
S.E.	0.0392	0.0554	0.0602	0.0578	0.0578
N	134	133	132	131	130

Parentheses represent t-test values with *, **, and *** for significance at 10%, 5%, and 1%, respectively.

3.4 Results

Results document varying and inconclusive findings on February 22, 2017, the flash crash event on KCHOL and TCEL stocks. For KCHOL, VPIN 1-50-50 starts to rise from 0.30 level to 0.40 accordingly to the event whereas TCELL's VPIN 1-50-50 displays an upward behavior prior to the event from approximately 0.40 level to 0.50. Each VPIN metric experiences an increase during the flash crash in the course of a 1-second period. Subsequently, they diverge in terms of the following behavior throughout the post-event period. Clearly, VPIN 1-50-50 shows the ability to detect such an order flow toxicity before the event in TCELL trading. However, VPIN 1-50-50 shows no activity prior to the event. Interpretation of VPIN behaviors is subject to acceptance of the existence of order flow toxicity due to the presence of a flash crash. In order to expound on the predictive power of VPIN, regression results should be addressed and discussed accordingly. Univariate regression models without considering additional variables but VPIN, regression outputs for t-1 to t-5 lead-lag relationship, the assessment yields statistically significant VPIN 1-50-50 predictive power on subsequent VWAP levels for both stocks. In contrast to what we have seen in Figures 3.3 and 3.4 where KCHOL (TCELL) shows VPIN reaction in the course of the event (pre-event) period, regression outputs imply increasing pre-event significance explanatory power of VPIN 1-50-50 for KCHOL. Even though TCELL has higher significance levels, its corresponding t-value levels show a decreasing behavior from t-2 to t-5 lagged time. As expected, for both stocks, VPIN coefficients have a negative sign which implies a negative directional relationship between lagged VPIN and following VWAP level, in parallel to the nature of the VPIN metric. On the other hand, by looking into multivariate regression outputs as in Tables 3.4 and 3.5 for VPIN 1-50-50, concrete differences can be viewed. After controlling various variables, no solid statistically significant VPIN explanatoriness can be mentioned for KCHOL. On the contrary, TCELL's VPIN shows absolutely robust significance even though controlling such variables. In terms of predictive power on subsequent VWAP, corresponding VPIN levels at times t-3 and t-4 show the highest significance degree. In parallel, MAX and MIN significance levels imply statistically meaningful predictive power on VWAP at time t-2 and t-3 periods. Additionally, volatility variable VOL shows another significance around t-2 and t-3 times. However, for KCHOL, no MIN variable effect can be mentioned on VWAP and only the highest prices' (MAX) prior to the event may be associated with VWAP. Moreover, significance levels start to drop from t-1 to t-5 time lags - more vividly for KCHOL. This observation also converges with the fact that higher R^2 levels for the most recent independent variable levels' predictive power. Also, checking the significance levels of control variables, insignificant (significant) relation with VWAP for KCHOL (TCELL). To sum them up, the aggregated observations and empirical findings may indicate that VPIN has quite limited predictive power on VWAP with respect to KCHOL trading, especially with the most recent VPIN measures at time t-1 to t-3. However, on the contrary, VPIN is actually able to effectively predict the following time's VWAP for TCELL even controlling after additional market quality proxy variables. Additionally, VPIN's early predictive power prior to the event is estimated by considering the pre-event period via dummy interaction on a separate regression structure. While VPIN has no predictive power prior to the event, for TCELL, quite significant early warning signal ability is observed.

Apart from VPIN 1-50-50 estimation, two more additional VPIN estimations have been also measured for the sake of the investigation's robustness - VPIN-1-25-50 and VPIN 1-75-50. Figures 3.5 to 3.8 show the respective VPIN levels' behavior in the course of pre-and post-event periods together with the event. For KCHOL and TCELL, both figures indicate almost the same reaction for VPIN measure with an increase at the event and prior to the event, respectively. Then, flattish (downward) trend afterward, for KCHOL (TCELL). Tables 3.8 to 3.11, which display the output of multivariate regression models for VPIN 1-25-50 and 1-75-50 on VWAP via controlling the same control variables suggest almost similar results for KCHOL and TCELL. While KCHOL VPIN significance implies weak predictive power on VWAP, TCELL VPIN implies total opposite where both VPIN 1-25-50 and VPIN 1-75-50 have statistically significant explanatory power on VWAP for all of the lagged time. Additionally, for the robustness purposes again, multivariate regression models on

VWAP are regressed via adding the Herfindahl-Hirschman Index. In order to stress VPIN 1-50-50 measure a bit more together with the rest of the control variables, VPIN has still statistically significant predictive power for TCELL on VWAP. Nevertheless, KCHOL shows exactly the same results for VPIN with no significance on VWAP prediction. As Easley et al. (2011a) suggest that VPIN may be used in order to predict impending price change, the estimated VPIN measure has robust predictive power on VWAP level for TCELL different than KCHOL. Also, even though controlling the HHI as a proxy for order flow concentration, VPIN for TCELL still maintain its explanatory power on the impending price level. Therefore, even after considering control variables, event-specific dummy interaction, and abnormal order flow proxy, VPIN measures suggest the effective usefulness of VPIN, for TCELL whereas vice versa for KCHOL.

As a result, empirical findings throughout the thesis investigation are consistent with the displayed observations on plotted figures. In order to stress and test the level of robustness, multivariate vs univariate regression models are run. Results document that VPIN measure as a proxy for order flow toxicity in the course of a flash event performs quite successfully for TCELL rather than KCHOL. Even though they both have experienced the same flash event at exactly the same time, empirical findings differ among KCHOL and TCELL. After running multivariate regressions with the purpose of robustness, KCHOL's results imply no evidence of VPIN's ability predictive power as a proxy for order flow toxicity. Inconclusive results for KCHOL and TCELL simultaneous flash events suggest no uniform role of VPIN as a possible early warning signal prior to such extreme market movements at the microstructure level. The main explanation is that lagged VPIN measures are not useful as contrary to expectations for KCHOL due to no significant contribution and relationship presence between toxic order flow and impending sudden price behavior. Naively, the level of volatility comprises the predictive power on VWAP change. Whereas, for TCELL, abnormal order imbalance actually has significant addition to the prediction of subsequent price volatility.

4. CONCLUSION

Since the BISTECH transformation, the number of extreme market movements has been causing abnormal events on BIST stocks, even on blue-chips. Accordingly, on February 22, 2017, two Turkish blue-chip stocks Koç Holding and Turkcell shares experienced sudden price fall approximately in the course of 1-second time period, simultaneously. Such an event on two stocks at the same time deserves attention to be observed. Therefore, in this thesis, the specific instant flash crash event on Koç Holding and Turkcell is investigated with respect to the information asymmetry perspective in terms of informed trading activity purposes. Literature covers relatively less academic research on emerging markets even though information asymmetry matters on such markets more than developed markets. Due to biased estimations of PIN due to unobserved prespecified parameters, in this study, the VPIN metric is employed instead. Empirical findings on Koç Holding and Turkcell stocks show partially similar implications on their intraday trading activity and VPIN relationship. For Koç Holding, VPIN shows consecutive increasing behavior in line with the event without any prior activity whereas for Turkcell, VPIN shows already upward direction in the course of the pre-event period as an informed trading measure. Koç Holding maintains high VPIN levels during the post-event period whereas Turkcell shows a reverting behavior in VPIN. In this study, VPIN levels and most importantly, its predictability try to be investigated in the course of the quest for detecting such toxic order flow action prior to the event. After controlling for various variables, the VPIN measure shows no reliable predictive power on impending price volatility as VWAP, for Koç Holding. Significance levels of VPIN for Koç Holding have vanished when considering such control variables to stress the level of efficiency of VPIN. However, empirical findings suggest exactly the opposite for Turkcell. Results vividly display quite robust VPIN measures on VWAP prediction, as a proxy for order flow toxicity. In parallel to the measure's original analogy, in this study, VPIN can be unambiguously declared as a proxy for toxic order flow environment besides its predictive power on VWAP. The respective implications stemmed from the statistically significant volatility and skewness variables on VWAP for Turkcell together with VPIN estimation. Apparently, abnormal order flow activity as implied by skewed trade price distribution refers to such an unbalanced trading environment prior to the flash crash event. Lagged variables at time t-1 to t-5 volatility and skewness results show supportive significance levels together with VPIN. However, there are still unanswered questions regarding differences between Koç Holding and Turkcell's empirical findings. Even though they both have experienced the exact flash event, how and why VPIN becomes useless for Koç Holding. Accordingly to the public disclosures by the official regularities back then, the same investors and algorithms are the subject of the respective event without any discrepancies. Therefore, one should naturally expect almost exact empirical findings for both of them. However, since indefinite results were obtained for Koç Holding and Turkcell (via both plotted charts and applied regression models), one should also suspect the structure of the respective algorithm run by related traders to the event. Post-event VPIN levels suggest further order imbalances on the Koç Holding whereas declining imbalance for Turkcell. As a result, the toxic order flow activity implied by VPIN keeps increasing even more strongly after the event on the Koç Holding trade. To sum them all up, one would suggest that ability of VPIN's predictive power may be varied for two blue-chip stocks' simultaneous flash events in the same market. Such a result may raise possible uncovered exogenous effects and/or variables, fast and slow traders' existence, and complex interaction between order flows.

Further research could implement an abnormal event study methodology by separating an event into two parts in order to investigate possible abnormal trading activities on various parameters. Additionally, higher frequency intraday data can be used on the investigation such as high-frequency trade, order, and quote data. Another extension of this study could be focusing on different asset classes and/or capital markets in order to investigate the presence of informed trading as a proxy for information asymmetry.

REFERENCES

Abad, D., Massot, M. and Pascual, R., 2018. Evaluating VPIN as a trigger for single-stock circuit breakers. *Journal of Banking Finance*, 86, pp.21-36.

Abad, D. and Yagüe, J., 2012. From PIN to VPIN: An introduction to order flow toxicity. *The Spanish Review of Financial Economics*, 10(2), pp.74-83.

Akay, O.O., Cyree, K.B., Griffiths, M.D. and Winters, D.B., 2012. What does PIN identify? Evidence from the T-bill market. *Journal of Financial Markets*, 15(1), pp.29-46.

Akins, B.K., Ng, J. and Verdi, R.S., 2012. Investor competition over information and the pricing of information asymmetry. *The Accounting Review*, 87(1), pp.35-58.

Aktas, N., De Bodt, E., Declerck, F. and Van Oppens, H., 2007. The PIN anomaly around MA announcements. *Journal of Financial Markets*, 10(2), pp.169-191.

Akyildirim, E., Altarovici, A., and Ekinci, C., 2015. Effects of Firm-Specific Public Announcements on Market Dynamics: Implications for High-Frequency Traders. *Handbook of High Frequency Trading*, Academic Press (Elsevier).

Alexander, S.S., 1964. Price movements in speculative markets: Trends or random walks. *Industrial Management Review (pre-1986)*, 2(2), pp.7.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1), pp.31-56.

Andersen, T.G., and Bondarenko, O., 2013. Assessing Measures of Order Flow Toxicity via Perfect Trade Classification. *Working Paper, School of Economics and Management.*

Andersen, T.G., and Bondarenko, O., 2014a. VPIN and the flash crash. *Journal of Financial Markets*, 17, pp.1-46.

Andersen, T.G., and Bondarenko, O., 2014b. Reflecting on the VPIN dispute. *Journal of Financial Markets*, 17, pp.53-64.

Andersen, T.G., and Bondarenko, O., 2015. Assessing measures of order flow toxicity and early warning signals for market turbulence. *Review of Finance*, 19(1), pp.1-54.

Ane, T. and Geman, H. 2000. Order flow, transaction clock, and normality of asset returns. *The Journal of Finance*, 55(5), pp.2259-2284.

Asness, C.S., Frazzini, A., and Pedersen, L.H., 2014. Quality minus Junk. *Technical Report, AQR Capital Management.*

Bachelier, L., 1900. Theorie de la speculation. *Annales Scientifiques de l'Ecole Normale Superieure Series*, 3(17), pp.21-86.

Ball, R., 1978. Anomalies in relationships between securities' yields and yield-surrogates. *Journal of Financial Economics*, 6(2-3), pp.103-126.

Barber, B., Lehavy, R., McNichols, M., and Trueman, B., 2001. Can investors profit from the prophets? Security analyst recommendations and stock returns. *The Journal of Finance*, 56(2), pp.531-563.

Baron, M., Brogaard, J., Hagströmer, B. and Kirilenko, A., 2019. Risk and return in high-frequency trading. *Journal of Financial and Quantitative Analysis*, 54(3), pp.993-1024.

Bethel, E.W., Leinweber, D., Rübel, O. and Wu, K., 2012. Federal Market Information Technology in the Post–FlashCrash Era: Roles for Supercomputing. *The Journal of Trading*, 7(2), pp.9-25.

Bhattacharya, N., Chakrabarty, B., and Wang, F.X., 2017. Do high frequency traders bring fundamental information into prices?. *Working Paper, Saint Louis University and Singapore Management University.*

Bhattacharya, U., Daouk, H., Jorgenson, B., and Kehr, C.H., 2000. When an event is not an event: the curious case of an emerging market. *Journal of Financial Economics*, 55(1), pp.69-101.

Biais, B., Foucault, T. and Moinas, S., 2015. Equilibrium fast trading. *Journal of Financial economics*, 116(2), pp.292-313.

Boehmer, E., Fong, K. and Wu, J., 2015, March. International evidence on algorithmic trading. *In AFA 2013 San Diego Meetings Paper.*

Boehmer, E., Li, D. and Saar, G., 2018. The competitive landscape of high-frequency trading firms. *The Review of Financial Studies*, 31(6), pp.2227-2276.

Braun, T., Fiegen, J.A., Wagner, D.C., Krause, S.M., and Guhr, T., 2018. Impact and recovery process of mini flash crashes: An empirical study. *PloS one*, 13(5), e0196920.

Bremer, M. and Sweeney, R.J., 1991. The reversal of large stock-price decreases.

The Journal of Finance, 46(2), pp.747-754.

Brennan, M.J. and Subrahmanyam, A., 1995. Investment analysis and price formation in securities markets. *Journal of Financial Economics*, 38(3), pp.361-381.

Brockman, P., and Yan, X.S., 2009. Block ownership and firm-specific information. *Journal of Banking and Finance*, 33(2), pp.308-316.

Brogaard, J. 2010. High frequency trading and its impact on market quality. *Northwestern University Kellogg School of Management Working Paper, 66.*

Brogaard, J., Carrion, A., Moyaert, T., Riordan, R., Shkilko, A., and Sokolov, K., 2016. High-frequency trading and extreme price movements. *Working Paper, University of Washington.*

Brogaard, J. and Garriott, C., 2019. High-frequency trading competition. *Journal of Financial and Quantitative Analysis*, 54(4), pp.1469-1497.

Brogaard, J., Hendershott, T., and Riordan, R., 2014. High-frequency trading and price discovery. *The Review of Financial Studies*, 27(8), pp.2267-2306.

Brogaard, J., Hendershott, T. and Riordan, R., 2017. High frequency trading and the 2008 short-sale ban. *Journal of Financial Economics*, 124(1), pp.22-42.

Brown, K.C., Harlow, W.V. and Tinic, S.M., 1988. Risk aversion, uncertain information, and market efficiency. *Journal of Financial Economics*, 22(2), pp.355-385.

Brown, S., and Hillegeist, S.A., 2007. How disclosure quality affects the level of information asymmetry. *Review of Accounting Studies*, 12(2), pp.443-477.

Brown, S., Hillegeist, S.A., and Lo, K., 2004. Conference calls and information asymmetry. *Journal of Accounting and Economics*, 37(3), pp.343-366.

Budish, E., Cramton, P. and Shim, J., 2015. The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 130(4), pp.1547-1621.

Carhart, M.M., 1997. On persistence in mutual fund performance. *The Journal of Finance*, 52(1), pp.57-82.

Carrion, A., 2013. Very fast money: High-frequency trading on the NASDAQ. *Journal of Financial Markets*, 16(4), pp.680-711.

Cartea, Á. and Penalva, J., 2012. Where is the value in high frequency trading?. *The Quarterly Journal of Finance*, 2(03), p.1250014.

Chaboud, A.P., Chiquoine, B., Hjalmarsson, E. and Vega, C., 2014. Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, 69(5), pp.2045-2084.

Chae, J., 2005. Trading volume, information asymmetry, and timing information. *The Journal of Finance*, 60(1), pp.413-442.

Chakrabarty, B., Pascual, R. and Shkilko, A., 2012. Trade classification algorithms: A horse race between the bulk-based and the tick-based rules. *Available at SSRN, 2182819*.

Chan, L.K., Jegadessh, N., and Lakonishok, J., 1996. Momentum strategies. *The Journal of Finance*, 51(5), pp.1681-1713.

Chan, K., Menkveld, A.J., and Yang, Z., 2008. Information asymmetry and asset prices: Evidence from the China foreign share discount. *The Journal of Finance*, 63(1), pp.159-196.

Chari, V.V., Jagannathan, R., and Ofer, A.R., 1988. Seasonalities in security returns: The case of earnings announcements. *Journal of Financial Economics*, 21(1), pp.101-121.

Choe, H., Kho, B.C. and Stulz, R.M., 1999. Do foreign investors destabilize stock markets? The Korean experience in 1997. *Journal of Financial economics*, 54(2), pp.227-264.

Chopra, N., Lakonishok, J., Ritter, J.R., 1992. Measuring abnormal performance: Do stocks overreact?. *Journal of Financial Economics*, 31(2), pp.235-268.

Christensen, K., Oormen, R.C., and Podolskij, M., 2014. Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, 114(3), pp.576-599.

Christophe, S.E., Ferri, M.G., and Hsieh, J., 2010. Informed trading before analyst downgrades: Evidence from short sellers. *Journal of Financial Economics*, 95(1), pp.85-106.

Cinier, C. and Karagozoglu, A.K., 2008. Information asymmetry, speculation and foreign trading activity: Emerging market evidence. *International Review of Financial Analysis*, 17(4), pp.664-680.

Clark, P.K. 1973. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1), pp.135-155.

Cohen, L., Malloy, C., and Pomorski, L., 2012. Decoding inside information. *The Journal of Finance*, 67(3), pp.1009-1043.

Collin-Dufrense, P. and Fos, V., 2015. Do prices reveal the presence of informed trading? *The Journal of Finance*, 70(4), pp.1555-1582.

Comerton-Forde, C. and Rydge, J., 2006. The influence of call auction algorithm rules on market efficiency. *Journal of Financial Markets*, 9(2), pp.199-222.

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

Cootner, P.H., 1962. Stock prices: Random vs. systematic changes. *Industrial Management Review*, 3(2), pp.24-45.

Corrado, C.J. and Jordan, B.D., 1997. Risk aversion, uncertain information, and market efficiency. *Review of Quantitative Finance and Accounting*, 8(1), pp.51-68.

Cowles, 3rd, A., 1944. Stock market forecasting. *Econometrica*, 12(3/4), pp.206-214.

De Bondt, W.F.M., and Thaler, R., 1985. Does the stock market overreact? *The Journal of Finance*, 40(3), pp.793-805.

DellaVigna, S., and Pollet, J.M., 2009. Investor inattention and Friday earnings announcements. *The Journal of Finance*, 64(2), pp.709-749.

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

Durnev, A.A., and Nain, A.S., 2007. Does insider trading regulation deter private information trading? International evidence. *Pacific-Basin Finance Journal*, 15(5), pp.409-433.

Easley, D., De Prado, M.M.L., and O'Hara, M., 2011a. The microstructure of the "flash crashes": flow toxicity, liquidity crashes, and the probability of informed trading. *The Journal of Portfolio Management*, 37(2), pp.118-128.

Easley, D., De Prado, M.M.L., and O'Hara, M., 2012a. Flow toxicity and liquidity in a high-frequency world. *The Review of Financial Studies*, 25(5), pp.1457-1493.

Easley, D., De Prado, M.M.L., and O'Hara, M., 2014. VPIN and the flash crash: A rejoinder. *Journal of Financial Markets*, 17, pp.47-52.

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

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

Easley, D. and O'Hara, M., 1992b. Adverse selection and large trade volume: The implications for market efficiency. *Journal of Financial and Quantitative Analysis*, 27(2), pp.185-208.

Easley, D., Hvidkjaer, S., and O'Hara, M., 2002. Is information risk a determinant of asset returns? *The Journal of Finance*, 57(5), pp.2185-2221.

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

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

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

Egginton, J.F., Van Ness, B.F. and Van Ness, R.A., 2016. Quote stuffing. *Financial Management*, 45(3), pp.583-608.

Einstein, A., 1905. Über die von der molekularkinetischen Theorie der Warme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322(8), pp.549-560.

Ekinci, C., Akyıldırım, E., and Corbet, S., 2019. Analysing the dynamic influence of US macroeconomic news releases on Turkish stock markets. *Finance Research Letters*, 31, pp.155-164.

Ekinci, C., and Ersan, O., 2018. A new approach for detecting high-frequency trading from order and trade data. *Finance Research Letters*, 24, pp.313-320.

Ersan, O., 2016. Multilayer Probability of Informed Trading. Available at SSRN 2874420.

Ersan, O. and Alıcı, A., 2016. An unbiased computation methodology for estimating the probability of informed trading (PIN). *Journal of International Financial Markets, Institutions and Money*, 43, pp.74-94.

Ersan, O., and Ekinci, C., 2016. Algorithmic and high-frequency trading in Borsa Istanbul. *Borsa Istanbul Review*, 16(4), pp.233-248.

Ersan, O., Dalgic, N., Ekinci, C., and Bodur, M., 2021. High-frequency trading and its impact on market liquidity: a review of literature. *Alanya Academic Review*, 5(1), pp.345-368.

Fama, E.F., 1965. The behavior of stock-market prices. *Journal of Business*, 38(1), pp.34-105.

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

Fama, E.F., and French, K.R., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), pp.3-56.

Fama, E.F., and MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3), pp.607-636.

Garfinkel, J.A., 1994. The importance of postevent insider trading: evidence from earnings announcements. *Doctoral dissertation, University of Florida.*

Gibson, G., 1889. The Stock Markets of London, Paris and New York. *G.P. Putnam's Sons.*

Golub, A., Keane, J., and Poon, S.H., 2012. High frequency trading and mini flash crashes. *Working Paper*, arXiv: 1211.6667.

Granger, C.W.J., and Morgenstern, O., 1963. Spectral analysis of New York stock market prices. *Kyklos*, 16(1), pp.1-27.

Grant, J., 2010. High-frequency trading: Up against a bandsaw. *Financial Times.*

Griffin, J.M., Harris, J.H. and Topaloglu, S., 2003. The dynamics of institutional and individual trading. *The Journal of Finance*, 58(6), pp.2285-2320.

Griffin, J.M., Hirschey, N.H., and Kelly, P.J., 2008. Why does the reaction to news announcements vary across countries. *Working Paper, University of Texas at Austin.*

Grossman, S.J., and Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3), pp.393-408.

Grullon, G. and Wang, F.A., 2001. Closed-end fund discounts with informed ownership differential. *Journal of Financial Intermediation*, 10(2), pp.171-205.

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

Haldane, A.G., 2010. Patience and finance. Speech to Oxford China business forum, *Bank of England*.

Haldane, A.G., 2012. The race to zero. *The Global Macro Economy and Finance - Palgrave Macmillan* pp.245-270.

Harry, V.R., 1959. Stock-market "patterns" and financial analysis: Methodological suggestions. *The Journal of Finance*, 14(1), pp.1-10.

Hasbrouck, J., 1991a. Measuring the information content of stock trades. *The Journal of Finance*, 46(1), pp.179-207.

Hasbrouck, J., 1991b. The Summary Informativeness of Stock Trades: An Econometric Analysis. *Review of Financial Studies*, 4, pp.571-595.

Hasbrouck, J., 2018. High-frequency quoting: Short-term volatility in bids and offers. *Journal of Financial and Quantitative Analysis*, 53(2), pp.613-641.

Hasbrouck, J. and Saar, G., 2013. Low-latency trading. *Journal of Financial Markets*, 16(4), pp.646-679.

Healy, P.M., and Palepu, K.G., 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31(1-3), pp.405-440.

Herfindahl, O.C., 1997. Concentration in the steel industry. Doctoral dissertation, Columbia University.

Hirschman, A.O., 1945. National power and the structure of foreign trade. *University of California Press*, pp.155-162.

Hoffmann, P., 2014. A dynamic limit order market with fast and slow traders. *Journal of Financial Economics*, 113(1), pp.156-169.

Huang, J. and Wang, J., 2009. Liquidity and market crashes. *Review of Financial Studies*, 22, pp.2607-2643.

Hughes, J.S., Liu, J., and Liu, J., 2007. Information asymmetry, diversification, and cost of capital. *The Accounting Review*, 82(3), pp.705-729.

Hwang, L.S., Lee, W.J., Lim, S.Y. and Park, K.H., 2013. Does information risk affect the implied cost of equity capital? An analysis of PIN and adjusted PIN. *Journal of Accounting and Economics*, 55(2-3), pp.148-167.

Ibikunle, G., Mare, D. and Sun, Y., 2020. The paradoxical effects of market fragmentation on adverse selection risk and market efficiency. *The European Journal of Finance*, 26(14), pp.1439-1461.

Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. *The Journal of Finance*, 45(3), pp.881-898.

Jegadeesh, N., and Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), pp.65-91.

Jin, M., Liu, Y., and Meng, J., 2019. Fat-finger event and risk-taking behavior. *Journal of Empirical Finance*, 53, pp.126-143.

Johnson, N., Zhao, G., Hunsader, E., Qi, H., Johnson, N., Meng, J., and Tivnan, B., 2013. Abrupt rise of new machine ecology beyond human response time. *Scientific reports*, 3, 2627.

Jones, C.M., 2013. What do we know about high-frequency trading?. *Columbia Business School Research Paper*, (13-11).

Kang, J., Kwon, K.Y., and 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(2), pp.164-191.

Ke, B., Huddart, S., and Petroni, K., 2003. What insiders know about future earnings and how they use it: Evidence from insider trades. *Journal of Accounting and Economics*, 35(3), pp.315-346.

Ke, Y. and Zhang, Y., 2019. Does high-frequency trading reduce market underreaction to earnings news?. *Finance Research Letters*, 34, 101239.

Keynes, J.M., 1923. Some aspects of commodity markets. *Machester Guardian Commercial: European Reconstruction Series*, 13, pp.784-786.

Kirilenko, A., Kyle, A.S., Samadi, M., and Tuzun, T., 2017. The flash crash: High-frequency trading in an electronic market. *The Journal of Finance*, 72(3), pp.967-998.

Kitamura, Y., 2017. Predicting a flash crash in the yen/dollar foreign exchange market. *Applied Economics Letters*, 24(14), pp.987-990.

Krishnaswami, S. and Subramaniam, V., 1999. Information asymmetry, valuation, and the corporate spin-off decision. *Journal of Financial Economics*, 53(1), pp.73-112.

Kyle, A., 1985. Continuous auctions and insider trading. *Econometrica*, 53, pp.1315-1336.

Lai, S., Ng, L. and Zhang, B., 2014. Does PIN affect equity prices around the world?. *Journal of Financial Economics*, 114(1), pp.178-195.

Lakonishok, J., and Lee, I., 2001. Are insider trades informative?. *The Review of Financial Studies*, 14(1), pp.79-111.

Lakonishok, J., Shleifer, A., and Vishny, R.W., 1994. Contrarian investment, extrapolation, and risk. *The Journal of Finance*, 49(5), pp.1541-1578.

Lee, C.C., Lee, J.D., and Lee, C.C., 2010. Stock prices and the efficient market hypothesis: Evidence from a panel stationary test with structural breaks. *Japan and the World Economy*, 22(1), pp.49-58.

Lee, D.W., and Liu, M.H., 2011. Does more information in stock price lead to greater or smaller idiosyncratic return volatility? *Journal of Banking and Finance*, 35(6), pp.1563-1580.

Lehmann, B.N., 1990. Fads, martingales, and market efficiency. *The Quarterly Journal of Economics*, 105(1), pp.1-28.

Lin, H.W.W., and Ke, W.C., 2011. A computing bias in estimating the probability of informed trading. *Journal of Financial Markets*, 14(4), pp.625-640.

Lo, A.W. and MacKinlay, A.C., 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies*, 1(1), pp.41-66.

Low, R.K.Y., Li, T. and Marsh, T., 2018. BV–VPIN: Measuring the impact of order flow toxicity and liquidity on international equity markets. *Working Paper*.

Lu, Y.C., Wong, W.K., and Wong, W.K., 2008. Probability of information-based trading as a pricing factor in Taiwan stock market. *Available at SSRN 1115419.*

MacCauley, F.R., 1925. Forecasting security prices. *Journal of the American Statistical Association*, 20(150), pp.244-249.

Madhavan, A., 2012. Exchange-traded funds, market structure, and the flash crash. *Financial Analysts Journal*, 68(4), pp.20-35.

Malinova, K., Park, A. and Riordan, R., 2013. Do retail traders suffer from high frequency traders. *Available at SSRN*, 2183806.

Malkiel, B. G., 2003. The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1), pp.59-82.

McInish, T., Upson, J. and Wood, R.A., 2014. The flash crash: Trading aggressiveness, liquidity supply, and the impact of intermarket sweep orders. *Financial Review*, 49(3), pp.481-509.

Mehdian, S., Nas, T. and Perry, M.J., 2008. An examination of investor reaction to unexpected political and economic events in Turkey. *Global Finance Journal*, 18(3), pp.337-350.

Menkveld, A.J., 2013. High frequency trading and the new market makers. *Journal of financial Markets*, 16(4), pp.712-740.

Mohanram, P., and Rajgopal, S., 2009. Is PIN priced risk? *Journal of Accounting and Economics*, 47(3), pp.226-243.

Morck, R., Yeung, B., and Yu, W., 2000. The information content of stock markets: why do emerging markets have synchronous stock price movements? *Journal of Financial Economics*, 58(1-2), pp.215-260.

Nyholm, K., 2002. Estimating the probability of informed trading. *Journal of Financial Research*, 25(4), pp.485-505.

Schaub, N., 2018. The role of data providers as information intermediaries. *Journal of Financial and Quantitative Analysis*, 53(4), pp.1805-1838.

Osborne, M.F.M., 1959. Brownian motion in the stock market. *Operations Research*, 7(2), pp.145-173.

Paparizos, P., Dimitriou, D., Kenourgios, D. and Simos, T., 2016. On high frequency dynamics between information asymmetry and volatility for securities. *The Journal of Economic Asymmetries*, 13, pp.21-34.

Pearson, K., 1905. The problem of the random walk. *Nature*, 72(1865), pp.294.

Poeppe, T., Moos, S., and Schiereck, D., 2014. The Sensitivity of VPIN to the Choice of Trade Classification Algorithm. *World Finance Conference Paper*, 2014.

Popper, N. 2012. High-speed trading no longer hurtling forward. *The New York Times*, 14 October. Available at: https://www.nytimes.com/2012/10/15/business/with-profits-dropping-high-speed-trading-cools-down.html.

Regnault, J., 1863. Calcul des Chances et Philosophie de la Bourse. *Mallet-Bachelier et Castel*, Paris.

Roberts, H., 1967. Statistical versus clinical prediction of the stock market. *Working Paper*.

Samuelson, P., 1965. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6, pp.41-49.

Scholes, M.S., 1972. The market for securities: Substitution versus price pressure and the effects of information on share prices. *The Journal of Business*, 45(2), pp.179-211.

Scholtus, M., Van Dijk, D., and Frijns, B., 2014. Speed, algorithmic trading and market quality around macroeconomic news announcements. *Journal of Banking and Finance*, 38, pp.89-105.

Schwert, G.W., 2003. Anomalies and market efficiency. *Handbook of the Economics of Finance*, 1, pp.939-974.

Shiller, R.J., 2000. Irrational Exuberance. *Princeton University Press*, Princeton, NJ.

Silva, F.B.G. and Volkova, E., 2018. Can VPIN forecast geopolitical events? Evidence from the 2014 Crimean Crisis. *Annals of Finance*, 14(1), pp.125-141.

Steiger, W., 1964. A Test of Nonrandomness in Stock Price Changes, *in P.H. Cootner, The random character of stock market prices, The MIT Press*, pp.303-312.

Stiglitz, J.E., 2014. Tapping the brakes: Are less active markets safer and better for the economy?. Presented at the Federal Reserve Bank of Atlanta, *Working Paper*.

Stoll, H.R., 1978. The pricing of security dealer services: An empirical study of NASDAQ stocks. *The Journal of Finance*, 33(4), pp.1153-1172.

Tinic, M., 2019. Informed trading in Borsa Istanbul. *Doctoral Dissertation, Bilkent University.*

Tinic, M., and Salih, A., 2020. Informed trading, order flow shocks and the cross section of expected returns in Borsa Istanbul. *Applied Economics*, 52(13), pp.1446-1459.

Tinic, M., Tanyeri, B., and Bodur, M., 2021. Who to trust? Reactions to analyst recommendations of domestic versus foreign brokerage houses in a developing stock market. *Finance Research Letters*, 101950.

Van Ness, B.F., Van Ness, R.A., and Yildiz, S., 2017. The role of HFTs in order flow toxicity and stock price variance, and predicting changes in HFTs' liquidity provisions. *Journal of Economics and Finance*, 41(4), pp.739-762.

Wang, J., 1993. A model of intertemporal asset prices under asymmetric information. *The Review of Economic Studies*, 60(2), pp.249-282.

Wei, W.C., Gerace, D. and Frino, A., 2013. Informed trading, flow toxicity and the impact on intraday trading factors. *Australasian Accounting, Business and Finance Journal*, 7(2), pp.3-24.

Wilson, E.J., and Marashdeh, H.A., 2007. Are co-integrated stock prices consistent with the efficient market hypothesis? *The Economic Record*, 83, pp.87-93.

Womack, K.L., 1996. Do brokerage analysts' recommendations have investment value? *The Journal of Finance*, 51(1), pp.137-167.

Working, H., 1934. A random-differences series for use in the analysis of time series. *Journal of the American Statistical Association*, 29(185), pp.11-24.

Working, H., 1949. The investigation of economic expectations. *The American Economic Review*, 48(2), pp.188-199.

Wu, K., Bethel, W., Gu, M., Leinweber, D. and Ruebel, O., 2013. Testing VPIN on Big Data–Response to'Reflecting on the VPIN Dispute'. *Available at SSRN 2318259.*

Yan, Y., and Zhang, S., 2012. An improved estimation method and empirical properties of the probability of informed trading. *Journal of Banking and Finance*, 36(2), pp.454-467.

Yang, J., 2003. Market segmentation and information asymmetry in Chinese stock markets: A VAR analysis. *Financial Review*, 38(4), pp.591-609.

Yildiz, S., Van Ness, R.A., and Van Ness, B.F., 2013. Analysis Determinants of VPIN, HFTs' Order Flow Toxicity and Impact on Stock Price Variance. *In FMA annual meetings*, October. Chicago, Illinois.

Zhang, F., 2010. High-frequency trading, stock volatility, and price discovery. *Working Paper.*

CURRICULUM VITAE

Personal Information

Name and Surname: Mehmet Bodur

Education

Bachelor of Arts: International Finance, Yeditepe University, 2016

Master of Arts: Financial Engineering, Kadir Has University, 2021

Master of Sciences: Banking and Finance, University of St. Gallen, 2018 (ABD)

Foreign Languages: English, German

Working Experience

08/2021 - Present: Analyst, QNB Finans Asset Management, Istanbul

02/2021 - 07/2021: Anayst, Alternatif Securities, Istanbul

07/2020 - 10/2020: Anayst, InvestAZ Securities, Istanbul

03/2017 - 10/2018: Analyst, Bayswater Capital Partners, Zurich