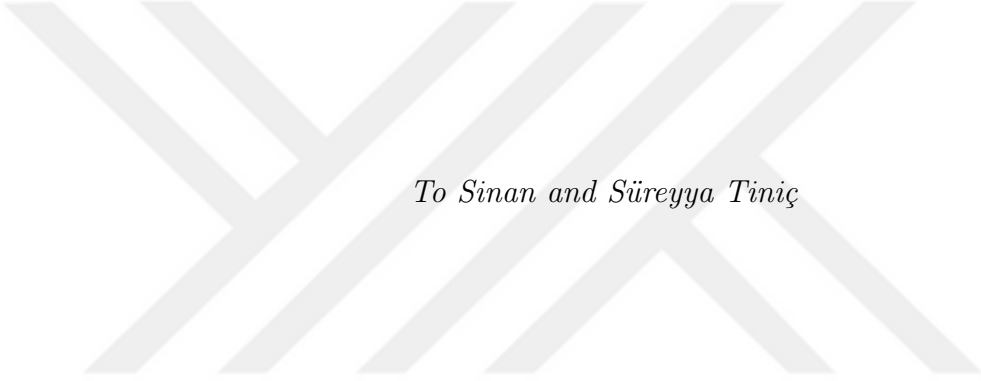




To Sinan and Süreyya Tiniç

INFORMED TRADING IN BORSA İSTANBUL

The Graduate School of Economics and Social Sciences
of
İhsan Doğramacı Bilkent University

by

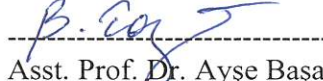
MURAT TİNİÇ

In Partial Fulfillment of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY IN MANAGEMENT

THE DEPARTMENT OF MANAGEMENT
İHSAN DOĞRAMACI BİLKENT UNIVERSITY
ANKARA

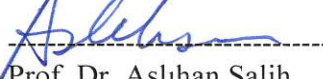
May 2019

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy in Finance.



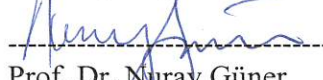
Asst. Prof. Dr. Ayşe Başak Tanyeri
Supervisor

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy in Finance.



Prof. Dr. Aslıhan Salih
Co-Advisor

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy in Finance.



Prof. Dr. Nuray Güner
Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy in Finance.



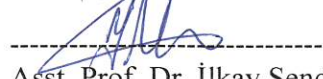
Assoc. Prof. Dr. Levent Akdeniz
Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy in Finance.



Prof. Dr. Kürşat Aydoğan
Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy in Finance.



Asst. Prof. Dr. İlkey Şendeniz Yüncü
Examining Committee Member

Approval of the Graduate School of Economics and Social Sciences



Prof. Dr. Halime Demirkan
Director

ABSTRACT

INFORMED TRADING IN BORSA İSTANBUL

Tiniç, Murat

Ph.D. in Department of Management

Advisor: Asst. Prof. Dr. Ayşe Başak Tanyeri

Co-Advisor: Prof. Dr. Aslıhan Salih

May 2019

This thesis investigates how information asymmetry affects asset prices in Borsa İstanbul. In the first chapter, we introduce the R package **InfoTrad** that estimates the probability of informed trading. Next, we examine the relationship between information asymmetry and stock returns in Borsa İstanbul. Firm-level cross-sectional regressions indicate an economically insignificant relationship between PIN and future returns. Moreover, univariate and multivariate portfolio analyses show that portfolios of stocks with high levels of informed trading do not realize significant return premiums. Consequently, our results, suggest that information asymmetry is a firm-specific risk and it can be eliminated with portfolio diversification. Finally, we compare the informational (dis)advantage of foreign investors trading in Borsa İstanbul. We first show that an average foreign trade creates buy pressure whereas an average local trade generates a sell pressure. The permanent impact of foreign investors over and above local investors is significant only for 24 stocks which correspond to 7% of our sample. Importantly, we show that the foreign price impact occurs primarily in a period of political instability which started with the Gezi Park protests in June 2013. In a panel setting, we also show that adverse selection cost due to foreign trading significantly increases even when we control for firm-specific

factors along with global and local macroeconomic conditions. Domestic investors with undiversified portfolios may be more risk-averse during periods of increased turmoil. This may enable foreign investors to have a better position to take advantage of potential price misalignments, especially for stocks of commercial banks.

Keywords: Asset Pricing, Market Microstructure, Political Turmoil, Probability of Informed Trading, R.



ÖZET

BORSA İSTANBUL'DA BİLGİYE DAYALI ALIM VE SATIMLAR

Tiniç, Murat

Doktora, İşletme

Tez Danışmanı: Dr. Öğr. Üyesi Ayşe Başak Tanyeri

2. Tez Danışmanı: Prof. Dr. Aslıhan Salih

Mayıs 2019

Bu tez, bilgi asimetrisinin Borsa İstanbul'da varlık fiyatlarını nasıl etkilediğini incelemektedir. Tezin ilk bölümünde, bilgiye dayalı alım satım oranı (BDASO) ölçütünün hesaplanması üzerine geliştirilen yöntemleri bir araya getiren, R diliyle hazırlanmış **InfoTrad** istatistiksel paketini tanıtmaktayız. Bir sonraki bölümde, Borsa İstanbul'da işlem gören tüm hisseler için, bilgiye dayalı alım satım oranı (BDASO) ile hisse senedi getirileri arasındaki ilişki incelenmektedir. İşlem gören her hisse için her ay BDASO değerlerini gün içi işlem defteri verilerini kullanarak hesapladık. BDASO değerleri ile hisse senetlerinin aylık getirileri arasındaki ilişkiyi, regresyon analizi, tek değişkenli portföy analizi ve çift değişkenli portföy analizleri ile inceledik. Sonuçlarımız Borsa İstanbul'da BDASO ile ölçülen bilgi asimetrisinin hisse senedi fiyatları üzerinde sistematik bir risk unsuru olmadığını göstermektedir. Bu kapsamda, sonuçlarımız, literatürde bilgi asimetrisi riskinin portföy çeşitlendirmesiyle ortadan kaldırılması gerektiğini savunan teorik çalışmalarla uyumludur.

Son bölümde, Borsa İstanbul'da işlem yapan yerli ve yabancı yatırımcılar arasındaki bilgi asimetrisini incelemekteyiz. İlk olarak, ortalama bir yabancı yatırımcı işleminin piyasada alım baskısı oluştururken, yerli yatırımcının satış baskısı yarattığını göstermekteyiz. Yabancı yatırımcıların fiyatlara kalıcı etkisini, veri kümемizin yüzde

7'sine tekabül eden, 24 hisse de gözlemlemekteyiz. Yabancı yatırımcıların fiyat etkisinin, Haziran 2013'te Gezi Parkı protestolarıyla başlayan siyasi olarak çalkantılı dönemde gerçekleştiğini göstermekteyiz. Panel regresyon analizlerimizde, firmaya özgü faktörler, yerel ve küresel makroekonomik koşulları kontrol ettiğimizde bile, yabancı yatırımcılara bağlı ters seçim maliyetlerinin, Haziran 2013'den sonra arttığını göstermekteyiz. Portföy çeşitliliği düşük olan yerli yatırımcılar artan karmaşa dönemlerinde riskten kaçabilirler. Bu durum, yabancı yatırımcıların, olası fiyat yanlışlıklarından kendilerine avantaj sağlamalarına neden olabilir.

Anahtar Kelimeler: Bilgiye Dayalı-Alım Satım Oranı, Piyasa Mikroyapısı, Politik Karışıklıklar, R, Varlık Fiyatlaması.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my advisors, Prof. Aslıhan Salih, Asst. Prof. Tanseli Savaşer and Asst. Prof. Başak Tanyeri for their constant support throughout my Ph.D. studies. I am thankful for the trust they have put on me and on my work, for their encouragement, valuable guidance, and immense knowledge. They are not only my advisors but also life-long mentors to me. Their support enables me to go forward, even when everything seems dark. It was an honor for me to have the chance to work with such great people. I sincerely hope that I continue to work with them long into the future.

Second, I would like to thank the members of my thesis committee Prof. Nuray Güner and Assoc. Prof. Levent Akdeniz for spending valuable time on the material I prepared for my committee meetings, for challenging me with their visionary questions and comments throughout my dissertation research. I am also very grateful to Prof. Kürşat Aydoğan and Asst. Prof. İlkey Şendeniz Yüncü for accepting to serve as members of my dissertation examination committee, taking the time off their busy schedules to read this thesis, and for their valuable suggestions for future directions.

My special appreciation is reserved for Prof. Albert Menkveld, who gave me the opportunity to visit Vrije Universiteit Amsterdam. I genuinely enjoyed every moment I spent there and it turned out to be a quite productive period for my Ph.D. studies. I thank Prof. Menkveld for devoting his valuable time and resources to enrich my visit. I genuinely hope that I can be as much an inspiration to my students some days as he has been to me. Moreover, I would like to thank Asst.

Prof. Günseli Tümer Alkan for her kindness and support throughout my visit and many delightful conversations we had. I would like to also thank Shihao Yu and Ekaterina Serikova for their friendship during my visit.

Special thanks to the Faculty of Business Administration at Bilkent University for giving me a solid background and the opportunity to carry out my doctoral research. I am grateful to each faculty member but I would like to express my wholehearted gratitude to our graduate student director Assoc. Prof. Ahmet Ekici for taking many steps to make our lives better in every direction. In addition, I am particularly thankful to Prof. Güliz Ger for inspiring me to work harder and for her guidance towards life after Ph.D. I would like to also thank Assoc. Prof. Sühelya Özyıldırım, for her encouragement and kind words at critical moments. Moreover, I would like to thank the members of our administrative staff, Dilek Akın, Fulya Akmut, Zeliha Baran, İsmail Çetin and Remin Tantoğlu for making me feel like I am part of a large family and easing my life with their constant assistance.

I have had the chance to work with various scholars outside Bilkent FBA throughout my Ph.D. study. My heartfelt thanks go to Dr. Canan Güneş Çorlu, who introduced me to finance during my undergraduate years, for her kindness, patience, and support. I also thank her for being an incredible person to work with. Her guidance started the fire that led me to where I am today. I would like to also thank my co-authors, Dr. Melike Meterelliyoz Kuyzu and Dr. Syed Mahmud, for being excellent mentors and patiently guiding me through the publication process in the early years of my studies. I hope our collaboration will continue in the future.

I have met many wonderful people during my time at Bilkent. I would like to thank İdil Ayberk for being the best colleague one can ever have. I had so much fun working with İdil, since day one. Our joint struggle drew us together and created a

bond that will last for a lifetime. I also thank Ecem Cephe, John Omole, Shahid Mahmud, Burçak Baş, Mubin Memon whom I shared great memories and many cups of coffee within a tiny office space called MAZ24-A. My sincere thanks to Duygu Çelik, for being a great co-author, and for her patience and determination.

In addition, I would like to thank Alp Yıldırım, Mehmetcan Kılınc, Gülce Çakman, Güray Kara, Zeynep Bora, Müge Demir, Merve Demirel, Sabeeh Iqbal, Pelin Sümer, Ezgi Arslan Alp, Gülay Taltekin, Zeynep Baktır, Emre Bulut, Bengisu Sert, Okan Dükkancı, Nihal Berktaş, Kamyar Kargar, Burak Paç, Esra Koca, Hatice Çalık, Meltem Peker, Merve Meraklı and Ece Zeliha Demirci for their invaluable friendship. A special thanks to my dear friend Emrehan Şenöz, who helped me go over my codes many times when I thought I was at a dead end.

There are no words to express my love for my mom Süreyya Tiniç and my dad Sinan Tiniç. I am forever grateful for their constant, unconditional love and support, and for encouraging me to pursue my interests. I would not be here if it were not for my family.

For the last ten years, my life has been transformed thanks to a true genius; Dr. Gizem Özbaygın. Gizem is not only the Northstar that I follow when I am lost; but also my humble travel companion who is constantly serving as a ladder that enables me to reach beyond my potential. Her love gives me purpose and I am eternally grateful for her mere existence. I cannot be more happy to spent the rest of my life, my career and everything else I have with her. I know that, together, we can achieve anything we strive for.

I gratefully acknowledge the financial support provided by The Scientific and Technological Research Council of Turkey (TÜBİTAK) with grant number 116K335 for funding part of this thesis.

TABLE OF CONTENTS

ABSTRACT	iii
ÖZET	v
ACKNOWLEDGEMENTS	vii
TABLE OF CONTENTS	x
LIST OF TABLES	xiii
LIST OF FIGURES	xiv
CHAPTER I: INTRODUCTION	1
1.1 Overview	1
CHAPTER II: InfoTrad: An R package for estimating the probability of informed trading	4
2.1 Introduction	4
2.2 PIN Model	7
2.2.1 Floating-Point Exception	8
2.2.2 Boundary Solutions	9
2.2.3 Clustering Approach	11
2.3 InfoTrad package	12
2.3.1 EHO() function	14
2.3.2 LK() function	16
2.3.3 YZ() function	18
2.3.4 GAN() function	20
2.3.5 EA() function	23

2.4	Simulations and Performance Evaluation	25
2.5	Conclusion	28
CHAPTER III: Informed trading, order flow shocks and the cross section of expected returns in Borsa Istanbul		30
3.1	Introduction	30
3.2	PIN model and its extensions	33
3.2.1	Probability of Symmetric Order-Flow Shock (PSOS)	34
3.3	Sample Formation	36
3.4	Methodology	39
3.4.1	Fama-MacBeth Regressions	39
3.4.2	Univariate Portfolio Analysis	42
3.4.3	Multivariate Portfolio Analysis	43
3.4.4	Factor Tests	44
3.5	Results	44
3.5.1	Regression analysis	44
3.5.2	Univariate Portfolio Analysis	45
3.5.3	Multivariate Portfolio Analysis	51
3.6	Conclusion	55
CHAPTER IV: Political turmoil and the impact of foreign orders on equity prices		57
4.1	Introduction	57
4.2	Data	63
4.3	Methodology	75
4.3.1	Estimating the price impact of foreign trades	75
4.3.2	Panel Regressions	79
4.4	Results	82
4.4.1	Average price impacts	82

4.4.2 Panel Regressions	86
4.5 Conclusion	91
CHAPTER V: CONCLUSIONS	94
REFERENCES	98
APPENDIX: Variable Definitions	105



LIST OF TABLES

2.1	Simulation results to examine the relationship between trade intensity and the estimates of PIN.	27
2.2	Simulation results to examine the performance of different methods in estimating the parameters for PIN	29
3.1	Descriptive statistics on number of buy and sell orders.	38
3.2	Descriptive statistics on the cross sectional distributions of monthly factors.	40
3.3	Factor Correlations	41
3.4	Fama-MacBeth Regressions	46
3.5	Descriptive Statistics for PIN portfolios-1	48
3.6	Return characteristics of univariate PIN portfolios	49
3.7	Characteristics and Returns of PIN-Size Portfolios	53
3.8	Fama-French Three Factor Regressions	54
4.1	Descriptive Statistics for firm specific factors	67
4.2	Factor Correlations	68
4.4	Full sample estimates for price impact	84
4.5	List of stocks that foreigners have significant price impact	85
4.6	Descriptive Statistics for the monthly coefficients	88
4.7	Panel regression results	89

LIST OF FIGURES

4.1	Time Series Variation in Daily Volume of Foreign and Domestic Net Flows	70
4.2	Time Series Variation in Daily Foreign Participation	71
4.3	Bloomberg Country Risk Political Score for Turkey	74

CHAPTER I

INTRODUCTION

1.1. Overview

This study focuses on the impact of information asymmetry on the asset prices in Borsa Istanbul. In recent years, the number of studies that examine the market microstructure of Borsa Istanbul increased significantly. Throughout this research, I employed a variety of tools from the field of market microstructure and empirical asset pricing to contribute to this growing literature.

In the next chapter of this thesis, we provide a survey of the existing literature and introduce an R package called **InfoTrad** for estimating the probability of informed trading (PIN). PIN is one of the most prominently used information asymmetry measures in the market microstructure literature. Estimates, obtained through maximum likelihood estimation, are prone to bias due to several problems. First, early studies using PIN were vulnerable to sample selection bias due to floating point exception in estimating the measure. Second, the estimates are susceptible to potential bias due to boundary solutions obtained in the optimization process.

InfoTrad is a platform that combines the alternative factorizations and methods that deal with the problems due to floating point exception along with issues regarding the boundary solutions. Moreover, **InfoTrad** provides recently proposed clustering algorithms that ensure fast and reliable estimates for the PIN.

In the third chapter, we examine the relationship between the information asymmetry, proxied by PIN, and the stock returns in Borsa Istanbul. To the best of our knowledge, this study is the first to examine the systematic impact of information asymmetry on returns of stocks that are traded on Borsa Istanbul. There are conflicting views on the relationship between information risk and securities' returns. One side of the argument asserts that uninformed investors are inferior in adjusting their portfolios quickly, in the presence of new information, and therefore require a premium for stocks that they are at a disadvantage. On the other hand, other studies argue that in a large economy where the number of securities is infinite, increase in the information asymmetry yields an increase in returns for the overall market and therefore, after controlling for systematic factors, information risk is idiosyncratic and fully diversifiable. We find an economically *insignificant* relationship between PIN and future returns of Turkish stocks over March 2005 and April 2017. Moreover, we provide univariate and multivariate portfolio analyses which suggest that investors who hold shares with high information asymmetry do not obtain significant future returns. Therefore, we conclude that information asymmetry is a firm-specific risk in Borsa Istanbul, which can be eliminated through portfolio diversification.

In the last chapter, we compare the informational (dis)advantage of foreign investors across all stocks that are traded in Borsa Istanbul between 2007 and 2015. Foreigners dominate the Turkish stock market, holding almost 65 percent of the publicly-traded shares in Borsa Istanbul. Therefore, examining the foreign investors' role is crucial for understanding the price discovery process in emerging economies like Turkey, which rely heavily on outside financing. To this end, we test the baseline hypothesis that local investors have no informational advantage over foreign investors. Our findings indicate that domestic investors have a significant price impact over 43 percent of the stocks in our sample whereas the price impact

of foreign investors is statistically significant for only seven percent of the entire cross-section. More importantly, in this study, we examine if and how the increased political uncertainty in Turkey affects the price impact of local and foreign investors. We argue that, due to inherent differences in domestic and international investors' exposure to Turkey-specific political uncertainty, the price impact of these investor groups is sensitive to the underlying political environment. Our subsample analysis indicates that foreign impact increases during times of the political turmoil, which started with the Gezi protests in June 2013. Our panel regression analysis also confirms these findings and suggests that the adverse selection component of the foreign trades significantly increases following June 2013, even after controlling for firm-specific and macroeconomic factors. Our results indicate that though foreigners on average are at a disadvantage in Borsa Istanbul, political uncertainty partially alleviates this disadvantage and creates an opportunity for foreign investors to exert price impact over a larger number of securities. This research contributes to a broad literature that examines the impact of foreign investor flows in equity markets which have grown significantly following the capital market liberalization wave of the 1990s. To the best of our knowledge, it is the first market microstructure study that explores the effect of political instability on the relative price impact of foreign and domestic investors.

CHAPTER II

INFOTRAD: AN R PACKAGE FOR ESTIMATING THE PROBABILITY OF INFORMED TRADING

2.1. Introduction

The main aim of this chapter is to present the **InfoTrad** package that estimates the probability of informed trading (PIN) initially proposed by Easley et al. (1996). PIN is one of the primary measures that proxy the information asymmetry in the market. The sequential model is driven from maximum likelihood estimation (MLE). Wide range of studies use PIN to answer questions in different fields of finance¹.

Although it is a heavily used measure in the finance literature, the development of applications that calculate PIN are quite slow. An initial attempt for **R** community is made by Zagaglia (2012). **FinAsym** package of Zagaglia (2012) and the **PIN** package of Zagaglia (2013) provide the trade classification algorithm of Lee and Ready (1991) which is an important tool for studies that use databases where a flag to denote the direction of the trade is not provided (e.g, TAQ database). Both

¹For instance, analyst coverage (Easley et al., 1998), stock splits (Easley et al., 2001), initial public offerings Ellul and Pagano (2006), credit ratings (Odders-White and Ready, 2006), M&A announcements (Aktas et al., 2007) and asset returns (Easley et al., 2002),(Easley et al., 2010).

packages also provide PIN estimates through *pin_likelihood()* functions. However, those estimates are prone to bias due to misspecification and other limitations.

InfoTrad package aims to overcome such limitations and provide users with a wide range of options when estimating PIN.

Due to the popularity of the measure, problems in estimating PIN recently gained attention in the finance literature. Easley et al. (2010) indicate that for stocks with a large trading volume, it is not possible to estimate the PIN due to floating-point-exception (FPE). Two different numerical factorizations are provided by Easley et al. (2010) and Lin and Ke (2011) to overcome the bias created due to FPE.

In addition, boundary solutions in estimating PIN are also shown to create bias in empirical studies. Yan and Zhang (2012) show that, independent of the type of factorization, the likelihood function can stuck at local optimum and provide bias in PIN estimates. They propose an algorithm (YZ algorithm) that spans the parameter space using 125 different initial values for the MLE problem and obtain the estimate that gives the highest likelihood value with non-boundary solutions. Although YZ algorithm provides estimates with higher likelihood and guarantees to obtain non-boundary solutions, the iterative structure makes this algorithm time-consuming especially for studies that use large datasets.

Considering the fact that recent studies that estimate PIN use large datasets, the effectiveness of the YZ algorithm is questioned. In recent years, clustering algorithms have become popular due to their efficiency in processing large sets of data. Gan et al. (2015) propose an algorithm that use hierarchical agglomerative clustering to estimate PIN. Ersan and Alici (2016) later extends this framework.

FPE and boundary solutions are *not* the only problems of PIN model. Duarte and Young (2009) indicate that the structural model of Easley et al. (1996) enforces a

negative contemporaneous covariance between intraday buy and sell orders, which is contrary to the empirical evidence for symmetric order flow shocks. In addition, they show that the PIN model fails to capture the difference in the volatility of buy and sell orders, through simulations. Moreover, Duarte and Young (2009) adjust PIN to take into account the liquidity impact and show that liquidity is more prominent on stock returns compared to information asymmetry. Finally it is important to note that PIN does not consider any strategic behaviour of investors such as order splitting. Order splitting can be more evident when a stock is jointly trading on multiple venues (Menkveld, 2008). Even for a stock that is traded on a single market, an informed investors may want to split her order to avoid revealing her private information too quickly (Foucault et al., 2013). PIN model, by construction, fails to attach multiple small orders to a single informed investor.

This chapter introduces and discusses the **R** R Core Team (2016) **InfoTrad** package for estimating PIN. **InfoTrad** provides users with necessary methods to *solely* address the problems of FPE and boundary solutions². The package contains the likelihood factorizations of Easley et al. (2010) and Lin and Ke (2011) as separate functions (*EHO()* and *LK()*, respectively) which provide likelihood specifications to avoid FPE. In addition, through *YZ()*, *GAN()* and *EA()* functions, PIN estimates can be obtained using the grid-search algorithm of Yan and Zhang (2012), and clustering algorithms of Gan et al. (2015) and Ersan and Alici (2016). For all algorithms, likelihood specification can be set to EHO or LK.

The chapter is organized as follows; Section 2.2 provides a brief description of PIN. Specifically, section 2.2.1 discusses the problem of FPE and the alternative factorizations EHO and LK. Section 2.2.2 reviews the problem of boundary solutions and the YZ algorithm. Section 2.2.3 describes the clustering algorithms of Gan et al. (2015) and Ersan and Alici (2016). Section 2.3 introduces the package

²As of version 1.2.

InfoTrad along with examples. Section 2.4 evaluates the performance of each method through simulations. Section 2.5 provides concluding remarks.

2.2. PIN Model

The structural model of Easley et al. (1996) and Easley et al. (2002) consists of three types of agents; informed traders, uninformed traders and market makers. On a trading day t , one risky asset is continuously traded. Market maker sets the price for a given stock by observing the buy orders (B_t) and sell orders (S_t). For that stock, an information event is assumed to follow a Bernoulli distribution with success probability α . This event reveals either a good or a bad signal for the underlying value of the stock. The event is assumed to provide a bad signal with probability δ . When informed traders observe a good (bad) signal, they are assumed to place buy (sell) orders at a rate of μ . Uninformed traders are assumed to place orders, independent of the information event or the signal. They arrive to market to place a buy (sell) order at a rate of $\epsilon_b(\epsilon_s)$. Orders of both informed and uninformed investors are assumed to follow independent Poisson processes.

The joint probability distribution with respect to the parameter vector $\Theta = \{\alpha, \delta, \mu, \epsilon_b, \epsilon_s\}$ and the number of buys and sells (B_t, S_t), is specified by

$$\begin{aligned}
 f(B_t, S_t | \Theta) = & \alpha \delta \exp(-\epsilon_b) \frac{\epsilon_b^{B_t}}{B_t!} \exp[-(\epsilon_s + \mu)] \frac{(\epsilon_s + \mu)^{S_t}}{S_t!} \\
 & + \alpha(1 - \delta) \exp[-(\epsilon_b + \mu)] \frac{-(\epsilon_b + \mu)^{B_t}}{B_t!} \exp(-\epsilon_s) \frac{\epsilon_s^{S_t}}{S_t!} \\
 & + (1 - \alpha) \exp(-\epsilon_b) \frac{\epsilon_b^{B_t}}{B_t!} \exp(-\epsilon_s) \frac{\epsilon_s^{S_t}}{S_t!}
 \end{aligned} \tag{2.1}$$

The estimates of arrival rates $\{\hat{\mu}, \hat{\epsilon}_b, \hat{\epsilon}_s\}$, along with the estimates of the probabilities, $\{\hat{\alpha}, \hat{\delta}\}$ can be obtained by maximizing the joint log-likelihood

function given the order input matrix (B_t, S_t) over T trading days. The non-linear objective function of this problem can be written as;

$$L(\Theta|T) \equiv \sum_{t=1}^T L(\Theta|(B_t, S_t)) = \sum_{t=1}^T \log(f(B_t, S_t|\Theta)) \quad (2.2)$$

The maximization problem is subject to the boundary constraints $\alpha, \delta \in [0, 1]$ and $\mu, \epsilon_b, \epsilon_s \in [0, \infty)$ ³. The PIN estimate is then given by;

$$\widehat{PIN} = \frac{\hat{\alpha}\hat{\mu}}{\hat{\alpha}\hat{\mu} + \hat{\epsilon}_b + \hat{\epsilon}_s} \quad (2.3)$$

2.2.1 Floating-Point Exception

PIN estimates are prone to selection bias, especially for stocks for which the number of buy and sell orders are large⁴. Lin and Ke (2011) show that the increase in the number of buy and sell orders for a given stock, significantly shrinks the feasible solution set for the maximization of the log-likelihood function in equation (2.2). To maximize the non-linear function (2.1), the optimization software introduces initial values for Θ . The numerical optimization method is applied after those initial parameters are introduced. Therefore, for large enough B_t and S_t , whose factorials cannot be calculated by mainstream computers (i.e. FPE), the optimal value for equation (2.2) becomes undefined. The FPE problem is therefore, more pronounced in active stocks.

To avoid the selection bias created due to FPE, one factorization of the equation

³Both **PIN** package of Zagaglia (2013) and **FinAsym** package of Zagaglia (2012) fail to acknowledge the boundary constraints on the arrival rates, $\mu, \epsilon_b, \epsilon_s$. Similar to event probabilities, they restrict these parameters to $[0,1]$ which forces the estimates for the arrival of informed and uninformed traders on a given day to take values at most one. This creates significant bias in PIN estimates.

⁴For example, Zagaglia (2012) provides a sample data to calculate PIN. In sample data, the maximum trade number is 19. If you multiply each observation in the sample data by 10, the *pin_likelihood()* function of **FinAsym** package fails to provide results with the sample initial parameter vector.

(2.2) is provided by Easley et al. (2010) as $L_{EHO}(\Theta|T) \equiv \sum_{t=1}^T L_{EHO}(\Theta|B_t, S_t)$ which is given by;

$$\begin{aligned} L_{EHO}(B_t, S_t) = & \log \left[\alpha \delta \exp(-\mu) x_b^{B_t - M_t} x_s^{M_t} + \alpha(1 - \delta) \exp(-\mu) x_b^{-M_t} x_s^{S_t - M_t} \right. \\ & \left. + (1 - \alpha) x_b^{B_t - M_t} x_s^{S_t - M_t} \right] + B_t \log(\epsilon_b + \mu) + S_t \log(\epsilon_s + \mu) \\ & - (\epsilon_b + \epsilon_s) + M_t [\log(x_b) + \log(x_s)] - \log(S_t! B_t!) \end{aligned} \quad (2.4)$$

where $M_t = \min(B_t, S_t) + \max(B_t, S_t)/2$, $x_b = \epsilon_b/(\mu + \epsilon_b)$ and $x_s = \epsilon_s/(\mu + \epsilon_s)$.

Lin and Ke (2011) introduce another algebraically equivalent factorization of the equation (2.1), $L_{LK}(\Theta|T) \equiv \sum_{t=1}^T L_{LK}(\Theta|B_t, S_t)$ which is given by;

$$\begin{aligned} L_{LK}(\Theta|B_t, S_t) = & \log \left[\alpha \delta \exp(e_{1t} - e_{\max t}) + \alpha(1 - \delta) \exp(e_{2t} - e_{\max t}) \right. \\ & \left. + (1 - \alpha) \exp(e_{3t} - e_{\max t}) \right] + B_t \log(\epsilon_b + \mu) + S_t \log(\epsilon_s + \mu) \\ & - (\epsilon_b + \epsilon_s) + e_{\max t} - \log(B_t! S_t!) \end{aligned} \quad (2.5)$$

where $e_{1t} = -\mu - B_t \log(1 + \mu/\epsilon_b)$, $e_{2t} = -\mu - S_t \log(1 + \mu/\epsilon_s)$, $e_{3t} = -B_t \log(1 + \mu/\epsilon_b) - S_t \log(1 + \mu/\epsilon_s)$ and $e_{\max t} = \max(e_{1t}, e_{2t}, e_{3t})$. The last term $\log(B_t! S_t!)$ is constant with respect to the parameter vector Θ , and is, therefore, dropped in the MLE for both factorizations.

2.2.2 Boundary Solutions

Another source of bias in estimating PIN arises from boundary solutions. Yan and Zhang (2012) indicate that in calculating PIN, parameter estimates $\hat{\alpha}$ and $\hat{\delta}$ usually fall onto the boundaries of the parameter space, that is, they are equal to zero or

one. PIN estimate presented in equation (2.3) is directly related to the estimate of $\hat{\alpha}$. Letting $\hat{\alpha}$ equal to zero will make sure that PIN is zero as well. This can create a sample selection bias in portfolio formation especially for quarterly estimations⁵. Yan and Zhang (2012) show that;

$$E(B) = \alpha(1 - \delta)\mu + \epsilon_b \quad (2.6)$$

$$E(S) = \alpha\delta\mu + \epsilon_s \quad (2.7)$$

Then they propose the following algorithm to overcome the bias created due to boundary solutions. Let $(\alpha^0, \delta^0, \epsilon_b^0, \epsilon_s^0, \mu^0)$ be the initial parameter function to be placed in the non-linear program presented in equation (2.2). In addition, let $\bar{B}$ and $\bar{S}$ be the average number of buy and sell orders. Then, set

$$\alpha^0 = \alpha_i, \quad \delta^0 = \delta_j, \quad \epsilon_b^0 = \gamma_k \bar{B}, \quad \mu^0 = \frac{\bar{B} - \epsilon_b^0}{\alpha^0(1 - \delta^0)}, \quad \text{and } \epsilon_s = \bar{S} - \alpha^0 \delta^0 \mu^0 \quad (2.8)$$

where $\alpha_i, \delta_j, \gamma_k \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. This procedure yields 125 different PIN estimates along with their likelihood values. In line with Yan and Zhang (2012), we drop any initial parameter vector having negative values for ϵ_s^0 . In addition, following Ersan and Alici (2016), we also drop any initial parameter vector with $\mu_0 > \max(B_t, S_t)$. Yan and Zhang (2012) then select the estimate with non-boundary parameters yielding the highest likelihood value. This method, by construction, spans the parameter space and tries to avoid local optima and provides non-boundary estimates for α .

⁵For quarterly estimations of PIN, one can be sure that there is at least one information event: earnings announcements. Therefore, $\hat{\alpha}$ cannot be equal to zero.

2.2.3 Clustering Approach

In recent years, clustering algorithms are increasingly becoming popular in estimating the probability of informed trading due to efficiency concerns. Gan et al. (2015) and Ersan and Alici (2016) use clustering algorithms to estimate PIN. Gan et al. (2015) introduce a method that clusters the data into three groups (good news, bad news, and no news) based on the mean absolute difference in order imbalance. Let $X_t = B_t - S_t$ be the order imbalance on day t computed as the difference between buy and sell orders. The clustering is then based on the distance function defined as $D(I, J) = |X_i - X_j|$, $1 \leq i, j \leq T$ where $i \neq j$. They use hierarchical agglomerative clustering (HAC) to group the data elements based on the distance matrix. Specifically, they use *hclust()* function of Müllner (2013) in R⁶. The algorithm sequentially clusters, in a bottom-up fashion, each observation into groups based on X_t and stops when it reaches three clusters. The theoretical framework of Easley et al. (1996) indicates that a stock has high (low) X_t on good (bad) days. Therefore, the cluster which has the highest (lowest) mean X_t is labelled as good (bad) news cluster. The remaining cluster is then labelled as no news. Once each observation is grouped into their respective clusters (good news, bad news, no news), $c \in \{G, B, N\}$, the parameter estimates for $\Theta \equiv \{\alpha, \delta, \mu, \epsilon_b, \epsilon_s\}$ are calculated simply by counting. Let ω_c be the proportion of cluster c occupying the total number of days T , such that $\sum_{c=1}^3 \omega_c = 1$. Similarly, let $\bar{B}_c$ and $\bar{S}_c$ be the average number of buys and sells on cluster c , respectively.

Then, the probability of information event is given by $\hat{\alpha} = \omega_B + \omega_G$. Moreover, the estimate for the probability of information event releasing a bad new is given by $\hat{\delta} = \omega_B / \hat{\alpha}$. The estimate for the arrival rate of buy orders of uninformed traders is represented by $\hat{\epsilon}_b = \frac{\omega_B}{\omega_B + \omega_N} \bar{B}_B + \frac{\omega_N}{\omega_B + \omega_N} \bar{B}_N$. Similarly, the estimate for the arrival

⁶We also use *hclust()* at its default setting in line with Gan et al. (2015).

rate of sell orders of uninformed traders is represented by

$\hat{\epsilon}_s = \frac{\omega_G}{\omega_G + \omega_N} \bar{S}_G + \frac{\omega_N}{\omega_G + \omega_N} \bar{S}_N$. Finally, the arrival rate for the informed investors is calculated as $\hat{\mu} = \frac{\omega_G}{\omega_B + \omega_G} (\bar{B}_G - \hat{\epsilon}_b) + \frac{\omega_B}{\omega_B + \omega_G} (\bar{S}_B - \hat{\epsilon}_s)$ where $(\bar{B}_G - \hat{\epsilon}_b)$ corresponds to the buy rate of informed investors $\hat{\mu}_b$ and $(\bar{S}_B - \hat{\epsilon}_s)$ corresponds to the sell rate of informed investors $\hat{\mu}_s$ ⁷.

Through simulations, Gan et al. (2015) show that estimates calculated as above are proper candidates for the initial parameter values to be used in the MLE process.

Ersan and Alici (2016) argue that the estimates for the informed arrival rate, μ , contains a downward bias with GAN algorithm⁸. In addition, they state that the GAN algorithm provides inaccurate estimates for δ . In order to overcome these issues, instead of using X_t , Ersan and Alici use absolute daily order imbalance, $|X_t|$, to cluster the data. They initially cluster, $|X_t|$ into two, again by using *hclust()*. The cluster with lower mean daily absolute order imbalance is labelled as “no event” cluster. Then, the formation of “good” and “bad” event day clusters are obtained through separating the days in the “event” cluster into two with respect to the *sign* of the daily order imbalances. The parameters estimates are then computed with the same procedure presented above⁹.

2.3. InfoTrad package

The R package **InfoTrad** provides five different functions *EHO()*, *LK()*, *YZ()*, *GAN()* and *EA()*. The first two functions provide likelihood specifications whereas the last three functions can be used to obtain the parameter estimates for Θ to

⁷Both Gan et al. (2015) and Ersan and Alici (2016) do not mention the case where $\hat{\mu}_b < 0$ (or $\hat{\mu}_s < 0$). It is fair to assume that in such cases informed investors are not present on the buy (sell) side. Therefore, we set μ_b, μ_s equal to zero when we obtain a negative estimate.

⁸We also show that estimates for μ contains a significant downward bias due to poor choice of initial parameter value μ^0 when GAN algorithm is used.

⁹Ersan and Alici (2016) also provide an iterative process in which they systematically update the clusters. We plan to introduce this methodology in the future versions of our package.

calculate the PIN in equation (2.3). All five functions require a data frame that contains B_t in the first column, and S_t in the second column. We create B_t and S_t for ten hypothetical trading days¹⁰. $EHO()$ and $LK()$ read (B_t, S_t) and return the related functional form of the negative log-likelihood. These objects can be used in any optimization procedure such as $optim()$ to obtain the parameter estimates $\hat{\Theta} \equiv \{\hat{\alpha}, \hat{\delta}, \hat{\mu}, \hat{\epsilon}_b, \hat{\epsilon}_s\}$, the likelihood value and other specifications, in one iteration with a pre-specified initial value vector, Θ_0 , for parameters. We define $EHO()$ and $LK()$ as simple likelihood specifications rather than functions that execute the MLE procedure. This is due to the fact that MLE estimators vary depending on the optimization procedure. Users who wish to develop alternative estimation techniques, based on the proposed likelihood factorization, can use $EHO()$ and $LK()$. This is the underlying reason why those function do not have built-in optimization procedures. By specifying $EHO()$ and $LK()$ as simple likelihood functions, we give developers the flexibility to select the most suitable optimization procedure for their application.

For researchers who want to calculate an estimate of PIN, $YZ()$, $GAN()$ and $EA()$ functions have built-in optimization procedures. Those functions read a likelihood specification value along with data. likelihood specifications can be set either to "LK" or to "EHO" with "LK" being the default. All estimation funtions use $neldermead()$ function of the **nloptr** package to conduct MLE with the specified factorization. $GAN()$ and $EA()$ functions also use $hclust()$ function of Müllner (2013) to conduct clustering. The output of these three functions is an object that provides $\{\hat{\alpha}, \hat{\delta}, \hat{\mu}, \hat{\epsilon}_b, \hat{\epsilon}_s, f(\hat{\Theta}), \widehat{PIN}\}$ where $f(\hat{\Theta})$ represents the optimal likelihood value given the parameter estimates $\hat{\Theta}$.

¹⁰The numbers are randomly selected. We set the numbers to be high enough so that the original likelihood framework presented in equation (2.1) cannot be used due to FPE. Easley et al. (1996) indicate that at least 60 days worth of data is required in order to obtain proper convergence for $\widehat{PIN}$. We use ten days for demonstration purposes.

2.3.1 EHO() function

An example is provided below for *EHO()* with a sample data and initial parameter values. Notice that the first column of sample data is for B_t and the second column is for S_t . Similarly, the initial parameter values are constructed as;

$\Theta = \{\alpha, \delta, \mu, \epsilon_b, \epsilon_s\}$. We use *optim()* with “Nelder-Mead” method to execute MLE, however the developer is flexible to use other methods as well.

```
library(InfoTrad)

# Sample Data
#   Buy Sell
#1  350  382
#2  250  500
#3  500  463
#4  552  550
#5  163  200
#6  345  323
#7  847  456
#8  923  342
#9  123  578
#10 349  455

Buy<-c(350,250,500,552,163,345,847,923,123,349)
Sell<-c(382,500,463,550,200,323,456,342,578,455)
data=cbind(Buy,Sell)

# Initial parameter values
# par0 = (alpha, delta, mu, epsilon_b, epsilon_s)
```

```

par0 = c(0.5,0.5,300,400,500)

# Call EHO function
EHO_out = EHO(data)
model = optim(par0, EHO_out, gr = NULL,
              method = c("Nelder-Mead"), hessian = FALSE)

## Parameter Estimates
model$par[1] # Estimate for alpha
# [1] 0.9111102
model$par[2] # Estimate for delta
# [1] 0.0001231429
model$par[3] # Estimate for mu
# [1] 417.1497
model$par[4] # Estimate for eb
# [1] 336.075
model$par[5] # Estimate for es
# [1] 466.2539

## Estimate for PIN
(model$par[1]*model$par[3])/
  ((model$par[1]*model$par[3])+model$par[4]+model$par[5])
# [1] 0.3214394
####

```

In this example, B_t and S_t vectors are selected so that the likelihood function cannot be represented as in equation (2.1). We set the initial parameters to be $\Theta_0 = \{0.5, 0.5, 300, 400, 500\}$. For the given (B_t, S_t) and Θ_0 , the PIN measure is

calculated as 0.32 with EHO factorization.

2.3.2 LK() function

An example is provided below for $LK()$ with a sample data and initial parameter values. Notice that the first column of sample data is for B_t and the second column is for S_t . Similarly, the initial parameter values are constructed as;

$\Theta = \{\alpha, \delta, \mu, \epsilon_b, \epsilon_s\}$. We use *optim()* with “Nelder-Mead” method to execute MLE, however the developer is flexible to use other methods as well.

```
library(InfoTrad)

# Sample Data
#   Buy Sell
#1  350  382
#2  250  500
#3  500  463
#4  552  550
#5  163  200
#6  345  323
#7  847  456
#8  923  342
#9  123  578
#10 349  455

Buy<-c(350,250,500,552,163,345,847,923,123,349)
Sell<-c(382,500,463,550,200,323,456,342,578,455)
data=cbind(Buy,Sell)

# Initial parameter values
```

```

# par0 = (alpha, delta, mu, epsilon_b, epsilon_s)
par0 = c(0.5,0.5,300,400,500)

# Call LK function
LK_out = LK(data)
model = optim(par0, LK_out, gr = NULL,
              method = c("Nelder-Mead"), hessian = FALSE)

## The structure of the model output ##
model

#$par
#[1] 0.480277 0.830850 315.259805 296.862318 400.490830

#$value
#[1] -44343.21

#$counts
#function gradient
# 502 NA

#$convergence
#[1] 1

#$message
#NULL

```

```
## Parameter Estimates

model$par[1] # Estimate for alpha
# [1] 0.480277

model$par[2] # Estimate for delta
# [1] 0.830850

model$par[3] # Estimate for mu
# [1] 315.259805

model$par[4] # Estimate for eb
# [1] 296.862318

model$par[5] # Estimate for es
# [1] 400.4908


## Estimate for PIN

(model$par[1]*model$par[3])/

((model$par[1]*model$par[3])+model$par[4]+model$par[5])

# [1] 0.178391

####
```

For the given (B_t, S_t) and Θ_0 , the PIN measure is calculated as 0.18 with LK factorization.

2.3.3 YZ() function

An example is provided below for $YZ()$ function with a sample data. Notice that the first column of sample data is for B_t and second column is for S_t . In addition, the first example is with default likelihood specification LK and the second one is with EHO. Notice that $YZ()$ function do not require any initial parameter vector Θ_0 .

```
library(InfoTrad)
```

```

# Sample Data
#   Buy Sell
#1  350  382
#2  250  500
#3  500  463
#4  552  550
#5  163  200
#6  345  323
#7  847  456
#8  923  342
#9  123  578
#10 349  455

Buy<-c(350,250,500,552,163,345,847,923,123,349)
Sell<-c(382,500,463,550,200,323,456,342,578,455)
data<-cbind(Buy,Sell)

# Parameter estimates using the LK factorization of Lin and Ke (2011)
# with the algorithm of Yan and Zhang (2012).
# Default factorization is set to be "LK"

result=YZ(data)
print(result)

# Alpha: 0.3999999
# Delta: 0
# Mu: 442.1667

```

```

# Epsilon_b: 263.3333
# Epsilon_s: 424.9
# Likelihood Value: 44371.84
# PIN: 0.2004457

# Parameter estimates using the EHO factorization of Easley et. al. (2010)
# with the algorithm of Yan and Zhang (2012).

result=YZ(data,likelihood="EHO")
print(result)

# Alpha: 0.9000001
# Delta: 0.9000001
# Mu: 489.1111
# Epsilon_b: 396.1803
# Epsilon_s: 28.72002
# Likelihood Value: Inf
# PIN: 0.3321033

```

For the given B_t and S_t vectors, PIN measure is calculated as 0.20 with YZ algorithm along with LK factorization. Moreover, PIN measure is calculated as 0.33 with YZ algorithm along with EHO factorization.

2.3.4 GAN() function

An example is provided below for $GAN()$ function with a sample data. Notice that the first column of sample data is for B_t and second column is for S_t . In addition, the first example is with default likelihood specification LK and the second one is with EHO. Notice that $GAN()$ function do not require any initial parameter vector

Θ_0 .

```
library(InfoTrad)
```

```
# Sample Data
```

```
#   Buy Sell
```

```
#1  350  382
```

```
#2  250  500
```

```
#3  500  463
```

```
#4  552  550
```

```
#5  163  200
```

```
#6  345  323
```

```
#7  847  456
```

```
#8  923  342
```

```
#9  123  578
```

```
#10 349  455
```

```
Buy<-c(350,250,500,552,163,345,847,923,123,349)
```

```
Sell<-c(382,500,463,550,200,323,456,342,578,455)
```

```
data<-cbind(Buy,Sell)
```

```
# Parameter estimates using the LK factorization of Lin and Ke (2011)
```

```
# with the algorithm of Gan et. al. (2015).
```

```
# Default factorization is set to be "LK"
```

```
result=GAN(data)
```

```
print(result)
```

```
# Alpha: 0.3999998
```

```

# Delta: 0
# Mu: 442.1667
# Epsilon_b: 263.3333
# Epsilon_s: 424.9
# Likelihood Value: 44371.84
# PIN: 0.2044464

# Parameter estimates using the EHO factorization of Easley et. al. (2010)
# with the algorithm of Gan et. al. (2015)

result=GAN(data, likelihood="EHO")
print(result)

# Alpha: 0.3230001
# Delta: 0.4780001
# Mu: 481.3526
# Epsilon_b: 356.6359
# Epsilon_s: 313.136
# Likelihood Value: Inf
# PIN: 0.1884001

```

For the given B_t and S_t vectors, PIN measure is calculated as 0.20 with GAN algorithm along with LK factorization. Moreover, PIN measure is calculated as 0.19 with GAN algorithm along with EHO factorization.

2.3.5 EA() function

An example is provided below for $EA()$ function with a sample data. Notice that the first column of sample data is for B_t and second column is for S_t . In addition, the first example is with default likelihood specification LK and the second one is with EHO. Notice that $EA()$ function do not require any initial parameter vector Θ_0 .

```
library(InfoTrad)

# Sample Data

#   Buy Sell
#1  350  382
#2  250  500
#3  500  463
#4  552  550
#5  163  200
#6  345  323
#7  847  456
#8  923  342
#9  123  578
#10 349  455

Buy=c(350,250,500,552,163,345,847,923,123,349)
Sell=c(382,500,463,550,200,323,456,342,578,455)
data=cbind(Buy,Sell)

# Parameter estimates using the LK factorization of Lin and Ke (2011)
# with the modified clustering algorithm of Ersan and Alici (2016).
```

```
# Default factorization is set to be "LK"

result=EA(data)
print(result)

# Alpha: 0.9511418
# Delta: 0.2694005
# Mu: 76.7224
# Epsilon_b: 493.7045
# Epsilon_s: 377.4877
# Likelihood Value: 43973.71
# PIN: 0.07728924

# Parameter estimates using the EHO factorization of Easley et. al. (2010)
# with the modified clustering algorithm of Ersan and Alici (2016).

result=EA(data,likelihood="EHO")
print(result)

# Alpha: 0.9511418
# Delta: 0.2694005
# Mu: 76.7224
# Epsilon_b: 493.7045
# Epsilon_s: 377.4877
# Likelihood Value: 43973.71
# PIN: 0.07728924
```

For the given B_t and S_t vectors, PIN measure is calculated as 0.08 with EA algorithm along with LK factorization. Moreover, PIN measure is calculated, again, as 0.08 with EA algorithm along with EHO factorization.

2.4. Simulations and Performance Evaluation

In this section, we investigate the performance of the estimates obtained for Θ and PIN using the existing methods. We evaluate the methods based on their accuracy proxied by mean absolute errors (MAE)¹¹. We first examine how the estimates vary under different trade intensity levels. To this end, we follow the methodology proposed in Gan et al. (2015). Let I be the set of trade intensity levels ranging from 50 to 5000 at a step size of 50, that is, $I = \{50, 100, 150, \dots, 5000\}$. We first set our parameters as $\Theta = \{\alpha = 0.5, \delta = 0.5, \mu = 0.2i, \epsilon_b = 0.4i, \epsilon_s = 0.4i\}$, where $i \in I$. For each trade intensity level, we generate $N=50$ random samples of $\tilde{\alpha}$ and $\tilde{\delta}$ that are binomially distributed with parameters α and δ , respectively. $\tilde{\alpha}$ and $\tilde{\delta}$ proxy the content of the information event. For each pair of $\tilde{\alpha}$, $\tilde{\delta}$ values, we generate buy and sell values (B_t, S_t) for hypothetical $T = 60$ days in the following manner;

- if $\tilde{\alpha} = 0$, then there is no information event, therefore, generate $B_t \sim Pois(\epsilon_b)$ and $S_t \sim Pois(\epsilon_s)$.
- if $\tilde{\alpha} = 1$, and $\tilde{\delta} = 1$, then there is bad news, therefore generate $B_t \sim Pois(\epsilon_b)$ and $S_t \sim Pois(\epsilon_s + \mu)$
- if $\tilde{\alpha} = 1$, and $\tilde{\delta} = 0$, then there is good news, therefore generate $B_t \sim Pois(\epsilon_b + \mu)$ and $S_t \sim Pois(\epsilon_s)$

We then form the joint log-likelihood function represented by equation (2.4) in EHO form or by equation (2.5) in LK form and obtain the estimates using $YZ()$,

¹¹All estimations are conducted on a 2.6 Intel i7-6700HQ CPU. We do not consider speed as a performance measure since the average processing time for each method is less than 10 seconds.

$GAN()$ and $EA()$ methods. The psuedo code for the first simulation method is given as follows;

Algorithm 1 Simulations under different trade intensity levels

```

1: Set  $\alpha = 0.5$ ,  $\delta = 0.5$ ,  $N = 50$ ,  $T = 60$ ;
2: for  $i \in I$  do
3:   Set  $\mu = 0.2i$ ,  $\epsilon_b = 0.4i$ ,  $\epsilon_s = 0.4i$ ;
4:   Generate  $N$  observations of  $\tilde{\alpha} \sim \text{Binom}(\alpha)$ ;
5:   Generate  $N$  observations of  $\tilde{\delta} \sim \text{Binom}(\delta)$ ;
6:   for  $n \in N$  do
7:     if  $\tilde{\alpha}[n] == 0$  then
8:       Generate  $T$  observations of  $B_t \sim \text{Pois}(\epsilon_b)$ ;
9:       Generate  $T$  observations of  $S_t \sim \text{Pois}(\epsilon_s)$ ;
10:    else if  $\tilde{\alpha}[n] == 1$  and  $\tilde{\delta}[n] == 1$  then
11:      Generate  $T$  observations of  $B_t \sim \text{Pois}(\epsilon_b)$ ;
12:      Generate  $T$  observations of  $S_t \sim \text{Pois}(\epsilon_s + \mu)$ ;
13:    else
14:      Generate  $T$  observations of  $B_t \sim \text{Pois}(\epsilon_b + \mu)$ ;
15:      Generate  $T$  observations of  $S_t \sim \text{Pois}(\epsilon_s)$ ;
16:    end if
17:    Call estimation function ( $YZ()$  or  $GAN()$  or  $EA()$ );
18:    Obtain the parameter estimates;
19:  end for
20: end for

```

The results are presented in Table 2.1¹² which indicates that $YZ()$ method with $LK()$ factorization provides the PIN estimates with lowest MAE. Although the clustering algorithms, especially $GAN()$ method, provide powerful estimates of $\hat{\alpha}, \hat{\delta}, \hat{\epsilon}_b, \hat{\epsilon}_s$, they fail to estimate the arrival rate of informed investors $\hat{\mu}$, accurately. This is in line with Ersan and Alici (2016). On the contrary, $YZ()$ method with $EHO()$ factorization provides the best estimates for $\hat{\mu}$, but fails to provide good estimates for other parameters.

¹²This table represents the mean absolute errors (MAE) of the parameter estimates obtained by a given method for a given factorization. Each row represents a different method with a different factorization. First two column represent the specification of method and factorization respectively. The last six columns represents the power of estimates of PIN along with the parameter space $\Theta \equiv \{\alpha, \delta, \mu, \epsilon_b, \epsilon_s\}$. MAE measures for the estimates calculated as $\sum_{i=1}^N \frac{|\hat{\Theta}_i - \Theta_i^{TR}|}{N}$ where $\hat{\Theta}$ represent the estimates and Θ^{TR} represents the true value.

Table 2.1: Simulation results to examine the relationship between trade intensity and the estimates of PIN.

Method	Factorization	$\widehat{PIN}$	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\mu}$	$\hat{\epsilon}_b$	$\hat{\epsilon}_s$
YZ	LK	0.075	0.199	0.059	415.2	104.3	109.0
YZ	EHO	0.134	0.428	0.310	154.6	288.3	247.4
GAN	EHO	0.101	0.087	0.083	479.4	124.1	117.3
GAN	LK	0.101	0.087	0.083	479.5	123.8	118.1
EA	LK	0.102	0.268	0.274	484.6	128.7	119.3
EA	EHO	0.102	0.270	0.275	483.1	128.5	107.8

A more general way of examining the accuracy of PIN estimates is proposed in several studies (e.g, Lin and Ke (2011), Gan et al. (2015), Ersan and Alici (2016)). In this setting, we fix the trade intensity, $I=2500$. The total trade intensity represents the overall presence of informed and uninformed traders, that is, $I=(\mu, \epsilon_b, \epsilon_s)$. We then generate three probability terms p_1, p_2, p_3 with $N=5000$ random observations that are distributed uniformly between 0 and 1. p_1 represents the fraction of informed investors in total trade intensity, that is, $\mu=p_1 * I$. The rest of the trade intensity is distributed equally to buy and sell orders of uninformed investors, that is, $e_b = e_s = (1 - p_1) * I/2$. p_2 represents the true parameter for the probability of news arrival, α , and p_3 is the true parameter for the content of the news, δ . We generate observations for $\tilde{\alpha}$ and $\tilde{\delta}$, as described earlier. For each pair of $\tilde{\alpha}$ and $\tilde{\delta}$, we generate buy and sell values (B_t, S_t) for hypothetical $T=60$ days, again, in the manner presented above; form the likelihood and obtain the parameter estimates. The psuedo code for the second simulation method is given as follows;

The results are presented in Table 2.2¹³. Similar to first simulation, $GAN()$

¹³This table represents the mean absolute errors (MAE) of the parameter estimates obtained by a given method for a given factorization. Each row represents a different method with a different factorization. First two column represent the specification of method and factorization respectively. The last six columns represents the power of estimates of PIN along with the parameter space $\Theta \equiv \{\alpha, \delta, \mu, \epsilon_b, \epsilon_s\}$. MAE measures for the estimates calculated as $\sum_{i=1}^N \frac{|\hat{\Theta}_i - \Theta_i^{TR}|}{N}$ where $\hat{\Theta}$ represent the estimates and Θ^{TR} represents the true value.

Algorithm 2 A simulation to examine performance of the parameter estimates for PIN

```

1: Set N=5000, T=60, I= 2500;
2: Generate N observations of  $p_1 \sim U(0, 1)$ ;
3: Generate N observations of  $p_2 \equiv \alpha \sim U(0, 1)$ ;
4: Generate N observations of  $p_3 \equiv \delta \sim U(0, 1)$ ;
5: Obtain  $\tilde{\alpha} \sim \text{Binom}(\alpha)$  and  $\tilde{\delta} \sim \text{Binom}(\delta) \quad \forall n \in N$ ;
6: Set  $\mu = p_1 * I$ ,  $\epsilon_b = \epsilon_s = (1 - p_1) * I/2$ ;
7: for  $n \in N$  do
8:   if  $\tilde{\alpha}[n] == 0$  then
9:     Generate T observations of  $B_t \sim \text{Pois}(\epsilon_b)$ ;
10:    Generate T observations of  $S_t \sim \text{Pois}(\epsilon_s)$ ;
11:   else if  $\tilde{\alpha}[n] == 1$  and  $\tilde{\delta}[n] == 1$  then
12:     Generate T observations of  $B_t \sim \text{Pois}(\epsilon_b)$ ;
13:     Generate T observations of  $S_t \sim \text{Pois}(\epsilon_s + \mu)$ ;
14:   else
15:     Generate T observations of  $B_t \sim \text{Pois}(\epsilon_b + \mu)$ ;
16:     Generate T observations of  $S_t \sim \text{Pois}(\epsilon_s)$ ;
17:   end if
18:   Call estimation function ( YZ() or GAN() or EA() );
19:   Obtain the parameter estimates;
20: end for

```

captures the true nature of $\hat{\alpha}$ and $\hat{\delta}$ better than any other method with both factorizations. *YZ()* method with *EHO()* factorization performs best when estimating the arrival of informed traders, $\hat{\mu}$. The importance of estimating $\hat{\mu}$ becomes quite evident in Table 2.2. Although other methods outperform *YZ()* method with *EHO()* factorization in estimating α, ϵ_b and ϵ_s , it provides the best estimate for PIN due to its performance on estimating $\hat{\mu}$.

2.5. Conclusion

This chapter provides a short survey on five most widely used estimation techniques for the probability of informed trading (PIN) measure. In this chapter, we introduce the R package **InfoTrad**, covering estimation procedures for PIN using EHO, LK factorizations along with YZ, GAN and EA algorithms

Table 2.2: Simulation results to examine the performance of different methods in estimating the parameters for PIN.

Method	Factorization	$\widehat{PIN}$	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\mu}$	$\hat{\epsilon}_b$	$\hat{\epsilon}_s$
YZ	LK	0.323	0.428	0.432	1,212.0	303.4	325.0
YZ	EHO	0.237	0.437	0.357	942.9	386.0	470.2
GAN	LK	0.348	0.380	0.410	1,218.7	314.5	323.3
GAN	EHO	0.347	0.357	0.397	1,216.2	328.5	339.5
EA	LK	0.348	0.437	0.421	1,224.0	325.1	336.3
EA	EHO	0.347	0.428	0.413	1,222.0	331.3	345.9

(*EHO()*, *LK()*, *YZ()*, *GAN()*, *EA()*). The functions *EHO()* and *LK()* read a (Tx2) matrix where the rows of the first column contains total number of buy orders on a given trading day t , B_t , and the rows of the second column contains the total number of sell orders on a given trading day t , S_t , where $t \in \{1, 2, \dots, T\}$. In addition, they also require an initial parameter vector in the form of, $\Theta_0 = \{\alpha, \delta, \mu, \epsilon_b, \epsilon_s\}$. Both functions produce the respective log-likelihood functions.

The functions *YZ()*, *GAN()* and *EA()* read (B_t, S_t) as an input along with a likelihood specification that is set to LK by default. These functions do not require initial parameter matrix to obtain the parameter estimates when calculating PIN. All three functions use *neldermead()* method of **nloptr** as built-in optimization procedure for MLE. *YZ()*, *GAN()* and *EA()* produce an object that gives the parameter estimates $\hat{\Theta}$ along with likelihood value and $\widehat{PIN}$.

CHAPTER III

INFORMED TRADING, ORDER FLOW SHOCKS AND THE CROSS SECTION OF EXPECTED RETURNS IN BORSA ISTANBUL

3.1. Introduction

There are conflicting views on the relationship between information risk which arises from the information asymmetry between informed and uninformed investors, and stock returns. On one hand, Easley and O'Hara (2004) construct a rational expectations model with information asymmetry to show that uninformed investors are inferior in adjusting their portfolios quickly, in the presence of new information. Therefore, *ceteris paribus*, uninformed investors demand a premium for holding stocks that have high number of informed traders which in turn implies that information risk is systematic and undiversifiable. On the other hand, Hughes et al. (2007) argue that in a large economy where the number of securities is infinite, increase in the information asymmetry yields an increase in returns for the overall market. Their claims indicate that after controlling for systematic factors, firm-specific information should not affect the cross-section of expected returns and thus the risk of information asymmetry is idiosyncratic and fully diversifiable.

These conflicting views have empirical implications. Easley et al. (1996)

cornerstone study that quantifies the information asymmetry for a stock trade. Easley et al. (1996) and Easley et al. (2002) derive a proxy called the probability of informed trading (PIN) which aims to measure the proportion of informed investors. Easley et al. (1996) show that transaction costs are, on average, larger for stocks with higher PIN. Similarly, Easley et al. (2002) and Easley et al. (2010) indicate that investors demand a premium for holding stocks that have higher PIN. On the contrary, Lai et al. (2014) show that the information is not a priced risk for 47 different countries. More importantly, Duarte and Young (2009) decompose PIN into two separate components: one related to adverse selection (adjusted PIN) and the other related to stocks liquidity. They adjust the PIN by introducing the probability of systematic order flow shocks (PSOS) that enables to model the positive correlation between buy and sell order flow. Their results also indicate that PSOS is a priced risk rather than the adjusted PIN.

There is a growing literature on the role of information asymmetry on the stocks traded in the Turkish stock market, Borsa Istanbul (BIST). Aktas (2016) indicate that informed trades in BIST tend to be large in size. In addition, he shows that short sales have higher price impact compared to an average trade. Ekinici and Ersan (2018) examines the involvement of high-frequency traders (HFT) in BIST, proposing a new method to capture HFT activity. They estimate the HFT involvement to be 1.23% in 2015 where BIST switch to BISTECH trading system, allowing HFT's for the first time. They argue that HFT participation is generally larger for large cap stocks. Simsir and Simsek (2018) show an abnormal trading volume around the first private notification times which is right before public disclosure announcements made public in BIST. They argue that the profits from the trades right before an announcement made public can sum up to 77 million \$. In the next chapter of this thesis, we examine the differences in information between foreign and domestic investors. We show that local investors have a price

impact over a broader array of stocks however the foreign price impact significantly increases after June 2013 with the increased political turmoil within the country. We also show that foreign impact clusters over stocks with larger market capitalization.

This chapter examines the systematic relationship between information asymmetry and stock returns in BIST. Due to the lack of studies that examine the stock mispricing and the microstructure of BIST, this paper fills an essential gap in the literature.

In this chapter, we analyze the nature of expected returns of high-PIN stocks traded in Borsa Istanbul to explore the preferences of investors towards the risk of trading with informed investors. We first examine this relationship through Fama-MacBeth regressions (Fama and MacBeth, 1973). Besides, similar to Easley et al. (2010), we examine the behavior of PIN-sorted portfolios through cross-sectional regressions.

Our results are in line with the theoretical expectations of Hughes et al. (2007) and the empirical findings of Lin et al. (2014). We fail to reject the hypothesis that information asymmetry is an idiosyncratic risk in BIST. Although we observe that PIN predicts next month returns, the predictive power disappears after we control for other systematic factors. Our univariate and multivariate portfolio analyses show that the investors who trade on BIST do not expect a premium for holding stocks that have high levels of information asymmetry. The results are free from any bias due to classification algorithms, floating point exception, boundary solutions and systematic order flow shocks in estimating PIN.

The structure of the paper is as follows. In the next section, we revisit the literature on the PIN model and its extensions. Section 3.3 describes how we form

our sample. Section 3.4 discusses the methodology. Section 3.5 presents the results and Section 3.6 concludes.

3.2. PIN model and its extensions

Chapter II of this thesis provides a detailed survey of the PIN model and its extensions. In Chapter II, we underline the problems due to floating point exception and boundary solutions, and present the different factorization and estimation methods to overcome these problems.

The likelihood factorization presented in equation (2.2) (EKOP factorization) is prone to selection bias. Lin and Ke (2011) show that the feasible solution set for the non-linear optimization problem shrinks significantly as the daily number of buy and sell orders increases. The optimal value for the log likelihood in equation (2.2) becomes undefined, for large enough (B_t, S_t) whose factorials cannot be computed by mainstream computers (floating point exception, (FPE)). Therefore, studies that use EKOP factorization may systematically exclude stocks that are actively traded.

To deal with this problem, Duarte and Young (2009) modifies the EKOP factorization by replacing $\exp(-\epsilon_b) \frac{\epsilon_b^{B_t}}{B_t!}$ with $\exp[-\epsilon_b + B_t \ln(\epsilon_b) - \sum_{i=1}^{B_t} \ln(i)]$ and other components of the likelihood with the similar modifications. In addition, two other factorizations that are algebraically equivalent to (2.1) are provided in order to avoid FPE. We present the first factorization (EHO factorization) of Easley et al. (2010) in equation (2.4) and the second factorization (LK factorization) of Lin and Ke (2011) in equation (2.5). Both studies modify the factorization such that $B_t!$ and $S_t!$ become constant with respect to the parameter vector Θ and dropped in the maximum likelihood procedure.

3.2.1 Probability of Symmetric Order-Flow Shock (PSOS)

The sequential trading model of Easley et al. (1996) enforces a negative contemporaneous correlation between the number of buy and sell orders (B, S). Informed trading can inflate the number of buy and sell orders, separately. Duarte and Young (2009) indicate that the contemporaneous covariance between the number of buy and sell orders in PIN model is given as follows;

$$\text{cov}(B, S) = (\alpha\mu)^2(\delta - 1)\delta \leq 0 \quad (3.1)$$

However, Duarte and Young (2009) argue that this aspect of the model is not observed empirically. They show a strong positive correlation between buy and sell orders and high levels of buy and sell volatility.

To account for the different levels of variance between buy and sell orders, they separate the arrival process for informed traders to place buy and sell orders from one another. They defined μ_b (μ_s) to be the arrival rate of informed traders to place a buy (sell) order. More importantly, they allow for increased buy and sell variation along with a positive correlation between buys and sells as they define another event that causes both buy and sell order flow to increase. They called this event *a symmetric order-flow shock*. The probability of such an event is assumed to follow a Bernoulli distribution with parameter γ conditional on the absence of private information. The probability of such an event conditional on the arrival of private information is also assumed to follow a Bernoulli distribution, but with a different parameter, γ' . In case of a symmetric order flow shock, the additional arrival rate of buys (sells) is defined as ϕ_b (ϕ_s).

Then the parameter vector of the extended model becomes,

$\Theta = \{\alpha, \delta, \gamma, \gamma', \mu_b, \mu_s, \epsilon_b, \epsilon_s, \phi_b, \phi_s\}$. The joint probability function in equation (2.1)

can then be updated as follows;

$$\begin{aligned}
f(B_t, S_t | \Theta) \equiv & (1 - \alpha)(1 - \gamma) \exp \left[-\epsilon_b + B_t \log(\epsilon_b) - \sum_{i=1}^{B_t} \log(i) \right] \\
& * \exp \left[-\epsilon_s + S_t \log(\epsilon_s) - \sum_{i=1}^{S_t} \log(i) \right] \\
& + (1 - \alpha)\gamma \exp \left[-(\epsilon_b + \phi_b) + B_t \log(\epsilon_b + \phi_b) - \sum_{i=1}^{B_t} \log(i) \right] \\
& * \exp \left[-(\epsilon_s + \phi_s) + S_t \log(\epsilon_s + \phi_s) - \sum_{i=1}^{S_t} \log(i) \right] \\
& + \alpha(1 - \delta)(1 - \gamma') \exp \left[-(\epsilon_b + \mu_b) + B_t \log(\epsilon_b + \mu_b) - \sum_{i=1}^{B_t} \log(i) \right] \\
& * \exp \left[-\epsilon_s + S_t \log(\epsilon_s) - \sum_{i=1}^{S_t} \log(i) \right] \\
& + \alpha(1 - \delta)\gamma' \exp \left[-(\epsilon_b + \mu_b + \phi_b) + B_t \log(\epsilon_b + \mu_b + \phi_b) - \sum_{i=1}^{B_t} \log(i) \right] \\
& * \exp \left[-(\epsilon_s + \phi_s) + S_t \log(\epsilon_s + \phi_s) - \sum_{i=1}^{S_t} \log(i) \right] \\
& + \alpha\delta(1 - \gamma') \exp \left[-\epsilon_b + B_t \log(\epsilon_b) - \sum_{i=1}^{B_t} \ln(i) \right] \\
& * \exp \left[-(\epsilon_s + \mu_s) + S_t \log(\epsilon_s + \mu_s) - \sum_{i=1}^{S_t} \log(i) \right] \\
& + \alpha\delta\gamma \exp \left[-(\epsilon_b + \phi_b) + B_t \log(\epsilon_b + \phi_b) - \sum_{i=1}^{B_t} \log(i) \right] \\
& * \exp \left[-(\epsilon_s + \phi_s + \mu_s) + S_t \log(\epsilon_s + \phi_s + \mu_s) - \sum_{i=1}^{S_t} \log(i) \right] \tag{3.2}
\end{aligned}$$

The structural model presented in equation (3.2), enables a positive covariance between buy and sell order. The probability of informed trading adjusted for the symmetric order flow shock is given by;

$$AdjPIN \equiv \frac{\alpha(\delta\mu_s + (1 - \delta)\mu_b)}{\alpha(\delta\mu_s + (1 - \delta)\mu_b) + (\phi_b + \phi_s)(\alpha\gamma_I + (1 - \alpha)\gamma) + \epsilon_b + \epsilon_s} \quad (3.3)$$

The probability of symmetric order flow shock (PSOS) is then given by;

$$PSOS \equiv \frac{(\phi_b + \phi_s)(\alpha\gamma_I + (1 - \alpha)\gamma)}{\alpha(\delta\mu_s + (1 - \delta)\mu_b) + (\phi_b + \phi_s)(\alpha\gamma_I + (1 - \alpha)\gamma) + \epsilon_b + \epsilon_s} \quad (3.4)$$

The problem of boundary solutions are also evident for estimating PIN when DY factorization is used. To this end, we choose 10 different random starting points for $(\alpha^0, \delta^0, \gamma^0, \gamma_I^0)$ from uniform distribution $U(0, 1)$ for parameters that are bounded from above and below. For parameters that are bounded only from above, $(\mu_A^0, \mu_S^0, \epsilon_A^0, \epsilon_S^0, \phi_A^0, \phi_S^0)$ we set the arrival rates to be the one third of the respective buy and sell values within the specified period.

3.3. Sample Formation

We collect the limit-order book for all stocks traded in BIST National Markets Listing, between March 2005 to April 2017. Our sample consists of 498 different stocks. We obtain price and accounting information from Bloomberg Terminal. We calculate the firm specific variables such as size (SIZE), book-to-market ratio (BTM), beta (BETA), illiquidity (ILLIQ), idiosyncratic volatility (IVOL), reversal (REV) and momentum (MOM) at monthly frequency as shown in the Appendix. Each order in the limit-order book contains the date, time, stock ticker, order ID, order type, quantity and price stamps. Order type enables us to track the order process for buys and sells, given a trading day. Therefore, we construct the order input matrix (B_t, S_t) from the actual number of buy and sell orders, as a panel

data which has trading days in the time dimension and each firm in BIST in the cross-section. We exclude all orders that arrive in the opening auction.

The observation of the actual order arrivals frees us from considering the trade initiation sequence. This freedom relieves us from choosing different trade classification algorithms to assign values for (B_t, S_t) . The literature on the accuracy of algorithms that classify trades as buyer or seller initiated is inconclusive. However, all classification algorithms are shown to create some bias in BIST (Aktas and Kryzanowski, 2014). Our results presented in the next section, are therefore robust to any biases that might arise due to improper selection of trade classification algorithms.

Table 3.1¹ presents the descriptive statistics on the daily number of buy and sell orders. Daily buy orders are more volatile compared to sell orders. More importantly, there is a positive correlation between buy and sell orders for BIST stocks. We show in Table 3.1 that the minimum cross sectional correlation between buy and sell orders is 10% throughout our sample period. The average correlation is 85%. Similar to Duarte and Young (2009), we do not observe any negative correlation among buy and sell orders. This implies that the Duarte and Young (2009) extension of the PIN model which takes the probability of symmetric order flow shock into consideration, is necessary when examining the pricing implications of information asymmetry in BIST.

¹This table presents the descriptive statistics for the cross-sectional distribution of the daily number of buy and sell observations, (B_t, S_t) in our sample. Our sample covers all stocks that are traded on BIST between March 2005 and April 2017, consisting 498 different stocks. For each stock, on each trading day, we compute the total number of buy and sell orders submitted to BIST from the intraday limit-order book. Then for each stock, we compute the mean and variance of the daily buy and sell orders throughout our sample period. First two columns report the descriptive statistics for the mean buy and sell values, respectively. Third and forth column report the descriptive statistics for the variance of the daily buy and sell values. Fifth column reports the descriptive statistics for the *Spearman* correlation coefficient between daily buy and sell orders.

Table 3.1: Descriptive statistics on number of buy and sell orders.

Statistics	Mean Buy	Mean Sell	Variance Buy	Variance Sell	Correlation
Minimum	18.56	14.11	17.71	13.53	0.10
25th Percentile	248.97	182.06	355.83	229.35	0.81
Median	411.84	290.13	562.06	343.73	0.88
Mean	521.07	398.34	733.13	487.61	0.85
75th Percentile	620.56	445.17	856.33	557.30	0.92
Maximum	4627.44	4223.77	5728.53	2945.22	0.98

Table 3.2² reports the descriptive statistics for the PIN and PSOS estimates calculated using the Duarte and Young (2009) factorization along with the descriptive statistics for firm specific variables. Table 3.3³ gives the pair-wise correlations among firm specific variables. In line with Alkan and Guner (2018), the highest correlation is between IVOL and MAX factors with 84%. This is followed by the correlation between RETURN and MAX with 34%. The largest negative correlation is among PIN and PSOS factors.

3.4. Methodology

3.4.1 Fama-MacBeth Regressions

To test the information asymmetry hypothesis, we estimate the cross-sectional regression presented in (3.5) in each sample month. For each month in our sample, the returns are regressed on PIN and various firm specific measures.

$$R_{i,t+n} = \gamma_{0,t} + \gamma_{1,t}PIN_{i,t} + \gamma_{2,t}BETA_{i,t} + \gamma_{3,t}SIZE_{i,t} + \gamma_{4,t}BTM_{i,t} + \gamma_{5,t}X_{i,t} + \epsilon_{i,t} \quad (3.5)$$

where $R_{i,t}$ is the return on stock i , $i \in \{1 \dots 498\}$ for holding periods of n months after month t . We try three different holding periods, one-month, three-months, six-months, in line with Atilgan et al. (2016) and Alkan and Guner (2018). The vector $X_{i,t}$ includes other firm specific factors such as MOM, MAX, IVOL, PSOS

²This table reports the descriptive statistics for monthly firm specific factors for sample stocks that are traded in BIST between March 2005 - April 2017. RETURN represents the monthly percentage returns. BETA is the systematic risk factor. SIZE is the logarithm of the end of month market capitalization. BTM is the book-to-market ratio. MOM is the momentum variable. REV represents the reversal variable. ILLIQ is the illiquidity measure. MAX is the maximum daily return within a month. IVOL is the idiosyncratic volatility. PIN is the probability of informed trading estimated using DY factorization. PSOS is the probability of systematic order flow shocks.

³This table reports the cross-sectional correlations among time-series averages of firm-specific factors. PIN is the probability of informed trading. PSOS is the probability of systematic order flow shocks. SIZE is the market capitalization. BETA is the market risk factor. MOM is cumulative return over the period $[t-7, t-1]$. MAX is the maximum daily return within the month t . IVOL is the idiosyncratic volatility. BTM is the book to market ratio. ILLIQ is the Amihud (2002) illiquidity measure.

Table 3.2: Descriptive statistics on the cross sectional distributions of monthly factors.

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
RETURN	44967	0.01	0.14	-0.89	2.94
BETA	44957	0.60	0.56	-13.07	23.33
SIZE	44967	5.25	1.95	0.29	10.68
BTM	44967	1.10	1.23	0.00	53.40
MOM	44226	-0.05	0.62	-73.01	0.95
REV	44501	0.01	0.14	-0.89	2.94
ILLIQ	44967	0.03	1.10	0.00	100.15
MAX	44967	0.06	0.04	-0.10	0.20
IVOL	44957	0.02	0.01	0.00	0.18
PIN	44966	0.16	0.17	0.00	1.00
PSOS	44966	0.24	0.25	0.00	1.00

Table 3.3: Factor Correlations

	RETURN	BETA	SIZE	BTM	MOM	REV	ILLIQ	MAX	IVOL	PIN	PSOS
RETURN	1										
BETA	-0.109	1									
SIZE	0.058	0.174	1								
BTM	-0.080	-0.011	-0.233	1							
MOM	-0.010	0.006	0.132	-0.210	1						
REV	0.032	0.007	0.035	-0.036	0.026	1					
ILLIQ	-0.006	0.009	0.070	0.001	-0.001	-0.002	1				
MAX	0.343	0.105	-0.164	0.008	-0.043	0.012	0.000	1			
IVOL	0.202	-0.013	-0.254	0.050	-0.050	-0.002	0.049	0.841	1		
PIN	0.027	0.017	0.034	-0.014	0.011	0.009	-0.004	0.029	0.024	1	
PSOS	-0.030	0.066	0.130	0.002	0.000	0.010	0.003	-0.054	-0.047	-0.217	1

and ILLIQ which are described in the Appendix. We form our first hypothesis as follows;

Hypothesis 1

$$H_0 : \gamma_{1,t} = 0$$

$$H_A : \gamma_{1,t} > 0$$

3.4.2 Univariate Portfolio Analysis

Easley and O'Hara (2004) suggest that uninformed investors demand compensation for holding stocks that have high information asymmetry as they cannot timely update their portfolios according to the new information arrive at the market. To examine the relationship between information asymmetry and stock returns, we form portfolio-deciles each month by sorting stocks based on their PIN estimate. The stocks that are in the lowest decile mimics an investor's portfolio which is predominantly composed of stocks that have low information asymmetry. The stocks that are in the highest decile mimics an investor's portfolio which is composed of stocks that have high information asymmetry. We then mimic an investors portfolio who has a long position in stocks with high information asymmetry and short position of equal size in stocks that have low information asymmetry and check whether the resulting zero investment portfolio gains significant return premiums. Following Easley et al. (2002), Easley and O'Hara (2004), and Easley et al. (2010) we expect that zero investment portfolio will earn a significant risk-adjusted positive return. Letting R^{High} and R^{Low} denote the returns of high-PIN and low-PIN stocks, our second hypothesis is as follows;

Hypothesis 2:

$$H_0 : R^{High} - R^{Low} = 0$$

$$H_A : R^{High} - R^{Low} > 0$$

3.4.3 Multivariate Portfolio Analysis

Easley et al. (2010) documents that informed trading is larger for small firms. In Table 3.3, we observe a positive correlation between PIN estimates and SIZE. Although this finding is contradicting with Easley et al. (2010), the positive relationship may arise from the fact that larger firms, due to their nature, produce more information through investment or research and development activity⁴.

To this end, in addition to univariate portfolio analysis, we apply multivariate portfolio analysis in which we consider the variation in firm size. Within the scope of multivariate portfolio analysis, first, we sort the stocks into deciles based on their market value at the end of previous month. Then, we divide each size decile into three groups based on the PIN estimates of the stocks that are in that decile. This procedure gives us 30 different portfolios. We follow the returns of these portfolios for one, three and six months. We then form zero investment portfolios R_p for each decile ($p = 1 \dots 10$) by subtracting the returns of the high-PIN portfolio (R_p^{High}) from the low-PIN (R_p^{Low}) portfolio. These returns provides us to examine the impact of information asymmetry controlling for firm value.

Similarly, controlling for firm value, we test the relationship between information asymmetry and stock returns as follows;

Hypothesis 3:

$$H_0 : R_p = 0$$

$$H_A : R_p > 0$$

⁴Presence of HFT's may be another reason why we observe higher informed trading for larger stocks (Ekinici and Ersan, 2018; Ersan and Ekinici, 2016). More over informed trading activity may also be attributed to the brokerage house activity (Aktas, 2016; İmişikler et al., 2015; Ozsoylev et al., 2013)

3.4.4 Factor Tests

Finally, we use factor tests to examine the relationship between information asymmetry and stock returns under the control of size and profitability factors that Fama and French (1993) introduced for U.S. markets. Akdeniz et al. (2000) show that size and profitability are systematic risk factors for BIST. The standard model is presented as below;

$$R_p = \vartheta_p^1 + \vartheta_p^2 R_m + \vartheta_p^3 SMB + \vartheta_p^4 HML + \epsilon_p \quad p = 1 \dots 10 \quad (3.6)$$

where R_m represents the market (BIST 100 Index) return series, SMB and HML are the size and value factors of Fama and French (1993) applied for all stocks traded in BIST. We update SMB and HML portfolios yearly, in order to be consistent with literature. We then test the hypothesis that whether *the constant terms* $\vartheta^1 \equiv (\vartheta_1^1 \dots \vartheta_{10}^1)$ is jointly equal to zero or not.

Hypothesis 4:

$$H_0 : \vartheta^1 = 0$$

$$H_A : \vartheta^1 > 0$$

To test this hypothesis, we use the F statistics of (Gibbons et al., 1989) (GRS test statistic).

3.5. Results

3.5.1 Regression analysis

We present Fama-MacBeth regression results in Table 3.4⁵. The impact of information asymmetry on stock returns is positive as expected. In the first

⁵Each month, we regress the stock returns on previous months PIN estimates along with the control factors BETA, SIZE, BTM, MOM, REV, ILLIQ, MAX, IVOL and PSOS. Each column

column, we present the impact of information asymmetry on returns without controlling for any other firm-specific factors. In this case, we observe that a 1% increase in the information asymmetry will result in 0.02% (2bps) increase, on average in next months returns. To this end, we argue that although information asymmetry has a statistically significant impact on stock returns in BIST, the impact is not economically significant. In bivariate regressions, the impact of PIN varies from 1.9bps to 0.6bps. The statistical significance of PIN in bivariate regression disappears when we control it with the MAX factor. This indicates that the impact of lottery stock preference in BIST, proxied by MAX (Alkan and Guner, 2018) is more prominent compared to the impact of information asymmetry. In multivariate Fama-MacBeth regressions the impact of PIN reduces down to 0.3bps and lose its statistical significance⁶. To this end, we fail to reject our first hypothesis which suggest that information asymmetry has no significant impact over the cross section of stock returns.

3.5.2 Univariate Portfolio Analysis

We sort stocks into deciles based on PIN values estimated using the Duarte and Young (2009) factorization, for each month between March 2005 and April 2017.

We form equally and value weighted portfolios using the stocks in each decile and follow the returns of these portfolios for one, three and six months. In table 3.5⁷,

corresponds to different specifications where the dependent variable is one-month ahead future returns $R_{i,t+1}$. Each row corresponds to firm-specific factors. Entries in the table are the time-series averages of the slope coefficients obtained from the cross-sectional regressions. Values in parenthesis presents t-statistics based on Newey and West (1987) standard errors.

⁶Interestingly, contrary to the standard literature, we observe a significant and negative impact of BETA and BTM over future returns, and the positive impact of SIZE.

⁷This table presents a descriptive statistics on the decile portfolios formed based on monthly PIN estimates. At the end of each month, we sort stocks based on PIN measures estimated on that month, and form decile portfolios based on those estimates. We follow the returns of stocks in portfolios next month and set the return of each portfolio to be the equal weighted average of the stock returns within those portfolios. The first column shows the average PIN estimate of each portfolio. The second column shows the average and the third column shows the median return

Table 3.4: Fama MacBeth Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PIN	0.017 (4.46)	0.019 (5.01)	0.016 (4.11)	0.017 (4.39)	0.017 (4.34)	0.016 (4.23)	0.017 (4.60)	0.006 (1.77)	0.010 (2.86)	0.013 (3.24)	0.004 (1.52)	0.003 (1.09)
BETA		-0.017 (-4.28)									-0.047 (-8.87)	-0.047 (-8.83)
SIZE			0.003 (4.94)								0.007 (12.50)	0.007 (12.52)
BTM				-0.011 (-10.10)							-0.003 (-4.22)	-0.003 (-4.22)
MOM					0.006 (1.31)						-0.009 (-2.40)	-0.009 (-2.43)
REV						0.012 (1.65)					0.007 (1.19)	0.007 (1.23)
ILLIQ							0.013 (1.79)				0.005 (1.19)	0.005 (1.19)
MAX								1.386 (30.72)			2.269 (39.75)	2.266 (39.64)
IVOL									2.883 (15.09)		-2.562 (-9.98)	-2.558 (-9.95)
PSOS										-0.014 (-5.93)		-0.003 (-1.66)

we present the descriptive statistics for decile portfolios obtained with one-month holding period⁸.

The portfolio statistics is set to be the equal-weighted averages of the stocks within that portfolio and the measures are calculated at the end of each month where the portfolios are formed. We observe that, by construction, the average PIN values increase from low-PIN decile to high-PIN decile. Other than PIN estimates, we do not observe a linear increase or decrease in any other measure. We observe the highest average monthly return in the ninth decile (average return 8%) and the second decile (average return is 4.5%). The stocks in the ninth-decile also has the highest daily return volatility. On average, the stocks in the fifth-decile has the highest market capitalization. To this end, we can say that PIN sorted portfolios in BIST does not have a monotonically increasing or decreasing feature which may suggest that information asymmetry does not have any systematic impact on stocks traded in BIST.

We examine whether the investors in BIST demand a premium for holding stocks that have high information asymmetry through examining the difference in returns of high-PIN and low-PIN portfolios ($R^{High} - R^{Low}$). In Table 3.6⁹, we show the time-series analysis of zero investment PIN portfolios.

respectively. Fourth column shows the historical volatility of the monthly portfolio returns whereas the fifth column show the skewness of returns. From sixth column to tenth column, we present the portfolios' SIZE, BETA, BTM, IVOL, MOM values, respectively.

⁸We observe similar values for three and six month holding periods. We do not report these tables to save space.

⁹In this table, we present the equally and value-weighted returns of decile portfolios. At the end of each month, we sort our stocks into decile portfolios based on monthly PIN estimates obtained using Duarte and Young (2009) factorization. We follow the portfolio returns for one, three and six months. Panel A, B and C present the findings for one, three and six months holding period, respectively. The three factor α is obtained through regressing the portfolio returns on the market return (BIST100), and size (SMB) and profitability factors (HML) of Fama and French (1993). High-Low represents the zero investment PIN portfolio which mimics the returns of an investor who has a long position in High-PIN portfolio and a short position of equal size in the Low-PIN portfolio. t-statistics are calculated with Newey and West (1987) adjusted standard errors.

Table 3.5: Descriptive Statistics for PIN portfolios

Deciles	Average PIN	Average Return	Median Return	Volatility	Skewness	SIZE	BETA	BTM	IVOL	MOM
Low-PIN	0.0000	0.015	-0.016	0.13	1.20	902	0.57	1.01	0.02	0.15
2	0.0004	0.045	0.013	0.13	1.78	1017	0.18	1.27	0.02	0.08
3	0.0038	0.027	0.001	0.10	1.19	1139	0.20	0.95	0.03	0.08
4	0.0472	0.017	-0.005	0.10	1.16	3515	0.36	1.24	0.02	0.09
5	0.1126	0.013	0.001	0.09	1.77	4938	0.48	0.99	0.02	0.10
6	0.1592	0.014	-0.006	0.10	1.16	2147	0.40	1.24	0.02	0.16
7	0.2083	0.026	-0.003	0.09	0.94	1723	0.30	1.18	0.02	0.13
8	0.2558	0.039	0.016	0.13	2.55	2150	0.36	1.05	0.02	0.12
9	0.3620	0.078	0.027	0.22	2.66	1296	0.31	1.12	0.03	0.18
High-PIN	0.5195	0.023	-0.002	0.09	0.83	1368	0.31	0.96	0.02	0.11

Table 3.6: Return characteristics of univariate PIN portfolios

Panel A: Holding Period= 1 month							
Decile	Equally Weighted Portfolios				Value Weighted Portfolios		
	Average Re- turn (%)	Re-	Three Fac- tor alpha		Average Re- turn (%)	Re-	Three Fac- tor alpha
Low-PIN	0.019		0.011		0.014		0.007
2	0.017		0.008		0.015		0.011
3	0.014		0.005		0.016		0.008
4	0.015		0.005		0.007		0.000
5	0.012		0.003		0.016		0.007
6	0.016		0.005		0.012		0.001
7	0.018		0.008		0.029		0.037
8	0.014		0.004		0.016		0.005
9	0.013		0.004		0.008		-0.003
High-PIN	0.014		0.006		0.009		-0.001
High-Low	-0.004		-0.005		-0.004		-0.008
t-statistics	(-1.16)		(-1.20)		(-0.73)		(-1.19)
Panel B: Holding Period= 3 months							
Decile	Equally Weighted Portfolios				Value Weighted Portfolios		
	Average Re- turn (%)	Re-	Three Fac- tor alpha		Average Re- turn (%)	Re-	Three Fac- tor alpha
Low-PIN	0.018		0.009		0.018		0.013
2	0.016		0.007		0.007		-0.001
3	0.026		0.017		0.042		0.057
4	0.014		0.004		0.013		0.005
5	0.017		0.006		0.015		0.004
6	0.014		0.002		0.010		-0.001
7	0.015		0.004		0.014		0.001
8	0.014		0.005		0.027		0.023
9	0.014		0.004		0.010		-0.001
High-PIN	0.014		0.004		0.004		-0.006
High-Low	-0.004		-0.005		-0.014		-0.019
t-statistics	(-1.14)		(-1.29)		(-2.06)		(-2.73)
Panel C: Holding Period= 6 months							
Decile	Equally Weighted Portfolios				Value Weighted Portfolios		
	Average Re- turn (%)	Re-	Three Fac- tor alpha		Average Re- turn (%)	Re-	Three Fac- tor alpha
Low-PIN	0.012		0.001		0.010		0.002
2	0.013		0.005		0.017		0.017
3	0.016		0.006		0.011		0.003
4	0.016		0.007		0.010		0.002
5	0.015		0.005		0.012		0.003
6	0.012		0.005		0.018		0.018
7	0.011		0.001		0.007		-0.003
8	0.016		0.006		0.011		0.002
9	0.013		0.004		0.014		0.004
High-PIN	0.019		0.010		0.016		0.006
High-Low	0.007		0.009		0.006		0.004
t-statistics	(1.85)		(2.22)		(0.80)		(0.49)

In Table 3.6, Panel A and B, we observe that the average returns on the zero investment PIN portfolios are negative for one-month and three-month holding periods, respectively. This finding contradicts with the theoretical expectations of Easley and O'Hara (2004) and the empirical findings of Easley et al. (2002), Easley et al. (2010) and Lai et al. (2014). However, we fail to reject the hypothesis that the returns of zero investment portfolios are, on average, different from zero. That is why, the importance on the sign of the coefficient largely disappears. In Table 3.6, Panel C, we observe that only the three-factor α for the equally weighted portfolio is positive and statistically significant. To this end, we can say that the six-month premium for zero investment PIN portfolios are 90bps, which is economically insignificant given the average monthly return is 1% in our sample.

In line with our regression analysis, our univariate portfolio analysis suggests that the information asymmetry proxied by PIN does not have any systematic impact on stock returns in BIST. We fail to reject the hypothesis that the return premium of zero investment PIN portfolios is equal to zero. In line with the expectations of Hughes et al. (2007), we show that the risk due to information asymmetry can be eliminated entirely within a one-month or three-month holding period, through diversification. For the six-month period, we show that the economic meaning of the premium due to information asymmetry largely disappears.

To a large extent, our univariate analysis show that risk due to information asymmetry is firm specific for stocks that are traded in BIST. Lastly, in line with Easley et al. (2010), we examine whether the firm size plays any role in explaining the impact of information asymmetry on stocks returns through multivariate portfolio analysis.

3.5.3 Multivariate Portfolio Analysis

We follow a similar methodology proposed in Easley et al. (2010) to examine the impact of informed trading on securities returns. We sort stocks with respect to their size at the beginning of each month, and form ten size portfolios. Then again for each month, we divide the portfolios into three with respect to previous month's *PIN* estimates. This yields us the time series of 30 different portfolio returns. We follow the portfolio returns for one, three and six months. Then, we subtract the low-PIN portfolio returns from high-PIN portfolio returns in each size decile and form ten zero investment PIN-Size portfolios (R_p $p = 1 \dots 10$). This gives us the mimicking portfolio returns of an investor who has a long position in high-PIN stocks and a short position with equal size in low-PIN stocks, controlling for size. The descriptive statistics on zero investment PIN-Size portfolios are provided in Table 3.7¹⁰.

In Table 3.7 Panel A, we observe that for each size decile, the difference between the market value of high-PIN portfolios and low-PIN portfolios monotonically increase, except for the last size decile. On the contrary, in Panel B, we observe that the difference in PIN values decrease as we move from small to big decile. Lastly, we share the results on Hypothesis 3 on Panel C. To this end, we fail to reject that the difference between high-PIN and low-PIN returns in each size decile is equal to zero. This finding implies that under the control of market capitalization, the investors who invest in stocks that have high information

¹⁰This table presents the characteristics of PIN-Size portfolios. At the beginning of each month, we divide the sample into size deciles based on end-of-month market capitalization of previous month. Then, we divide each decile into three PIN groups, based on previous month's PIN estimates. In panel A, we represent the time-series averages of the equally-weighted means of PIN for each portfolio. Panel B contains the Turkish Lira denominated time series averages of the mean market capitalization for each portfolio. Panel C represents the time-series average of average monthly returns in each portfolio. PIN-Size portfolios are constructed based on previous months market capitalization and PIN estimates. DIF is the average difference between High and Low PIN portfolios. $t(\text{DIF})$ is the corresponding t-statistics.

asymmetry do not obtain a return premium.

Finally, we examine the systematic impact of information asymmetry on stock returns through equation (3.6). Akdeniz et al. (2000) show that the size and profitability factors of Fama and French (1993) corresponds to a systematic risk in BIST. The results on Fama and French (1993) regressions provided in Table 3.8¹¹, below.

In line with our previous findings, results presented in Table 3.8 indicate that information asymmetry does not provide a systematic return premium for stocks that are trading in BIST. The coefficient estimates for the constant terms, $\hat{\vartheta}_p^1$ are mostly negative and insignificant. The values range between -1bps to 0.5bps. We also fail to reject the hypothesis that the constant terms are jointly equal to zero (pvalue(GRS) = 0.12). We then conclude that zero investment PIN-Size portfolios do not bring any premium after controlling for other systematic factors such as market return, size and profitability. With these results, our study provides strong empirical support to the theoretical expectations of Hughes et al. (2007). We show that information risk is idiosyncratic in nature and can be eliminated with diversification. Our results are robust to any possible bias due to trade classification, floating point exception, boundary solutions and systematic order flow shocks.

¹¹We run the following regression on the time series of monthly returns of ten zero investment PIN-Size portfolios. *PIN* estimates are obtained through DY algorithm. $R_p = \vartheta_p^1 + \vartheta_p^2 R_m + \vartheta_p^3 SMB + \vartheta_p^4 HML + \epsilon_p$. The sample period is between March 2005 - April 2017. We provide the coefficient estimates and their respective t-statistics. For goodness of fit, we provide adjusted R^2 measures along with F statistics. Last two rows, provide the Gibbons et al. (1989) test statistics for the null hypothesis that $\hat{\vartheta}^1 = 0$.

Table 3.7: Characteristics and Returns of PIN-Size Portfolios

Panel A: SIZE				
	Low	2	High	DIF
Small	2.383	2.532	2.459	0.077
2	3.191	3.402	3.271	0.080
3	3.717	3.986	3.828	0.111
4	4.189	4.486	4.326	0.137
5	4.652	5.027	4.859	0.207
6	5.134	5.593	5.409	0.275
7	5.687	6.180	6.000	0.313
8	6.263	6.774	6.575	0.312
9	6.926	7.638	7.387	0.461
Large	8.915	9.445	9.218	0.303

Panel B: PIN				
	Low	2	High	DIF
Small	0.004	0.117	0.374	0.369
2	0.004	0.116	0.373	0.369
3	0.004	0.119	0.372	0.368
4	0.004	0.120	0.371	0.367
5	0.004	0.121	0.364	0.359
6	0.004	0.121	0.359	0.355
7	0.004	0.122	0.365	0.360
8	0.004	0.121	0.361	0.357
9	0.004	0.125	0.355	0.351
Large	0.005	0.128	0.350	0.345

Panel B: RETURN					
	Low	2	High	DIF	t(DIF)
Small	0.020	0.022	0.013	-0.008	0.846
2	0.027	0.011	0.015	-0.013	0.988
3	0.012	0.016	0.010	-0.002	0.617
4	0.016	0.011	0.010	-0.006	0.913
5	0.015	0.019	0.019	0.004	0.228
6	0.014	0.018	0.014	0.000	0.467
7	0.016	0.016	0.013	-0.004	0.773
8	0.013	0.012	0.017	0.004	0.082
9	0.011	0.011	0.012	0.001	0.120
Large	0.017	0.020	0.014	-0.003	0.543

Table 3.8: Fama-French Three Factor Regressions

Zero Investment PIN-Size Portfolios										
	Small	2	3	4	5	6	7	8	9	Large
$\hat{\vartheta}_p^1$	-0.004 (-0.58)	-0.009 (-1.55)	0.000 (-0.09)	-0.008 (-1.61)	0.005 (1.00)	-0.003 (-0.50)	-0.004 (-0.82)	0.005 (1.31)	-0.002 (-0.46)	-0.010 (-1.83)
$\hat{\vartheta}_p^2$	-0.241 (-1.24)	-0.164 (-1.12)	0.287 (2.29)	-0.037 (-0.30)	-0.045 (-0.33)	-0.257 (-1.98)	-0.083 (-0.72)	-0.090 (-0.85)	0.050 (0.48)	0.026 (0.19)
$\hat{\vartheta}_p^3$	(-0.36)	(-0.27)	(-0.33)	(0.00)	(-0.16)	(0.19)	(0.00)	(-0.05)	(0.19)	(0.47)
$\hat{\vartheta}_p^4$	-1.48 (-0.75)	-1.45 (-0.83)	-2.09 (1.35)	0.01 (1.89)	-0.89 (0.34)	1.17 (1.68)	0.02 (0.18)	-0.39 (-0.51)	1.43 (1.07)	2.80 (1.63)
	0.077 (0.75)	-0.065 (-0.83)	0.090 (1.35)	0.126 (1.89)	0.025 (0.34)	0.116 (1.68)	0.011 (0.18)	-0.029 (-0.51)	0.059 (1.07)	0.115 (1.63)
F-stat	2.21	2.38	2.31	1.54	0.47	3.58	0.24	0.56	1.97	5.76
Adj. R ²	0.02	0.03	0.03	0.01	-0.01	0.05	-0.02	-0.01	0.02	0.09
GRS Test Statistics	1.59									
p-value(GRS)	0.12									

3.6. Conclusion

Standard asset pricing models ignore the impact of information asymmetry on equity returns. In finance literature, significant number of studies focus on asset pricing anomalies that standard models cannot explain. Previous work on the role of information asymmetry provide conflicting results.

The rational expectations model of Easley and O'Hara (2004) expects uninformed investors demand premium for holding stocks that have high information asymmetry. This is due to the fact that uninformed investors are unable to adjust their portfolios quickly to the new information arrive at the market. On the contrary, Hughes et al. (2007) argue that, in a market with sufficient number of securities, only the market-wise information asymmetry will be priced and therefore firm specific information risk is idiosyncratic in nature.

Numerous empirical studies have been conducted to test the hypotheses of these two opposing views. Easley et al. (2002) and Easley et al. (2010) find that the information asymmetry proxied by PIN is a systematic risk factor for U.S. markets. On the contrary, Duarte and Young (2009) and Lai et al. (2014) questions the systematic impact of PIN factor.

Recently, the studies that examine the role of information asymmetry in BIST significantly increased. For instance, Aktas (2016) show that the short sellers have higher price impact for BIST30 firms between 2008-2009. Similarly, Simsir and Simsek (2018) documents an abnormal volume of trades right before an announcement made public in public disclosure platform. In the next chapter of this thesis, we focus on the information asymmetry between domestic and foreign investors in BIST and underline informational advantage of domestic investors over the foreigners. This chapter, is the first attempt in exploring the *systematic impact* of informed trading in equity returns in BIST.

We obtain the PIN estimates for all stocks that are traded in BIST between March 2005 and April 2017. To estimate the PIN, we use the factorization provided by Duarte and Young (2009) and a grid-search algorithm similar to Yan and Zhang (2012). We examine the relationship between returns and PIN in a Fama and MacBeth (1973) framework, along with univariate and multivariate portfolio analysis.

Our results are inline with the theoretical expectations of Hughes et al. (2007) and the empirical findings of Lai et al. (2014). We find that the information asymmetry proxied by PIN, is an idiosyncratic risk and therefore it can be eliminated with portfolio diversification. Even though we observe a predictive power of PIN on future returns, the impact is economically insignificant. The predictive power of PIN disappears in bivariate regressions controlling for other factors such as preference towards lottery stocks documented by Alkan and Guner (2018). With univariate and multivariate portfolio analysis, we mimic the portfolio of an investors that has long position in high-PIN stocks and a short position of equal size in low-PIN stocks. We show that investors do not obtain any premium for investing in these zero investment PIN portfolios. Our results do not depend on trade classification algorithms, different factorizations in estimating PIN. Similarly, our estimates are not effected by any bias due floating point exception, boundary solutions and systematic order flow shocks.

CHAPTER IV

POLITICAL TURMOIL AND THE IMPACT OF FOREIGN ORDERS ON EQUITY PRICES

4.1. Introduction

This chapter investigates the effect of political instability on the information content of foreign and local investor trades in the Turkish stock market. Foreign investors are crucial, especially for emerging economies which have large current account deficits. According to the IMF, in 2017, top twenty countries that have the highest current account deficits include fifteen emerging economies¹. During that year, Turkey had a current account deficit of about \$39 billion, ranking highest among the emerging markets (International Monetary Fund, 2017).

Recent changes in the Turkish political landscape make the economy vulnerable to sudden foreign outflows. Acemoglu and Ucer (2015) indicate that the weakening institutional strength is the underlying reason behind the lower-paced and lower-quality growth since 2007. They argue that the weakening independence of economic institutions and the judiciary system is a consequence of the changing political dynamics in the country.

¹In that list, top three countries are developed economies. U.S. has the highest current account deficit with almost half a trillion dollars, four times the size of U.K which has the second largest current account deficit. Canada has the third place.

The first signs of turmoil in Turkish politics came to the forefront with the June 2013 Istanbul Gezi Park protests². Since then, credit rating agencies such as S&P, Moody's and Fitch continually downgraded either the outlook or the rating of Turkish sovereign debt, mainly on the basis of increased concerns over continuous political instability in foreign and domestic affairs. More recently, in March 2018, Moody's downgraded the credit rating of Turkish sovereign debt and emphasize the increase in the probability of sudden and disruptive foreign outflow due to faltering institutional strength³.

Foreign portfolio flows into the Turkish equity market, Borsa Istanbul (BIST), constitute 6 percent of the average annual total inflows into the economy and, as of 2016, foreigners dominate the Turkish stock market, holding almost 65 percent of the publicly traded shares in BIST (Borsa Istanbul, 2016). Consequently, examining foreign investors' role is crucial for understanding the price discovery process in emerging economies like Turkey, which heavily rely on outside financing.

This chapter is motivated by the academic debate on whether foreign investors possess an informational advantage over domestic investors. On the one hand, foreign investors are usually large institutional investors with significant amount of investment experience and expertise in international markets. Therefore, they may be better informed compared to domestic investors (Froot and Ramadorai, 2008). On the other hand, foreign investors face with linguistic and cultural barriers and may be at a disadvantage compared to local investors due to differences in accounting standards and country-specific legal environments (Choe et al., 2001).

²For the time line of Gezi Park Protests, see " Turkey protests: How events unfolded", *BBC News*, June 2013, <https://bbc.in/2tC5Du0>.

³See "Moody's downgrades Turkish debt", *Financial Times*, March 2018, <https://on.ft.com/2N6d6JR>. Recently on May 2018, in an interview with <https://www.youtube.com/watch?v=nLkpolAXtpUBloomberg>, President Recep Tayyip Erdogan argue that he wants to increase his role in monetary policy even further. In this interview, he also assert his desire to lower the interest rates. However, couple of days later Turkish Central Bank had to increase the interest rates by 300bps (see, *Reuters*, <https://reut.rs/2IwKAXy>) to stop the sharply declining Turkish Lira from falling further.

Therefore, domestic investors may be better at acquiring information about the underlying value of the asset.

To test the validity of these arguments in an emerging economy context, we build on the Glosten and Harris (1988) framework and estimate the price impact specific to each investor type. Glosten and Harris (1988) model separates price changes into a transitory component (e.g. order processing and inventory management costs) and an adverse selection component (capturing permanent price changes and informed trading). We further disaggregate the two components of the model to provide separate estimates for the transitory and permanent impact associated with local and foreign trades.

Using the intraday order book data for all stocks listed on BIST between January 2007 and November 2015, we test the baseline hypothesis that local investors have no informational advantage over foreign investors. Based on our analysis, we reject the null hypothesis. In particular, we find that domestic investors have a significant price impact on 153 stocks, which corresponds to 43 percent of our sample firms. Foreign investors, on the other hand, have a significant price impact on 24 stocks, corresponding to seven percent of the sample firms. Taken together, these findings suggest that foreign investors are at an informational disadvantage in Borsa Istanbul in terms of the number of stocks they impact. In addition, our results show that, on average, foreign investors tend to generate buy pressure on Turkish stocks, whereas domestic investors tend to generate sell pressure during our sample period.

In the second part of our analysis, we examine if and how the increased political uncertainty in Turkey affects the price impact of local and foreign investors. This part of our investigation is motivated by prior literature, which examines the role of political uncertainty in financial markets. In particular, Pástor and Veronesi (2013) demonstrate that uncertainty about government's actions can lead to

increased risk premiums even if the uncertainty is unrelated to economic shocks. Other studies show that increased political instability raises the cost of equity (Brogaard and Detzel, 2015; Çolak et al., 2017), the cost of debt (Francis et al., 2014; Gao and Qi, 2013) and leads businesses to delay new investments (Julio and Yook, 2012). It also restricts firms' access to new capital, distorts firms' information environment, leading to lower investor demand and higher intermediation costs in equity markets⁴.

Based on this background, we argue that, due to inherent differences in local and foreign investors' exposure to Turkey-specific political uncertainty, the price impact of these investor groups is sensitive to the underlying political environment. Consequently, in the second part of our analysis, we test our second null hypothesis that the increase in political turmoil has no effect on the adverse selection component of foreign and domestic trades. We test this hypothesis first through sub-sample analysis and then using a panel regression framework, which controls for firm-specific factors as well as local and global macroeconomic factors that may affect the price impact of both investor types.

Our results suggest that foreigners' price impact increases during the period of political turmoil, which started with the Gezi protests in June 2013. Through sub-sample analysis, we find that, after June 2013, foreigners have a significant impact on 28 stocks (corresponding to 23 percent of the monthly stock market cap) whereas prior to June 2013, foreign investors significant impact is limited to only seven stocks. Our panel regression analysis also confirms these findings and suggests that the adverse selection component of foreign trades significantly increases following June 2013, even after controlling for firm-specific and macroeconomic factors. The adverse selection component of domestic trades, on

⁴See, Durnev (2010), Boutchkova et al. (2011), Pastor and Veronesi (2012), Pástor and Veronesi (2013), Gungoraydinoglu et al. (2017) among others.

the other hand, does not exhibit a significant sensitivity to political turmoil over our sample period.

Finally, using the panel regression setting, we investigate the contribution of firm specific factors to the information asymmetry between foreign and domestic investors. Our results show that firm specific factors such as size and profitability have a significant role in explaining the adverse selection component of foreign trades. Specifically, we find that foreign investors face smaller adverse selection costs for larger and more profitable firms. However, thin trading, idiosyncratic volatility, systematic risk, global and local economic uncertainty, do not have a significant effect on the temporary or permanent impact of foreign trades. With regards to domestic investors, we find that firm specific characteristics, such as size, profitability, PIN, thin trading or systematic risk do not significantly affect the adverse selection component of local trades. The transitory impact of local investors, however, is larger for stocks that have high idiosyncratic volatility.

Overall, our main results are consistent with the previous findings documented by Choe et al. (2001) and Chan et al. (2008) which argue that the foreign investors are at a disadvantage when trading with domestic investors. We find that political uncertainty partially alleviates the informational disadvantage of foreign investors and creates an opportunity for foreign investors to exert price impact over a larger number of securities. This finding may be attributed to the differences in local and foreign investors' exposure to political uncertainty. Foreign investors are usually better positioned to take advantage of price misalignments that occur during periods of political turmoil since they tend to have more diversified portfolios relative to local investors, who are typically more exposed to domestic shocks and hence more constrained in their ability to take advantage of profitable trading opportunities. This asymmetry may lead to an increase in foreigners' relative price impact during periods of heightened political uncertainty.

Our research contributes to a large literature that examines the impact of foreign flows in equity markets, which has grown significantly following the capital market liberalization wave of the late 1990's. In particular, Froot et al. (2001) investigate the relationship between daily foreign equity flows and market index returns of 44 different countries between 1994 and 1998. In a vector autoregression setting, they show that future returns have positive impulse responses to flows and conclude that foreign flows contain more information about the future value of the equity. Many other studies closely follow this methodology⁵. However, none of these studies provides a direct test for the differences in the information content of the foreign and domestic investor's trades. To the best of our knowledge, this paper is the first microstructure study that explores the effect of political instability on the relative price impact of foreign and domestic orders.

The closest study in spirit to our analysis in terms of utilizing a microstructural approach is Chan et al. (2008). The paper uses the segmented nature of the Chinese stock market, which provides a useful environment for testing the informational advantage (disadvantage) of foreign investors⁶. Until 2001, China had two different equity markets: Shanghai market where A-type shares are traded among domestic investors and Shenzhen market where B-type shares are traded among foreign investors⁷. Chan et al. (2008) find that, on average, B-type shares were traded at 70 percent discount compared to A-type shares⁸. The authors

⁵See Bekaert et al. (2002) for 20 different emerging markets, Dahlquist and Robertsson (2004) for Swedish Market, Ülkü and İkizlerli (2012) for BIST, among many others.

⁶When the markets are segmented, there are two different order books which are unique to their investor groups (foreign or domestic) Chan et al. (2008) exploit this advantage and estimate information asymmetry measures for the two investor types separately. Since in BIST, there is no segmentation, we update the Glosten and Harris (1988) framework to decompose the price impact measure into foreign and domestic components.

⁷When a company went public, a share would be issued in both markets. A-type and B-type shares were identical in every aspect except the investor groups that are trading those shares.

⁸In the segmented market setting, various other studies also show a significant discount in the price of B-type shares, especially for emerging economies. See for example, Bailey and Jagtiani (1994) for stocks traded in Thailand, Domowitz et al. (1997) for stocks traded in Mexico and Chan et al. (2008) for stocks traded in China.

interpret the significant discount in B-type shares as evidence suggesting that foreigners are uninformed traders as they seem to avoid trading in segmented markets⁹. We complement Chan et al. (2008) by examining the research question in a non-segmented market setting in a country that is subject to significant political uncertainty, both of which are common to many emerging economies.

This chapter is organized as follows; Section 4.2 presents the data. Section 4.3 discusses the methodology. Section 4.4 presents our results. Section ?? concludes.

4.2. Data

BIST is a fully computerized order driven market in which the buy and sell orders are matched by price and time priority (Aktas and Kryzanowski, 2014). The opening session is a call auction in which for each stock the limit orders are processed to maximize the executable amount. Aktas and Kryzanowski (2014) indicate that opening prices do not generate new base prices and price limits in the continuous auction market. If there is no opening price for a given stock, all orders transfer to continuous auction session.

We collect intraday quotes and trades of all stocks traded in BIST during January 2007 to November 2015 in one-second frequency. In our quote data set, each incoming order has a date, time, ticker, order ID, order type, quantity and price stamps. In addition, we observe that each order has an intermediary ID which provides information about the brokerage firm which sent the order to BIST¹⁰.

Order ID is not unique. An order can get the same ID, for the same stock or, for different stocks with different order specifications even if they are sent by different

⁹The study further shows that the information asymmetry variables such as price impact (Hasbrouck, 1991), adverse selection component of the bid-ask spread (Glosten and Harris, 1988) and the probability of informed trading (PIN) (Easley et al., 1996) can explain the cross sectional differences in the B-share discount in Chinese stocks.

¹⁰This information is shared in the limit order book data until November 2015.

intermediary, as long as those orders are sent to exchange at the same second.

Order type shows whether the order is a buy or a sell order. This enables us to track the order process for each stock objectively. We label an order as foreign if it is submitted to BIST through foreign intermediaries that have no retail banking operation in Turkey. The list of foreign intermediaries are as follows; BGC Partners (BGC), Credit Suisse (CSM), Deutsche (DSI, DBI), JP Morgan (JPM), JP Morgan Chase Bank (TCM), Merrill Lynch (MLB), Societe Generale (SOG) and UBS (UBS). An average retail investor in Turkey has significantly limited access to these intermediaries as they charge the highest commissions (Borsa Istanbul, 2016).

Local institutional investors such as pension funds and mutual funds that operate in Turkey typically prefer domestic intermediaries as most of them also have a brokerage license or have a sister company that has a brokerage license. Overall, our final dataset includes 480 stocks over our sample period, but our analysis is based on 353 stocks for each there is at least one foreign trade. Foreign orders on average correspond to 5 percent of the aggregate daily volume.

Our trade dataset contains intraday trades for all stocks in BIST. Each trade consists of date, time, ticker, quantity and price stamps. In addition, each trade entry has the ID's of an order on the buy side and an order on the sell side. For each side of the trade, we match the intermediaries who sent the order to the exchange through order ID, stock ticker and order type. Moreover, each trade contains a flag that shows the active and passive side. A side of the trade is flagged as active if the order on that side initiated the trade. Hence, we do not depend on any classification algorithm to assess the direction of the trade¹¹. We classify a trade as a buy trade (sell trade) if the buy side (sell side) of the order is active. Similarly, we label a trade as foreign if a foreign order is on the active side.

Multiple orders that are sent by different intermediaries can be matched with a

¹¹Trade classification algorithms are shown to create bias for BIST (Aktas and Kryzanowski, 2014).

single trade as the order ID's are not unique. In that case, we choose the order that provides the highest volume. In rare instances, we observe multiple orders with the same volume, in which case we assign a random intermediary out of potential candidates¹².

For our panel regressions, we collect firm specific variables such as price, number of shares outstanding, total liabilities, total equity, as well as market index (BIST100) levels from Bloomberg Terminal at monthly frequency. For each month and for each stock, we calculate the firm-specific variables, such as daily return, size, book-to-market ratio, beta, idiosyncratic volatility, Amihud (2002) illiquidity measure, momentum, short-term reversal, total spread and probability of informed trading of Easley et al. (1996). The definitions of these variables are provided in the Appendix¹³. We remove stocks that have missing size or book value of equity, or if it has negative book value of equity for at least one month in our sample.

In addition, we remove infrequently traded stocks from our sample, in particular, stocks for which the median number of daily trades is less than 5 in any given month. We account for ticker changes¹⁴. We exclude all orders that are submitted in the opening session and use the orders that are submitted only within the continuous trading session to avoid any possible heterogeneity in the level of information contained in the orders. Tables 4.1¹⁵ and 4.2¹⁶ provide the descriptive

¹²We observe this problem on average for 0.1 percent of our sample.

¹³In this chapter, we use EHO factorization and YZ algorithm to compute the PIN via InfoTrad package presented in Chapter 2 (Celik and Tiniç, 2016) and (Celik and Tiniç, 2018). This is different that the PIN estimates used in Chapter 3, which use the DY factorization.

¹⁴Ticker changes can be found at <http://www.borsaistanbul.com/en/data/data/equity-market-data/equity-based-data>.

¹⁵This table represents the descriptive statistics for the firm specific factors. We calculate monthly factors for each stocks that are traded in BIST between January 2007 and November 2015. Size is the logarithm of the firm market value. BTM is the book to market ratio. PIN is the probability of informed trading measure of Easley et al. (1996), MAX is the maximum daily return. ILLIQ is the Amihud (2002) illiquidity measure, MOM is the momentum, REV is the short-term return reversal, IVOL is the firm specific risk proxied by the idiosyncratic volatility, and BETA is the systematic risk of a stock proxied by the slope coefficient of the market model. Details about the calculations of firm specific factors provided in the Appendix.

¹⁶This table represents the correlations among firm specific factors. We calculate monthly factors

statistics for the firm specific factors and correlations among them, respectively. The highest positive correlation is 24.4 percent and is observed between momentum and reversal factors. The highest negative correlation is -26.6 percent between momentum and book-to-market ratio.

Table ??¹⁷ provides descriptive statistics for the daily order flows along with the correlations. We observe a 98 percent correlation among total buy and total sell flows. This seems to be driven from strong positive correlation in domestic buy and sell flow. Total buy and total sell flows are almost perfectly positively correlated with domestic flows. Foreign buy and sell flows are also positively correlated. We observe a 63.5 percent correlation among daily foreign buy and sell order flows. These observations are strongly in line with our findings in the previous chapter and the arguments of Duarte and Young (2009) which emphasize the importance of symmetric order flows shocks that move the buy and sell orders together in equity markets.

Figure 4.2¹⁸ provides the time series variation in *net* foreign and domestic flows.

We calculate net flow volume for a day as the difference between buy order volumes

for each stocks that are traded in BIST between January 2007 and November 2015. Size is the logarithm of the firm market value. BTM is the book the market ratio. PIN is the probability of informed trading measure of Easley et al. (1996), MAX is the maximum daily return. ILLIQ is the Amihud (2002) illiquidity measure, MOM is the momentum, REV is the short-term return reversal, IVOL is the firm specific risk proxied by the idiosyncratic volatility, and BETA is the systematic risk of a stock proxied by the slope coefficient of the market model. Details about the calculations of firm specific factors provided in the Appendix.

¹⁷Panel A represents the time series averages of daily transactions of foreign and domestic investors. We calculate the aggregate foreign and domestic inflows and outflows for each stock in each day between January 2007 and November 2015, from the order book. Then, we sum the transaction values in our cross section to obtain the aggregate daily transaction values. All of the values in the table are denominated in millions of Turkish Lira. Panel B represents the pairwise correlations.

¹⁸Net Flow Volume for a given day is calculated as the difference between buy order volume and sell order volume. Panel A represents the time series variation in daily net flows of foreign investors. Panel B represents the time series variation in daily net flows of domestic investors. We calculate the net flow volumes to each stock that is traded on BIST and sum across stocks to calculate the total volume for a given day. Values in the y-axis are in millions of TL. The vertical line on June 2013 divides the sample into a politically stable and a politically unstable periods

Table 4.1: Descriptive Statistics for firm specific factors

Variable	Observation	Mean	Standard Deviation	Minimum	Maximum
SIZE	33947	5.224	1.962	0.157	10.680
BTM	33947	1.103	1.161	0.003	53.398
MAX	33947	0.057	0.038	- 0.114	0.200
ILLIQ	33946	0.066	0.638	0.000	51.300
MOM	33335	- 0.062	0.667	- 73.007	0.953
REV	33522	0.009	0.144	- 0.885	2.941
BETA	33943	0.615	0.528	- 13.068	5.843
IVOL	33943	0.021	0.013	0.000	0.185
PIN	33947	0.137	0.058	0.000	0.450

Table 4.2: Factor Correlations

	SIZE	BTM	MAX	ILLIQ	MOM	REV	BETA	IVOL	PIN
SIZE	1								
BTM	-0.248	1							
MAX	-0.166	0.008	1						
ILLIQ	0.010	-0.018	0.015	1					
MOM	0.134	-0.266	-0.053	-0.012	1				
REV	0.060	-0.101	0.027	-0.015	0.244	1			
BETA	0.155	-0.006	0.141	-0.028	0.006	-0.052	1		
IVOL	-0.258	0.054	0.833	0.036	-0.058	0.096	0.018	1	
PIN	0.006	-0.025	0.151	0.008	0.018	0.070	-0.009	0.200	1

and sell order volumes. Figure 4.2 ¹⁹ provides the time series variation in foreign participation which is calculated as the ratio of foreign volume over total volume for a given day.



¹⁹Daily foreign participation is calculated as the ratio of daily foreign volume over daily total volume. Similarly, foreign buy (sell) participation is calculated as the ratio of daily foreign buy (sell) over daily total buy (sell) volume. The vertical line on June 2013 divides the sample into a politically stable and a politically unstable periods.

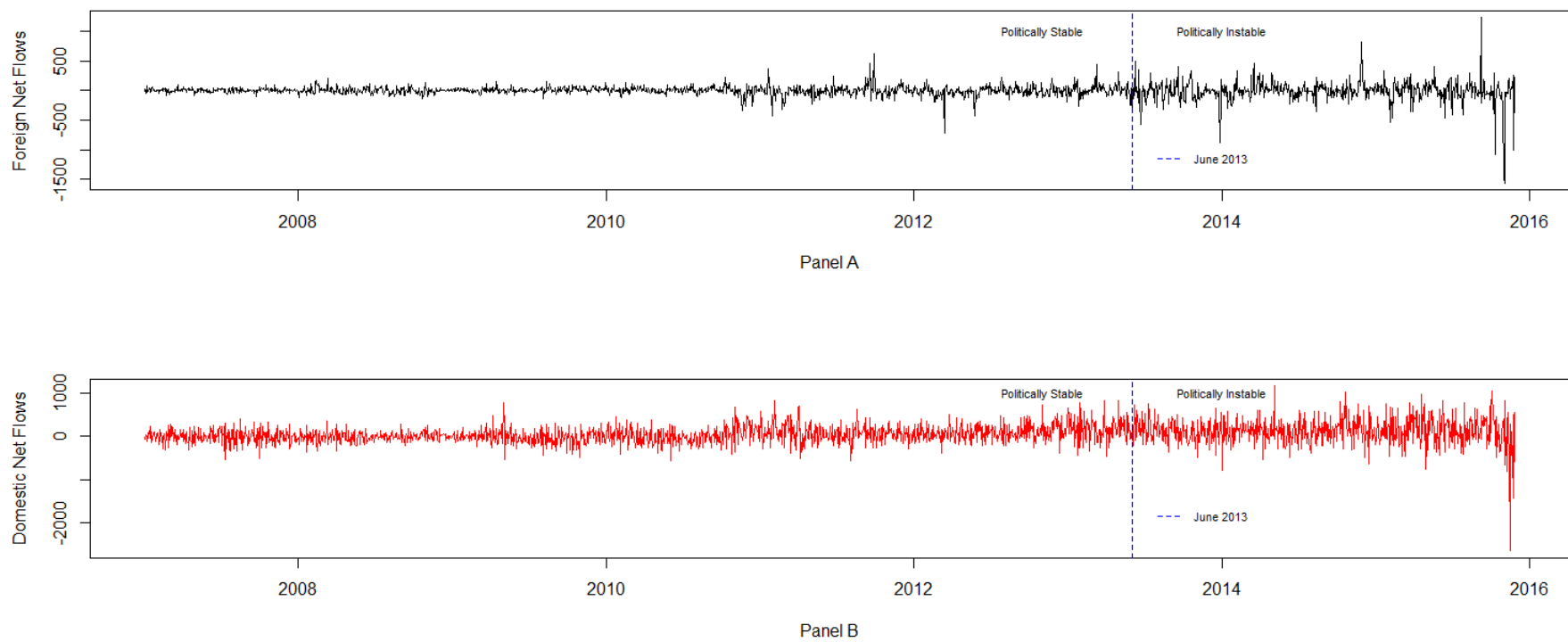


Figure 4.1: Time Series Variation in Daily Volume of Foreign and Domestic Net Flows

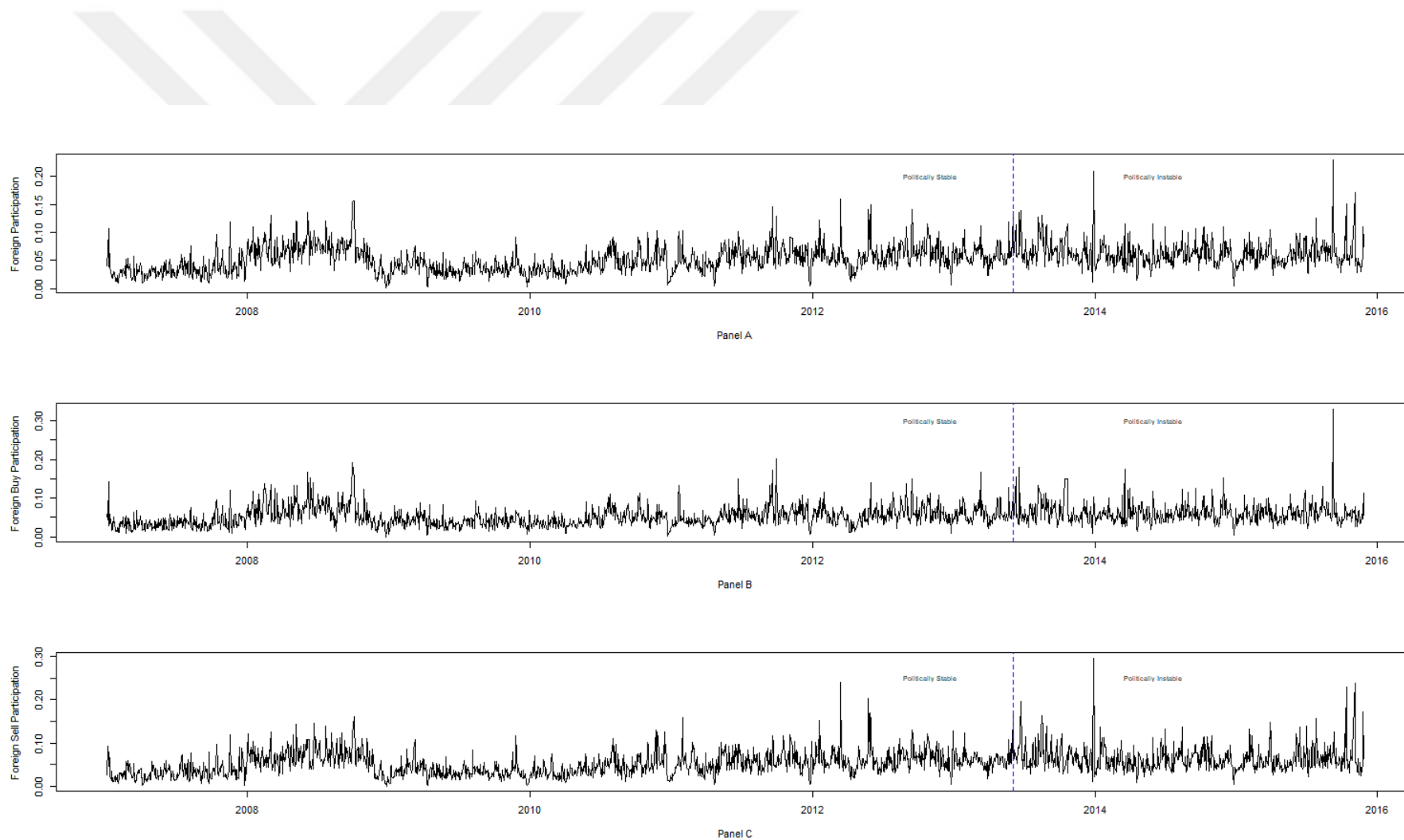


Figure 4.2: Time Series Variation in Daily Foreign Participation

Figures 4.1 and 4.2 provide an interesting observation. Foreign participation in BIST increase when the overall uncertainty increased in the economy. For example, the foreign *buy* participation in BIST reaches its maximum to 33 percent on 8th of September 2015, two days after a terrorist attack in Daglica/Hakkari, the south eastern part of Turkey. The day after the attack, daily foreign *buy* participation also hits 15 percent, second highest throughout the full sample. In addition, on 20th of September 2011, foreign *buy* participation once again hits 14 percent, the third highest, when there is another terrorist attack in the city center of Ankara, the Turkish Capital, during the trading hours. These events seem to lead domestic investors to sell their holdings whereas foreigners enter into BIST to buy. The tendency of foreigners to buy when uncertainty is high might indicate that foreigners are better in assessing the impact of an exogenous event on the intrinsic value of Turkish stocks. It may also indicate that they are better positioned to take advantage of price misalignment during periods of uncertainty as their portfolios tend to be more diversified compared to local investors who are typically more exposed to domestic shocks. Similarly, foreign sell participation reaches its maximum on 27th of December 2013, at a time when the country was going through a significant political crisis due to corruption allegations. Observing maximum foreign inflow and outflow on days where there is high political uncertainty, may also indicate that foreign investors do not shy away from the uncertainty in the Turkish market. It may also be the case that foreigners may exploit their advantage in assessing the impact of exogenous events on firm fundamentals and time their enter and exit in and out of BIST more rationally compared to local investors.

To examine the effect of political turmoil on how foreign and domestic trades impact equity prices, we divide our sample into two: politically stable period (January 2007 - May 2013), and politically unstable period (June 2013 - November

2015). On May 28, 2013, in the midst of the Arap Spring events in the Middle East, Turkey faced its own big civil unrest that started as a protest against the urbanization of Istanbul's Taksim Gezi Park. These protests, later named as "Gezi Park Protests" was the start of a period of increased political instability in Turkey. Between 2013 and 2017, the country faced several political crises both in domestic and in foreign affairs. These crises along with the deteriorating macroeconomic performance (Acemoglu and Ucer, 2015), led Moody's, S&P and Fitch to downgrade the credit ratings of Turkish sovereign debt due to a potential risk of a sudden outflow of foreign investors. After June 2013, the strong influence of the Turkish government over media outlets, the judiciary system, and economic institutions has become evident in the world. From the frequent social media bans to the questions regarding the independence of the Turkish central bank, a wide range of problems related to political oppression has become the center of discussions.

In Figure 4.3²⁰, we provide the Bloomberg Country Risk Political Score for Turkey which supports this claim. Bloomberg Country Risk score measures a country's overall risk across financial, economic and political sectors relative to the performance of other emerging and developed countries. Specifically, Bloomberg provides a political risk score for each country based on news stories about civil unrest through its extensive news analytics platform. We observe a permanent drop in the political index right after the Gezi Park Events which indicate that political risk rose significantly. This sudden change in the political landscape provides an environment in which we can observe the changes in the impact of foreign orders on securities' prices.

²⁰The Bloomberg Country Political Risk Score measures countries political risk relative to the performance of other emerging and developed countries. Political risk score is calculated using a quarterly, equally-weighted, percent-rank model and is measured on a scale of 0-100. Higher scores indicate more stability and less risk. With the extensive news analytics platform, Bloomberg construct this index through news stories about civil unrest.

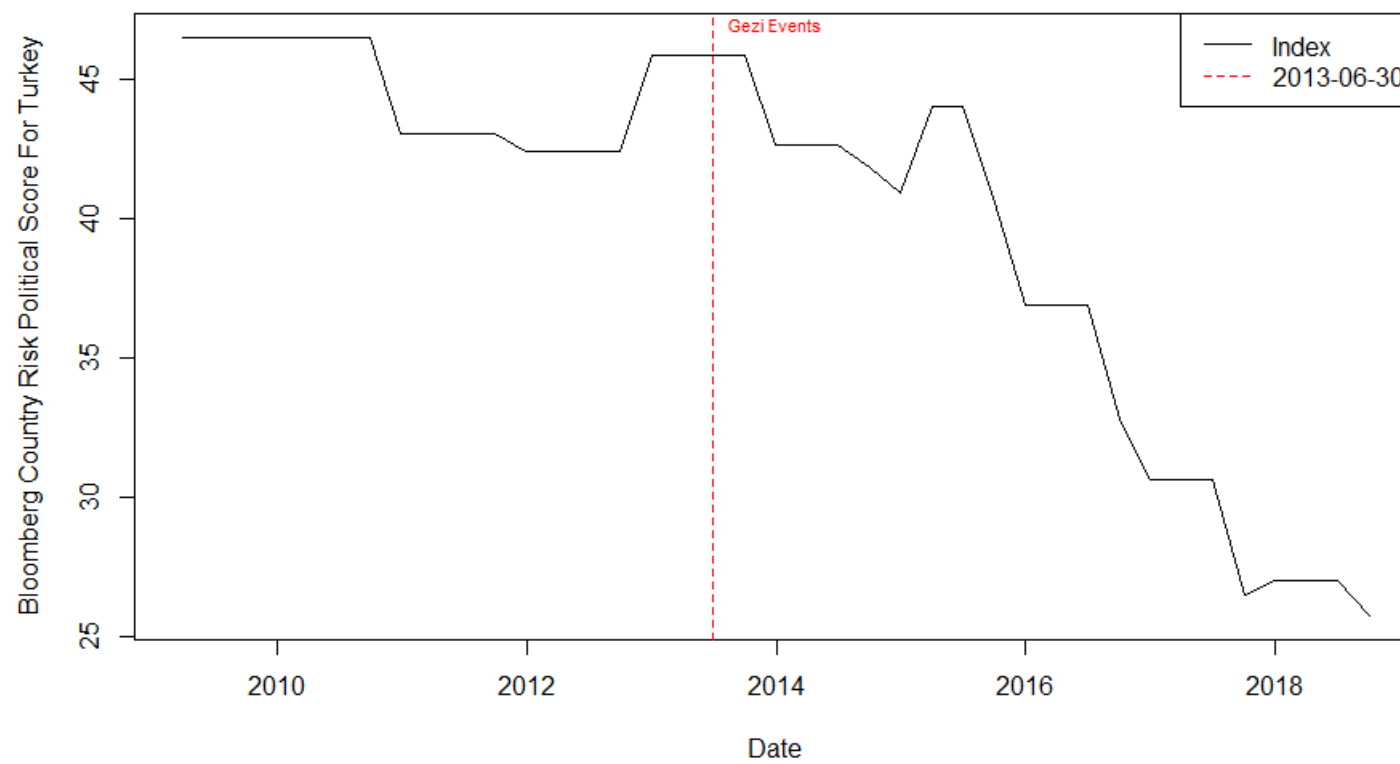


Figure 4.3: Bloomberg Country Risk Political Score for Turkey

Although we label the period between January 2007 and May 2013 as *politically stable*, we are aware of the fact that there were significant global events that may affected the investment behavior of international investors. The global financial crisis has started with the collapse of Lehman Brothers in September 2008 which is followed by a European crisis which ended in May 2013. To account for the effect of the global financial crisis and the following events in the Euro zone, we use the CBOE Volatility Index (VIX), as a measure of global economic uncertainty.

Similarly, Pástor and Veronesi (2013) argue that political uncertainty is state dependent and closely linked with the overall economic uncertainty. To control for the possible economic uncertainty that is generated within Turkey, we use the OECD based Recession Indicators for Turkey. Finally, since elections can generate significant political uncertainty²¹, we also control for local and general elections by generating a dummy in our analysis²².

4.3. Methodology

4.3.1 Estimating the price impact of foreign trades

To estimate the price impact of foreign trades, we closely follow the seminal work of Glosten and Harris (1988). The model uses the trade indicator sequence to separate price changes into a “transitory” component (e.g. order processing and inventory management costs) and an “adverse selection” component associated with permanent price changes and informed trading. We define m_t as the true

²¹For example, as a source of political uncertainty, see the impact of elections on stock prices Santa-Clara and Valkanov (2003), return volatility (Boutchkova et al., 2011), corporate investment cycles (Julio and Yook, 2012), initial public offerings activity (Çolak et al., 2017), financial intermediation costs (Gungoraydinoglu et al., 2017), among others.

²²Between 2007 and 2015, there were six elections in total. 2 of those are local elections which are held on 22nd of March 2009 and on 30th of March 2014. The remaining are general elections which were held on 22nd July 2007, 12th of June 2011, 7th June 2015 and 1st of November 2015, respectively.

value of the underlying security conditional on the information available at time t . m_t changes with the information in the new trade, $Q_t V_t$ where Q_t denotes the direction of the trade, V_t is the trade size in terms of number of shares traded. $Q_t = 1$ if the trade is initiated on the buy side and $Q_t = -1$ if the trade is initiated by the sell order. ϵ_t is the public information innovation. The specifications of Glosten and Harris are provided below;

$$m_t = m_{t-1} + Q_t Z_t + \epsilon_t \quad (\text{True price process}) \quad (4.1)$$

$$P_t = m_t + Q_t C_t \quad (\text{Observed price process}) \quad (4.2)$$

$$Z_t = z_0 + z_1 V_t \quad (\text{Adverse selection component}) \quad (4.3)$$

$$C_t = c_0 + c_1 V_t \quad (\text{Transitory component}) \quad (4.4)$$

where Z_t is the adverse selection component and C_t is the transitory component. The adverse selection component Z_t has a permanent impact on prices since it changes the expectations about the future value of the asset whereas the transitory component C_t has a temporary impact on prices. Specifically, the estimate for z_1 determines the price impact of the trade. The system of equations (4.1) - (4.4) cannot be estimated directly as m_t is not observable. However, using the price difference $\Delta P_t = P_t - P_{t-1}$ we can write the system as;

$$\begin{aligned} \Delta P_t &= m_t + Q_t C_t - m_{t-1} - Q_{t-1} C_{t-1} \\ &= m_{t-1} + Q_t Z_t + \epsilon_t + Q_t C_t - m_{t-1} - Q_{t-1} C_{t-1} \\ &= Q_t Z_t + Q_t C_t - Q_{t-1} C_{t-1} + \epsilon_t \\ &= Q_t(z_0 + z_1 V_t) + Q_t(c_0 + c_1 V_t) - Q_{t-1}(c_0 + c_1 V_{t-1}) + \epsilon_t \\ &= z_0 Q_t + z_1 Q_t V_t + c_0(Q_t - Q_{t-1}) + c_1(Q_t V_t - Q_{t-1} V_{t-1}) + \epsilon_t \\ &= z_0 Q_t + z_1 Q_t V_t + c_0(\Delta Q_t) + c_1(\Delta Q_t V_t) + \epsilon_t \end{aligned} \quad (4.5)$$

We divide the adverse selection and transitory components, Z_t and C_t , into two sub components foreign and domestic by introducing a dummy D_t . $D_t = 1$ if the trade is initiated by a foreign order and $D_t = 0$ if the order is a domestic order. The updated equations are presented below;

$$Z_t = z_0 + z_1 D_t + z_2 V_t + z_3 V_t D_t \quad (\text{Adverse selection component with sub components}) \quad (4.6)$$

$$C_t = c_0 + c_1 D_t + c_2 V_t + c_3 V_t D_t \quad (\text{Transitory component with sub components}) \quad (4.7)$$

Using (4.1), (4.2), (4.6) and (4.7), ΔP_t can be re-written as;

$$\begin{aligned} \Delta P_t &= m_t + Q_t C_t - m_{t-1} - Q_{t-1} C_{t-1} \\ &= m_{t-1} + Q_t Z_t + \epsilon_t + Q_t C_t - m_{t-1} - Q_{t-1} C_{t-1} \\ &= Q_t Z_t + Q_t C_t - Q_{t-1} C_{t-1} + \epsilon_t \\ &= Q_t(z_0 + z_1 D_t + z_2 V_t + z_3 V_t D_t) + Q_t(c_0 + c_1 D_t + c_2 V_t + c_3 V_t D_t) \\ &\quad - Q_{t-1}(c_0 + c_1 D_{t-1} + c_2 V_{t-1} + c_3 V_{t-1} D_{t-1}) + \epsilon_t \\ &= z_0 Q_t + z_1 Q_t D_t + z_2 Q_t V_t + z_3 Q_t V_t D_t \\ &\quad + c_0(\Delta Q_t) + c_1(\Delta Q_t D_t) + c_2(\Delta Q_t V_t) + c_3(\Delta Q_t V_t D_t) + \epsilon_t \end{aligned} \quad (4.8)$$

Equation (4.8) decomposes the aggregate price impact measure into two.

Specifically, in equation (4.8), z_0 and z_1 correspond to the constant term in the adverse selection component. Similarly, z_2 corresponds to the price impact of domestic trades and z_3 corresponds to the price impact of foreign trades. That is, we separate the adverse selection component into two parts; $Z_t = Z_t^D + Z_t^F$, where $Z_t^D = z_0 + z_2 V_t$ and $Z_t^F = z_1 D_t + z_3 V_t D_t$. In addition, equation (4.8) also enables us to disaggregate the transitory component of foreign and domestic flows, which in turn helps us to analyze the per-share cost of providing liquidity for different investor groups. c_0 is the baseline spread for domestic orders whereas c_1

corresponds to the baseline spread for the foreign orders. c_2 and c_3 correspond to the influence of order size of domestic and foreign orders, respectively, which are not driven by the adverse selection. That is, we separate the transitory component into two parts; $C_t = C_t^D + C_t^F$, where $C_t^D = c_0 + c_2V_t$ and $C_t^F = c_1D_t + c_3D_tV_t$. Notice that total spread for a transaction still equals to $2C_t + 2Z_t$.

Glosten and Harris (1988) argue that the constant term, z_0 in the adverse selection components presented in equation (4.3) should be zero since small orders are unlikely to change the expectations about the underlying value of an asset and therefore they are unlikely to be initiated by informed investors. The empirical findings of Glosten and Harris (1988) and Lin et al. (1995) confirm this expectation for U.S. equity market. Osler et al. (2011) also take the constant term of the adverse selection component of the spread to be zero for different types of investors in currency markets. These findings imply that the constant terms z_0 and z_1 in our model should be expected to equal zero. However, one might also expect z_0 and z_1 to be significantly different from zero, as argued by O'Hara et al. (2014). This is because odd lots (trades of size less than 100 stocks) correspond to 35 percent of the price discovery in the U.S and informed investors may use odd lots to avoid detection.

With regards to the expected signs of quantity related variables, Kyle (1985), Glosten and Milgrom (1985) and Easley and O'Hara (1987) suggest that the adverse selection component of an order, Z_t , should be increasing with the quantity of orders. This is because the informed traders want to maximize their returns since their informational advantage disappears overtime as more information becomes available to the public. Therefore, these studies predict that slope terms z_2 and z_3 should be positive.

Finally, equation (4.8) enables us to compare per share cost of supplying liquidity for domestic and foreign investors, separately. If liquidity supplying cost is constant

for both investor groups, we expect c_0 and c_1 to be positive and c_2 and c_3 to be zero. If, on the other hand, the cost is increasing (decreasing) with the transaction size, then the slope terms c_2 and c_3 should be positive (negative).

4.3.2 Panel Regressions

In this section, we examine the time series and cross sectional variation in the estimates of transitory and adverse selection components. To this end, we estimate equation (4.8), separately for each stock in each month. That is, we obtain 107 estimate for each coefficient in (4.8), if a stock has a foreign trading in each of the months in our sample. Our framework consists of three simultaneous panels. The first two regression equations explain the transitory and adverse selection components of domestic and foreign investor trades. In the third regression, we also estimate the number of daily trades for each stock in our sample.

To determine the explanatory variables in our model, we rely on prior literature. Ho and Stoll (1981) argue that trading frequency can be an important determinant of the inventory costs associated with a trade. According to the model to mitigate these costs, the transitory spreads are expected to increase with the overall risk of a stock and the time between trades. Therefore, for a given stock i , in a given month t , we use the inverse of the median values of the number of daily trades (INVNT) as a proxy for trading frequency, similar to Glosten and Harris (1988). We use idiosyncratic volatility (IVOL) to proxy for firm's idiosyncratic risk and beta (BETA) to proxy for firm's systematic risk.

To capture the impact of political uncertainty, we introduce two dummies. The first dummy is $GEZI_t$ which takes the value 1 for each month after the Gezi Park Events in June 2013. In addition, we use $ELECTION_t$ which takes the value 1 for every month in an election year. We also control for local and global economic

uncertainty. For local economic uncertainty we use the OECD recession indicator for Turkey; $RECESSION_t$ (Federal Reserve Bank of St. Louis, 2018). For global economic uncertainty, we use monthly levels of VIX . Hence our model for the transitory component takes the following form;

$$C_{it}^{F,D} = \gamma_1 + \gamma_2 INVNT_{it} + \gamma_3 IVOL_{it-1} + \gamma_4 BETA_{it-1} + \gamma_5 VIX_t + \gamma_6 RECESSION_t + \gamma_7 GEZI_t + \gamma_8 ELECTIONS_t + \epsilon_{it}^1 \quad (4.9)$$

To identify the determinants of the adverse selection component, we rely on Glosten (1987) which suggests that adverse selection is a function of the transitory component of the spread, the liquidity trading, and the percentage of informed trading.

Glosten and Harris (1988) argue that the transitory spread and the adverse selection spread should be positively related. The increase in the transitory spread will crowd out the liquidity traders and hence increase the proportion of informed traders which will result in an increase in the adverse-selection component. Therefore we expect a positive relationship between the adverse selection component and transitory component.

To see the impact of the percentage of informed trading on adverse selection component, we use PIN (Easley et al., 1996). Fundamental microstructure studies show that informed investors face with larger transaction costs in equity markets (Easley and O'Hara, 1987; Glosten and Milgrom, 1985). Therefore, we expect a positive relationship between the adverse selection component and PIN.

Moreover, Aslan et al. (2011) show that the firm specific variables such as size and profitability can significantly capture the variation in the information asymmetry proxied by PIN. However, in addition to this indirect effect through PIN, there

may be a direct effect of firm size and profitability on spreads. This is why, we control for the firm specific variables such as market capitalization (SIZE) and book to market ratio (BTM). For example, Easley et al. (2010) and Aslan et al. (2011) show that the proportion of informed trading decreases uniformly with the market value of the firm. Therefore we expect adverse selection component to decrease as the firms market value increase. Moreover, Aslan et al. (2011) documents a negative relationship between profitability and informed trading. Therefore, we expect adverse selection component to be smaller for profitable firms. The model for the adverse selection component is then given as below;

$$Z_{it}^{F,D} = \nu_1 + \nu_2 C_{it}^{F,D} + \nu_3 SIZE_{it-1} + \nu_4 BTM_{it-1} + \nu_5 PIN_{it-1} + \epsilon_{it}^2 \quad (4.10)$$

Finally, we also take the median number of trades for a given stock in a given month as endogenous. We expect the median number of trades to be negatively related with the total spread (TOTS), since the large spread reduces the attractiveness to invest in a security. In addition, we consider the overall illiquidity of the stock by using the Amihud (2002) measure. We expect the number of trades to increase as the stock becomes more and more liquid, indicating a negative relationship. We also consider the past return characteristics by examining the previous months return (REV), the maximum daily return in the previous month (MAX) and the momentum of returns (MOM) as in Jegadeesh and Titman (1993). We expect a positive relationship between return characteristics and the number of trades, assuming that the recent stock performance will attract investors attention in BIST (Alkan and Guner, 2018). We simultaneously estimate this framework using three stage least squares.

$$NT_{it} = \alpha_1 + \alpha_2 TOTS_{it} + \alpha_3 MAX_{it-1} + \alpha_4 ILLIQ_{it-1} + \alpha_5 MOM_{it-1} + \alpha_6 REV_{it} + \epsilon_{it}^3 \quad (4.11)$$

4.4. Results

4.4.1 Average price impacts

We start by estimating equation (4.8) for the entire sample that covers the period between January 2007 and November 2015. Table 4.4²³ - Panel A, provides the cross sectional distribution of each of the coefficient estimates obtained through full sample analysis. During this period, our sample consists of 480 stocks. We observe a foreign order for 353 different stocks. The median price impact of foreign (domestic) orders, z_3 (z_2), is positive (negative). This indicates that on average foreign (domestic) investors generate buy (sell) pressure in BIST. The coefficient that captures the impact of order quantity z_2 is significant for 153 stocks in our sample which corresponds to 43 percent.

The impact of foreign investors is limited as they have a significant price impact, z_3 , for 24 stocks, which corresponds to 7 percent of the sample firms. List of these

²³This table represents the cross-sectional mean of the estimated adverse-selection and transitory components in our sample. Our sample consist of 480 stocks that are traded in BIST over January 2007 and November 2015. Foreign trading is observed for 353 stocks during this period. The estimates are obtained from the following model: $\Delta P_t = z_0 Q_t + z_1 Q_t D_t + z_2 Q_t V_t + z_3 Q_t V_t D_t + c_0(Q_t - Q_{t-1}) + c_1(Q_t D_t - Q_{t-1} D_{t-1}) + c_2(Q_t V_t - Q_{t-1} V_{t-1}) + c_3(Q_t V_t D_t - Q_{t-1} V_{t-1} D_{t-1}) + \epsilon_t$ where Q_t is the order sign which takes a value of 1 if the order is a buy order and -1 if the order is a sell order, V_t is the order size, ΔP_t is the order price change and ϵ_t is the unobserved changes in price due to public information. z_0 and z_1 correspond to domestic and foreign constant terms for the adverse selection component, respectively. z_2 and z_3 are the respective domestic and foreign slope terms for the adverse selection component. c_0 and c_1 are the foreign and domestic influence of order size on spreads that are not driven by adverse selection. c_2 and c_3 slope terms of the transitory component of the spread. We ignore rounding error due to price discreteness. Panel A shows the results for the full sample. Panel B shows the mean estimates for the stable period. Panel C shows the mean estimates for the politically unstable period.

stocks can be found in Table 4.5²⁴. Commercial banks account for about a quarter of this list (Akbank (AKBNK), Garanti Bank (GARAN), Halkbank (HALKB), İş Bank (ISCTR), Vakıfbank (VAKBN) and YapıKredi Bank (YKBNK)). Foreign investors significant price impact on commercial bank stocks may arise due to these banks' foreign equity ownership or lending relationship with their foreign counterparts.

Table 4.4 - Panel B, provides the cross sectional mean of each of the coefficient estimates obtained for the politically stable period. In this period, 443 different stocks have been traded in BIST. We observe a foreign order for 312 different stocks. Unlike the full sample, between January 2007 and June 2013, foreigners generate sell pressure similar to domestic investors. Foreign investors have a significant price impact for only 7 stocks out of 312 which correspond to 2.2 percent of the overall sample. That is, on average, the impact of foreign investors is less prominent during the politically stable period. Those 7 stocks on average correspond to 10 percent of the average monthly market capitalization between January 2007 and May 2013.

Table 4.4 - Panel C, provides the cross sectional mean of the coefficient estimates obtained for the politically unstable period. The overall number of stocks traded is 455. We observe a foreign order for 277 different stocks. During this period, the signs of the coefficients for domestic and foreign investors, z_2 and z_3 , are same with the full sample. During the period of political instability, on average, foreigners tend to generate buy pressure whereas domestic investors generate sell pressure.

²⁴This table represents the list of stocks that foreigners have significant price impact. Foreign price impact is given by the coefficient z_3 in equation (4.8). First column represents the full sample between January 2007 and December 2015. Second column represents the politically stable period between January 2007 and May 2013. Final column represents the politically unstable period between June 2013 and December 2015. Last row shows the average monthly market share of stocks. Average monthly market share is the time series average of each months market share which is calculated as ratio between corresponding stocks market value over total market value.

Table 4.4: Full sample estimates for price impact

Panel A: (January 2007 - November 2015)									
	z0	z1	z2	z3	c0	c1	c2	c3	
Median	0.00064	-0.00004	-0.00283	0.00257	0.00482	-0.00206	-0.06598	0.03945	
NSignificant	351	63	153	24	353	194	320	79	
N Positive	353	163	117	195	351	70	40	251	
N Available	353	353	353	353	353	353	353	353	
N Total	480	480	480	480	480	480	480	480	
Panel B: (January 2007 - May 2013)									
	z0	z1	z2	z3	c0	c1	c2	c3	
Median	0.00053	0.00016	-0.00071	-0.00639	0.00487	-0.00248	-0.08007	0.05486	
NSignificant	310	36	112	7	312	181	294	62	
N Positive	312	208	135	125	311	43	32	237	
N Available	312	312	312	312	312	312	312	312	
N Total	443	443	443	443	443	443	443	443	
Panel C: (June 2013 - November 2015)									
	z0	z1	z2	z3	c0	c1	c2	c3	
Median	0.00067	-0.00043	-0.00021	0.00642	0.00439	-0.00019	-0.07018	0.02791	
NSignificant	274	103	105	28	277	118	250	57	
N Positive	277	58	131	162	275	128	34	186	
N Available	277	277	277	277	277	277	277	277	
N Total	455	455	455	455	455	455	455	455	

Table 4.5: List of stocks that foreigners have significant price impact

	Full Sample	Pre Gezi	Post Gezi
1	AKBNK	DAGHL	AKGUV
2	BAKAN	EGYO	AKSEN
3	BANVT	GARAN	BAKAN
4	BIZIM	HALKB	BOLUC
5	BOLUC	OTKAR	CLEBI
6	DAGHL	TMSN	DGGYO
7	DGATE	TUMTK	EKGYO
8	EGYO		EREGL
9	GARAN		FFKRL
10	HALKB		GARAN
11	ISCTR		HALKB
12	KRDMD		INDES
13	LOGO		ISGYO
14	NTHOL		KOZAA
15	ODAS		MNDRS
16	OTKAR		ODAS
17	PNSUT		PARSN
18	TMSN		PETKM
19	TUKAS		SISE
20	TUMTK		SNGYO
21	ULUSE		TACTR
22	VAKBN		THYAO
23	VERUS		TIRE
24	YKBNK		TUKAS
25			ULUSE
26			VAKBN
27			VERUS
28			YKBNK
Average Market Share			0.23

Foreign orders generate significant price impact for 28 different stocks. We observe no significant changes in the transitory impact of both domestic and foreign orders.

Overall, our analysis suggests that the foreign (domestic) investors on average generate buy (sell) pressure as they have positive (negative) price impact during the period of political stability. Domestic investors are the main driver of the adverse selection component of the spread. However, we observe a significant foreign impact during times of political instability where foreign investors on average generate buy pressure whereas domestic investors enter into market to sell. This finding may seem counter-intuitive, at first, from an information asymmetry standpoint. Foreigners, who are, on average, at an informational disadvantage relative to local investors, may be expected to avoid further uncertainty and shy away from BIST ((Chan et al., 2008),(Choe et al., 2001)). However, BIST may provide diversification benefits (costs) to international (local) investors since domestic investors are more exposed to country specific risks due to their inherent home bias. Therefore, domestic investors may be more inclined to sell when they face a downturn in the political landscape. As the country specific shocks become more prominent due to increased political uncertainty, BIST can serve as a diversification tool for foreign investors. Therefore, foreigners may be better positioned to exploit any price misalignments that may result from the overreaction of domestic investors during politically unstable periods.

4.4.2 Panel Regressions

We construct a panel regression setting to test the effect of increased political uncertainty over the price impact of foreign and domestic trades. Table 4.6²⁵

²⁵Descriptive statistics for monthly coefficients obtained from estimating the following equation; $\Delta P_t = z_0 Q_t + z_1 Q_t D_t + z_2 Q_t V_t + z_3 Q_t V_t D_t + c_0(Q_t - Q_{t-1}) + c_1(Q_t D_t - Q_{t-1} D_{t-1}) + c_2(Q_t V_t - Q_{t-1} V_{t-1}) + c_3(Q_t V_t D_t - Q_{t-1} V_{t-1} D_{t-1}) + \epsilon_t$ where Q_t is the order sign which takes a value of 1 if the order is a buy order and -1 if the order is a sell order, V_t is the order size, ΔP_t is the order price

provides the descriptive statistics for the monthly estimates of (4.8). This setting enables us to control for firm specific factors such as size, profitability, informed trading, liquidity and recent return performance. In addition, we control for Turkey's overall macroeconomic conditions through OECD's recession dummy. Moreover, we proxy global uncertainty through VIX. The results are presented in Table 4.7²⁶.

Our results indicate that the political turmoil has a significant effect on the temporary impact of foreign investors in BIST. In fact, only the political uncertainty measures seem to have a significant impact over transitory component of foreign trades in BIST. Even after controlling for global and local macroeconomic

change and ϵ_t is the unobserved changes in price due to public information. z_0 and z_1 correspond to domestic and foreign constant terms for the adverse selection component, respectively. z_2 and z_3 are the respective domestic and foreign slope terms for the adverse selection component. c_0 and c_1 are the foreign and domestic influence of order size on spreads that are not driven by adverse selection. c_2 and c_3 slope terms of the transitory component of the spread. We ignore rounding error due to price discreteness. We estimate the model for each stock on each month between January 2007 and November 2015.

²⁶This table represents the estimates obtained from the panel regressions of the 3-equation model. The model concern the time-series estimates transitory and adverse selection spread components to a set of proxies for security risk, adverse selection risk and trading activity. First two equations are for the transitory components of domestic and foreign trades, respectively. Third and forth equations are for the adverse selection components of domestic and foreign trades, respectively. The transitory component expected to increase with the security risk and with thin trading. We proxy the security risk in two different form; idiosyncratic risk (idiosyncratic volatility) and systematic risk (monthly beta estimated from daily returns). We capture the overall political instability with Gezi and Election dummies. $Gezi_t$ is a dummy variable that takes value 1 for months after May 2013. Similarly $Elections_t$ is a dummy variable that takes a value 1 for months that are in years of local and general elections. We capture the overall economic stability with Recession dummy. $Recession_t$ is the OECD based recession indicator. We proxy the thin trading with the inverse of median daily trading in a given month. The adverse selection component is expected to increase with the risk of informed trading proxied by the probability of informed trading (Easley et al., 1996), decrease with the extend of liquidity trading controlled for firms market value and profitability, and increase with the transitory component. The last equation is to model joint dependency. The median number of daily trades is expected to decrease with the overall liquidity and increase with the previous months return and return momentum. We proxy overall liquidity with Amihud (2002) illiquidity measure. We proxy the return characteristics with the return reversal, momentum measure of Jegadeesh and Titman (1993), maximum daily return of the previous month. The system is estimated separately for domestic and foreign measures using three stage least squares with fixed effects. The sample consist of 350 stocks over January 2007 and November 2015. We ignore price discreteness when calculating the transitory and adverse selection components. The values in parenthesis correspond to t-statistics computed from cluster-robust Huber/White standard errors. *** indicates significance at 1% level, ** indicates significance at 5% level, * indicates significance at 10% level.

Table 4.6: Descriptive Statistics for the monthly coefficients

Coefficients	Observation	Mean	Standard Deviation	Minimum	Maximum
z_0	33947	0.00244	0.011	- 0.138	0.531
z_1	13880	-0.00090	0.078	- 3.782	4.307
z_2	33947	0.00004	0.001	- 0.008	0.067
z_3	13485	-0.00008	0.006	- 0.483	0.092
c_0	33946	0.01471	0.067	- 2.881	1.921
c_1	13880	-0.00387	0.049	- 0.981	3.438
c_2	33947	-0.00005	0.001	- 0.063	0.019
c_3	13337	0.00005	0.002	- 0.031	0.233
Z^D	33947	0.00250	0.012	- 0.058	0.533
Z^F	13485	-0.00128	0.055	- 3.357	1.535
Z	33947	0.00199	0.037	- 3.356	1.540
C^D	33946	0.01449	0.066	- 2.878	1.914
C^F	13337	-0.00363	0.040	- 0.959	2.778
C	33947	0.01306	0.069	- 2.878	3.016
$TOTS$	33947	0.03011	0.156	- 6.683	4.016
Other Variables	Observation	Mean	Standard Deviation	Minimum	Maximum
Median Price	33947	11.00	46.88	0.11	1630
Median Trade (NT)	33947	621.74	934.44	6	15937
Median Volume	33947	377.52	528.08	1	21664

Table 4.7: Panel regression results

	$C_{i,t}^D$	$C_{i,t}^F$	$Z_{i,t}^D$	$Z_{i,t}^F$	$NT_{i,t}$
Endogenous Variables	<i>Constant</i>	0.0101*** (2.98)	-0.0179* (-1.8)	-0.0162 (-1.16)	0.0236** (2.02)
	$INVNT_{i,t}$	1.05 (0.85)	7.30* (1.66)		454.27*** (35.65)
	$C_{i,t}^D$		0.1044 (1.16)		
	$C_{i,t}^F$			0.9665* (1.96)	
	$TOTS_{i,t}$				-123.61 (-0.79)
Exogenous Variables	$IVOL_{i,t-1}$	0.4049*** (4.22)	0.1324 (0.74)		
	$BETA_{i,t-1}$	-0.0007 (-0.27)	-0.0025 (-0.81)		
	$SIZE_{i,t-1}$			0.0032 (1.35)	-0.0030** (-2.01)
	$BTM_{i,t-1}$			0.0006 (0.96)	-0.0016** (-2.03)
	$PIN_{i,t-1}$			-0.0016 (-0.85)	0.0005 (0.06)
	$MAX_{i,t-1}$				3097.76*** (14.6)
	$ILLIQ_{i,t-1}$				-43.70** (-2.53)
	$MOM_{i,t-1}$				46.30** (2.25)
	$REV_{i,t}$				201.41*** (4.34)
	VIX_t	-0.00005 (-0.54)	0.00003 (0.72)		
	$GEZI_t$	-0.0068* (-1.96)	0.0061** (2.44)		
	$ELECTIONS_t$	0.0058 (1.26)	-0.0051** (-2.04)		
	$RECESSION_t$	-0.0017 (-1.13)	0.0013 (0.81)		
	$GEZI_t * IVOL_{i,t-1}$	-0.0694 (-1.28)	-0.0346 (-0.41)		
	$GEZI_t * BETA_{i,t-1}$	-0.0001 (-0.03)	-0.0007 (-0.4)		
	$ELECTIONS_t * IVOL_{i,t-1}$	-0.2775*** (-2.76)	-0.0078 (-0.06)		
	$ELECTIONS_t * BETA_{i,t-1}$	-0.0022* (-0.62)	0.0054 (1.81)		
	$RECESSION_t * IVOL_{i,t-1}$	-0.0871 (-1.60)	-0.1681 (-1.38)		
	$RECESSION_t * BETA_{i,t-1}$	0.0009 (0.54)	0.0015 (0.59)		
Interactions	Observations	32911	13066	32898	13212
	Fixed Effects	YES	YES	YES	YES
	R Squared	0.0074	0.0057	0.0116	0.0004
					0.0189

conditions, overall increase in the political turmoil seems to increase the temporary impact of foreign investors. Specifically, after the Gezi Park Protests in June 2013, transitory component of foreign trades increases on average by 0.0061(61bps). The permanent impact of foreign trades on securities' prices is also larger after June 2013. In particular, the adverse selection component of an average foreign trade increases by 59 bps ($61\text{bps} * 0.9665$) after June 2013. The permanent price impact of foreign trades also decrease with firm size and profitability. Specifically, a 1million increase in firm's market capitalization will result in a 29bps decrease in the permanent price impact of an average foreign trade. Similarly, one percent increase in firm's book-to-market ratio results in 16bps decrease in the permanent price impact of an average foreign trade. On the other hand, during election years, we observe that the transitory and adverse selection components of the spread for foreign investors, on average, decrease by 51bps. Given that the average total spread in BIST were around 300bps over our sample period, these magnitudes are not only statistically significant, but also economically meaningful as well.

We also show that political turmoil only affects the transitory component of the spread for domestic investors. On average, the temporary impact of local investors decreases by 68bps after June 2013. We do not observe a significant effect of political turmoil over the permanent impact of local trades. Firm specific factors such as thin trading and systematic risk of a stock seem to have no significant impact on transitory component of the spread for domestic investors. Similarly, temporary impact of domestic investors does not significantly depend on local or global economic conditions. On the other hand, the temporary impact of domestic investors is higher for firms that have higher idiosyncratic volatility. One percent increase in firms idiosyncratic volatility increases the transitory component of the spread for domestic investors by 40 percent. Hence, stocks that have high firm specific risk tend attract domestic uninformed traders. On the contrary, the

temporary impact of domestic investors decreases significantly after June 2013²⁷.

In sum, based on the average price impact analysis presented in the previous sub-section, we reject the baseline null hypothesis that domestic investors do not have an informational advantage in BIST. Our results suggest that while locals have information on a wider set of stocks, foreigners' price impact is concentrated on a limited number of stocks. Through sub-sample analysis, we also show that the political uncertainty matters for the price impact both investor types. In particular, the permanent price impact of foreign investors increases during the periods of political turmoil. Finally, using a panel regression framework, we test our main hypothesis and reject the null hypothesis that political uncertainty have no effect on the adverse selection component of foreign trades' impact. We show that even after controlling for global and local macroeconomic conditions, the adverse selection component of foreign trades increases by 59 bps, which confirms our earlier finding that foreign trades contain more information during periods of increased political turmoil. Moreover, we show that firm-specific characteristics such as size, profitability, liquidity and past return performance are important determinants of the adverse selection component of foreign trades. Specifically, foreign investors face smaller adverse selection costs for larger, profitable, liquid stocks that have recently provided positive returns.

4.5. Conclusion

In this chapter, we examine the informational asymmetry between foreign and domestic investors in Borsa Istanbul. To conduct our analysis, we decompose the

²⁷It is interesting to note that the political uncertainty due to elections and due to Gezi Park protests have different impact over foreign trades. As a robustness check, we add another interaction term $GEZI_t * ELECTION_t$ to our model. Although not presented, the coefficients of these interaction terms are positive and insignificant for both local and foreign investors. This indicates that the impact of political uncertainty due to elections disappears after June 2013.

price discovery process into foreign and domestic components using the Glosten and Harris (1988) model. Our results suggest that median price impact for foreign trades is positive whereas the median price impact for domestic orders is negative. Hence, foreigners on average generate buy pressure whereas domestic investors generate sell pressure over equity prices. In addition, we show that domestic investors have a permanent price impact for 43 percent of the stocks in our sample. We also show that foreign investors have a significant permanent price impact for 24 stocks in our sample, corresponding to seven percent of the sample firms, quarter of which are commercial banks. More importantly, we show that the significant price impact of foreign investors mostly stems from the period after June 2013, during which Turkey has faced a significant increase in political turmoil.

Analyzing the effect of increased political instability on the price impact of foreign trades through panel regressions, we find that foreign trades have a larger adverse selection component during times of political turmoil even after controlling for global and local macroeconomic conditions. We show that the firm specific characteristics such as size and profitability also significantly affect the adverse selection component of foreign trades. Specifically, foreign investors face smaller adverse selection costs for larger and more profitable firms in BIST over our sample period.

With regard to local investors, firm specific factors do not have a significant effect on the permanent impact of domestic trades. However, the transitory price impact of domestic orders is higher for stocks that have high firm specific risk. That is, companies that have high firm specific risk attract local uninformed investors. Our results show that these local uninformed investors tend to avoid investing in stocks that have high firm specific risks during election years in which political uncertainty is higher.

Overall, our findings are in line with the predictions of Choe et al. (2001) which suggest that domestic investors are better equipped in acquiring information about the underlying value of an asset as foreign investors are facing certain obstacles when investing abroad. However, we also show that informational advantage of foreign investors tends to increase with increased political instability.

Diversification concerns might be the underlying reason behind this finding. As Turkey's country specific risks become more isolated from the rest of the world, foreign investors who are more exposed to international risks can invest in Turkish stocks to diversify their global portfolio. Finally, the informational advantage of foreign investors on the stocks of commercial banks may arise due to large equity ownership and/or due to lending relationship.

CHAPTER V

CONCLUSIONS

This study focuses on the impact of information asymmetry on the asset prices in Borsa Istanbul. Standard asset pricing models ignore the impact of information asymmetry on equity returns. In finance literature, significant number of studies focus on asset pricing anomalies that standard models cannot explain. Previous work on the role of information asymmetry provide conflicting results.

The rational expectations model of Easley and O'Hara (2004) expects uninformed investors demand premium for holding stocks that have high information asymmetry. This is due to the fact that uninformed investors are unable to adjust their portfolios quickly to the new information arrive at the market. On the contrary, Hughes et al. (2007) argue that, in a market with sufficient number of securities, only the market-wise information asymmetry will be priced and therefore firm specific information risk is idiosyncratic in nature.

Numerous empirical studies have been conducted on hypotheses presented by these two opposing views. Easley et al. (2002) and Easley et al. (2010) find that the information asymmetry proxied by PIN is a systematic risk factor for U.S. markets. On the contrary, Duarte and Young (2009) and Lai et al. (2014) questions the systematic impact of PIN factor.

The first chapter of this thesis provides a short survey on five most widely used

estimation techniques for PIN measure. In this chapter, we introduce the R package **InfoTrad**, covering estimation procedures for PIN using EHO, LK factorizations along with YZ, GAN and EA algorithms (*EHO()*, *LK()*, *YZ()*, *GAN()* *EA()*). The functions *EHO()* and *LK()* read a (Tx2) matrix where the rows of the first column contains total number of buy orders on a given trading day t , B_t , and the rows of the second column contains the total number of sell orders on a given trading day t , S_t , where $t \in \{1, 2, \dots, T\}$. In addition, they also require an initial parameter vector in the form of, $\Theta_0 = \{\alpha, \delta, \mu, \epsilon_b, \epsilon_s\}$. Both functions produce the respective log-likelihood functions.

The functions *YZ()*, *GAN()* and *EA()* read (B_t, S_t) as an input along with a likelihood specification that is set to LK by default. These functions do not require initial parameter matrix to obtain the parameter estimates when calculating PIN. All three functions use *neldermead()* method of **nloptr** as built-in optimization procedure for MLE. *YZ()* *GAN()* and *EA()* produce an object that gives the parameter estimates $\hat{\Theta}$ along with likelihood value and $\widehat{PIN}$. Simulation results indicate that YZ algorithm with EHO factorization provide the most accurate estimates with respect to mean absolute error.

In the third chapter, we examine the relationship between the information asymmetry, proxied by PIN, and the stock returns in Borsa Istanbul. To the best of our knowledge, this study is first the examine the systematic impact of information asymmetry on returns of stocks that are traded solely on Borsa Istanbul. We obtain the PIN estimates for all stocks that are traded in BIST between March 2005 and April 2017. Since we observe a significant and positive covariance between buy and sell order flows, we use the factorization provided by Duarte and Young (2009) and a grid-search algorithm similar to Yan and Zhang (2012), to estimate the PIN. We examine the relationship between returns and PIN in a Fama and MacBeth (1973) framework, along with univariate and multivariate portfolio analysis.

Our results are inline with the theoretical expectations of Hughes et al. (2007) and the empirical findings of Lai et al. (2014). We find that the information asymmetry proxied by PIN, is an idiosyncratic risk and therefore it can be eliminated with portfolio diversification. Even though we observe a predictive power of PIN on future returns, the impact is economically insignificant. The predictive power of PIN disappears in bivariate regressions controlling for other factors such as preference towards lottery stocks documented by Alkan and Guner (2018). With univariate and multivariate portfolio analysis, we mimic the portfolio of an investors that has long position in high-PIN stocks and a short position of equal size in low-PIN stocks. We show that investors do not obtain any premium for investing in these zero investment PIN portfolios. Our results do not depend on trade classification algorithms, different factorizations in estimating PIN. Similarly, our estimates are not effected by any bias due floating point exception, boundary solutions and systematic order flow shocks.

In the last chapter, we examine the informational asymmetry between foreign and domestic investors in Borsa Istanbul. To conduct our analysis, we decompose the price discovery process into foreign and domestic components using the Glosten and Harris (1988) model. Our results suggest that median price impact for foreign trades is positive whereas the median price impact for domestic orders is negative. Hence, foreigners on average generate buy pressure whereas domestic investors generate sell pressure over equity prices. In addition, we show that domestic investors have a permanent price impact for 43 percent of the stocks in our sample. We also show that foreign investors have a significant permanent price impact for 24 stocks in our sample, corresponding to seven percent of the sample firms, quarter of which are commercial banks. More importantly, we show that the significant price impact of foreign investors mostly stems from the period after June 2013, during which Turkey has faced a significant increase in political turmoil.

Analyzing the effect of increased political instability on the price impact of foreign trades through panel regressions, we find that foreign trades have a larger adverse selection component during times of political turmoil even after controlling for global and local macroeconomic conditions. We show that the firm specific characteristics such as size and profitability also significantly affect the adverse selection component of foreign trades. Specifically, foreign investors face smaller adverse selection costs for larger and more profitable firms in BIST over our sample period.

With regard to local investors, firm specific factors do not have a significant effect on the permanent impact of domestic trades. However, the transitory price impact of domestic orders is higher for stocks that have high firm specific risk. That is, companies that have high firm specific risk attract local uninformed investors. Our results show that these local uninformed investors tend to avoid investing in stocks that have high firm specific risks during election years in which political uncertainty is higher.

Overall, our findings are in line with the predictions of Choe et al. (2001) which suggest that domestic investors are better equipped in acquiring information about the underlying value of an asset as foreign investors are facing certain obstacles when investing abroad. However, we also show that informational advantage of foreign investors tends to increase with increased political instability.

Diversification concerns might be the underlying reason behind this finding. As Turkey's country specific risks become more isolated from the rest of the world, foreign investors who are more exposed to international risks can invest in Turkish stocks to diversify their global portfolio. Finally, the informational advantage of foreign investors on the stocks of commercial banks may arise due to large equity ownership and/or due to lending relationship.

REFERENCES

- Acemoglu, D. and Ucer, M. (2015). The Ups and Downs of Turkish Growth, 2002-2015: Political Dynamics, the European Union and the Institutional Slide. Working Paper 21608, National Bureau of Economic Research.
- Akdeniz, L., Altay-Salih, A., and Aydogan, K. (2000). A cross-section of expected stock returns on the Istanbul Stock Exchange. *Russian & East European Finance and Trade*, 36(5):6–26.
- Aktas, N., De Bodt, E., Declerck, F., and Van Oppens, H. (2007). The PIN anomaly around M&A announcements. *Journal of Financial Markets*, 10(2):169–191.
- Aktas, O. U. (2016). *Three Essays on the Microstructure of the BIST*. PhD thesis, Concordia University, Concordia University, Montreal, Quebec, Canada.
- Aktas, O. U. and Kryzanowski, L. (2014). Trade classification accuracy for the BIST. *Journal of International Financial Markets, Institutions and Money*, 33:259–282.
- Alkan, U. and Guner, B. (2018). Preferences for lottery stocks at Borsa Istanbul. *Journal of International Financial Markets, Institutions and Money*.
- Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.
- Aslan, H., Easley, D., Hvidkjaer, S., and O’Hara, M. (2011). The characteristics of informed trading: Implications for asset pricing. *Journal of Empirical Finance*, 18(5):782–801.
- Atilgan, Y., Demirtas, K. O., and Gunaydin, A. D. (2016). Liquidity and equity

- returns in Borsa Istanbul. *Applied Economics*, 48(52):5075–5092.
- Bailey, W. and Jagtiani, J. (1994). Foreign ownership restrictions and stock prices in the Thai capital market. *Journal of Financial Economics*, 36(1):57–87.
- Bekaert, G., Harvey, C. R., and Lumsdaine, R. L. (2002). The dynamics of emerging market equity flows. *Journal of International Money and Finance*, 21(3):295–350.
- Borsa Istanbul (2016). 2016 Borsa Istanbul Annual Report. <http://www.borsaistanbul.com/docs/default-source/yayinlar/borsa-istanbul-annual-report-2016.pdf?sfvrsn=4/>. [Online].
- Boutchkova, M., Doshi, H., Durnev, A., and Molchanov, A. (2011). Precarious politics and return volatility. *The Review of Financial Studies*, 25(4):1111–1154.
- Brogaard, J. and Detzel, A. (2015). The asset-pricing implications of government economic policy uncertainty. *Management Science*, 61(1):3–18.
- Celik, D. and Tiniç, M. (2016). Package “InfoTrad”. <https://CRAN.R-project.org/package=InfoTrad>.
- Celik, D. and Tiniç, M. (2018). InfoTrad: An R package for estimating the probability of informed trading. *The R Journal*, 10(1):31–42.
- 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):159–196.
- Choe, H., Kho, B.-C., and Stulz, R. M. (2001). Do domestic investors have more valuable information about individual stocks than foreign investors? Working Paper 8073, National Bureau of Economic Research.
- Çolak, G., Durnev, A., and Qian, Y. (2017). Political uncertainty and IPO activity: Evidence from US gubernatorial elections. *Journal of Financial and Quantitative Analysis*, 52(6):2523–2564.

- Dahlquist, M. and Robertsson, G. (2004). A note on foreigners' trading and price effects across firms. *Journal of Banking & Finance*, 28(3):615–632.
- Domowitz, I., Glen, J., and Madhavan, A. (1997). Market segmentation and stock prices: Evidence from an emerging market. *The Journal of Finance*, 52(3):1059–1085.
- Duarte, J. and Young, L. (2009). Why is pin priced? *Journal of Financial Economics*, 91(2):119–138.
- Durnev, A. (2010). The real effects of political uncertainty: Elections and investment sensitivity to stock prices. Available at SSRN: <https://ssrn.com/abstract=1549714>.
- Easley, D., Hvidkjaer, S., and O'Hara, M. (2002). Is information risk a determinant of asset returns? *The Journal of Finance*, 57(5):2185–2221.
- Easley, D., Hvidkjaer, S., and O'Hara, M. (2010). Factoring information into returns. *Journal of Financial and Quantitative Analysis*.
- 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):1405–1436.
- Easley, D. and O'Hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics*, 19(1):69–90.
- Easley, D. and O'Hara, M. (2004). Information and the cost of capital. *Journal of Finance*, 59(4):1553–1583.
- Easley, D., O'Hara, M., and Paperman, J. (1998). Financial analysts and information-based trade. *Journal of Financial Markets*, 1(2):175–201.
- Easley, D., O'Hara, M., and Saar, G. (2001). How stock splits affect trading: A microstructure approach. *Journal of Financial and Quantitative Analysis*, 36(01):25–51.
- Ekinci, C. and Ersan, O. (2018). A new approach for detecting high-frequency

- trading from order and trade data. *Finance Research Letters*, 24:313–320.
- Ellul, A. and Pagano, M. (2006). IPO underpricing and after-market liquidity. *Review of Financial Studies*, 19(2):381–421.
- Ersan, O. and Alici, A. (2016). An unbiased computation methodology for estimating the probability of informed trading (PIN). *Journal of International Financial Markets, Institutions and Money*, 43:74–94.
- Ersan, O. and Ekinici, C. (2016). Algorithmic and high-frequency trading in Borsa Istanbul. *Borsa Istanbul Review*, 16(4):233–248.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.
- Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636.
- Federal Reserve Bank of St. Louis, O. (2018). Oecd based recession indicators for turkey from the peak through the trough. <https://fred.stlouisfed.org/series/TURRECM>. Accessed: 2018-06-01.
- Foucault, T., Pagano, M., and Röell, A. (2013). *Market liquidity: theory, evidence, and policy*. Oxford University Press.
- Francis, B. B., Hasan, I., and Zhu, Y. (2014). Political uncertainty and bank loan contracting. *Journal of Empirical Finance*, 29:281–286.
- Froot, K. A., O’connell, P. G., and Seasholes, M. S. (2001). The portfolio flows of international investors. *Journal of Financial Economics*, 59(2):151–193.
- Froot, K. A. and Ramadorai, T. (2008). Institutional portfolio flows and international investments. *The Review of Financial Studies*, 21(2):937–971.
- Gan, Q., Wei, W. C., and Johnstone, D. (2015). A faster estimation method for the probability of informed trading using hierarchical agglomerative clustering. *Quantitative Finance*, 15(11):1805–1821.

- Gao, P. and Qi, Y. (2013). Political uncertainty and public financing costs: Evidence from us gubernatorial elections and municipal bond market. *Available at SSRN: <https://ssrn.com/abstract=1992200>*.
- Gibbons, M. R., Ross, S. A., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, pages 1121–1152.
- Glosten, L. R. (1987). Components of the bid-ask spread and the statistical properties of transaction prices. *The Journal of Finance*, 42(5):1293–1307.
- Glosten, L. R. and Harris, L. E. (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics*, 21(1):123–142.
- Glosten, L. R. and Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1):71–100.
- Gungoraydinoglu, A., Çolak, G., and Öztekin, Ö. (2017). Political environment, financial intermediation costs, and financing patterns. *Journal of Corporate Finance*, 44:167–192.
- Hasbrouck, J. (1991). Measuring the information content of stock trades. *The Journal of Finance*, 46(1):179–207.
- Ho, T. and Stoll, H. R. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial economics*, 9(1):47–73.
- Hughes, J. S., Liu, J., and Liu, J. (2007). Information asymmetry, diversification, and cost of capital. *The Accounting Review*, 82(3):705–729.
- İmişikler, S., Özcan, R., and Taş, B. K. O. (2015). Price manipulation by intermediaries. *Emerging Markets Finance and Trade*, 51(4):788–797.
- International Monetary Fund (2017). 2017 IMF DataMapper - Current account balance U.S. dollars. <http://www.imf.org/external/datamapper/BCA@WEO/OEMDC/ADVEC/WEOWORLD>.

[Online].

- Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of Finance*, 45(3):881–898.
- Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.
- Julio, B. and Yook, Y. (2012). Political uncertainty and corporate investment cycles. *The Journal of Finance*, 67(1):45–83.
- Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335.
- Lai, S., Ng, L., and Zhang, B. (2014). Does pin affect equity prices around the world? *Journal of Financial Economics*, 114(1):178–195.
- Lee, C. and Ready, M. J. (1991). Inferring trade direction from intraday data. *The Journal of Finance*, 46(2):733–746.
- Lin, H. and Ke, W.-C. (2011). A computing bias in estimating the probability of informed trading. *Journal of Financial Markets*, 14(4):625–640.
- Lin, J.-C., Sanger, G. C., and Booth, G. G. (1995). Trade size and components of the bid-ask spread. *The Review of Financial Studies*, 8(4):1153–1183.
- Menkveld, A. J. (2008). Splitting orders in overlapping markets: A study of cross-listed stocks. *Journal of Financial Intermediation*, 17(2):145–174.
- Müllner, D. (2013). fastcluster: Fast hierarchical, agglomerative clustering routines for r and python. *Journal of Statistical Software*, 53(9):1–18.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society*, pages 703–708.
- Odders-White, E. R. and Ready, M. J. (2006). Credit ratings and stock liquidity. *Review of Financial Studies*, 19(1):119–157.
- O’Hara, M., Yao, C., and Ye, M. (2014). What’s not there: Odd lots and market

- data. *The Journal of Finance*, 69(5):2199–2236.
- Osler, C. L., Mende, A., and Menkhoff, L. (2011). Price discovery in currency markets. *Journal of International Money and Finance*, 30(8):1696–1718.
- Ozsoylev, H. N., Walden, J., Yavuz, M. D., and Bildik, R. (2013). Investor networks in the stock market. *The Review of Financial Studies*, 27(5):1323–1366.
- Pastor, L. and Veronesi, P. (2012). Uncertainty about government policy and stock prices. *The Journal of Finance*, 67(4):1219–1264.
- Pástor, L. and Veronesi, P. (2013). Political uncertainty and risk premia. *Journal of Financial Economics*, 110(3):520–545.
- R Core Team, R. C. T. (2016). *R: A language and environment for statistical computing*. Vienna, Austria.
- Santa-Clara, P. and Valkanov, R. (2003). The presidential puzzle: Political cycles and the stock market. *The Journal of Finance*, 58(5):1841–1872.
- Simsir, S. A. and Simsek, K. (2018). Measuring the Market Impact of Private Information: The Curious Case of Corporate Announcements in Turkey. *Working Paper, Sabanci University, Sabanci School of Management*.
- Ülkü, N. and İkizlerli, D. (2012). The interaction between foreigners’ trading and emerging stock returns: Evidence from Turkey. *Emerging Markets Review*, 13(3):381–409.
- Yan, Y. and Zhang, S. (2012). An improved estimation method and empirical properties of the probability of informed trading. *Journal of Banking & Finance*, 36(2):454–467.
- Zagaglia, P. (2012). *FinAsym*. R package version 1.0.
- Zagaglia, P. (2013). Pin: Measuring asymmetric information in financial markets with r. *The R Journal*, 5(1):80–87.

APPENDIX: Variable Definitions

SIZE: Firm size for a given month is taken to be the natural logarithm of the firms' market value, that is, the end of month price times the number of shares outstanding.

BOOK TO MARKET: Book to market ratio for a firm is taken to be the ratio of the firms end of month total equity value to its market value.

BETA: We estimate the market risk of a given stock i , on a given month t , from the market model given below;

$$R_{id} = \alpha_i + \beta_i R_{md} + e_{id} \quad d = 1 \dots D_t \quad (5.1)$$

where D_t is the number of trading days on month t . R_{id} is the daily return on the stock i on a given day d . Similarly we take R_{md} to be the return of the BIST100 index of a given day d . We estimate equation (5.1) for each stock using daily returns within a month.

ILLIQUIDITY: We measure stock illiquidity for each month, t , with the ratio of the absolute monthly stock return to its trading volume, similar to Amihud (2002).

$$ILLIQ_{it} = |R_{it}|/VOL_{it} \quad (5.2)$$

IDIOSYNCRATIC VOLATILITY: We measure the idiosyncratic risk of a stock from the standard deviation of the daily residuals, e_{id} , presented in (5.1). That is,

$$IVOL_{it} = \sqrt{var(e_{id})} \quad (5.3)$$

REVERSAL: We define the reversal variable for each stock in month t as the return

on the stock over previous month, similar to Jegadeesh (1990).

MAX: We define the maximum daily return within a month as:

$$MAX_{it} = \max(R_{id}), \quad d = 1 \dots D_t \quad (5.4)$$

where D_t is the number of trading days on month t . R_{id} is the daily return on the stock i on a given day d .

MOMENTUM: We define the momentum variable for each stock in month t as the cumulative percentage return over the last six months. That is,

$$MOM_{it} = \frac{P_{it-1} - P_{it-7}}{P_{it-7}} \quad (5.5)$$

where P_{it} is the price of the stock at the end of month t .

TOTAL SPREAD: Following Glosten and Harris (1988), we define the total spread for a stock i on month t to be the twice the sum of transitory component C_t and the adverse selection component of the stock Z_t .

$$TOTS_{it} = 2 * (C_{it} + Z_{it}) \quad (5.6)$$